University of the
West of England

3D Face Recognition Using Photometric Stereo

Mark F. Hansen

A thesis submitted in partial fulfilment of the requirements of the University of the West of England, Bristol for the degree of Doctor of Philosophy

Faculty of the Environment and Technology

University of the West of England, Bristol

May 2012

Abstract

Automatic face recognition has been an active research area for the last four decades. This thesis explores innovative bio-inspired concepts aimed at improved face recognition using surface normals. New directions in salient data representation are explored using data captured via a photometric stereo method from the University of the West of England's "Photoface" device. Accuracy assessments demonstrate the advantage of the capture format and the synergy offered by near infrared light sources in achieving more accurate results than under conventional visible light. Two 3D face databases have been created as part of the thesis – the publicly available Photoface database which contains 3187 images of 453 subjects and the 3DE-VISIR dataset which contains 363 images of 115 people with different expressions captured simultaneously under near infrared and visible light. The Photoface database is believed to be the first to capture naturalistic 3D face models. Subsets of these databases are then used to show the results of experiments inspired by the human visual system. Experimental results show that optimal recognition rates are achieved using surprisingly low resolution of only 10×10 pixels on surface normal data, which corresponds to the spatial frequency range of optimal human performance. Motivated by the observed increase in recognition speed and accuracy that occurs in humans when faces are caricatured, novel interpretations of caricaturing using outlying data and pixel locations with high variance show that performance remains disproportionately high when up to 90% of the data has been discarded. These direct methods of dimensionality reduction have useful implications for the storage and processing requirements for commercial face recognition systems. The novel variance approach is extended to recognise positive expressions with 90% accuracy which has useful implications for human-computer interaction as well as ensuring that a subject has the correct expression prior to recognition. Furthermore, the subject recognition rate is improved by removing those pixels which encode expression.

Finally, preliminary work into feature detection on surface normals by extending Haar-like features is presented which is also shown to be useful for correcting the pose of the head as part of a fully operational device. The system operates with an accuracy of 98.65% at a false acceptance rate of only 0.01 on front facing heads with neutral expressions. The work has shown how new avenues of enquiry inspired by our observation of the human visual system can offer useful advantages towards achieving more robust autonomous computer-based facial recognition.

Contents

List of Figures

Acknowledgements

This document would not have been possible without the help and support of many people on both academic and personal levels. My supervisory team of Dr. Gary Atkinson, Prof. Melvyn Smith and Dr. Lyndon Smith have always been extremely helpful and inspirational; they are all remarkable at finding time to answer my questions or provide constructive feedback; I don't remember any of them once asking me to come back later.

I'd also like to thank the late Prof. Tom Troscianko, whose lectures on vision and perception during my first degree were inspirational, and for asking one of his usual insightful questions during a Bristol Vision Institute seminar, which helped steer the early stages of my research.

I am extremely grateful to my family for allowing me to drop out from a mainstream career to study something I find extremely interesting. My thanks go especially to my loving and lovely partner Suzanne for allowing me to leave a well paid job and become a student again and to my son Joel for teaching me the true meaning of tiredness and just how important a good night of sleep can be.

The Ph.D. as a whole would not have been possible without the funding of the EPSRC. Participating in a number of international conferences has been a highlight of the project and attending these would not have been possible without additional funding from the British Machine Vision Association, The University of the West of England and the National Science Foundation in the USA.

List of Publications Arising from this Work

The importance of publications cannot be underestimated, both for the importance of having the work peer reviewed, but also as a means of publicity. This Ph.D. has given me a wealth of experience in these areas as detailed below.

Journal Publications

[1] M. F. Hansen, G. A. Atkinson, L. N. Smith, and M. L. Smith, "3D face reconstructions from photometric stereo using near infrared and visible light," Computer Vision and Image Understanding, vol. 114, no. 8, pp. 942-951, Aug. 2010.

[2] M. F. Hansen, G. A. Atkinson, M. L. Smith, L. N. Smith, "Psychologically inspired dimensionality reduction for 2D and 3D Face Recognition," The Visual Computer, [submitted]

[3] S. Zafeiriou, M. F. Hansen, G. A. Atkinson, W. A. P. Smith, M. L. Smith, M. Petrou, "The effect of different photometric stereo and reconstruction algorithms on face recognition accuracy," Pattern Recognition, [to be submitted]

Conference Publications

[1] G. A. Atkinson, M. F. Hansen, M. L. Smith, and L. N. Smith, "A efficient and practical 3D face scanner using near infrared and visible photometric stereo," in Proc. International Conference and Exhibition on Biometric Technology, Procedia Computer Science, vol. 2, pp. 11–19, 2010.

[2] S. Zafeiriou, M. F. Hansen, G. A. Atkinson, M. Petrou, and M. L. Smith, "Baseline face recognition using photometric stereo data," in Proc. International Conference and Exhibition on Biometric Technology, Procedia Computer Science, vol. 2, pp. 20–25, 2010.

[3] M. F. Hansen and G. A. Atkinson, "Biologically inspired 3D face recognition from surface normals," in Proc. International Conference and Exhibition on Bio-

metric Technology, Procedia Computer Science, vol. 2, pp. 26–34, 2010.

[4] S. Zafeiriou, M. F. Hansen, G. A. Atkinson, V. Argyriou, M. Petrou, M.L. Smith, L.N. Smith, "The PhotoFace Database," in Proc. Biometrics Workshop of Computer Vision and Pattern Recognition, Colorado Springs, USA, 2011, pp. 161-168.

[5] V. Vijayan, K. Bowyer, P. Flynn, D. S Huang, L. Chen, M. F. Hansen, O. Ocegueda, S. Shah, I. Kakadiaris "Twins 3D Face Recognition Challenge," in International Joint Conference on Biometrics, Arlington, Virginia, 2011.

[6] M. F. Hansen, G. A. Atkinson, L. N. Smith, and M. L. Smith, "Psychologically inspired dimensionality reduction for 2D and 3D face recognition," in Proc. of the 3rd British Machine Vision UK Student Workshop, Dundee, 2011, pp.15-26.

Other

Seminars and Lectures

[1] "3D face reconstruction and recognition using photometric stereo", Ph.D. Showcase Lecture, 2010.

[2] "3D face reconstruction and recognition using photometric stereo", Bristol Vision Institute Seminar, 2010.

[3] "Photometric stereo face recognition", Invited Speaker, Association of Biometrics, London, 2010

[4] "Face recognition: an introduction to 2D, 3D and human approaches", Lecture to 2nd Year Undergraduates, University of the West of England, 2009

Posters

[1] "3D face reconstruction and recognition using Photometric Stereo", Bristol Vision Institute Student Colloquium, 2010.

[2] "Computationally efficient bio-inspired dimension reduction for 3D faces", selected for Doctoral Consortium, International Conference on Automatic Face and Gesture Recognition, California, 2011

Miscellaneous

[1] “FaceTime”, Photonics Spectra, 2010, Interviewed for feature article (http://www.photonics.com/Article.aspx?AID=43362)

[2] “Daily Planet” 2011, Discovery Channel television program featured the work of the Machine Vision Laboratory

Reviews

[1] Two journal articles for Computer Vision and Image Understanding

[2] One journal article for the International Journal of Computer Vision and Signal Processing (on which this author sits on the board of editors)

[3] One conference paper for the Ninth International Conference of Computational Methods in Sciences and Engineering 2011

[4] Three conference papers for the International Conference and Exhibition on Biometric Technology 2010.

[5] Twelve conference papers for the International Conference on Informatics, Electronics & Vision 2012.

Chapter 1

Introduction

> It is the common wonder of all men, how among so many millions of faces, there should be none alike.
>
> Thomas Browne, Religio Medici

Humans see faces in just about everything. Two eyes, a nose and a mouth in near identical configuration and yet we are able to distinguish between thousands and readily recognise familiar faces in an instant.

Automated face recognition has been the subject of considerable research effort for four decades. It offers many perceived security benefits and as a biometric does not require such levels of cooperation as iris recognition nor physical contact that is necessary for most fingerprint readers. We as humans believe we are very good at it – it is arguably a primary mechanism for recognising people. Holding a BSc in Psychology and an MSc in Computer science it offers a fascinating area of study for the author.

Face recognition and pattern recognition in general can be split into two discrete areas of research: feature extraction and classification. The first concerns itself with extracting the important information from a raw signal leaving obfuscating information behind and the latter finds ways of reliably labelling the features. Even though both have been areas of research for many decades they are still

rich subject areas. This work focuses on the former and attempts to use psychologically inspired ideas to improve upon existing research on 3D face recognition, ultimately leading to the proposal of a system which performs landmark localisation, feature extraction and recognition automatically using surface normal data generated from a 3D Photometric stereo (PS) capture device. The device was developed prior to the start of this project and while its operation is described in some detail, the design and construction of the device itself is not part of this Ph.D..

At the beginning of this Ph.D. a colleague questioned why I would choose automated face recognition as a research topic as it is already *saturated* with research, and it is often stated that simple face recognition is a solved problem. While there is undoubtedly a level of truth to these statements the author does not wholly agree with them. It is very hard to find novel areas of research in an area with such a seemingly simple purpose but in this thesis I have attempted to explore under-represented areas such as caricaturing and low resolution imagery to enhance face recognition. Countering the argument that the face recognition is a solved problem, one only has to look at the recently released work by some of the best known names in the area on an old dataset as proof that it is by no means solved (even in the simplest of cases – expressionless and frontal) [151]. Here Phillips *et al.* use the Face Recognition Grand Challenge (FRGC) database to create a new set of test faces that have been shown to be difficult to recognise by the best algorithms.

Face recognition as a technology has the possibility of becoming ubiquitous in the near future as long as the public are aware of its pitfalls in return for the convenience it offers. Already in the latest version of the Google Android ^{TM} operating system, 'Ice Cream Sandwich', an option which allows a device to be unlocked via face recognition has been included. At the time of writing, Google have not explicitly confirmed whether or not it can be spoofed into false verification by using a photograph, only stating "Give us some credit!". Even though research

into automated face recognition is in its fifth decade, it is still current, and this is what makes it such an appealing research topic.

Another aspect of faces which makes them an intriguing study subject are the idiosyncrasies associated with human processing. They all (usually) are in a very similar configuration; having two eyes above a nose above a mouth and yet we can make judgements about a person in an instant based on very superficial differences. Even though the variations in these configurations are slight, we have no problems instantaneously recognising familiar faces even after long periods or when the appearance has altered through spectacles or facial hair. The human ability to recognise familiar faces is remarkable under normal conditions although we tend to overestimate our ability with unfamiliar faces [83] which can have serious consequences in police line-ups. We also recognise familiar faces which have been cartooned or caricatured better than veridical images. However if faces are inverted either in orientation or in contrast polarity our ability to process them is extremely diminished. Damage to a small area of the brain (Fusiform face area (FFA)) can lead to a condition where the person is unable to recognise faces at all, but recognition of other objects remains intact (prosopagnosia). The fact that a process so inherent in human nature is still so poorly understood makes it a fascinating area to study and it may be that as we learn more of the human processes we can use them to augment automatic systems.

While automated 2D face recognition has been studied for decades, using 3D scans is a far more recent topic due to the availability of suitable scanning devices. As well as providing 3D morphology of the face 3D scans help to overcome the problems of pose and illumination that have been found in 2D recognition. Throughout this thesis I refer to the surface normal data captured using PS as 3D data. I am aware that the data is not truly 3D as we do not capture the full 3D shape, only the part facing the camera, but the alternative 2.5D is clumsy and slightly ambiguous and does nothing to aid in the readability of the manuscript.

With over 40 years of research, the variety of approaches is vast and many

build on the successes of previous attempts. This often means that methods become more and more specific and complex to cater for special cases. This is how science works and I am in no way criticizing it. Working towards a Ph.D. gives one the unique opportunity to explore areas which perhaps have been neglected or overlooked or come at the problem with a different viewpoint and the novel methods presented here reflect this rather than extend specific aspects of previous successes (of which there are many!). Discussing this with a colleague I was pointed towards a paper which takes a similar standpoint, and very successfully demonstrates that for texture classification simple measures of intensity in a number small patches gives better performance than using filter banks [185].

The journey of work in this Ph.D. has not only allowed me study a fascinating subject area for three years, but has also allowed me to travel and meet with highly respected scholars in the area. I hope that the reader enjoys and is stimulated by the findings reported here as much as I have been.

Contributions

The major original contributions of this project are:

1. The use of Near Infrared (NIR) light sources provide a more accurate PS reconstruction than visible light sources

2. Providing a publicly available database and detailed metadata of 3187 sessions of 453 subjects

3. That Low resolution images (10×10 px) give the best rates of recognition for our dataset of frontal, expressionless faces contained in our 3D database using raw surface normals. Similarly low resolutions are also optimal for 2D images.

4. The finding that pixel locations with statistically outlying values, or high vari-

ance rates contain disproportionately high amounts of discriminatory information which is useful for face recognition

5. Presenting a variance based approach which is used to identify pixel locations which encode expressions. Removing pixels at these locations is shown to improve face recognition performance.

6. A proposed fully automatic face recognition system incorporating all of the above findings with a empirically determined threshold allowing a 98.65% accuracy at a False Acceptance Rate of 0.01.

Applications

Apart from the academic interest in understanding face recognition and developing robust algorithms there are many real-world applications of the technology. The advantages of face recognition being used as a biometric over other commonly used methods such as iris recognition or fingerprints are that it requires significantly less cooperation and, in comparison to fingerprint biometrics, no physical contact is required. The applications can broadly be categorised into physical security, computer security, surveillance and the computer game industry.

Physical security is perhaps the most obvious application. Within this category, face recognition can be used to control access to buildings and access to secure areas within buildings instead of using cards and combination codes which can be forgotten or copied. Already in use in certain airports, passport verification using face recognition is used at border control and extending this to 3D is likely to bring additional robustness.

Instead of using passwords for computer access, face recognition can be used. As mentioned earlier, Google are incorporating it into their Android smartphone operating system. It is useful when users cannot touch a keypad for ex-

ample in industrial environments. Additionally it could serve as an extra level of security at a document level such that a screen becomes obscured if an unauthenticated person appears in view of the monitor. 3D systems offer the advantage of being harder to spoof – a 2D recognition system can often be fooled by a photograph – but it is far harder to create an accurate 3D mask of an individual. An attacker would be more likely find a simpler weakness to exploit.

It is harder to incorporate 3D face recognition into surveillance due to the fact that some cooperation is required to capture a 3D scan. Theoretically though, a long range NIR PS system could be used to capture the images and then used to covertly monitor and flag persons on a watch list e.g. at football games during entry to a stadium.

Computer games are already using face recognition in the form Microsoft's Kinect^{TM} 3D camera to recognise individual players. Another use more suited to the high resolution PS 3D capture would be for incorporating an accurate 3D face of a player into the game.

There is also interest from advertising companies for tailoring advertisements to the demographic of a person e.g. a male teenager would likely be interested in computer games and music, therefore display this kind of marketing material. If a database of regular shoppers and shopping habits (from receipts) could be built up, a far more accurate system could be capable of matching products and offers to individuals. While such concepts challenge notions of privacy and data protection, they also challenge the common idea that marketing and advertising are mere nuisances as they could become more of a service if they actually provided accurate, relevant and useful information.

There are undoubtedly many more such applications, but these give an idea of the diversity of applications and benefits of using 3D face recognition. The work presented here lends itself well to biometric uses for security purposes. It is probably best suited to commercial systems for verification and authentication in SME sized companies with up to a few hundred employees for authentication

purposes, or verification systems for granting access to restricted areas.

Thesis Outline

The literature review in the next chapter aims to provide the context of this work from an historical as well as state-of-the-art point of view. 2D and 3D automated methods are covered, followed by an overview of certain aspects of human processing of faces and the visual system as a whole. Then a critique of the currently available face databases follows in order to highlight the contributions of the ones produced during this Ph.D., before methods of 3D acquisition are compared and contrasted.

Chapter 3 assesses the accuracy of the Photoface PS capture device using visible and NIR light sources by comparison with those captured using the commercially available 3dMD system. It shows that using NIR light sources is slightly more accurate than visible light, probably due to skin reflectance being more Lambertian under NIR. The overall errors of the PS reconstructions are sufficiently low and are judged to suitable for face recognition.

Once the suitability of the data has been demonstrated, Chapter 4 introduces the two databases which have been developed and used for experiments in this thesis together with baseline results using some common data representations and algorithms. This allows us to select the raw surface normals as a suitable representation which offers very good recognition performance without the need for the additional processing required for other representations. One of the databases (the Photoface Database) is now publicly available to other researchers on request.

Chapters 5 and 6 introduce direct methods of dimensionality reduction. The experimental results show that the recognition rate is unaffected when the images are resized to as little as 10×10 px which seems counter-intuitive but, interestingly, is comparable to spatial frequencies favoured by humans in face recognition

studies. Chapter 6 implements a percentile based theory to find support for Unnikrishnan's [182] hypothesis that the outlying percentiles contain more discriminatory information than central percentiles. The basis for this work comes from the fact that humans can recognise face caricatures as quickly or often faster than veridical images. The percentile approach is extended to finding pixel locations which vary the most and using these for recognition. It is shown that by only using the 10% most varying pixels locations, and discarding the rest, that recognition rates decrease by a disproportionate amount which is taken as indication of the discriminatory power of this direct method. It may be possible to incorporate these findings into a weighting system for different parts of the face make recognition more robust to expression.

A likely shortcoming of the variance approach is that it will fail in the case of data containing expressions. This is examined in Chapter 7 where it is shown that the most varying pixel locations encode expression and that by removing them, subject recognition rates can be improved. Although the data are limited, the results also show that it is easiest to discriminate positive expressions from neutral or negative expressions (with a 90% accuracy).

In attempting to define an operational system, the final experimental chapter investigates two areas which are still absent. The databases used in this research have all been manually labelled to allow accurate alignment. An automated system would be required to automatically extract the features and align them. Inspired by the work of Viola and Jones [187] on using Haar-like features, new features are investigated for use with surface normal components to locate the lateral canthi (the outer corner of the eye where the upper and lower eyelids meet) and nose tip. Additionally the threshold required for a verification rate of 98.65% with only one false positive in every 100 trials is determined empirically.

The concluding chapter brings the major findings of the work in this thesis together and discusses their implications in terms of face recognition applications as a whole as well as the limitations specific to this project and those pertaining

to biometrics in general.

Chapter 2

Literature Review

Face recognition is one of the most active research areas in machine vision. After over 40 years of study into the problem, great progress has been made but a universally accurate computer based system is still elusive. Unless affected by prosopagnosia (also known as face blindness), humans are able to recognise familiar faces effortlessly with great accuracy from a very early age [45]. Most computer based systems are not based on any biologically plausible model of the Human Visual System (HVS) – this is not a criticism as the processes involved are poorly understood, but it is a possible reason for the disparity in performance. There have been improvements in our understanding of the HVS in recent years and it may be possible to adapt some of these in order to improve face recognition rates.

It is not only of academic interest; accurate automated face recognition is also of potentially great commercial value in security areas such as access to buildings, border control and surveillance. The vast majority of this research has used 2D data because until recently the use of 3D data for this task was prohibitively expensive or of poor quality. However it is now both affordable and available in the form of Photometric stereo (PS), and the use of 3D potentially overcomes illumination and pose problems associated with 2D recognition.

The University of the West of England, Bristol has developed PhotoFace, a

four source PS face capture device capable of generating 3D models of faces. It is expected that by utilising the surface normals inherent to this methodology together with information known about the HVS and higher level processes involved in recognition, that novel methods which aid recognition will be found.

This literature review will therefore review the history and current state of research into 2D and 3D face recognition. An overview of 3D acquisition techniques will then be given, showing that PS is a particularly valid technology for this process. The key properties of the HVS and the higher level processes involved in human face recognition will then be presented, with particular attention to those that may lend themselves well to the problem but appear not to have been fully explored in the literature.

2.1 Automatic Face Recognition

This section outlines some of the most successful approaches and directions of research in both 2D and 3D automatic face recognition together with their limitations.

2.1.1 Early Methods

The first research into computer based face recognition began in the 1960s in which a simple system was developed by Bledsoe based on 20 manually extracted measurements of facial features from photographs [25]. A database of these measurements was then used to return the closest matching records given another set of (manually extracted) measurements of a person. The system worked to a limited extent but recognition accuracy depended greatly on the pose of the subject. To overcome this problem, Bledsoe normalized the photographs by building up a 3D model of an average head in order to carry out accurate transformations. Using a 3D model in a similar way has been used successfully to limit

the affects of pose in recent research [24], [12], which will be discussed more fully in Section 2.1.3.

The work of Kelly [105] helped Kanade [100] to produce the first automatic face recognition system. The device extracted feature points using edge detection and the geometric measurements of these features (16 parameters – 11 distance ratios and a combination of angles and curvatures) for recognition. The system correctly identified 15 out 20 people, marking the *'... first success of machine identification of human-face photographs'* [100].

A milestone in the history of facial recognition took place with the introduction of eigenfaces [181] which was inspired by the earlier work of Sirovich and Kirby [168]. Instead of treating the task as a geometric problem and measuring distances between fiducial features, Turk and Pentland represented the problem as one of statistical variance motivated by information theory. The underlying theory behind this concept is that, given sufficient images of faces, Principal Components Analysis (PCA) is performed and the dimensions containing the major variations between the images are extracted. If the images are then projected into the new eigenspace composed of the eigenvectors with the most significant eigenvalues, they ideally form clusters for each individual (alternatively the mean face image for an individual can be used). If a probe image of an individual is projected into the eigenspace, then it is possible to see which cluster it is nearest to (e.g. by Nearest Neighbour classification), and for that image to be classified as a particular person.

The eigenface approach means that once the images are normalized and registered, fiducial points are not required for recognition. However it is very sensitive to illumination and pose. Using a database of 2500 face images of 16 individuals, correct recognition rates of 64% to 100% were reported under various conditions [181].

PCA is the most commonly used technique in face recognition literature from its first application in eigenfaces [181] to its use in many more recent papers as a

baseline performance measure [15], [28], [74], [37], [134], [155]. [103]. It has also been generalized to Principle Geodesic Analysis (PGA) allowing shape analysis for 3D data [170].

However, the method is not without its limitations. When new faces are needed to be identified, they must be added to the initial training set and PCA run over the entire set of images in order to generate a new face space. Illumination and pose have a great effect on the images themselves and thus on the statistical results. Variations in illumination have been found to correspond to the three eigenvectors which have the largest eigenvalues. Disposing of these helps reduce the effect of illumination [17], but as these contain the most variation, they also discard plenty of information which may have been useful in face discrimination. Even if the images are normalized to reduce the effects of scale and orientation, they introduce artefacts of the scaling or transformation process e.g. scaling is the same as changing a camera's focal length not the camera's position [180]. Having to choose how many eigenvectors to use is another weakness – too many, and there is the risk that there will be too much noise and that computation will be more intense, too few, and the face space will be too small and faces of different individuals will be too close together for discrimination. Using PCA for face recognition also suffers from an *own-race effect* whereby it is better at discriminating faces from other races [67], contrary to human face recognition which displays the *other race effect* [124].

With the success of this milestone into using statistical methods, more recent advances in 2D and 3D face recognition will now be discussed.

2.1.2 2D Face Recognition

The vast majority of research into face recognition has focused on 2D images. This is due to the fact that 2D images (e.g. photographs) are more readily available and easier to obtain than 3D data of faces. Key research in this area is doc-

umented here, together with some promising recent directions. The main focus of this thesis is 3D face recognition, but it is important to understand the historical importance of the research in terms of 3D developments and also the limitations of using 2D information, some of which can be overcome using 3D data.

Continuing in the success of using statistical tools for face recognition and in an attempt to overcome some of the shortcomings of using PCA, Belhumeur *et al.* [17] developed what they termed *Fisherfaces*. This technique uses a derivative of Linear Discriminants Analysis (LDA) called Fisher's Linear Discriminant (FLD) which although similar to PCA explicitly attempts to find the dimensions which limit the intra-class scatter and maximise the inter-class scatter rather than finding the dimensions along which most variation occurs. In this case the classes correspond to different people and since the data set for face recognition is already labelled it makes sense to use it in order to reduce the problem space.

To test the efficacy of their approach two databases were used – one constructed by the Harvard Robotics Laboratory which contained images with well constrained illumination variations and the publicly available Yale Face B database. Under the most extreme illumination, the Fisherface technique performed at an error rate of 4% compared to 42% for the eigenface technique and in those images with different facial expressions and illumination the error rate is as low as 0.6% compared to 19.5% for eigenfaces.

LDA is similar to PCA in that it transforms the data onto another coordinate system to describe variance. The main difference comes from the fact that no assumptions for the data are necessary for PCA (known as unsupervised), LDA requires that the data be separated into classes (supervised). This is why it is so attractive (and successful) when applied to face recognition, as the classes correspond to different people.

A graph showing how the two methods treat the same data differently can be seen in Figure 2.1 (reproduced from [17]) which plots the data onto the principal dimension in PCA and LDA space. In the example given, it can be seen that LDA

has separated the two classes more successfully than PCA.

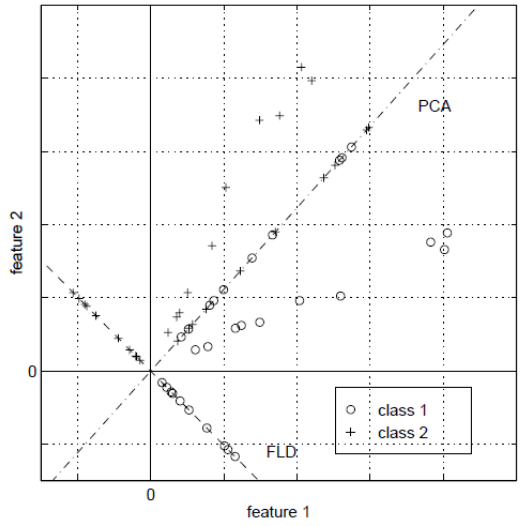


Figure 2.1: A comparison of principal component analysis (PCA) and Fisher's linear discriminant (FLD) for a two class problem where data for each class lies near a linear subspace. *Reproduced from [17]*. While both methods have reduced the dimensionality of the data to one, the FLD projection clearly demonstrates a linear separation between the two classes (in this case at the origin) while the PCA has not.

There are many papers which compare the effectiveness of eigenfaces to Fisherfaces e.g. [17, 127, 162], and although the assumption is that because it deals with classes it should lead to better performance on face recognition tasks, this has been shown to not always be the case i.e. with small training sets [127], known as the small sample size problem.

The Fisherface technique provided error rates so low that one might think that the problem of face recognition had effectively been solved. However, this view was not confirmed by a comparison of eigenfaces, Fisherfaces and FaceIT (a commercial system based on Local Feature Analysis) using data from three publicly available databases was performed by Gross *et al.* [80]. The databases contain far more variation in terms of illumination, pose (viewpoint), expression, ageing and occlusion. They concluded that although recognition algorithms were robust against illumination and to a certain extent expression (except for extremes such as a scream), significant decreases in performance are caused by pose,

ageing and occlusion. Throughout the tests, the FaceIt systems performs best, followed by Fisherfaces and then eigenfaces. All techniques showed a marked deterioration when the pose was greater than 32° from full frontal, when parts of the face were occluded, and when images from the AR database were taken two weeks apart. Interestingly they also report on a gender effect where better results are consistently achieved by all algorithms for recognising women.

Another statistical tool which has been used for face recognition is Independent Components Analysis (ICA). Bartlett *et al.* compared ICA against PCA as a benchmark on a subset of the Facial Recognition Technology (FERET) database and found it to be statistically better under certain conditions [15]. They used two architectures for implementing ICA and boosted performance by using a combined classifier recording a maximum recognition rate of 91% (compared with 85% using PCA). While PCA seeks to separate correlations in the data, ICA seeks to recover statistically independent sources of variations via non-orthogonal transforms (see Fig. 2.2). It has been used successfully in the cocktail party phenomenon, whereby individual conversations can be separated out from the ambient noise. The two architectures used in [15] arise as the first treats images as random variables and pixels as outcomes and the second treats pixels as the random variables and images as outcomes. The architectures produce very different outcomes – the first produces spatially localized images (i.e. mouth, eye, nose regions) and the second produce non-localized images (more like eigenfaces), and lead to better performance on identification and expression classification respectively [57].

More detail and analysis of these statistical methods can be found in Section 4.4. Now some commonly used non-statistical methods will be reviewed.

A Gabor filter is a sinusoid windowed with a Gaussian. As such, when convolved with fiducial features or the face as a whole they can provide a compressed representation of the information in what is termed *Gabor space*. The HVS has been shown to have primitive cortical regions sensitive to specific spatial fre-

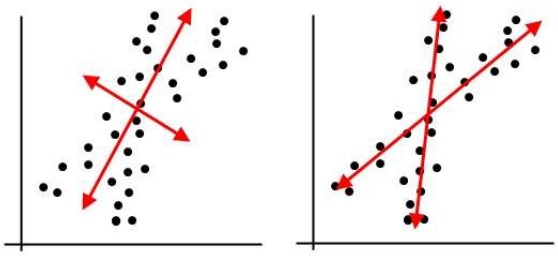


Figure 2.2: Example 2D data and co-ordinates found via PCA and ICA. The data points (black spots) are best represented along non-orthogonal axes. Applying PCA would have results similar to the plot on the left while ICA (right) would be able to select non-orthogonal axes thereby representing the data more accurately.

quency and orientations [95] and that these can be modelled using 2D Gabor functions [50]. Using Gabor filters could therefore be said to have a neurological basis when used in face recognition to extract salient features from the image of a face.

Manjunath *et al.* [126] discovered that the features highlighted by the Gabor filters often corresponded to features such as the nose and eyes. The feature vectors and spatial locations for each feature are used to build up a topological graph for each face. The graphs are then used to carry out the recognition which was reported to be with an accuracy of 86% for the correct person to be returned and 94% for the person to be in the top three candidate matches on a database of 300 images of 86 persons. Their approach to pose and illumination issues was to capture images of people under as many conditions as possible.

An extension to using Gabor filters is to use Gabor jets which are a stack of Gabor filters of differing orientations, frequencies and scales. They are used in successful techniques pioneered in Dynamic Link Architecture (DLA) [110] and Elastic Bunch Graph Matching [191]. The Gabor jets are applied to fiducial points on a face such as the corners of the eyes and the mouth for three pose images (full frontal, half-profile and profile) to create a Face Bunch Graph. A face can then be compared to any other using a similarity metric that takes into account the appearance and spatial configuration of the fiducial features. Performance using data from the FERET database was 98% using full frontal images, and 57%

for half-profile images. This novel technique demonstrated robust recognition for pose up to a 22° angle and differences in facial expression. The success of such algorithms is cited in Shen and Bai's 2006 review of Gabor wavelets [163], in which algorithms using Gabor wavelets came top in the FERET tests [148], and held the top two positions in the Face Authentication Test 2004 [129].

A different approach is used by Cootes *et al.* in the development of Active Shape Model (ASM) [42] and Active Appearance Model (AAM) [41]. The impetus for this research was to be able to parameterise how the shape of objects can change between examples and as such it has been found to be a robust metric in face recognition research as well. Fiducial features are marked on images of many examples of the object in question (in the case of [42] the objects were resistors, hands and heart scans). The positions between these markers vary, and these variations are analysed via PCA in order to provide information about how the shape of a particular object can change between examples. Once these variations are learned, it is possible to use the model in order to find examples of the objects in previously unseen images. An AAM extends on this to include the variation parameters of greyscale or texture information. Using this approach in face recognition Edwards *et al.* [59] achived an accuracy rate of 88% on 200 images of 20 individuals.

AAMs have also been used to match and remove expressions from face images to create neutral expression images which are then used for recognition. Lee and Kim [112] used 1280 face images of 80 subjects with four facial expressions in the frontal pose under four moderate illuminations, and compared performance of unprocessed images to those that had been transformed to a neutral expression using Nearest Neighbour, Linear Discriminant Analysis + Nearest Neighbour and Generalized Discriminant Analysis + Nearest Neighbour. The highest recognition rate reported for images that had been transformed to neutral expressions 96.7% compared to the highest for unprocessed images of 79.2%. The idea that a neutral expression results in better recognition seems intuitive.

However, research on the Face Recognition Grand Challenge (FRGC) questions this and has found that better results occur when the subject smiles, at least when a single face image is enrolled [21].

All of the above research has focused on using faces captured in visible light. As has been shown in numerous studies, variations in ambient lighting produces significant degradation in recognition performance [206]. Apart from those approaches above which aim to minimize this problem, using light outside of the visible spectrum has also been researched. Thermal infrared (thermal IR) has been used in facial recognition systems with some success especially as using thermal IR is robust against ambient illumination, some disguises and is useful for face detection and ensuring that a subject is alive. However, the affordability of sensors, its inherent inability to cope with spectacles (glass is opaque to thermal IR) [166] and its sensitivity to the changes in the thermal appearance of a face e.g. after exercise or due to fever means that it currently does not offer a viable solution on its own. However, there has been some success of combining the benefits of thermal IR with visible light techniques [171].

Near Infrared (NIR) has the potential to overcome the problems associated with visible and thermal IR face recognition: it is more robust against illumination variations, is useful for face detection [56], sensing equipment is far cheaper and it is not sensitive to changes in facial appearance caused by heat. NIR is useful for face detection for two reasons: the *bright eye* effect [137] allows the eyes to be localized and skin reflectance properties at just above and below 1.4 microns allows face regions to be clearly highlighted [56]. This phenomenology also has the potential to make a system more robust to disguise. At least three fully automated (face detection, feature extraction and recognition) NIR face recognition systems are present in the literature ([211, 204, 120]) and all report high recognition rates 99.7%, 79.6%, 86.8% respectively (and even higher if manual intervention is involved). The algorithms employed to perform the recognition itself vary (e.g. Discrete Cosine Transform (DCT) with Support Vector Machine (SVM), Lo-

cal Binary Patterns (LBP) and then adaboost classifiers, LDA with SVM) but the high accuracy, especially with manual intervention, indicate how robust NIR is for face recognition, although applications will be limited to controlled environments (near an NIR source).

An emerging area in 2D face recognition research has been to use the theory of sparse representation and compressed sensing [193]. Although at present the research has been limited to full frontal face images, the results are impressive, especially those concerning the resiliency against disguise and occlusion whereby 100% recognition rate is recorded when the 30% of the face is occluded and still achieving a rate of 65.3% when half of the face is occluded. On experimentation with natural occlusions e.g. sunglasses and scarves the rates are somewhat lower but still far better than competing algorithms such as PCA. By adding a simple partitioning algorithm, accuracy on the scarf occlusion which covers about 40% of the face is raised to 93.5%.

This technique has some parallels with HVS processes although this is not mentioned as a motivation. The visual cortex has been shown to be organised in such a way so that visual perceptions are a result of sparse representation of visual patterns [141]. What this means is that although a very large number of neurons are fed the visual signal, very few respond to it, but those few that do are sufficient for an accurate representation from which perception and recognition can occur. In the same way, when a human sees a face, they don't compare it to every face they have seen before (as a computer would in a database), but based on all the faces they have seen before, the brain is able to produce a response which gives rise to recognition.

Chapter 6 introduces two methods of dimension reduction which could be seen as a form of compressive sensing. They show that certain pixel locations contain proportionately more discriminatory information. In finding these locations, the other pixels can either be discarded or ignored - a method of extracting the important from the background information has been found, resulting in a

relatively sparse representation from direct methods.

2.1.2.1 Limitations of 2D Approaches

The previous section, while certainly not exhaustive, demonstrates there have been many successful developments in the field of 2D face recognition. However, a lack of standardised measurements means that it is hard to compare the efficacy of one study to another. Differently sized datasets, different levels of difficulty of the datasets, different experimental methodologies (including statistical analysis) all make it extremely difficult to compare performance of different approaches accurately.

The US Government has been instrumental in attempting to set up standardised empirical testing environments for facial recognition applications and methodologies for assessing performance. Starting with the FERET program [148] which ran from 1993 to 1997 and provided a large set of face data (14,126 images from 1,199 subjects) and continuing to the FRGC and the Face Recognition Vendor's Test (FRVT) which last ran in 2006 [149]. Databases are publicly available from these programs, the use of which are encouraged in current research.

Regardless of the limitations contained in the research methodologies, the problems which still present problems to 2D face recognition systems are:

- Illumination
- Pose
- Expression
- Occlusion
- Ageing

In the next section, 3D recognition approaches are reviewed. 3D inherently allows pose correction and should be illumination invariant

2.1.3 3D Face Recognition

A key advantage of using 3D data is that proper morphology analysis is possible instead of what amounts to texture analysis from 2D data. Subsequently, a great deal of research into 3D face recognition uses the surface curvatures or contours to aid recognition. It is worth noting that although it is often stated that the use of 3D models for recognition are robust against illumination, this assumption is not entirely accurate. While the model itself will be illumination invariant, artefacts can be introduced at the acquisition stage.

One of the first papers on 3D face recognition was by Lee and Milios [113] in which convex features of the range imaged face e.g. cheeks, eyebrows, forehead, chin, are segmented and represented by an Extended Gaussian Image (EGI). The EGI represents surface normal information mapped onto a unit sphere. A weight is associated with each point representing the total area of surfaces with that vector. Recognition occurs by the use of a graph matching algorithm on correlations between EGIs. An interesting aspect of this paper is that they suggest that the convex regions of a face are less susceptible to changes caused by expressions. The results are promising, although a very small sample size is used. Using the EGI for face recognition was pioneering and was extended by Tanaka *et al.* [175].

Gordon investigated supplementing frontal views with profile views to extract pseudo 3D features [75]. Using a simple feature weighting and correlation system, it was found that adding profile information significantly increased recognition performance and, if manual intervention was permitted on marking the features, an accuracy of 98% was achieved. Gordon also investigated using depth and curvature information for face recognition [76] of range images stating that *'curvature descriptors: (1) have the potential for higher accuracy in describing surface-based events, (2) are better suited to describe properties of the face in areas such as the cheeks, forehead, and chin, and (3) are viewpoint invariant'*. Range images were

segmented by the sign of the Gaussian and the mean curvature to produce a set of features from which additional descriptors were drawn. The vector of these descriptors places a given face in the space of all possible faces – so a simple cluster analysis is used for recognition. Using the optimal feature set, which was determined experimentally, 100% recognition accuracy is reported for a relatively small 24 cases.

Extracting feature sets composed of the curvature of the face at different segments is a common method of reducing the dimensionality of the data; its discriminatory properties are still the focus of more recent papers e.g. [136], [116].

Colbry and Stockman introduce the 3DID system [37] which uses the Root Mean Squared (RMS) error returned from using Iterative Closest Point (ICP) matching to measure the difference between a probe head and each head in a gallery. They discovered that there is a threshold that below which the RMS error indicates that the faces belong to the same person, and above which they belong to different persons. It also handles pose variations of 30 $^\circ$ of yaw and 15 $^\circ$ of roll and pitch. Initial anchor point detection is performed using shape index on models captured by a laser scanner. Shape index defines the curvature of a surface in terms of an index ranging from -1 to 1 where -1 is a spherical cup and +1 is a spherical cap and 0 is a saddle point. The paper supplements their own data capture with data from the FRVT and compares performance of the proposed algorithms with PCA as a baseline using presenting results in the same way as the FRVT. However, they make the point that their approach only works well for neutral expressions with frontal poses. The repeated use of the computationally expensive ICP algorithm also limits the real world applicability of the system.

Bardsley *et al.* [12] overcome the computational inefficiencies of Colbry and Stockman's approach by registering the probe head against an average head (instead of against every head in the gallery) using ICP and take the average point-to-plane error as the recognition metric. They report an accuracy of 98.2% on 58 subjects with relatively unconstrained poses and expressions. They also

propose a novel correspondence measure when resolving stereo images – Gabor wavelets, which they show to be valid by comparison with two other techniques using the commercial 3DMD [3] capture system as ground truth. They suggest that future work should look into using 3D Gabor wavelets as a recognition metric.

One method of tackling pose differences between gallery and probe images in 2D recognition is to store images of the same head in a variety of poses ([22, 70]). An alternative is to use what is known as a 3D Morphable Model (3DMM) [24] to generate synthesised views of the face from different viewpoints. Using a generic 3D model of a head, it is possible to morph the model so that it produces a close approximation to that represented by a 2D image. This may then be rotated to generate novel images of that face; 2D images of which are used in recognition i.e. rotate the head to a frontal position before comparing with other faces in the database. This morphable model technique is used with ten face recognition algorithms in the FRVT and it was found that it significantly improved the performances of non-frontal views [152]. Additionally using a 3D model of a head offers robustness against illumination variations as any illumination can be simulated on the model.

The 3DMM can also be used for direct 3D comparison rather than 2D synthetic image generation and subsequent recognition [24]. In the same way as discussed in the previous paragraph, a generic 3D morphable head comprising of 100 3D scans is used to generate an accurate 3D model of a head from one or more 2D images by iteratively comparing an image generated by the model to the 2D image and gradually changing the model until it matches the image. After fitting the model, coefficients describing the way in which the shape and texture of the generic model was altered to match the probe face are stored. All gallery images are also analysed in the same way and the coefficients stored. The probe face model's coefficients are then compared to the coefficients of those models in the gallery – the nearest neighbour being the match. Very high recognition accuracy was achieved on the FERET database averaging 95.9% across all pose variations

up to an estimated 38.9° which demonstrates the robustness of this approach to pose variations.

Bronstein *et al.* [28] have developed an expression invariant representation of a 3D model called a *canonical representation* in which the 3D face captured via photometric stereo is treated as a deformable object in Riemannian geometry. The geodesic measurements are much less sensitive to expression than Euclidean measurements and facial expressions can be modelled as isometries (geodesic distances are preserved) of a neutral expression. An example of what this means can be seen in Figure 2.3. The technique has (at least) two interesting features: (1) no surface reconstruction of surface normals is required, the canonical form can be generated directly from the surface metric (or surface normals) and (2) they claim that the system can distinguish between twin brothers (the authors). However, the performance of this method has not been tested against any common dataset and the claim of being expression invariant is probably limited to cases where the mouth is not ajar and where the elastic properties of the skin are not shown.

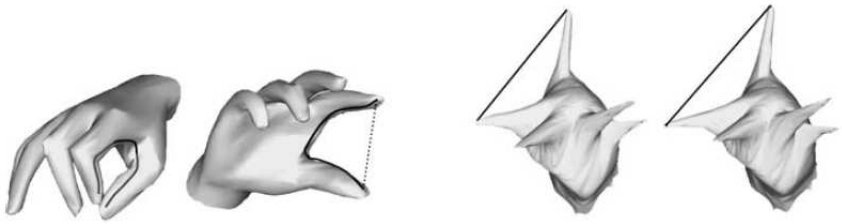


Figure 2.3: Isometric representations are useful as they preserve geodesic distances. (a) shows isometric transformations of a hand and their equivalents in Bronstein *et al.*'s canonical form (b). Note the Euclidean distance between thumb and forefinger is very different in (a) but identical in (b). *reproduced from [28]*

100% accuracy is reported for rank-1 matches of a probe face (with a variety of expressions) to a gallery of 65 templates of 30 subjects with neutral expressions, and the approach is shown to outperform classic surface matching (based on high order moments rather than ICP for simplicity) and classical 2D eigenface approaches [28].

Another approach is to use the 3D data to generate a 2D depth map and apply classical classification techniques. The motivation for this is that the depth information will be illumination invariant and pose correction can be performed before generation leading to more reliable data than the 2D texture information. In part of their paper, Chang *et al.* [33] compare intensity versus depth 2D images using a standard PCA algorithm and report better performance on the depth maps. Chen *et al.* used three different classifiers to assess the viability of wavelet transforms on depth maps and found a 5% performance increase over 2D images alone [34]. Pan *et al.* sought to limit the affect of pose variance by parameterizing the 3D face into an isomorphic 2D circle and then remapping the depth map accordingly, so that intrinsic geometric properties were kept [144]. Using the FRGC database, they report rank-1 accuracy at 95%, 4% higher than a baseline PCA comparison.

An extension of purely 3D recognition methods is to combine both morphological information from 3D with texture information from 2D. On the FRGC dataset, Mian *et al.* [131] reported the best results using such a fusion technique together with a hybrid (statistical and feature) classifier. Chang *et al.* [33] compare PCA recognition levels on 2D, 3D depth maps and combined 2D and 3D depth maps and report that 3D outperforms 2D data and again that combined data outperforms either 2D or 3D alone. Further support comes from Hüsken *et al.* [96] who also performed matching experiments on FRGC data. An interesting finding of this paper, that is in contradiction to Chang *et al.* , is that recognition on 2D data outperforms 3D data. Their explanation for this anomaly is that their Hierarchical Graph Matching based technique fully exploits the information in 2D, while other approaches do not.

Similar to the analysis of 2D algorithms carried out by Gross *et al.* [80], Gökberk *et al.* [73] implemented some popular shape-based representations of the publicly available 3DRMA database: ICP-based point cloud representation, surface normal-based representation, profile-based representation, and depth image. Surface normals are shown to provide the best recognition results both in terms of

direct surface normal comparison and via LDA, which they say *'is valuable since many of the previously proposed 3D face recognition systems make use of point cloud features'*.

Even though 3D face recognition has received far less attention than 2D face recognition, there are a large and ever increasing number of approaches. Typical 2D methods are often applied to the problem e.g. PCA, LDA, Gabor filters, and exclusively 3D approaches e.g. ICP, canonical forms, curvature, shape index and depth maps offer new ways of approaching the problem. 3D data certainly seems to offer the promised robustness to illumination, and certain methods provide additional robustness to expression and to pose. In almost all cases when performance on 3D data is compared to performance on 2D data, performance on 3D data is better. Ageing and occlusion are two areas which to date have received little or no attention in the literature due to difficulties in capturing the required data or defining useful occlusions.

2.1.3.1 Limitations of 3D Approaches

While similar problems concerning methodology highlighted with 2D recognition research exist for 3D work, again the research submitted which uses the FRGCv2.0 has done a great deal to address this. Additionally while 3D can certainly help solve problems with pose and illumination, it does little to solve the remaining issues of expression, occlusion or ageing.

Another limitation is that because most 3D experiments rely on data from a laser scanner, the individual is not in a natural environment. They will be in a laboratory setting, sitting very still and will probably have been told to pose in a certain way. Face recognition needs to be able to work in natural conditions as this is where the commercial interest lies. As explained in Section 2.6, this is one of the motivations of this thesis. Therefore, any method that increases the practical and ecological validity of the research is very important. Along with this and the cost of the scanners, what is needed is a fast, cheap and unobtrusive

capture device in a suitable environmental so that the advantages of 3D can be realised.

2.2 The Human Visual System and Face Recognition

James Reason famously described humans as *"furious pattern matchers"* [153] and our visual system has evolved to exploit these patterns in everything we see. A quick Google search reveals plenty of anecdotes in which people report seeing faces in objects where they exist only by chance (e.g. clouds, toast, fruit and vegetables). This is known as pareidolia, and has been exploited by artists such as Giuseppe Arcimboldo who constructed faces out of fruit and flower arrangements in the 16th century. Humans are also able to identify that a drawing represents a face given the most simple configuration of stimulus (e.g. an arrangement of three blocks into the positions of the eyes and mouth surrounded by an oval) and that this stimulus arrangement is preferential over an inverted version to newborn babies [77]. It is clear that humans are gifted at face processing and over recent decades evidence of brain regions that specifically process face type stimuli has been found. Are there certain properties which we can tease out of the evidence which can motivate and improve the performance of automated face recognition systems?

The HVS starts with light entering the eye and being detected by light sensitive cells (the rods and cones) in the retina. The output of these cells is collated by retinal ganglion cells, which play an important part in encoding spatial frequency of the stimulus. The signal is then propagated via the optic nerve to the striate cortex of the occipital lobe which is located at the posterior of the brain. From

here the signal moves through the extrastriate cortex and then is analysed by various regions of the brain. It is believed that the processing of the signal goes from simpler properties (e.g. colour, edges and orientation in the retina and striate cortex) through more complex properties (e.g. geometric shapes, motion and attentional selection in the extrastriate cortex) before higher level functions such as object recognition takes place in the temporal lobe. Broadly speaking there are two paths; the dorsal path is concerned with ”where/how” type processing involving spatial locations and motion, and the ventral path is concerned with ”what” type processing associated with object recognition.

Nobel prizewinning scientists Hubel and Wiesel [95] suggested that the striate cortex is organised in hypercolumns which fire preferentially when presented stimuli at particular orientations and are a topographic representation of the retina (i.e. areas which are close together in the retina are also close in the striate cortex). Combining the functionality of the retinal ganglion cells and the hypercolumn cells of the striate cortex leads to similar processing to a Fourier transform as the input signals get converted into spatial frequency and orientation.

There is still much uncertainty about the exact roles and processes involved in the HVS, even in the early stage processes up to and including the striate cortex. Arguably, these are still far better understood than the regions which follow in the extrastriate region. Nonetheless areas in the temporal lobe (more specifically Fusiform face area (FFA) in the fusiform gyrus) have been found to respond to faces preferentially over other classes of objects e.g. houses and hands using functional Magnetic Resonance Imaging [102]. Further evidence comes from neuropsychological studies in which patients show a separation in face and object processing due to some neurological impairment. Moscovitch *et al.* presented their findings on Patient CK whose object recognition was severely impaired, but his face processing was intact [138]. Interestingly, he failed to recognise the constituent vegetables in Arcimboldo’s paintings mentioned earlier but could see the faces!

Prosopagnosics demonstrate the opposite symptoms of Patient CK in that they are unable to recognise faces but perform at a normal level in object recognition for other classes. This is again more evidence of the special role that face processing has been granted in our visual systems. In contrast to the normal population, prosopagnosics do not exhibit the characteristic decrease in recognition performance when faces are inverted and the explanation offered for this is that they have come to rely on the non-face recognition system (i.e. to recognise other objects such as buildings) which has been shown not to be affected by these changes, but also does not process faces as well as the specifically tuned FFA [61].

While many earlier automated recognition systems relied on measuring the distance between fiducial features, evidence from human studies point to the fact that statistical techniques which deal with the face as a whole may be a more accurate model. Evidence that humans process faces holistically comes from the fact that when presented with face parts in their usual positions or in a scrambled order, the ability to identify a particular part is better when the faces are presented with features in their usual positions [176]. Furthermore, when two familiar faces are divided in half, and then combined so that the top and bottom halves are misaligned, the ability of a subject to recognise the two individual improves over when they are aligned [198]. This is interpreted as showing that when the composite halves are aligned, the face is convincing enough to interfere with the recognition process needed to identify the two individuals. Perhaps the most famous experiment demonstrating holistic face processing in humans (as well as the special processing that is invoked by an upright face) is the Thatcher illusion first published in a refreshingly short paper [177]. The eyes and mouth of a familiar face are inverted but when the face is presented inverted, it is not immediately obvious to the viewer that anything is amiss. However when the face is then presented the correct orientation, the effect is grotesque and the inverted features are obvious. The same features are present under both inverted and normal features, but

whether or not the inverted features are immediately apparent or not depends on the orientation of the face as a whole (Fig. 2.4).

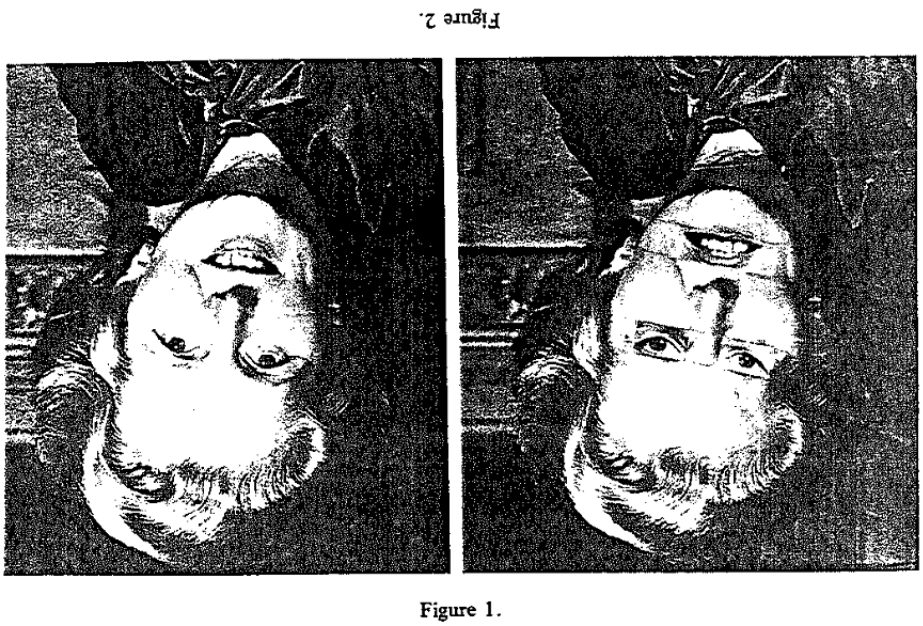


Figure 2.4: The original image demonstrating the Thatcher Illusion with interactive labelling included (reproduced from [177]). Both images look similar when presented upside down (Figure 1.), but it becomes grotesquely obvious that the eyes and mouth have been inverted when the face is upright (Figure 2.).

Caricaturing is another somewhat unexpected feature of the HVS. Traditionally, caricaturing is the process of exaggerating features that deviate from a norm for the purposes of cartooning an individual. Many examples can be found on a daily basis in any broadsheet paper where politicians are satirically lampooned. Often these types of politically motivated caricatures allude to non-physical characteristics of the individual as well as the purely physical. In terms of research into caricaturing, the purer physical renderings are used. The earliest research into the curious property of caricaturing on human face recognition was performed by Rhodes *et al.* [154] in which faces were both caricatured and anti-caricatured by using an automatic caricature generator. A line drawing of an individual is represented by 37 lines (made up of 169 points). The difference between these and a norm are then extended or reduced to render a caricature or anti-caricature. Rhodes *et al.* reported that caricatures were recognised faster than veridical line

drawings which in turn were recognised more quickly than anti-caricatures.

This finding has been repeated many times since ([128, 114]) in photographs [18] and video [66]. Caricaturing has also been extended into 3D [52, 143]. While caricaturing improves face recognition in humans there has been little reported investigation into its effects in automated systems. Zou *et al.* [212] suggest using 3D caricaturing on a morphable model to improve recognition on 2D photographs although no empirical results are reported. Wang *et al.* [188] develop a system for matching artistic sketches to photographs by converting photographs to sketches computationally. Nejati & Sim [139] show that caricaturing a photograph of a face leads to best recognition rate for a given probe sketch (although the recognition rate is only 28% at best).

An inherent difficulty with applying caricaturing in the traditional sense is that a good caricature is created by a talented artist. Different artists will create caricatures which vary wildly from one another depending on the artists own style. Attempts at automating this have met with varying success from Brennan's pioneering 2D line caricature system which generated reasonably human-like cartoons to Deffenbacher *et al.* 's grotesque appearing 3D models (this description is not meant to belittle their work – the generated 3D models were very impressive but the caricature effect tended to age and distort the individual in unconvincing ways). Another approach relies on image statistics rather than the features themselves. Sirovich and Kirby [168] refer to caricatures in their seminal paper which directly inspired the more familiar eigenfaces work by Turk and Pentland [181]. They define one of the initial steps of PCA, the subtraction of the population mean from the individual images to leave the residual differences as a caricature, and state that:

> It seems reasonable to assume that an efficient procedure for recognizing and storing pictures concentrates on departures from the mean.

Unnikrishnan [182] defines a different statistical method of caricaturing for the purposes of recognition and introduces the idea of using just the most outlying data. He presents the problem in terms of only using features which differentiate an individual from the norm. More specifically using only the 10% of data which deviates from the norm specified by the 5^{th} and 95^{th} percentile values, but offers no empirical evidence for his hypothesis. He clearly illustrates his ideas by example and then applies the interpretation to existing research. Chapter 6 adapts his ideas to the use on surface normals with encouraging results.

Sinha *et al.* [167] present a compelling overview of face recognition in humans, summarising the idiosyncrasies (such as caricaturing effects, inverted faces, contrast polarity inversion) in a way which is hoped is of use to computer vision researchers. Amongst the 19 findings reported, they highlight one which is intuitively at odds with automated face recognition. It has been reported that humans are good at recognising faces at low resolutions, even to as low as 7×10 pixels. If this finding can be generalised to automated face recognition, then it represents a direct way to reduce the dimensionality of the face so potentially increasing computational efficiency and reducing hardware costs. Most work into resolution effects in face recognition concern themselves with increasing the resolution of low quality images e.g. from CCTV video capture (either by hallucinating additional pixels from individual images or by extracting data from multiple frames). It should be stressed that low quality does not exclusively mean low resolution, compression artefacts are often present as well. In terms of the affects of reducing resolution on automated systems there has been little comparative work in 2D (isolated results can be found in [31]) and only one study in 3D [33]. Intuitively in computational pattern recognition, it seems odd to bluntly remove data that could aid in discrimination, but human studies reveal that low frequency data has a different role in face recognition to high frequency data [159, 43], and that high frequency images (drawings) are insufficient on their own for accurate recognition

[51]. It has been suggested that this represents two modes of sequential face processing, a more general face localization and early recognition process using the low frequencies followed by use of the higher frequencies for more detailed analysis. Recently, fMRI data has confirmed that the frequencies are processed temporally from coarse-to-fine frequencies [72]. Parker and Costen [145] provide an interesting critique of findings before running their own experiments in which they conclude that middle frequencies provide the best information to humans with performance declining as the frequency bands are moved lower or higher.

There is some conflict across the experimental findings but this is likely due to the different stimuli used and what is recorded rather than a real effect. An interesting study comparing the different processing associated with different frequencies shows that high and low frequencies may play different roles in face processing. Schyns and Oliva [158] showed that different frequencies code different properties of a face by superimposing a high frequency angry male face over a low frequency female face. There was no frequency bias for gender selection, but asked to judge the expressiveness of the face, the high frequency face dominated, but when asked what the particular expression was, the low frequency face dominated. There is sufficient evidence to suggest that lower frequencies play an important role in human face recognition and that this could provide an effective means of data reduction as reducing resolution effectively smooths the image, removing higher frequencies. This is examined in detail in Chapter 5.

Perhaps the best known model in the psychology literature of human face processing is by Bruce and Young [30]. Although the model has been adapted since it first appeared it defines two stages: the first is recognition which is a sequential process involving perceptual (visual) input which is compared with stored faces via face recognition units which in turn trigger semantic processing (e.g. what sort of job they do) and person identity nodes to recall information about the person and then finally (if sufficient activation has been achieved) the name retrieval stage. The second stage which proceeds independently of recognition,

is the extraction of other information from the face e.g. expression. While this model and its revisions have firm groundings on experimental evidence and provide a testable framework for cognitive psychologists, they provide little concrete specification for a computational implementation. For example, how faces are compared in the first stage (using Face Recognition Units) is not specified. It is therefore very difficult, if not impossible, to incorporate what this model tells us about human vision into an automated system.

Another aspect which is seemingly impossible to incorporate into an automated system involves the distinction between familiar and unfamiliar face processing. We tend to think of ourselves as being very capable of recognising faces, even if friends have aged, grown beards, are wearing glasses or are disguised in fancy dress, we can still usually recognise their face easily. Although our ability to process unknown faces in terms of gender, expression, age etc. at a glance is remarkable, our ability to recognise unfamiliar faces is surprisingly poor (at least under experimental conditions). One such situation where a person must rely on their memory of an unfamiliar face is in an identity parade. Bruce *et al.* [29] presented subjects with a probe face image together with a gallery of ten possible matches. The subject had to select the match for the probe (or say that it was not present). Error rates were around 30% when the probe was absent from the gallery in half the trials and 20% when the probe was always present. Given that the probe and gallery images were presented together and were taken in similar lighting and poses, this represents a large weakness in the human face recognition process for unfamiliar faces. While the distinction between familiar and unfamiliar faces is irrelevant for a computer (unless we can actually model what it means to be a familiar face), caution should be taken when considering the HVS to be the perfect model to base an automatic face recognition system on, although incorporating certain aspects can be beneficial as we will see in the next section.

2.3 Bio-inspired Face Recognition/Vision Systems

The HVS is an extremely complex system but research highlighted above has allowed researchers to model the various aspects and apply them successfully to a range of computer vision tasks. This section describes some of the most relevant research in this area.

Perhaps the most influential finding comes from modelling the early stages of the visual process. Combining the functionality of the striate cortex hypercolumns with the retinal ganglion cells which encode spatial frequency, Daugman [50] realised that their responses could accurately be mimicked by a bank of Gabor filters. The power of using such features for face recognition has already been discussed in Section 2.1.2 and they are still an active area of research with recent extensions into 3D [194].

As stated previously, Gabor filters represent a sinusoid windowed by a Gaussian function and performing a convolution over an image with banks of filters can be computationally intensive. A far simpler alternative exists – a bank of Haar-like features. The Haar wavelet is a square wave with no Gaussian function. By using a set of four such Haar-like features, Viola and Jones [187] were able to develop a system to rapidly detect faces in an image by using the *integral image* representation and by boosting the simple features in order to form a cascade of feature detectors. The learned cascade of Haar-like features are tuned to respond to regions of the image which have combinations of frequencies and orientations to provide a similar outcome to a Gabor filter bank. This is achieved by moving the features across the integral image at a range of resolutions (thus mimicking the spatial resolution) and due to the horizontal, vertical and diagonal patterns of the features, a range of orientations, albeit limited, are represented.

Another paper which models the early stages of the visual system is [53] which used facial texture and colour codes modelled on retinal, lateral geniculate nu-

cleus and striate cortex and reported state-of-the-art improvements.

Convolutional neural networks have been used for object recognition [160] and very successfully in digit recognition [111] and also for face verification [35]. Loosely interpreting the simple and complex cells in the striate cortex [94], the convolutional network consists of alternating sub-sampling and convolutional layers which gives the system some robustness to scale, orientation and noise. Using back-propagation the most useful features are extracted from the input image before being fed into a more conventional feed-forward classifier network.

Inevitably, perhaps the best known approach to face recognition, eigenfaces, has also been compared with the mechanisms behind human face recognition. However, although correlations were found to exist between human distinctiveness ratings and PCA eigenvalues, it seems an unlikely model [84, 83] and Dailey *et al.* suggest that a Multi-Dimensional Scaling (MDS) model is more accurate [47].

Burton *et al.* [32] use PCA as a pre-process to an implementation of a face recognition model evolved from Bruce and Young's work [30]. While the implementation is unrealistic in certain respects (e.g. in terms of how semantic information about the faces is set up) the results are promising, and it represents an interesting attempt at making cognitive models into concrete implementations.

Mimicking the HVS is an active topic of research. A recent paper [44] modelled the striate cortex using two previously developed models to classify the Labelled Faces in the Wild database [93] and report state-of-the-art recognition performance.

Improving our knowledge of understanding how the human brain processes 3D shape information is still the focus of cutting edge research as demonstrated by the recent grant awarded to Yale University [189] with an aim of providing models for computer vision and graphics as well as strategies for rehabilitation of patients with visual deficits. The research will cover processing of the striate, extrastriate and temporal regions of the brain – areas that we have seen to be

important to face processing.

This highlights an important reason for research in human/biologically motivated approaches. Apart from the hope of improving existing algorithms a computer vision scientist may also be able to feedback knowledge of the results of implementations to the psychological and biological scientists.

To date there has been little work in applying what is understood about HVS to 3D models, and to this author's knowledge, none on surface normal representations.

This is by no means an exhaustive treatment of face recognition research but demonstrates the wealth of approaches and highlights the areas which are of importance for this thesis. 3D data is preferable to 2D, surface normals are a good representation of such data for face recognition and the HVS appears to offer some useful features and processes that might be beneficial to an automated system.

There now follows a review on publicly available databases which are used for testing algorithms and providing a means for comparison.

2.4 Databases

The purpose of a database of faces is to assess how accurate a recognition algorithm is and therefore it logically makes sense to make it as realistic as possible in terms of the eventual deployment environment of a system.

Face recognition researchers have been collecting databases of face images for several decades now [119, Chapter 13]. While some databases can be regarded as superior to others, each of them are designed to test different aspects of recognition and have their own strengths and weaknesses. One of the largest databases available is the FERET database [150]. This has a total of 1199 subjects with up to 20 poses, two expressions and two light source directions. The FERET database was originally acquired using a 35mm camera. Others concen-

trate more on varying the capture conditions such as pose and illumination contrast e.g. the widely used CMU PIE database [165] or the Harvard RL database [82]. Another popular database collected for the purpose of face verification under well-controlled conditions is the XM2VTS database [130].

The PIE database is one of the most extensively researched. This is due to the fact that the faces are captured under highly controlled conditions involving 13 cameras and 21 light sources. The Yale B database [70] offers similar advantages to the PIE databases except with an even larger number of lighting conditions (64) using nine poses. However, the Yale B database includes just ten subjects. The original Yale database [17] was designed to consider facial expressions, with six types being imaged for 15 subjects. Finally, the extended Yale B database was published which contains 28 subjects with 9 different poses and 64 illumination conditions [115].

Even though the PIE [165], Yale [70] and extended Yale [115] databases provide facial samples taken under different illumination directions they contain very few subjects. More recently, the CMU Multi-PIE database [79] has been constructed with the aim of extending the image sets to include a larger number of subjects (337) and to capture faces taken in four different recording sessions. This database was recorded under controlled laboratory conditions, as with the others mentioned above. In contrast another trend in face recognition is totally unconstrained face matching and a database for this task called Labelled Faces in the Wild has been recently collected [93].

The recent trend in face recognition research has been to incorporate three-dimensional information into the recognition process has naturally led to the collection of databases with 3D facial samples. This was the motivation for the FRGC2.0 database [147], which consists of a multi-partition 2D and 3D database including a validation set of 4007 scans of 466 subjects. A Minolta Vivid 900/910 series laser range finder was used for data capture.

Up until this project, no large scale PS database was publicly available. Al-

though the Multi-PIE allows 3D reconstruction via PS, the images are captured under very controlled conditions. The Photoface database (generated for this thesis) is captured in a realistic workplace with no supervision and also provides detailed metadata for each captured session. More information on this database is provided in Section 4.1 In order to collect the images for a database, a suitable capture device must be used. The next section provides details on face 2D and 3D capture technologies used in face recognition.

2.5 2D and 3D Capture Techniques

This section outlines the most common methods for capturing 2D and 3D faces. Different approaches have associated advantages and disadvantages which are discussed below.

2.5.1 Light Sources

2.5.1.1 Visible Light

The most common way of capturing an image of a face is via visible light which lies in the 0.4-0.7 micron wavelength region of the electromagnetic spectrum. Although the cost of cameras sensitive to this bandwidth is low (and falling) there are numerous issues with using visible light; the main one that is repeatedly raised being illumination variations. Use of visible light is also susceptible to disguise e.g. make-up or prosthetics (which will have been made to be as inconspicuous as possible under normal lighting).

Visible light is most commonly used in 2D recognition experiments i.e. photographs but is also used for PS, stereo and structured light reconstructions of 3D faces.

2.5.1.2 Thermal and Near Infrared (NIR)

Infrared light has a wavelength of between 700nm and 1mm, and is commonly divided up into bands. Thermal infrared corresponds to the mid-wave infrared MWIR and long-wave infrared LWIR bands (approximately 3-5 microns and 8-14 microns respectively). The human body and face emits radiation in both of these bands, although LWIR is more commonly used for face recognition as a great deal of MWIR is absorbed by the atmosphere. Thermal radiation has the potential benefit for use in face recognition in that it is entirely illumination independent i.e. the body emits the radiation rather than reflects it. This means that it can be used just as successfully in brightly lit and dark environments. It can also detect the presence of prosthetics e.g. a false nose, as the thermal signature for the face will be altered significantly. Amongst its drawbacks however are that the sensors are extremely expensive, glasses are opaque to thermal IR and variations in ambient or body temperature significantly alter the thermal signature [107] i.e. from exercise or fever.

The near range of infrared light is often termed the reflected IR band as it contains no information about that thermal properties of materials and comprises the NIR and short-wave infrared (SWIR) wavelengths (approx 0.7-1 micron and 1-2.4 microns respectively). Most face recognition experiments use the NIR band as its proximity to the visual spectrum means that it is likely to behave in a similar manner on skin, and that silicon sensors that are sensitive to NIR are widely available and reasonably priced. It also has the benefit of being more robust to ambient illumination variations than visible light and the NIR lighting source is largely covert to a human observer.

As mentioned in Section 2.1.2, NIR has useful properties for face detection (the *bright eye* effect [137]) and skin reflectance to NIR rapidly diminishes at 1.4 microns [56]. NIR based face recognition has been shown to be a valid approach by means of three fully automated systems ([211, 204, 120]). It does however

suffer from the same pose limitations as to visible light when used for 2D recognition, but its use in 3D recognition has not been researched. Chapter 3 looks at this further by using NIR light sources for PS.

2.5.2 Existing 3D Capture Techniques

2.5.2.1 3D Laser Scanners

The most common method of 3D acquisition are laser scanners. While these systems are commonly considerably more expensive than the other systems mentioned in this section, this is the capture format used in the FRGCv2.0 database. Using a Minolta Vivid 900/910 Series sensor [1], one 3D image consisting of a range and a texture image is captured for each session. The acquisition itself is via laser stripe and triangulation. Three pieces of information allow the corners of the triangle to be fully determined, and the depth calculated: the distance between the laser projector and sensor, the angle at which the laser is shone and the position of the laser in the camera's field of view. The benefit of laser scans is that they offer extremely high accuracy; the disadvantage of using them is that range data is typically captured in about 2.5 seconds [26] and the texture image is captured after the range data. This means that discrepancies due to movement are common. There can also be considerable artefacts present (e.g. caused by specularities) in the reconstructions which present as spikes and holes on the surface, so before any recognition can be attempted, these must be removed via some form of preprocessing.

2.5.2.2 Stereo

Perhaps the most well known method of 3D capture is to use two images of an object or scene taken from slightly different viewpoints. By analysing the disparity between the objects in the two pictures, the relative depth can be calculated. Stereograms, which enjoyed immense popularity in Victorian times, presented

images taken in this way to each eye – the brain then performed the necessary computations to allow the scene to be seen in 3D. Calculating the disparity is not so straightforward for a computer. The correspondence problem is well known in the machine vision literature and refers to the difficulty in locating matching points in the two images. To simplify this, commercial systems commonly project a known pattern onto the objects during capture to allow easier point matching.

Indeed, the second most commonly used method for 3D face capture is via such a method, known as a projected pattern stereo device. An example of a system that employs this method is the 3dMD device [3]. The advantages of such systems are high accuracy (reported as $<$0.2mm) and fast acquisition times (1.5ms) which freeze motion. However, the processing time is approximately 90s to reconstruct a face on a modern desktop computer. This type of system is also expensive, requires a time consuming calibration procedure, and omits fine details such as wrinkles and pores as well as struggling with hair.

2.5.2.3 Structured Light

A third way to capture 3D information uses the deformations in a projected pattern to estimate the surface shape. This is known as structured light (or light coding) capture, and typically horizontal lines are projected onto an object. It has famously been used recently in Microsoft's Kinect accessory which uses a projection of NIR dots instead of stripes to allow real time 3D capture, albeit at a relatively low resolution of 640$\times$480 pixels with a depth resolution of only *"a few centimetres"* [164]. A higher resolution device using projected stripes was recently developed by a European Project (3D Face [2]) for use in automated border control. It captures a 3D model and a 2D texture image in 0.2s at an impressive depth resolution of 0.1mm and point spacing of 0.5mm and can also be used for video capture at 20 frames per second [190]. While the Kinect is low-cost (around GBP100), the captured depth map is too noisy for face recognition. Unless someone is moving very fast, the device used by the 3D Face project, provides very

good accuracy but although not stated in the project details, is likely to be very expensive.

2.5.2.4 Time of Flight

Time of flight systems operate by measuring or inferring the distance that a beam of reflected light has travelled. Two approaches can be adopted: intensity modulation or optical shutter, with the former being the most common. The interested reader can find more information on the approaches in [106]. They have the benefit of being fast with little processing power required so are well suited to real time applications. They are also generally robust to environmental illumination changes as they project their own light source (often in NIR). However they are costly, and to date have a low spatial resolution (e.g. current time-of-flight cameras have a maximum of 64 $\times$ 48 to 204 px [106]) and depth resolution is only sub-centimetre at optimal ranges. Additionally the data that is captured tends to be noisy.

2.5.2.5 Shape from Shading (Photoclinometry)

Shape-from-Shading (SFS) uses shading from an individual image in order to estimate the surface orientation and was first researched by Horn for his Ph.D thesis [90]. It is a popular area of research due to its obvious applications and capture simplicity – the goal is to recreate an accurate 3D model from a 2D photograph and so removes the need for expensive and/or complex capture devices. An inherent difficulty in using a single image is that it is extremely difficult to separate the gradient from colour or texture information, therefore there will always be an ambiguity present as to whether an intensity gradient is due to a slope or some colour, pattern change or shadowing.

The Bas-Relief ambiguity as described by Belhumeur [16] also poses a problem. Examples of bas-relief can be found in many stone wall carvings and the

technique is used for representing heads on coins. It is a means of tricking the eye into seeing a 3D representation from a far flatter relief by representing the lower levels of relief as being more extreme than they are to heighten the shading. They produce the appearance of an accurate 3D model when viewed from the correct angle, but the true (flatter) representation becomes clear as the viewing angle moves. Presented with a bas-relief viewed from the correct angle, a SFS algorithm will likely construct a 3D model similar to that which a human perceives rather than the veridical, flattened relief.

Another motivation for believing that SFS can provide accurate 3D models from photographs is that the HVS must incorporate some sort of mechanism for interpreting 2D shading in terms of 3D shape. This is not to say that humans employ similar algorithm to computation SFS e.g. they neither require the assumption that the object has Lambertian reflectance nor need to know the direction of a light source [132], but it is evidence that useful 3D models can be created from 2D images. The likelihood that humans employ some sort of SFS type processing is supported by the unresolvable ambiguity of a concave surface lit from above or a convex surface lit from below – either way, we still perceive a tangible 3D object from shading. The fact that make-up is used to trick human perception of faces is additional evidence of SFS's importance.

With particular reference to face processing, shading information is also used as shown by the almost complete disruption to recognition when contrast polarity is reversed i.e. recognising a face in a negative is extremely hard [68, 98]. The reason for this is not entirely clear unless the shading information is viewed as being extremely important. All of the fiducial features are in their proper places, the ratios between them maintained, the spatial frequencies are preserved but the shading information has been completely reversed. We are still able to recognise the object as a face correctly, but the ability to identify the face is decreased. This does not occur with other types of objects [173] lending support to the theory that humans perform SFS-like processing when recognising faces.

SFS could be said to be the holy grail of 3D capture – the ability to reconstruct accurate and high resolution 3D models from affordable and off-the-shelf cameras. However, due to the problems with ambiguities, its accuracy and reliability still have some way to go. The use of a single image means the problem is under-constrained without simplifying assumptions i.e. there is not enough information in a single image.

2.5.2.6 Photometric Stereo (PS)

PS is an enhanced SFS method which aims to resolve the ambiguities associated with the traditional SFS approach which uses one image to estimate 3D shape by separating the 3D morphology from the 2D texture. It constructs a 3D form from three or more images of the same object each lit from a different and known direction and estimating surface normals at each pixel coordinate [192]. A thorough mathematical description of PS is given in Section 3.1 but a schematic can be seen in Fig. 2.5.

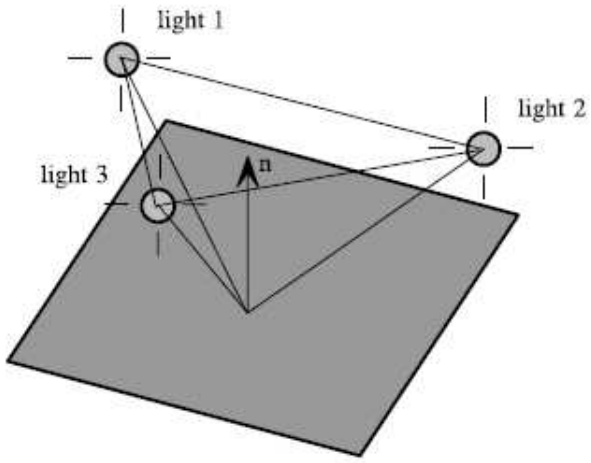


Figure 2.5: An example three-light PS system (from [169]).

Integration by a method such as Frankot and Chellappa method [64] across the resultant surface normals can then be used to reconstruct the surface.

Georghiades extended PS beyond Lambertian surfaces to incorporate the Torrance and Sparrow model of reflectance [179] and created very accurate re-

constructions [69]. However, a large number of images were required for the reconstruction which significantly increases the surface computation and image acquisition times. Sun *et al.* [174] use five lights to handle shadows and specularities on non-Lambertian surfaces and show that a minimum of six lights are required in order fully realise any convex surface using photometric stereo. Using 20 images of a face, Ghosh *et al.* [71] build up a very detailed model of the skin's reflectance taking into account specular reflection and single, shallow and deep scattering. However, the images are captured over "a few seconds" which makes this approach unsuitable for practical applications. Also, their method would add a large amount of complexity for relatively little gain as skin follows Lambert's Law reasonably well, as the results of this chapter demonstrate.

Of the vast amount of research into automatic face recognition during the last two decades [205], relatively little work has involved PS. Kee *et al.* [104] investigate the use of 3-source PS under dark room conditions. They were able to determine the optimal light source arrangement and demonstrate a working recognition system. Zhou, Chellappa and Jacobs apply rank, integrability and symmetry constraints to adapt PS to face-specific applications [208]. Zhou *et al.* extended a PS approach to unknown light sources [207]. Georghiades, Belhumeur and Kriegman show how reconstructions from PS can be used to form a generative model to synthesise images under novel pose and illumination [70].

Comparing point clouds, shape index, depth maps, profiles and surface normals in terms of face recognition performance, Gökberk *et al.* [73] concluded that surface normals provide the best features for face recognition. It is surprising therefore, that so few applications to date utilise PS, which inherently generates surface normals. The reason for this is likely to be that the availability and affordability of cameras with high enough frame rates and sensitivity for PS have only reached the market in recent years. Such cameras are necessary in commercial and industrial applications to effectively freeze the motion of the person if they are moving by capturing several images in a short burst.

The novel PS databases produced as part of this Ph.D. are therefore particularly useful to the research community as they are the first large scale collections of faces captured using PS. Uniquely, they facilitate face recognition research in a number of different modalities: 3D faces, surface normal representations, 2D faces using albedo images and 2D faces from different illuminations (four illuminations are used for each session to estimate the surface normals). The reasons for no substantial PS face database existing prior to this Ph.D. are likely due a lack of awareness of PS relative to other more commonly used 3D capture techniques combined with the necessary hardware only becoming affordable in recent years.

PS offers similar advantages to the standard SFS approach of high resolution and potentially fast capture using an off the shelf camera and overcomes the ambiguities associated with determining shading from texture and bas-reliefs. It is affordable and the fact that surface normals are an inherent product lends itself well for use in face recognition as some authors have found them to be the best data representation [73] in comparison with other commonly used formats. In the next chapter the accuracy of the captured 3D face models by PS is determined empirically in order to assess their suitability for practical face recognition.

2.5.3 Comparison of Capture Techniques

Table 2.1 shows a qualitative comparison of the different capture techniques. It should be stressed that the ratings are those of the author and are not the result of empirical work.

2.6 Thesis Context

This concludes the literature review which has provided an overview of important research in the face recognition literature for 2D and 3D data, some of the better understood aspects of the HVS which may be incorporated into an automated

Technique \ Category	Cost	Computation	Accuracy	Resolution	Calibration	Practicality
Laser Scanner	Average	Good	Average	Average	Good	Poor
Stereo	Good	Good	Average	Variable	Good	Average
Projected Pattern	Poor	Poor	Good	Good	Poor	Average
Structured Light	Good	Good	Poor	Poor	Good	Good
Time of flight	Poor	Good	Poor	Poor	Good	Average
Shape-from-Shading	Good	Average	Poor	Good	Good	Good
Photometric Stereo	Good	Good	Average	Good	Good	Average

Table 2.1: An qualitative comparison of different 3D capture techniques. Practicality refers to the ease by which the methods can be commercially deployed. The ratings for the Structured Light row are for an MS Kinect as it is commercially available, and are almost the opposite of those for the 3D Face project structured light device discussed in Section 2.5.2.3.

system and technologies used in the capture of 2D and 3D faces. As a result of this literature review certain areas suggest promising avenues of research which will form the basis of the contributions of this thesis. These are summarised below given in the context of the literature which motivates them.

- Section 2.1.3 states that Gökberk *et al.* [73] have found that surface normals provide the best representation for face recognition. Surface normals are an inherent product of photometric stereo. Therefore there is a need to test that the Photoface device is accurate enough and that surface normals captured in this way are suitable for face recognition (Gökberk *et al.* generated theirs by differentiating a surface). These are two of the aims of Chapter 3.

- Section 2.5.1.2 discussed some of the uses for NIR which has also been used in a variety of other biometric applications (iris and vein recognition), but to date it has not been used for photometric stereo face capture. Chapter 3 also explores this and suggests that the improved Lambertian reflectance of skin under NIR is a result of increased sub-surface scattering.

- Section 2.2 covered research showing that humans are excellent at recognising familiar faces and can do so from an early age. Humans have specific brain regions which are specialised for processing faces. Therefore the question arises whether incorporating aspects of the HVS into automatic face recognition can yield improvements. Two areas which would appear to lend themselves well to being incorporated are low resolution/low spatial frequency capture and caricaturing which may allow us to capture only the most important information of a face. Both provide a direct means of dimensionality reduction which means reduced computation and storage needs, and thus reduced cost. Unnikrishnan [182] has suggested a method of caricaturing which may be adapted to use on surface normal data. Chapters 5 and 6 explore these aspects and show that the face can be resized to as little as 10×10 px without seeing an decrease in recognition performance, and

that outlying pixels selected according to Unnikrishnan's hypothesis contain disproportionately high levels of discriminatory information.

- As stated in the limitations of 3D recognition (Section 2.1.3.1), along with occlusion and ageing, expression remains one of the problems affecting 3D recognition. The problem of expression is investigated through the use of novel photometric stereo database (3DE-VISIR) and Chapter 7 shows Happy expressions can be distinguished from other types accurately and goes on to use the variance and resolution findings of previous chapters to show that the most expression variant pixels can be removed to improve the recognition robustness.

- There currently exists no commercially available PS automated face recognition system in the world. The above findings represent a significant contribution to a fully developed system but Chapter 8 identifies that automatic alignment and thresholding for validation are two areas required in a complete system. One of the most successful methods of face detection is the rapid cascade of boosted features developed by Viola and Jones [187] as discussed in Section 2.3. The features for this are 2D, so in Chapter 8 a proposed extension of these Haar-like features is developed specifically for surface normal feature detection to allow automatic alignment. Additionally a threshold is determined using the Photoface database, allowing recognition performance of 98.65% accuracy at a False Acceptance Rate of 0.01. In this way the two identified deficiencies of the system are addressed.

Chapter 3

The Accuracy and Suitability of Photometric Stereo for Face Recognition

As stated in the literature review of the previous chapter, surface normals represent a good (if not the best according to [73]) representation for face recognition. This chapter introduces the Photoface device which was developed prior to the commencement of this Ph.D. and was funded by an EPSRC grant (EP/E028659/1) in collaboration with Imperial College London, the Home Office Scientific Development Branch and General Dynamics Ltd.. The project delivered Photoface, a four-source photometric stereo capture system, specifically designed for face capture. This Ph.D. uses data captured by the device and the purpose of this chapter is to assess its capture accuracy as well as to see if NIR light sources offer any significant benefit to the original rig. For clarity, it is should be stated that it is only the hardware that is not a part of this Ph.D.. Much of this chapter has been published in the Computer Vision and Image Understanding journal [85] which also contains additional research into optimising which light sources should be used to mitigate shadowing.

This chapter makes significant contributions to 3D face capture and process-

ing by presenting a novel PS hardware device and a detailed set of experiments to assess the accuracy and practicality of the device. Additionally it is shown for the first time that faces can be accurately reconstructed using NIR light. This offers several benefits to existing methods including exploiting skin phenomenology, creating a more covert capture system and making the system less intrusive. Extensive experimental results of these proposed advances are presented, including an analysis of skin reflectance qualities under NIR and visible light in terms of the Lambertian assumption.

In summary, the contributions of this chapter are threefold:

1. The introduction to and explanation of 3D data capture hardware suitable for practical face recognition environments.

2. Detailed experiments to test the accuracy of the device on a variety of faces under visible and NIR light sources in terms of ground truth reconstructions and the Lambertian assumption.

3. Detailed experiments to assess the validity of the Lambertian assumption and a test to determine any possible improvements that may be possible using the Oren-Nayar reflectance model [142].

A thorough description of PS is given next followed by the method section which details the operation of the Photoface device. The results section follows which provides examples and accuracy measurements of the basic reconstructions followed by a more in-depth analysis of the reflectance properties of skin under visible and NIR light. The chapter then concludes that the device produces data that are suitable for face recognition and that NIR used as a light source is more accurate than visible light.

3.1 Photometric Stereo

A key assumption of PS is that the surface reflectance obeys Lambert's Law: that light is reflected by a surface equally in every direction. While skin is not perfectly Lambertian, it is a close enough approximation as will be shown in due course. We can state that:

$$E = \rho L \cos\theta \tag{3.1}$$

where E is the emittance (reflected power per unit area) from the surface, ρ the albedo (ratio of reflected to incident irradiance at normal incidence), L is the irradiance (incident power per unit area) and θ is the angle between the light source vector and the surface normal. For a surface $z = f(x, y)$, the surface normal, $\mathbf{n}$, can be written:

$$\mathbf{n} = \left[\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}, -1\right]^T = [p, q, -1]^T \tag{3.2}$$

If the light source vector is denoted $\mathbf{n_s} = [p_s, q_s, -1]^T$, then we can write

$$\mathbf{n}.\mathbf{n_s} = pp_s + qq_s + 1 = \sqrt{p^2 + q^2 + 1}\sqrt{p_s^2 + q_s^2 + 1}\, cos\theta \tag{3.3}$$

Substituting Eqn. 3.1 and Eqn. 3.3 gives:

$$E = \rho L \frac{pp_s + qq_s + 1}{\sqrt{p^2 + q^2 + 1}\sqrt{p_s^2 + q_s^2 + 1}} \tag{3.4}$$

Which is often represented in Computer Vision as:

$$I = \rho \frac{pp_s + qq_s + 1}{\sqrt{p^2 + q^2 + 1}\sqrt{p_s^2 + q_s^2 + 1}^2} \tag{3.5}$$

where I is the measured pixel intensity. Here, a linear camera response is assumed and the incident light irradiance and camera response constant have been 'absorbed' into the albedo. For an 8-bit image, this means that both ρ and I fall

into the interval [0,255]. Typically, it is assumed that the light source vector $\mathbf{n_s}$ is known, meaning that from a single pixel measurement, I, we have one equation with three unknowns (i.e. p, q and ρ).

Usually, PS represents Lambert's Law as a vector equation so that:

$$I = \rho \mathbf{s}.\mathbf{N} = \rho \begin{bmatrix} s_x \\ s_y \\ s_z \end{bmatrix}^T \begin{bmatrix} N_x \\ N_y \\ N_z \end{bmatrix} \tag{3.6}$$

Where $\mathbf{N}$ is the unit surface normal (as opposed to $\mathbf{n}$ which is the surface derivatives representation). A three source PS can then be written in a matrix equation:

$$\begin{bmatrix} s_x^1 & s_y^1 & s_z^1 \\ s_x^2 & s_y^2 & s_z^2 \\ s_x^3 & s_y^3 & s_z^3 \end{bmatrix}^{-1} \begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} = \rho \begin{bmatrix} N_x \\ N_y \\ N_z \end{bmatrix} = \begin{bmatrix} m_x \\ m_y \\ m_z \end{bmatrix} \tag{3.7}$$

Performing substitutions in the above equations for the three unknowns in Eqn. 3.5 results in the following solutions:

$$p = -\frac{m_x}{m_z}, \quad q = -\frac{m_y}{m_z}, \quad \rho = \sqrt{m_x^2 + m_y^2 + m_z^2} \tag{3.8}$$

The intensity values and light source positions are known, and from these the albedo and surface normal components can be calculated by solving Eqn. (3.8). The resultant dense field of surface normals can then integrated to form height maps using the well-known Frankot and Chellappa method [64].

Figure 3.1 shows four raw images of an individual captured by the Photoface device (described in Section 3.2.1) operating with the visible light sources. The person was slowly ($\approx$1m/s) but casually walking through the device. Each image has pixel dimensions of 500$\times$400 and there are typically just a few pixel lengths misalignment between the first and last images. The face detection method of Lienhart and Maydt [121] is used to extract the face from the background of the image.

a b

Figure 3.1: Examples of photometric stereo inputs and output. (a) Four raw differently illuminated images. (b) Reconstructions using standard PS.

The four intensity images are processed using a MATLAB® implementation of a standard PS method [63, §5.4]. In our case of using four light sources instead of three, the system is overdetermined (i.e. there are more equations than unknowns). Our implementation uses the method of least squares to provide an approximate solution. For the system $Ax = b$, the least squares approximation is given by $x = (A^T A)^{-1} A^T b$.

For accurate reconstructions from PS, the following are assumed:

1. The object's surface reflection is Lambertian. This means that light hitting a point on the surface will be scattered equally in every direction. In reality a human face is not a Lambertian surface but is close enough under normal conditions to be a good approximation.

2. Light sources and viewing point are distant from the object. If this is not true then the incidence and reflectance angles from the same light source at two distant points on the surface will not be the same. The further the sources and camera are from the object the smaller these differences become.

3. The light source is a collimated point source. Collimated light has rays which are almost parallel meaning that the light disperses minimally with distance. This combined with a point source means that the direction of the light hitting the object can be specified very accurately. In reality there will be some divergence of the rays, but the truer this assumption is in practice, the better

the results.

4. There are no cast shadows nor inter-reflections.

The first assumption is an important one and increasing how Lambertian a surface is will lead to better surface normal estimation. It is possible that using NIR instead of visible light would increase how Lambertian the skin is, as more is likely to be absorbed and the scatter made more diffuse. Grease or sweat on the skin will decrease how Lambertian the reflectance of the face is and increase the likeliness of specularities (highlights) appearing on the captured images. Along with these, cast shadows (e.g. from the nose) are potential problems with using PS for 3D face capture. There are many pieces of research which aim to mitigate these problems e.g. [172, 13, 170, 88, 39, 104, 70] but for the purposes of this research a "pure" PS approach is adopted as cast shadows are minimal and skin is unlikely to be sweaty in the environment the data was obtained.

The acquisition of multiple images of the same object under different lighting conditions can be done in one of two ways: temporal or spectral multiplexing. Temporal multiplexing captures the images in a timed sequence, as the different light sources are switched on and off. Obviously if the duration between images is too large and the image is not static then the reconstruction will not be accurate as the images will be significantly different. One method to overcome this is to use different frequencies of light at the point sources and capture these at the same time either by splitting the light via a prism onto sensors with different sensitivities, or by using specific band filters in front of the sensors. For the purposes of this research the former is used, as the total capture time is $\approx$15ms, the motion is effectively frozen so it offers the same functionality as spectral multiplexing without the additional complexities. The specifics of the actual device used can be found in Section 3.2.1.

The majority of past work on PS has been conducted using visible illumination. Studies into the optical properties of skin have shown it to be increasingly

reflective in the NIR light band up to wavelengths of about 1.1μm [9]. This suggests that NIR, which is more covert and less intrusive, is a viable alternative to visible light. Furthermore, NIR can be used as a replacement for visible light because its proximity to the visual spectrum means that it is likely to behave in a similar manner on skin. It might be expected that some fine surface detail would be lost due to sub-surface scattering as reported by Zivanov *et al.* [210], but this is unlikely to affect overall face shape estimation. In addition to this work, infrared light has been used previously in 2D face recognition to control for ambient illumination [118, 107] and to aid eye detection algorithms using the "bright eye" effect [137]. NIR has also been used for biometrics outside of face recognition. It is frequently used for iris recognition as it has the benefit over visible light of recovering the details of darkly pigmented iris' [49]. Also, because it penetrates skin more than visible light and it is readily absorbed by dexoxygenated haemoglobin, it can be used to enhance vascular structures which can then be used as a biometric e.g. recognition via vein pattern in hands and fingers [133].

3.2 Methods and Data

This section first outlines the overall PS image acquisition hardware, before moving on to describe the reconstruction process. The differences between the use of visible and NIR light sources are also discussed.

3.2.1 Hardware

This section details the acquisition device hardware. Although not a novel contribution of the thesis, it is worthwhile to describe the hardware in some detail as it was used to capture the data used throughout. The device, shown in Fig. 3.2, is designed for practical 3D face geometry capture and recognition. The presence of an individual is detected by an ultrasound proximity sensor placed before

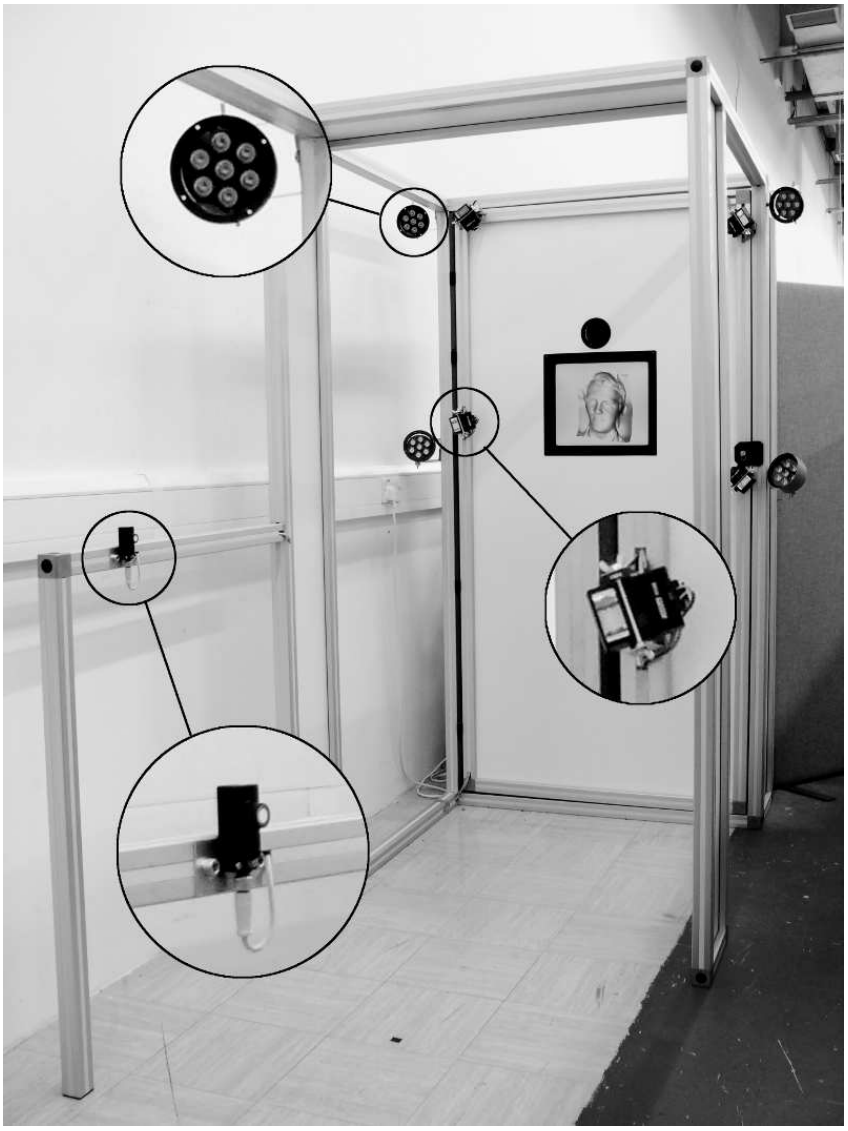


Figure 3.2: The Photoface photometric stereo capture device. Enlarged areas from top to bottom: NIR light source, a visible light source and an ultrasound trigger. The camera can be seen on the back panel above the monitor display.

the archway. This can be seen in Fig. 3.2 on the horizontal beam towards the left-hand side of the photograph. The sensor triggers a sequence of high speed synchronised frame grabbing and light source switching.

The aim is then to capture five images at a high frame rate: one control image with only ambient illumination and four images illuminated by the main light sources in sequence. Either one image per visible light is captured, or one image per NIR source. Note that the ambient lighting is uncontrolled (for the experiments presented in this thesis, overhead fluorescent lights are present). The four visible

light sources are low-cost Jessops M100 flashguns (colour temperature 5600K), while the NIR lights are stripped down X-vision VIS080IR lensed 7-LED clusters, which emit light at ≈850nm. The light sources are located approximately 75cm from the head of the subject at evenly spaced angles. The camera is 2m away from the head.

It was found experimentally that for people walking through the device, a minimum frame rate of approximately 150fps was necessary to avoid significant movement between frames. The device currently operates at 200fps, and it should be noted that it is only operating for the period required to capture the five images. That is, the device is left idle until it is triggered. A monitor is included on the back panel to show the reconstructed face or to display other information.

For visible light, the following sequence of events takes place to capture the five images as an individual passes through the device.

1. Await signal from ultrasound sensor.
2. Send trigger to camera.
3. Await integration enabled signal from camera.
4. Discharge first flashgun.
5. Await end of integration enabled signal.
6. Repeat from step 2 for the remaining light sources.
7. Capture control image with ambient lighting only.

All interfacing code is written in NI LabVIEW™. The ultrasonic sensor is a highly directional Baumer proximity switch. When its beam is broken within a distance of 70cm, it transmits a signal to an NI PCI-7811 DIO card fitted to a computer. When this signal is received, a trigger is sent to the camera. This is a Basler 504kc camera with a 55mm, f5.6 Sigma lens. As with many silicon-based

CCD sensors, the Basler chip is responsive to both visible and NIR irradiance. The trigger is transmitted to the camera from a frame grabber via Camera Link®. The frame grabber is an NI PCIe-1429, which communicates with the DIO card via a RTSI bus for triggering purposes.

To ensure that the signal has reached the camera, and that the camera has commenced frame capture (i.e. is integrating), a second connection from the camera to the DIO card as added. This connection is TTL-high while the camera is integrating. When the computer receives this signal, the first light source is to be immediately illuminated. A flashgun is discharged by making a short circuit between its input pins. This is achieved here by sending a short pulse from the DIO card to the input pins via a phototransistor opto-isolator IC. This electrically isolates the sensitive DIO card from the high voltages of the flashgun terminals. Finally, the DIO card awaits the falling edge of the camera integration enabled signal before moving on to the next light source.

For NIR light, a slightly different procedure is adopted whereby synchronous TTL signals are sent to the camera and LEDs. This is because the LEDs can be illuminated for the duration of the camera exposure, while the flashguns only last for a small fraction of the exposure. The NIR LEDs are powered independently from the DIO card and interfaced via a simple transistor circuit. As the LEDs are illuminated for only 5ms, it is possible to overpower them, in order to increase their brightness without causing damage. Therefore, 20V is applied across the LEDs, compared to the recommended 12V.

3.2.2 Visible and NIR Comparison

One possibly negative aspect of the visible light set-up is that the firing of flashguns is obvious to the subject and possibly intrusive to any surrounding people. Another possible advantage of NIR is that there may be additional subcutaneous or vascular structures present in the raw images taken under NIR light which may

be used to aid recognition. Unfortunately, such features were not visible in the wavelength band considered in this paper, but this could be an area to study in future work. NIR light is also more covert for a face recognition environment and subjects are less inclined to “pose” for the camera, meaning that more neutral expressions are likely. Finally, it is worth noting the advantage that many CMOS camera sensors are inherently more sensitive to NIR light.

There is also no colour information captured under the NIR arrangement, in contrast to the visible light set-up. However this is of no consequence to these experiments, as PS relates to monochromatic pixel intensities. If a hybrid system were to be developed for recognition purposes, which combined texture information with PS information, then using the visible lights sources may be necessary.

One disadvantage of NIR illumination is the relative difficulty in obtaining the necessary brightness for the required short exposure times. While the flashguns were easily bright enough with an exposure time of 1ms, an exposure of 5ms was needed for the NIR LEDs (i.e. the maximum possible exposure for the given frame rate). Although this was adequate for these experiments, the LED lenses employed provided a narrow divergence angle, meaning that the face had to be more precisely positioned to obtain full illumination. For the visible light sources, the images were bright enough even for large diversion angles, removing the need for accurate positioning of apparatus and allowing subjects to pass through the archway without having to consider their exact location with respect to the camera.

To account for ambient illumination, the control image is subtracted from the other four images. These images are then normalised in terms of intensity before reconstruction takes place. This was done by linearly scaling the greylevels of each image so that the mean intensity was equal for each image. A detailed comparison of the resulting reconstructions is presented in Section 3.3.

3.3 Results

3.3.1 Basic Reconstructions

Figure 3.3 shows a series of reconstructions from the method described in Section 3.2 using visible light. The device was placed at the entrance to a workplace to ensure casual (and thus realistic) usage. The general 3D structures of the faces have clearly been well estimated. Note, however, that the spectacles of one of the subjects have been "blended" into the face. This is a combined consequence of the rim of the spectacles being highly specular and the surface being non-integrable for this region of the image [64]. Although, ideally the shape of the spectacles would be estimated accurately, the blending effect can potentially be beneficial to face recognition algorithms because it means that such details have a lesser impact on the overall reconstruction. A set of images and reconstructions using both visible and NIR light sources can be seen in Fig. 3.4. It is clear that NIR is also capable of providing good estimates of the 3D geometry of the face.

The accuracy of the face reconstructions against ground truth data are now compared. To do this, eight different faces were scanned using a commercial 3dMD projected pattern range finder [3]. The 3dMD models were rescaled so that the distance between tear ducts was the same as in the visible PS reconstruction. All reconstructions are then cropped to 160$\times$200px regions centred on the nose tip that encompass the eyebrows and mouth. Part of the forehead is omitted by this choice of cropping region as it is frequently occluded by hair and is therefore deemed unreliable for face recognition. An example of the face regions used for comparison can be seen in Fig. 3.5, which also shows a ground truth reconstruction acquired using a 3dMD scanner [3]. The face regions from visible and NIR light sources are then aligned to ground truth using the ICP algorithm [19][1].

[1] MATLAB® implementation by Ajmal Saeed Mian (ajmal@csse.uwa.edu.au), Computer Science, The University of Western Australia.

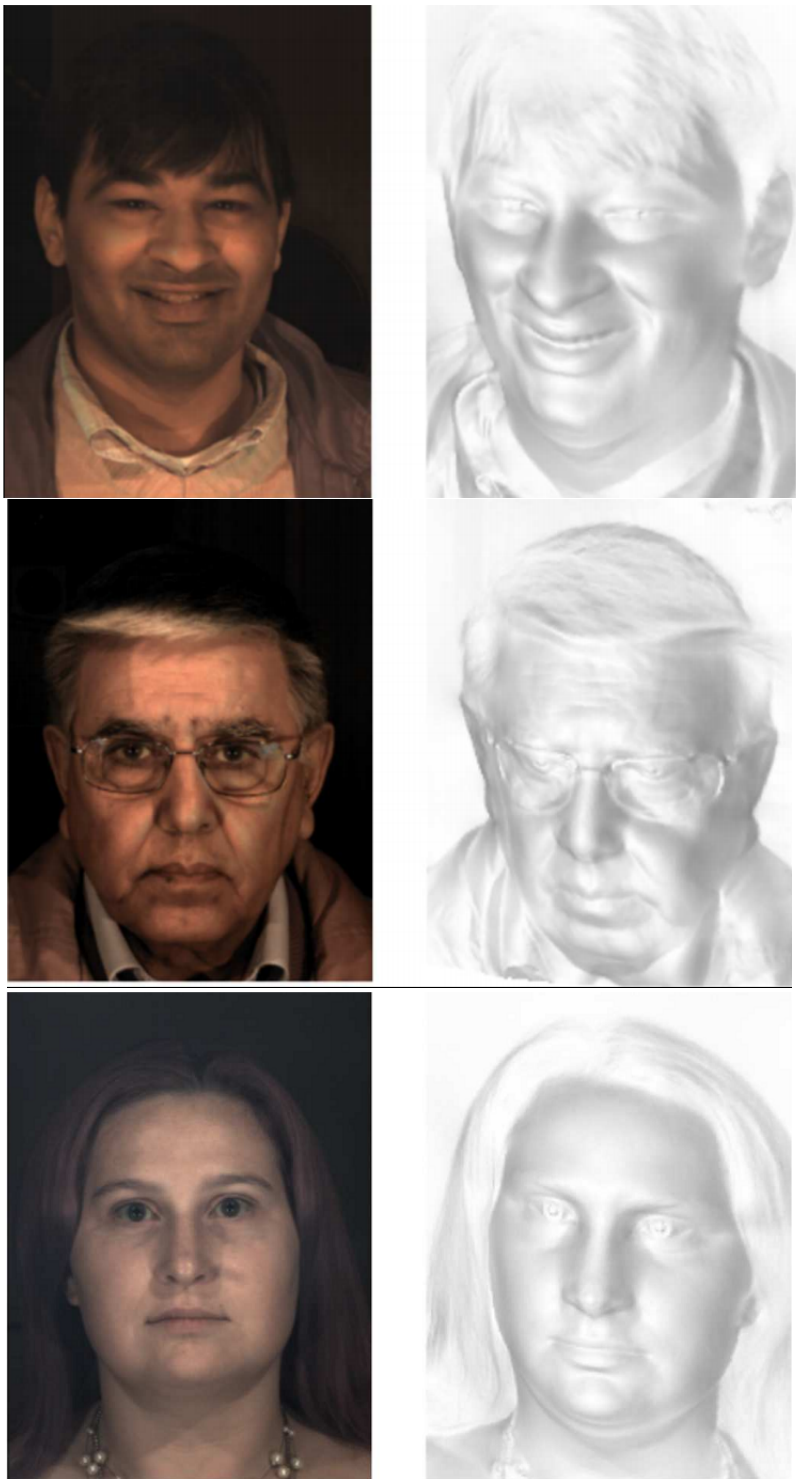


Figure 3.3: Estimated geometry of three different subjects using visible light sources.

Figure 3.4: Example raw images and reconstructions using visible (top) and NIR light sources for four subjects. For these experiments only, the subjects were asked to rest their chin on a support in order to ensure that all subjects are compared to each other in fair conditions. Note that hair is well recovered (other methods e.g. 3dMD tend to reconstruct hair poorly).

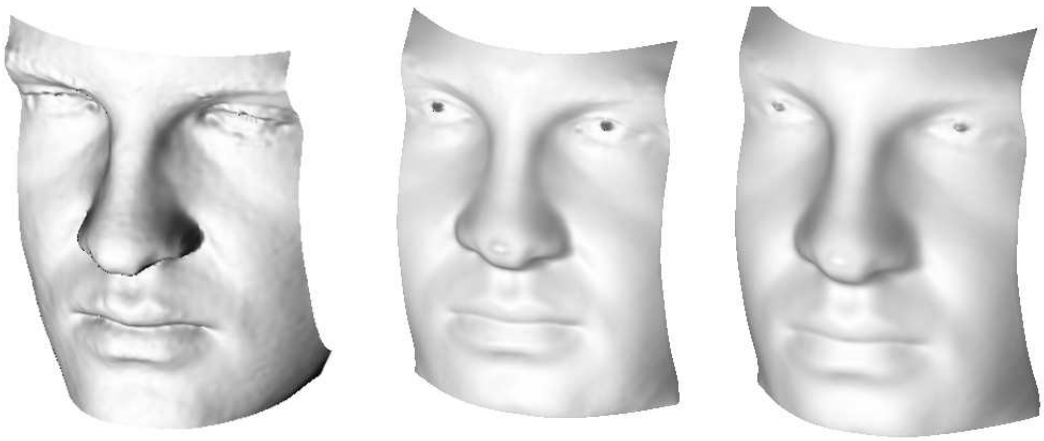


Figure 3.5: 3D Reconstructions for one subject from a 3dMD scanner (left) which is used as ground truth, PS using visible light sources (middle), and PS using NIR sources (right).

Individual RMS errors on surfaces and ℓ_2-norm errors on surface normals between the reconstructions and ground truth are displayed in Fig. 3.6. The eight subjects consist of 6 males and 2 females and a mixture of Caucasian and Asian ethnicities. The variations in residual errors and ℓ_2-norm distances between visible and NIR reconstructions are significant according to paired t-tests ($p = 0.05$). This demonstrates that PS using NIR as a light source is a perfectly valid approach and leads to more accurate reconstructions.

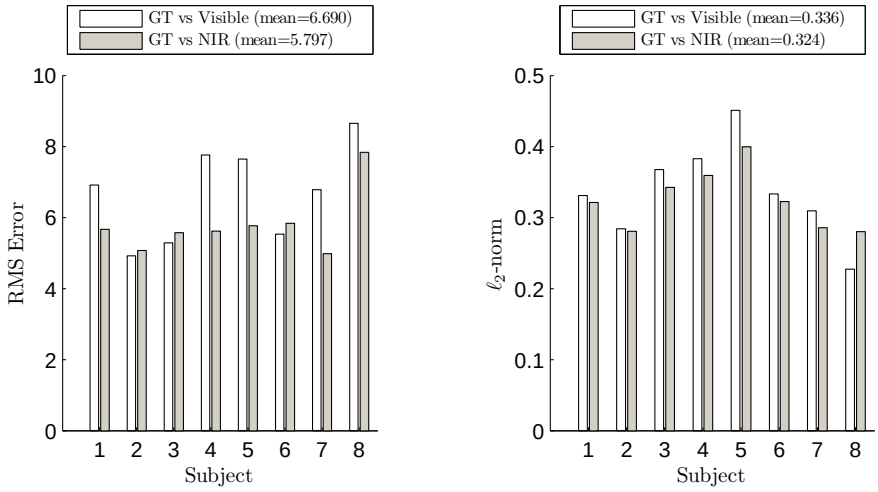


Figure 3.6: RMS (left) and ℓ_2-norm (right) errors between Ground Truth (GT) and visible PS and NIR PS for each subject. NB the order of subjects is arbitrary, i.e. there is no significance to the pattern that can be inferred from the ℓ_2-norm errors figure.

In order to get an indication of the regions where the greatest differences occur between ground truth and PS reconstructions, the residuals and ℓ_2-norm errors at each pixel are plotted in Fig. 3.7. Typically, the largest variations occur in regions with the highest curvatures, such as eye sockets, nose tips and the sides of the nose.

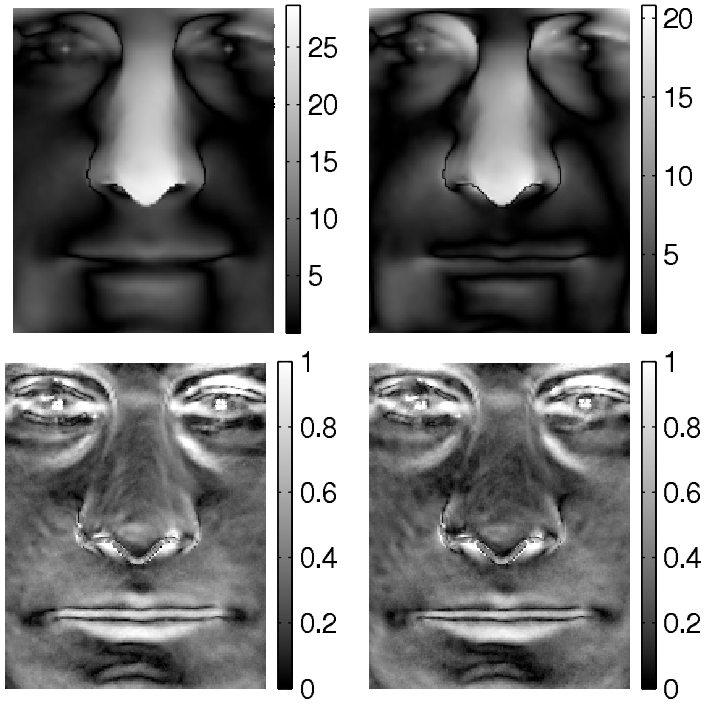


Figure 3.7: Representative examples of the residuals and the ℓ_2-norm errors at each pixel. Left to right: residuals for visible and NIR respectively, ℓ_2-norm errors for visible and NIR respectively. Lighter areas represent larger errors. The largest RMS errors appear around the nose which is near a discontinuity which the integration algorithm cannot cope with well. The ℓ_2-norm errors are highest around non-Lambertian surfaces such as eyes and lips and the shadowed area under the nose.

In attempting to produce the most accurate reconstructions possible via PS, it was found that the estimated surface normals could be enhanced by using normals acquired by re-differentiating the reconstructed height map estimate. It is unclear as to why this should be the case but preliminary analysis indicates that the reason may be due to the imposition of integrability constraints and the fitting of limited basis functions in the Fourier domain [64], as required by the adopted integration method. These factors may cause errant normals to be “smoothed

out" leading to a more accurate reconstruction. However, if this method of improving reconstructions is used, a second integration step would be needed thus removing one of the benefits of PS for face recognition: that the surface normals (and hence distinctive differential surface features) are recovered directly. More research is required into this area in order to confirm that the improvements are down to integrability constraints and if so, whether these constraints can be enforced without resorting to a full integration method such as that of Frankot and Chellappa. Using the Fisherface algorithm it was found the more accurate reconstructions produced using this method did not lead to any improvement in recognition accuracy, presumably because the effects were so small.

3.3.2 Reflectance Analysis

A reason for better reconstructions using NIR is that skin reflection is more Lambertian than under visible light. To confirm this, graphs of I/ρ against θ, the angle between the light source and the normal vector, have been plotted. For a purely Lambertian surface, the relationship between the two should follow a cosine law. The results can be seen in Fig. 3.8. To generate the graph, values of I, ρ and θ were estimated for each pixel of each image for each of eight faces. The angle θ is calculated for each point of the face from the 3dMD scan data and the known light source vectors. The average values of I/ρ are used for each 1° increment in θ. The line at $\theta = 60^\circ$ indicates a reasonable cut-off point after which data points become too sparse to be significant. The RMS difference between the measured curves and the cosine curve in the range of $0 \leq \theta \leq 60$ is 0.04 (s.d. 0.11) for NIR light and 0.06 (s.d. 0.12) for visible. For completeness, the RMS difference across the whole curve is 0.11 (s.d. =0.13) for NIR light and 0.17 (s.d. =0.12) for visible. The figure demonstrates that skin under NIR light is marginally more Lambertian than under visible light. It should be noted that the standard deviation across the whole range of θ remains approximately the same.

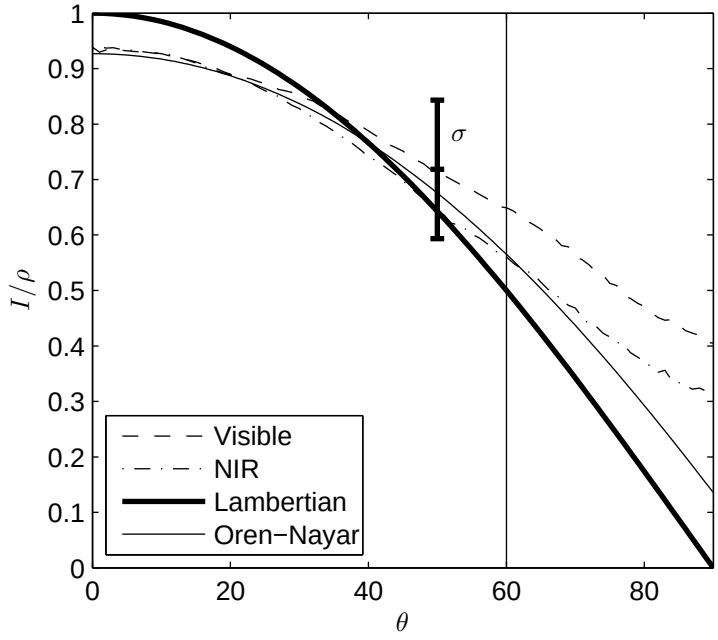


Figure 3.8: Plot showing that skin reflectance under NIR is more Lambertian. Mean I/ρ values averaged over 8 subjects against θ (the orientation of the skin at a pixel). To the right of the vertical line at $\theta = 60°$, data were too sparse to be of significance. For reference *one* standard deviation is shown to give an indication of spread (and is approximately the same through the range of θ, for both visible and NIR light sources). The Oren-Nayer plot is provided as an alternative reflection model to Lambert's Law and is discussed further in Section 3.3.3.

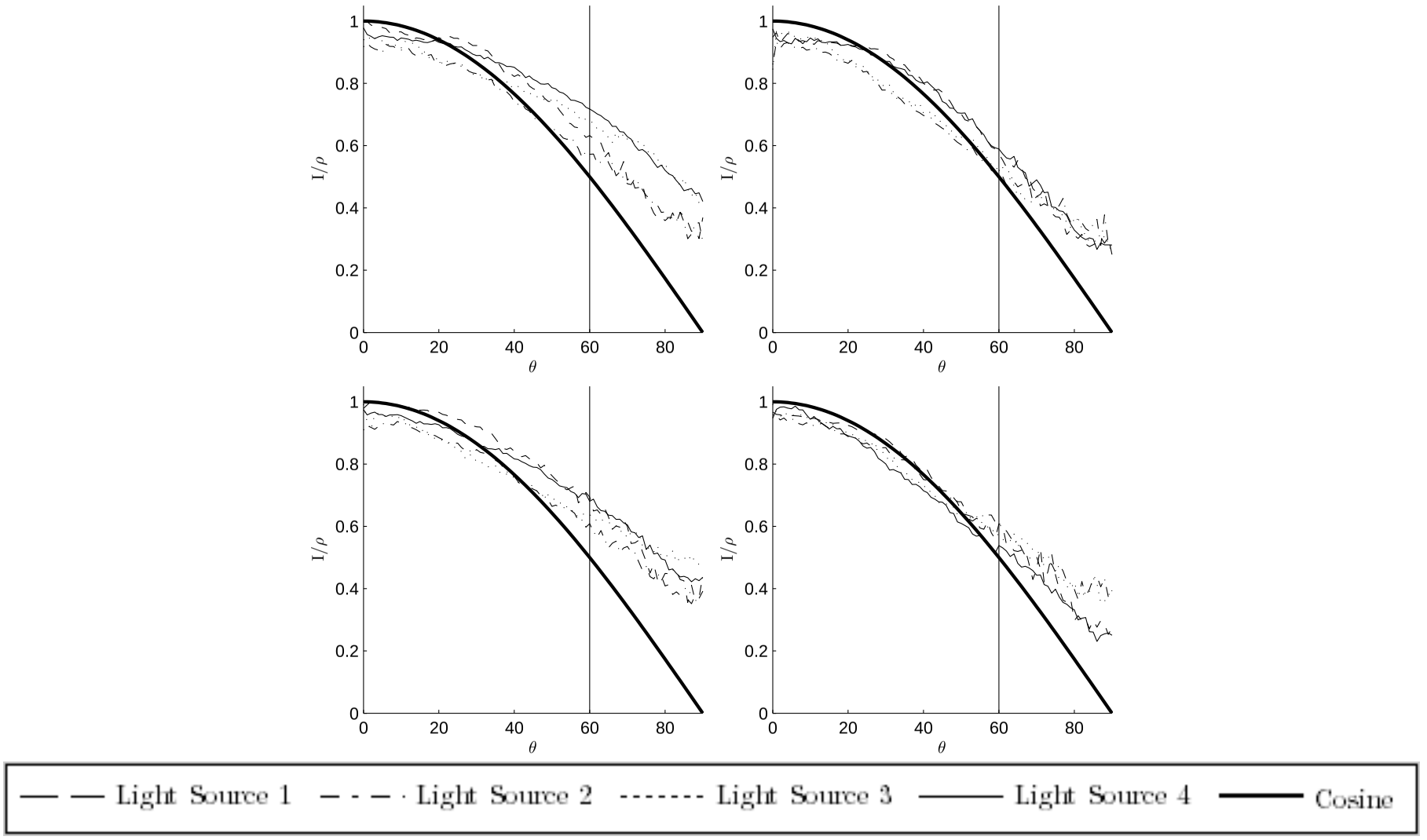


Figure 3.9: Reflectance of skin from individual light sources. I/ρ values from individual light sources plotted against θ for the first two reconstructions shown in Fig. 3.4. The top row shows results for Subject 1 under visible (left) and NIR (right) light sources, and the bottom row shows results for Subject 2 under visible (left) and NIR (right). The light sources are labelled clockwise from the bottom-left in Fig. 3.2.

Although the data suffers from significant noise levels (as indicated by a standard deviation exceeding 10% of the range for both conditions), the NIR condition has a lower RMS error and is therefore closer to the Lambertian curve than for visible light. This difference is significant given the large numbers of pixels and subjects used in the trials. This represents an average pixel intensity error of 10 grey levels for NIR and 15 for visible light across the image, assuming a maximum of 256 grey level intensities. This supports the hypothesis that skin is more Lambertian under NIR illumination. This result is likely related to the fact that NIR light penetrates the skin further than visible light [65], which facilitates a more uniform scattering than surface reflection. Note however, that neither the Lambertian model nor the Oren-Nayar model (see below) take account of internal scattering or Fresnel effects. The results in Section 3.3.1 demonstrate that the more Lambertian behaviour associated with NIR light also leads to more accurate reconstructions.

A more detailed analysis for two subjects is shown in Fig. 3.9 and Table 3.1. What can be noted immediately is the similarity across the plots. There are small differences in I/ρ caused by different light sources but this appears to have little negative impact on the reconstructions and is likely to be due to environmental effects. The figure suggests that PS using both visible and NIR is robust to different skin types and light intensities. A more thorough analysis of the effects of gender and race on reflectance properties could be the subject of future work.

	Visible		NIR	
	RMS, $\theta \leq 60^\circ$	RMS, overall	RMS, $\theta \leq 60^\circ$	RMS, overall
All Faces	0.06, ($\sigma = 0.11$)	0.16 ($\sigma = 0.12$)	0.04 ($\sigma = 0.12$)	0.11 ($\sigma = 0.13$)
Subject 1	0.07, ($\sigma = 0.09$)	0.16 ($\sigma = 0.18$)	0.05 ($\sigma = 0.12$)	0.10 ($\sigma = 0.22$)
Subject 2	0.07, ($\sigma = 0.10$)	0.17 ($\sigma = 0.18$)	0.04 ($\sigma = 0.13$)	0.12 ($\sigma = 0.21$)

Table 3.1: The results show consistency across different subjects and different types of light source. The RMS collective error across all eight reconstructions and for the first two reconstructions shown in Fig. 3.4 separately. The standard deviations are shown in brackets.

3.3.3 Comparison to the Oren-Nayar Model

A comparison is given of the recorded reflection measurements to the Oren-Nayar reflectance model [142], as shown in Fig. 3.8. The Oren-Nayar model represents the reflecting surface as an array of V-shaped groves of random orientation, commonly called “microfacets”. The distribution of microfacet orientations is characterised by a roughness parameter and each facet is assumed to act as perfect Lambertian reflector. This model is able to account for the common feature of limb-brightening and is itself based on the earlier Torrance-Sparrow model [179] where each microfacet is assumed to be mirror-like.

The Oren-Nayar model was chosen as a comparator, as skin is not a smooth surface (especially on older people) and the model has been shown previously to be successful on a range of materials of varying degrees of roughness [142]. It is not the case that the microscopic structure of skin closely matches the Oren-Nayar model, but it is used here for demonstrating how alternate methods for reflection may improve the framework in future work. Investigating the various degrees of freedom of the Bidirectional Reflectance Distribution Functions (BRDFs) is also reserved for future work. Furthermore, there are additional models for skin reflectance which take account of a huge range of physical phenomena [55, 117], but these are out of the scope of this thesis.

The Oren-Nayar curve in Fig. 3.8 represents an example intensity profile for reference with a roughness parameter of 0.2. This value for this parameter was chosen, not because of any research suggesting that it would model the subjects’ skin most accurately, but because the generated curve is similar to the observed data and offers an alternative to Lambert’s law. Clearly, this model fits the measured reflectance data significantly more accurately than the Lambertian curve, suggesting that the model could be incorporated into the method in the future. This will however, add significant complexity and computation time to the algorithm. This is because a minimisation method must be implemented in order to

recover all the model parameters and to accommodate the increased number of angular degrees of freedom in the model.

3.4 Discussion

Photoface is capable of good quality reconstructions with reasonably small errors compared with ground truth. Ground truth in these experiments was a 3D capture using the commercial 3dMD system which reproduces large scale geometry excellently, but is not capable of reproducing the fine details of the face. Some of errors may therefore not actually be present but occur when the finer details which have been captured by PS are compared to the 3dMD capture where they are not present. It is not an ideal ground truth but it does allow a consistent approach to quantifying the accuracy of the reconstructions. Using visible light sources results in a 15 pixel RMS error and using NIR gave a slightly better 10 pixel RMS error and it is suggested that this could be attributable to greater sub-surface scattering of the higher penetrating NIR light.

The results presented in Section 3.3.1 demonstrate that PS is an effective method for producing 3D facial reconstructions in terms of quality. This method also requires a relatively short computation time. Using the device with standard PS, LabVIEW™ interfacing, non-optimised MATLAB® processing and a typical modern PC, the time between device trigger and the reconstructed height map was approximately eight seconds. The construction of the hardware also lends itself well to relatively unobtrusive data capture with a minimum amount of effort from the subject. Of particular interest are the following points:

1. The PS technique offers a valid alternative to existing, more expensive and processor intensive, 3D face capture methods.

2. The PS technique is robust to common facial features such as spectacles, makeup and facial hair (see also [10]).

3. NIR light sources produce reconstructions that are more accurate than visible light sources.

In terms of assessing the suitability of the Photoface device, the system offers several benefits over commonly used existing laser triangulation and projected pattern 3D shape capture devices:

1. It is significantly cheaper to construct.
2. Acquisition time is shorter than laser triangulation systems.
3. Data processing time is shorter than projected pattern systems.
4. The method is robust to typical ambient illumination conditions.
5. It is very robust against accidental collisions (because it is tolerant to errors in the light source vectors).
6. Very fine details of the face can be reconstructed.
7. Calibration is very quick and simple and only needs to be performed after the initial light source positioning.
8. Although the Photoface system cannot reconstruct hair with high levels of accuracy, it can at least provide some details of its overall shape (see Fig. 3.3, for example). In contrast, laser triangulation and projected pattern systems usually fail completely with hair.

At present, the 3D reconstructions are not yet as accurate as those from projected pattern range finders. The reconstructions tend to be flatter than their real-world counterparts, with most protrusions understated. They do however provide extremely fine detail of a face such as wrinkles and pores. Even though the reconstructions suffer from a flattening of the features, they would still appear to be

viable for recognition purposes (each reconstruction is clearly of a distinct identity) and the additional fine details could potentially be used as supplementary information to aid recognition.

The reconstructions under NIR were shown to be more accurate than those under visible light, but provided no additional 2D texture information. They also diminish the need for flashing lights, making the system less intrusive compared to visible light.

Zivanov *et al.* [210] offer an alternative argument to ours, stating that shorter wavelength light gives better results. Their justification is that shorter wavelengths undergo less internal scattering and thus provide a crisper, more defined reconstruction. It would appear therefore that a compromise must be reached in deciding between fine detail (using Zivanov's short wavelength suggestion) and overall geometry and covertness (using the NIR method). If fine surface textures such as wrinkles and pores are of interest, then using shorter wavelength light would be most beneficial. However, for a covert system where fine details are of less importance than general shape, then NIR offers the best solution.

It is clear that the face reconstructions are of different individuals just from the naked eye and the low pixel inaccuracies are a good indication of the suitability of this device for face recognition. Subsequent chapters provide further evidence that Photoface is good for recognition and not just reconstruction. In order to be confident of any recognition results there needs to be a suitably large dataset. In the next chapter the publicly availably Photoface database is introduced together with baseline recognition performance results on the data using common algorithms and representations.

Chapter 4

Photometric Stereo Databases and Baseline Experiments

The previous chapter demonstrated the suitability of the Photoface device for face recognition in terms of its capture accuracy. This chapter goes on to describe the creation of two databases, one large scale using visible light sources in relatively unconstrained conditions and another smaller database which additionally includes using NIR light sources to capture a range of expressions.

In order to compare algorithms it is necessary to perform some baseline experiments on data that could be described as standards. This also provides an additional mechanism to assess the suitability of photometric stereo data. Standard 2D techniques are used and extended for use on 3D data (Eigenfaces, Fisherfaces), along with an examination of the effects of different representations e.g. shape index, depth map, albedo etc. on recognition performance.

The contributions of this chapter are:

1. The Photoface database, a publicly available PS database with rich metadata for each session and a Query Tool application for straightforward extraction of datasets.

2. The 3D Expression-VISible and near-InfraRed database (3DE-VISIR) con-

sisting of three expressions per subject captured using visible and NIR light sources.

3. Baseline experiments on Photoface data using five commonly used data representations and three algorithms showing the suitability of the data for recognition.

4.1 The Photoface Database

There are many existing face databases which can be used for face recognition research. The majority are 2D but there are an increasing number of 3D datasets (including one consisting solely of identical twins [186] and even 4D (3D surface video). Major databases and their key properties are given in Section 4.3. Here, a novel 3D database is presented which is the largest captured in terms of subjects and sessions using PS. Additionally the capture is relatively unconstrained (certainly in comparison with other 3D face databases): the Photoface device was left unattended in a workplace corridor and employees walked through it. This is a very useful feature and is unique to the Photoface database as it allows recognition algorithms to be tested on more naturally captured data. The database is freely available to the research community [199].

4.1.1 Collection and statistics

The Photoface database consists of 3187 captured sessions of 453 subjects. The data was collected from the offices of General Dynamics and from the DuPont Building of the University of the West of England. The majority of the data was collected from General Dynamics in two recording periods of approximately 6 months each, separated by approximately one year (1764 sessions between February 2008 - June 2008, 603 sessions between June 2009 - October 2009). The Photoface device was located in an unsupervised corridor and subjects were

Figure 4.1: A session captured by Photoface consisting of four differently illuminated photographs. The faces are detected using a MATLAB®implementation of the Viola And Jones face detection cascade and cropped to around 550×750 px.

free to choose whether or not to pass through it. The only incentive was that a reconstruction of their face was displayed immediately after capture. A more detailed breakdown of the captured session statistics can be seen in Table 4.1.

Each captured session consists of four differently illuminated images of the subject as they pass casually through the Photoface device. A publicly available MATLAB®implementation [108] of the well known Viola and Jones [187] face detection algorithm is used to extract faces from the photographs which are then saved to disk as bitmaps. The size of captured photograph varies from session to session depending on various factors such as the size of the face but are typically about 550px × 700px. Each session is saved in a directory named with a time stamp in the format `<YYYY-MM-DD_HH-MM-SS>`. An example session can be seen in Fig. 4.1.

Consent from the subjects to publish their images for research purposes (e.g. in journal or conference articles) was only granted explicitly by some subjects, others preferred not to give this consent but were happy to have their images used as data for research. Those that gave permissions are found in the subject range between 1001 and 1222, those that did not are found in the subject range between 2001 and 2231. The End User Licence Agreement that researchers must sign before getting access to the Photoface database, clearly states this and no images may be published from the range above 2001.

Once the capture period was over, the sessions were manually sorted into subject directories. Some sessions were discarded for various reasons e.g. false trigger, subject walking backwards through the device, subject holding a piece of paper with a face drawn onto it in front of his own face etc. Each session directory contains:

- four photographs
- a file that codes the capture location
- a file containing light source directions (these vary slightly between capture locations)
- two files of metadata: one containing the coordinates of 11 fiducial features and the other containing various other information such as gender, facial hair, pose, quality of capture. More about this can be found in Section 4.1.2.

Of the subjects, 381 subjects and 2891 sessions are male and 72 subjects and 296 sessions are female. Fig. 4.2 shows the number of sessions per subject – 157 subjects used the device only once, but 131 have used the device five times or more and 70 used the device at least 10 times.

Some examples of the captured sessions can be seen in Fig. 4.3.

4.1.2 Metadata and The Photoface Query Tool application

In order to align the images or to assess the accuracy of any landmark detection algorithms and to be able to extract subsets with certain properties it is necessary to manually annotate the data. For the Photoface data, two sets of metadata were created:

1. x and y coordinates of 11 fiducial features as shown in Fig. 4.4
2. Subject and session information: gender, facial hair, whether the subject is wearing glasses, the expression (positive, neutral or negative), occlusions

Location	Full Consent (sessions)	Partial consent(sessions)	Total (sessions)
UWE I (Dec 2007)	58 (530)	69 (246)	127 (776)
GD I (Feb 2008-Jun 2008)	80 (1209)	162 (555)	242 (1764)
GD II (Jun 2009-Oct 2009)	66 (603)	0 (0)	66 (603)
UWE II (Nov 2009-Feb 2010)	18 (44)	0 (0)	18 (44)
Totals	222 (2386)	231 (801)	453 (3187)

Table 4.1: Photoface distribution by date and location. Full consent indicates subjects who are happy to have their images published, partial consent refers to those that are only happy to have their data used in the research itself.

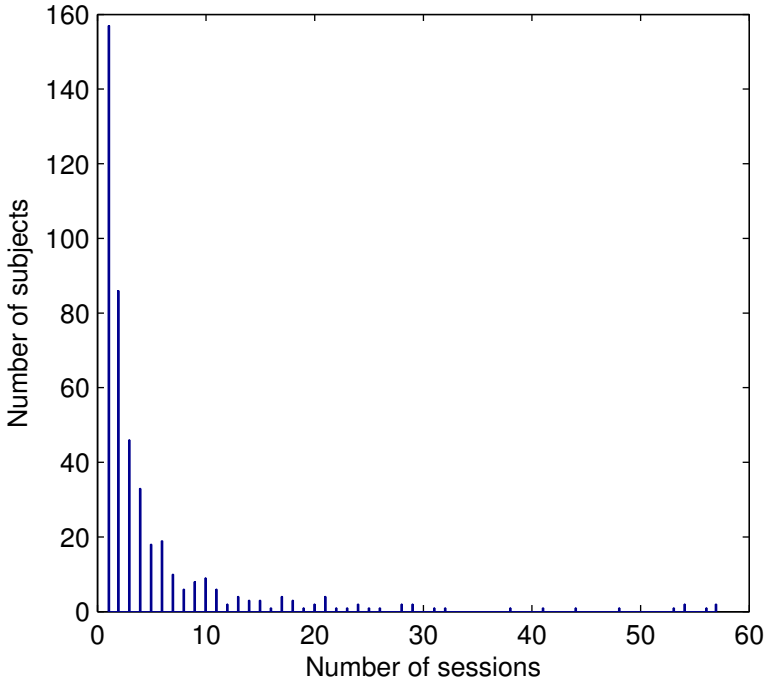


Figure 4.2: Photoface session distribution. The number of sessions per subject in the Photoface database. Each bar represents the number of subjects that used the device a certain number of times e.g. the first bar shows that 157 subjects only used the device once. To improve readability, six outliers have been omitted who used the device more than 60 times, the maximum of which visited it 290 occasions.

(e.g. a mobile phone, hair), pose (1-5 where 1 is frontal and 5 is extreme pitch, roll or yaw e.g. profile, quality (blurry capture, under exposed), other (eyes closed and mouth ajar).

To facilitate the marking of these features, two applications were written in MATLAB®. The first is for marking the fiducial features of each session, the other allows the secondary metadata (gender, occlusions etc.) to be recorded. To reduce the effects of extreme illumination present in individual photographs, the maximum value for each pixel location of the four captures of each session was calculated. This helped minimise errors by eliminating dark regions in the captured images. In cases where subjects have moved very quickly through the device, and there is disparity between the images, and thus blur in the image used for marking fiducial features, an estimate of the position is provided. There are few such captures and the disparity is minor. If one wished to improve on

Figure 4.3: Representative examples taken from the Photoface database.

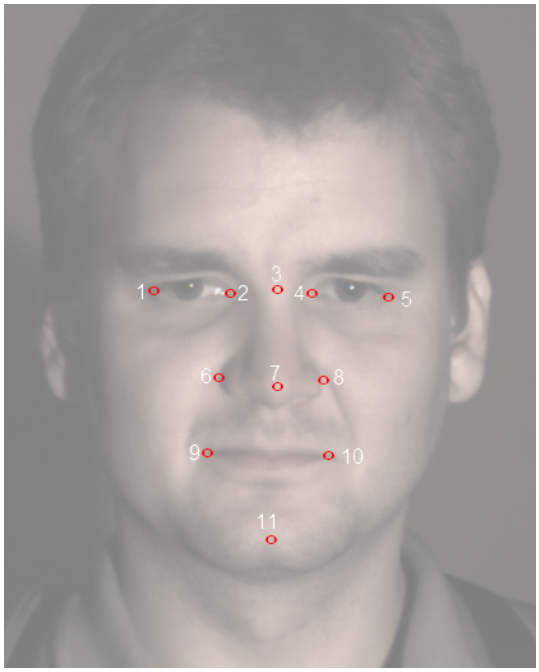


Figure 4.4: Manually marked fiducial points. 1) Left Lateral Canthus, 2) Left Medial Canthus, 3) Nasion, 4) Right Medial Canthus, 5) Right Lateral Canthus, 6) Left Nose, 7) Nose tip, 8) Right Nose, 9) Left Lip, 10) Right Lip, 11) Chin pogonion.

this an alignment algorithm could be employed to correct the differences caused by motion. However this is non-trivial in the case of PS as disparities caused by the motion will most likely be along the focal plane. Great care was taken to minimise inaccuracies, however due to the subjective nature of certain fiducial points (e.g. the exact tip of a nose) and the repetitiveness of the task, certain errors will be present in the database. The author encourages researchers who find such errors to inform the author so that the database can be updated. To reduce the errors in the first instance, it would have been useful to employ a number of others to mark the positions and take a median of the results. Unfortunately due to time constraints, this was not done. In cases where a fiducial point was occluded or out of frame, the relevant fiducial position is recorded as the top left hand corner of the image. When these values are loaded into the database, they are given a value of -1 to show that they are invalid. Screenshots of the two applications can be seen in Fig. 4.5.

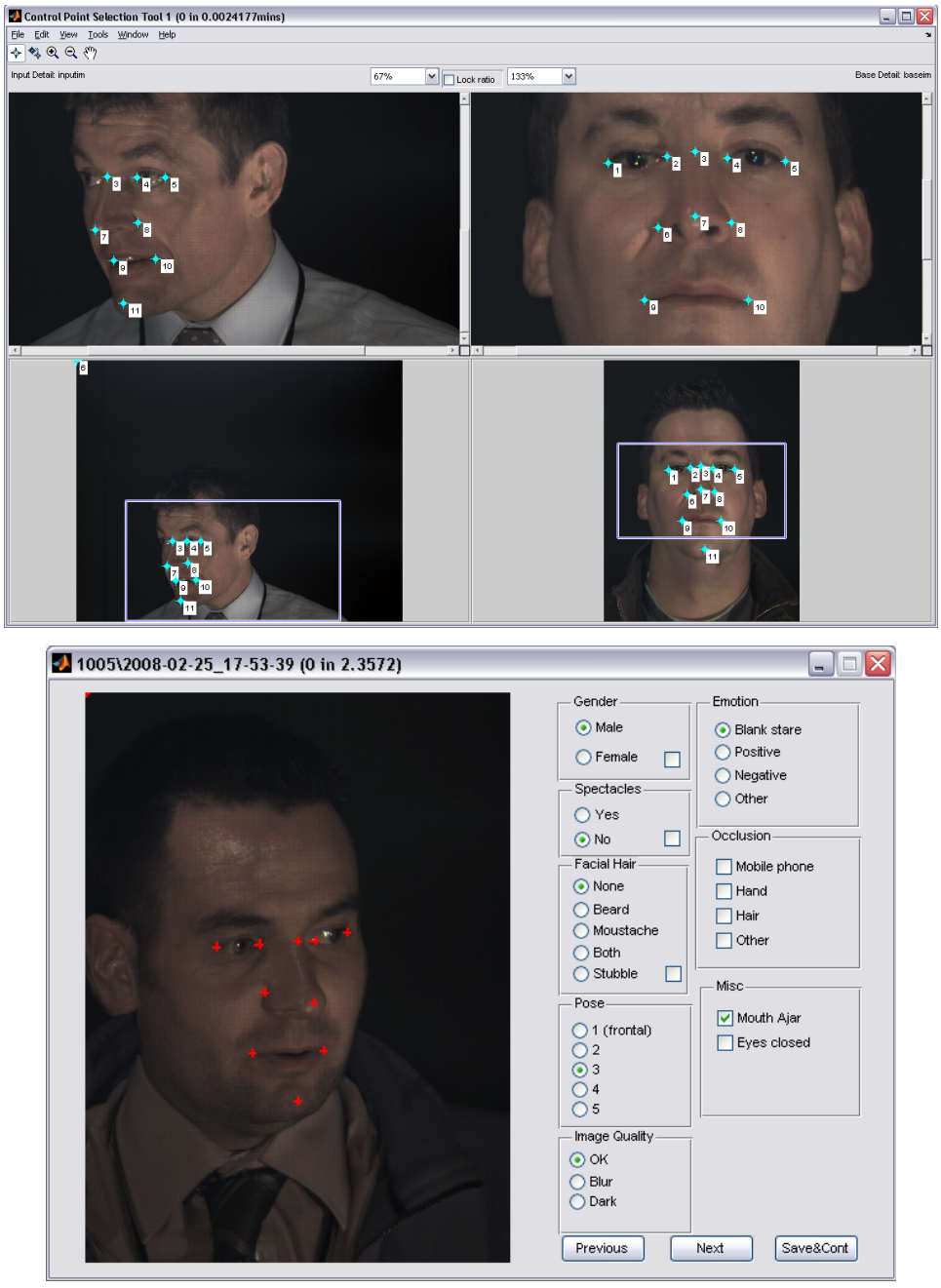


Figure 4.5: Screenshots of the applications for marking fiducial features (top) and additional metadata (bottom)

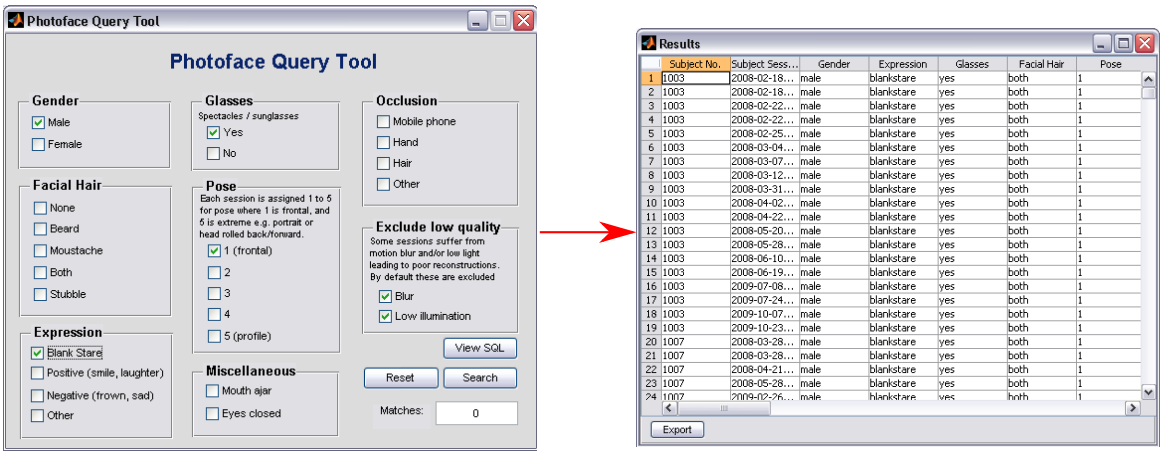


Figure 4.6: Screenshots of the Photoface Query Tool. The screen on the left allows selection of certain parameters and the matching records are shown in the screen on the right.

Both sets of metadata are stored in the session folder. Once all sessions had been processed, all the metadata was collated into one file stored in the root of the database. This was then converted to a set of Structured Query Language (SQL) statements for importing into a database. Storing the metadata on a SQL database allows efficient interrogation of the data. PostgreSQL was chosen for the database implementation as it is open source and free for non-commercial use and a proven technology used by Yahoo, Skype and the International Space Station.

The database can be queried in the usual way via SQL, but in order to facilitate searching a GUI based MATLAB®application was developed, called the Photoface Query Tool. This allows the metadata to be searched and results to saved to a comma-separated-value file which can then be loaded into MS Excel if any tweaks are required or used directly to load specific subsets of data e.g. all frontal images of males wearing spectacles. Screenshots of the application can be seen in Fig. 4.6.

The Photoface database and Photoface Query Tool have been made available to research community and were presented to the community at the Computer Vision and Pattern Recognition 2011 conference [199].

4.2 The 3D Expression - VISible and InfraRed (3DE-VISIR) Database

In addition to the Photoface database which contains relatively unconstrained sessions, a second database was also captured: the 3DE-VISIR database. This database contains 363 sessions of 115 people captured in five periods under more constrained conditions than the Photoface database. The novelty of this database is that:

1. For each capture there is one session captured using near NIR and another using visible light. These are captured in rapid succession, effectively freezing any motion making comparison between the lighting modes possible.

2. The aim was that for each subject there are at least three captures of different expressions, positive, negative and neutral.

As demonstrated in Chapter 3, NIR gives more accurate reconstructions than visible and is less intrusive. It was therefore logical to capture subjects using NIR as well as visible light. For practical reasons we were unable to catch multiple expressions per session in every case. The relatively ambiguous classifications of the expressions (positive, negative and neutral) were deliberately chosen as it was found that prompting a subject to look sad or miserable, or happy led to unrealistic and forced expressions which rarely looked like a typical sad or happy face. Positive and negative classes allow the subject a wider range and it is likely that they felt less judged on their attempt, resulting in a more natural expression. Inevitably the positive class of expressions contains almost exclusively smiles, while the negative contains mostly frowns but in some cases a scream or an angry face resulted.

Because the NIR lights are very directional, the subjects were made to rest their chin on a tripod to ensure that they were positioned correctly within the light sources.

In total there are 363 sessions of 115 people of which there are 100 subjects for which at least 3 expressions were captured in 322 sessions under both visible and NIR light. Of those 100 subjects there are 44 females and 56 males.

Although written consent has been given by all subjects involved, the database has not yet been made publicly available due to time constraints.

Examples of raw captures and resultant 3D surface under visible and NIR light can be seen in Fig. 4.7

4.3 Comparison with Existing 3D Databases

Table 4.2 compares the Photoface database and 3DE-VISIR database with the most commonly cited (publicly) available databases. In terms of subject/session numbers ours are clearly comparable, with the Photoface database being amongst the largest. The CMU Multi-PIE is a very comprehensive database in terms of pose, illumination and expression which can be used to generate 3D surface models via PS as the lighting angles are known, but it contrasts with the Photoface database in terms of it being highly constrained. Indeed, Burton *et al.* [97] stress this very point, stating that while the use of common datasets is important for benchmarking algorithm performance, they do not necessarily give a good indication of the likely performance in the real world. They state that the within-person variability will be less when the photographs are posed. The novelty of the Photoface database comes from the relatively unconstrained capture environment. While the 3DE-VISIR database is not the largest database it is the only to offer direct comparison of 3D faces and expressions captured under NIR and visible light.

Figure 4.7: Representative examples taken from the 3DE-VISIR database, showing raw images captured under visible light (top row), NIR (second row) with positive (first column), neutral (second column) and negative expressions (third column).

Name	Subjects (sessions)	Capture technology	Comments
FRGCv2.0 [147]	466 (4007)	Minolta Vivid 900/910 laser scanner	Released in 2004
CMU Multi-PIE [79]	337 (755,370)	Each subject is captured under 15 view points and 19 illumination conditions in four recording sessions with up to 6 expressions	Not strictly a 3D database, but it can be used to generate 3D from photometric stereo as the illumination angles are known
Bosphorous [156]	105 (4666)	Inspeck Mega Capturor II 3D structured light capture device	Up to 35 expressions per subject, coded using Facial Action Coding System (FACS)
3D RMA [20]	120 (206 usable)	Structured lighting device	Frontal / Left-right / Up-down. 120 subjects captured two months apart
3DFRD [81]	105 (1149)	High resolution stereo (0.32mm)	Highly constrained, but well registered with manually marked features
BJUT-3D Chinese Face Database [78]	500 (500)	CyberWare 3D laser scanner	High constrained, no variance. Half male, half female
FRAV3D [40]	106 (1696)	Minolta VIVID 700 laser scanner	Pose (and lighting conditions). Currently not being distributed
GavabDB [135]	61 (549)	Minolta VIVID 700 laser scanner	Pose and expression
The UOY 3D Face Database [89]	350 ($>$5000)	Projected pattern stereo camera	Pose and expression
Photoface [199]	453(3187)	Photometric stereo	Unconstrained capture in office environment
3DE-VISIR database	115(367)	Photometric stereo (visible and NIR)	Multiple expressions of subjects captured in controlled conditions

Table 4.2: A comparison of the Photoface database and 3DE-VISIR database databases with frequently cited 3D databases

4.4 Baseline Methods

In the previous section, the databases which have been created as part of this project were described. In this section, the recognition accuracy on the databases is measured using some common methods. To date there has been very little published research using PS generated faces so here experiments are performed on some different data representations using a variety of common techniques. These not only serve to justify the use of the Photoface database and 3DE-VISIR database databases, but also as a benchmark results against which novel approaches can be compared to in later chapters. It also proves that the performance of algorithms, as generally reported in the literature, is similar on this dataset. Detailed descriptions of the algorithms are now provided (a glossary of symbols can be found in Glossary of Symbols at the end of this thesis) before the results of applying the methods to Photoface database are given.

4.4.1 Principle Components Analysis (PCA) & eigenfaces

PCA was invented by Karl Pearson [146] in 1901 and is also known as the Karhunen-Loeve transform. As stated previously, PCA defines the dimensions along which the most variation in data occurs. It is an orthogonal linear transform which takes a set of possibly correlated variables and transforms them into uncorrelated variables; those with the lowest eigenvalues (a measure of how much variance they account for) can then be discarded. This highlights a very important property of PCA and a reason for its popularity in image processing: dimensionality reduction.

Use in face recognition

The idea of using PCA to describe faces came from Sirovich and Kirby [168]. They were able to show that a 128$\times$128 px grayscale image (so 2^{14} dimensional

vector) could be characterised by 40 coefficients to within 3% error. These coefficients are weights given to linear combinations of basis images (eigenfaces) which provide a description of the original face. These basis images are are linearly independent of one another (i.e. none can be formed from a linear combination of the others) but are sufficient to represent every face in the given face space. As an analogy, if we imagine that there is a recipe for each face, then the ingredients are the eigenfaces and the quantity of each ingredient is the coefficient. Given a set of M images $\mathbf{x}(a, b)$ where the dimensionality $N = a \times b$, the ensemble of images can be represented as an $N \times M$ matrix when the rows of each image are concatenated and transposed to form a vector. The covariance matrix **C** would normally be an $N \times N$ matrix. The key contribution of their paper was to realise that if the number of images M was less than the dimensionality of **C** then **C** is singular (or non-invertable) and cannot be of order greater than M. This simplifies the subsequent eigen-decomposition greatly as the dimensionality of **C** becomes $M \times M$.

The basic steps of PCA are outlined next. Let **X** be a matrix representation of a set of face images:

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3 \ldots \mathbf{x}_m] \tag{4.1}$$

where **X** is of dimension $N \times M$ where N is the number of pixels in an image and M is the number of images and $x_{1\ldots m}$ are vector representations of the (2D) photographs. Because we are interested in modelling the variation, let **A** be a matrix of the difference between each image and the mean image **u** (the average intensity at each pixel across all images, so **u** will be the same size as **x**)

$$\begin{aligned} \mathbf{A} &= [\mathbf{x}_1 - \mathbf{u}, \mathbf{x}_2 - \mathbf{u}, \mathbf{x}_3 - \mathbf{u} \ldots \mathbf{x}_m - \mathbf{u}] \\ &= [\mathbf{x}'_1, \mathbf{x}'_2, \mathbf{x}'_3 \ldots \mathbf{x}'_m] \end{aligned} \tag{4.2}$$

The principal components are the eigenvectors $\mathbf{e}_i$ of the covariance matrix $\mathbf{C}$:

$$\mathbf{C} = \sum_{i=1}^{M} x_i' x_i'^T = \mathbf{A}\mathbf{A}^T \tag{4.3}$$

$\mathbf{A}\mathbf{A}^T$ leads to the $N \times N$ problem, but it can be seen that the eigenvectors of $\mathbf{A}\mathbf{A}^T$ are equivalent to the eigenvectors of $\mathbf{A}^T\mathbf{A}$ pre-multiplied by $\mathbf{A}$:

$$\begin{aligned} \mathbf{A}^T\mathbf{A}\mathbf{e} &= \lambda\mathbf{e} \\ \mathbf{A}\mathbf{A}^T\mathbf{A}\mathbf{e} &= \lambda\mathbf{A}\mathbf{e} \qquad \text{Premultiply by } \mathbf{A} \\ \mathbf{C}(\mathbf{A}\mathbf{e}) &= \lambda(\mathbf{A}\mathbf{e}) \qquad \mathbf{C} = \mathbf{A}\mathbf{A}^T \text{ (Eqn. (4.3))} \end{aligned} \tag{4.4}$$

where $\mathbf{e}$ are the eigenvectors and λ the eigenvalues of $\mathbf{A}^T\mathbf{A}$

The eigenvectors are then sorted in descending order of their corresponding eigenvalues; the larger the eigenvalue, the more variance is described by the eigenvector. Typically, one is only interested in the eigenvectors which capture between 80%-90% of variance cumulatively – the others can be discarded. For the results reported in Section 4.4 the number of components is chosen which accounts for 85% of variance. Belhumeur *et al.* [17] confirm a finding by Sirovich and Kirby that recognition performance levels off at about 45 components. The number of eigenvectors kept will be referred to as P.

This has described how to find the principle components or face space describing the data. The next step is to use these for the purposes of recognition. A face image can be projected into this face space by the simple operation

$$\boldsymbol{\omega}_p = \mathbf{e}_p^T(\mathbf{x} - \mathbf{u}) \tag{4.5}$$

where p is $1 \ldots P$ and represents the set of operations for each selected eigenvector resulting in a P-dimensional vector $\boldsymbol{\Omega} = [\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \ldots \boldsymbol{\omega}_p]$. If we plot all the $\boldsymbol{\Omega}$ vec-

tors into this space, clustering should occur around identities e.g. the eigenspace representation of different faces of the same person should lead to a cluster of points. Alternatively, the mean of these clusters can be used to represent a person (also known as a prototype). The recognition method used by Turk and Pentland [181] is to project a new face into the eigenspace and see which of these k cluster means it is closest to in a Euclidean sense by choosing the lowest ε_k: $\varepsilon_k = \|(\mathbf{\Omega} - \mathbf{\Omega}_k)\|^2$

4.4.2 Fisher's Linear Discriminant (FLD) & Fisherfaces

FLD is named after Robert Fisher who developed the technique for taxonomic classification in 1936 [62]. The key aspect to this technique is that it uses *labelled* data and seeks to minimise intra-class scatter and maximise inter-class scatter (rather than merely maximising total scatter as PCA does). In PCA, it is expected that clusters in the face space are formed for images of the same person. The concept of clusters is important in FLD, and ideally the cluster for a given label (or class) is compact (small intra-class scatter) and distant from other clusters (large inter-class scatter). This lends itself well to face recognition as faces are labelled as being that of a certain person.

Use in face recognition

Belhumeur *et al.* [17] exploit the observation that:

> '...for a Lambertian surface without shadowing, the images of a particular face lie in a 3D linear subspace of the high dimensional image space.'

This observation suggests that linear methods are appropriate for the problem and because the data is labelled, a *class-specific method* such as FLD is applicable. Where symbols have already been defined, the same notation will apply.

The total scatter is the sum of the inter-class (B - between) and intra-class (W - within) scatter:

$$\mathbf{S}_T = \mathbf{S}_B + \mathbf{S}_W \tag{4.6}$$

The intra-class scatter matrix $\mathbf{S}_W$ is defined as:

$$\mathbf{S}_W = \sum_{i=1}^{c} \sum_{x_j \in K_i} (x_j - \mathbf{u}_i)(x_j - \mathbf{u}_i)^T \tag{4.7}$$

The inter-class scatter matrix $\mathbf{S}_B$ is defined as:

$$\mathbf{S}_B = \sum_{i=1}^{c} B_i(\mathbf{u}_i - \mathbf{u})(\mathbf{u}_i - \mathbf{u})^T \tag{4.8}$$

where c is the number of labels, or in this case, identities, $\mathbf{u}_i$ is the mean image for a class i and B_i is the number of samples in class i and K is the set of all classes. The B_i term in (4.8) is often referred to as $1/B_i$. However when Eqn. (4.8) is substituted into Eqn. (4.6) together with the covariance matrix definition of $\mathbf{S}_T$ and worked through to give $\mathbf{S}_W$, the inverse term is cancelled out (for more information see [58]). The goal is to find the projection W_{opt} which maximises the inter-class measure $\mathbf{S}_B$ while minimising the intra-class measure $\mathbf{S}_W$, e.g. to create tight clusters of data with the same label. The way Belhumeur *et al.* do this is via the determinants of the scatter matrices:

$$\mathbf{W}_{opt} = \arg\max_{\mathbf{W}} \frac{\left|\mathbf{W}^T \mathbf{S}_B \mathbf{W}\right|}{\left|\mathbf{W}^T \mathbf{S}_W \mathbf{W}\right|} \tag{4.9}$$

$$= [\mathbf{e}_1, \mathbf{e}_2 \ldots \mathbf{e}_P] \tag{4.10}$$

where $\mathbf{e}_i$ are the generalized eigenvectors of $\mathbf{S}_B$ and $\mathbf{S}_W$ corresponding to the P largest generalized eigenvalues λ_i. The form of the generalized eigenvalue problem is (note the addition of matrix $\mathbf{G}$ compared with the standard eigen-problem

$\mathbf{Fe}=\lambda\mathbf{e}$):

$$\mathbf{Fe} = \lambda\mathbf{Ge} \tag{4.11}$$

Substituting into Eqn. (4.11) and rearranging allows the problem to be solved in the standard eigenproblem way (as long as the S_W matrix is non-singular):

$$\mathbf{S}_B\mathbf{e}_i = \lambda_i\mathbf{S}_W\mathbf{e}_i \tag{4.12}$$
$$\mathbf{S}_W^{-1}\mathbf{S}_B\mathbf{e}_i = \lambda_i\mathbf{e}_i \tag{4.13}$$

A major assumption here is that $\mathbf{S}_W$ is nonsingular. However, when $N>>c$ which is the case when used for face recognition (i.e. the number of pixels is far greater than the number of identities), the likelihood of the matrix being singular is extremely high (especially given the nature of images – the faces mean that there is very likely to be high correlation). To get around this Belhumeur *et al.* use PCA on the image set as a means of reducing the dimensionality first before projecting into FLD space. It is this additional step which allows them to identify their approach as *Fisherfaces*.

In their paper, Belhumeur *et al.* compare performance of PCA (eigenfaces) and FLD (Fisherfaces) and report far better results for Fisherfaces under varying illumination and expression (error rates half that of any other method and a third of eigenface). Interestingly they confirm a hypothesis that the first three principle components (which describe the most variation) are a direct result of illumination, and that by removing them, the projections will be more illumination invariant leading to a 10% increase in performance.

4.4.3 Independant Components Analysis (ICA)

ICA is a generalisation of PCA which is not forced to describe the variance along orthogonal axes and is sensitive to higher order image statistics. The most frequently cited use of the technique is to separate mixed signals, an example of which is the 'Cocktail Party Problem' where it can be used to separate individual speech signals from the general hubbub in the room.

Use in face recognition

ICA was first used for face recognition in [15] in which Bartlett *et al.* compared performance with PCA. If the assumption that a face is composed of many independent signals is correct, then ICA should provide excellent performance. However when comparisons between PCA, FLD and ICA are performed, the results are often ambiguous with different algorithms performing differently on particular data ([15, 57, 162]). Generally however, ICA provides better performance than PCA or FLD, but it does so at the cost of computational intensity.

In Eqn. (4.14) $\mathbf{x}$ is the face image which is assumed to consist of a set of separable signals ($\mathbf{s}$) multiplied by some mixing matrix ($\mathbf{D}$).

$$\mathbf{x} = \mathbf{Ds} \tag{4.14}$$

ICA attempts to estimate what those separable signals are assuming that the mixing matrix $\mathbf{D}$ is invertible to a separating matrix $\mathbf{V}$:

$$\mathbf{v} = \mathbf{Wx} = \mathbf{VDs} \tag{4.15}$$

Finding the separating matrix $\mathbf{V}$ is computationally expensive as it results from a function to maximize the independence of the resulting matrix $\mathbf{v}$ which can only be solved iteratively.

Two architectures are referred to when ICA is applied to faces. The first as-

sumes that face images are a linear combination of basis images combined by an unknown mixing matrix. Face images are variables and pixel intensities are observations. A face can be represented as a vector of coefficients of the basis images. In the second architecture, the pixels are the variables and the faces are the observations and instead of finding independent basis images, this architecture seeks statistically independent coefficients for the input data. Architecture I produces spatially localised basis images which clearly represent a certain facial feature e.g. nose, left and right eyebrows, lips, whereas Architecture II leads to more global descriptions which resemble eigenfaces.

Of the three papers mentioned above all report the different architectures perform differently depending on the data, but [57] and [162] both report Architecture II tends to outperform Architecture I.

Unlike the eigenface and Fisherface code, the ICA code used in this thesis was not developed by the author but was made publicly available by [14].

4.4.4 Pearson's Moment Correlation Coefficient (PMCC)

Pearson's product-moment correlation coefficient (PMCC) provides a measure of how correlated two sets of variables are from -1 (negatively correlated) to +1 (positively correlated) with 0 representing no correlation between the variables. It is the covariance of the samples divided by their standard deviations as shown in Eqn 4.16.

$$\rho = \frac{\sum_{i=1}^{M} \sum_{j=1}^{M} (x_i - \mathbf{u}_i)(x_j - \mathbf{u}_j)}{\sqrt{\sum_{i=1}^{M} (x_i - \mathbf{u}_i)^2} \sqrt{\sum_{j=1}^{M} (x_j - \mathbf{u}_j)^2}} \tag{4.16}$$

It is not commonly used as a classifier in the face recognition literature so cannot be considered a *benchmark* method in the same way as eigenfaces, Fisherfaces or ICA. It is however used throughout this thesis and for this reason its performance is included in the results section for comparison with other methods. Empirically it was found to produce similar results to the Fisherface algorithm but

is approximately eight times faster.

One apparent reason why it is not used in the literature is because although the Rank-1 recognition rate (i.e. selecting the closest gallery image to the probe) is excellent, it produces a very poor Receiver Operating Characteristic (ROC) as can be seen in the penultimate chapter (Fig. 8.10).

4.5 Data Representation

The above recognition methods have been applied to a range of modalities as there are a variety of data representations that can be used for face recognition. This section briefly describes the five representations extracted from the data for the benchmark tests presented here: texture, surface normals, depth map, shape index and Local Binary Patterns (LBP). Texture and depth map are chosen as they represent the most common 2D and 3D data representations, surface normals are chosen as they are intrinsic to PS and have been found to be the best representation for face recognition and LBP and shape index are also effective representations for face recognition [8, 73]. In each case the 2D matrix (or matrices in the case of surface normals) of data are reshaped into a vector for each session.

4.5.1 Texture (2D)

Texture, when used in the face recognition (and more broadly in computer vision) literature, refers to the pixel intensities of an image. Essentially it is a 2D monochrome representation similar to a black and white photograph. The experiments in this chapter use the estimated albedo at each pixel location which is calculated by the PS equation and represents the intensity of the object independent of any shading caused by 3D shape or illumination variation. Examples of albedo images can be seen in Fig. 4.8.

Figure 4.8: Examples of cropped 2D texture images (albedo).

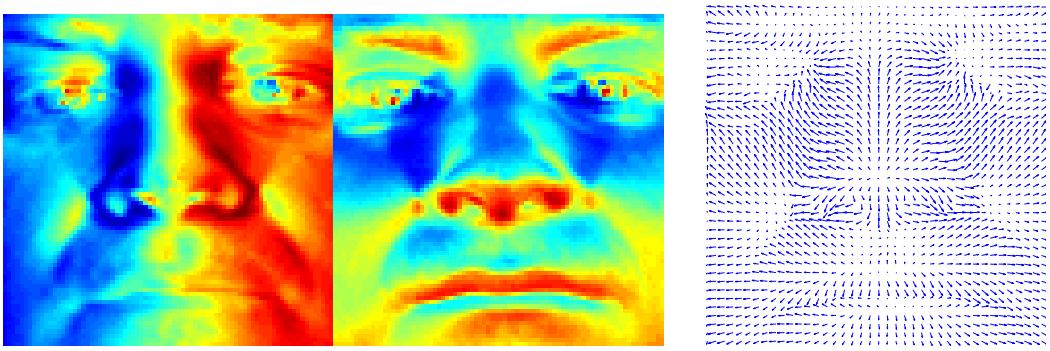
4.5.2 Surface normals

The surface normals are an intrinsic output of photometric stereo. At each pixel, the x, y and z components are estimated which can be visualized as a needle map. The vectors of components at each pixel are normalized to unit magnitude. In this way the z-component is made redundant as it can always be calculated from the remaining x and y components. In these experiments, the vector of y-components is concatenated onto the vector of x-components. An example of surface normal data for one subject can be seen in Fig. 4.9.

Different representations of the surface normals were also investigated as shown in Table. 4.3. The raw format as described above provides the highest performance and is used throughout the thesis.

4.5.3 Depth map

The depth map is a 2D representation of a 3D surface. Pixels with a high intensity are closer to the camera than those with a low intensity. The depth map



a b c

Figure 4.9: Surface normal representations. (a) Surface normal x-components. (b) Surface normal y-components. (c) A representation of the surface normals as a needle map.

	ℓ_1-norm	ℓ_2-norm	ℓ_∞-norm	Raw
% Correct	93	92.7	93	97.7

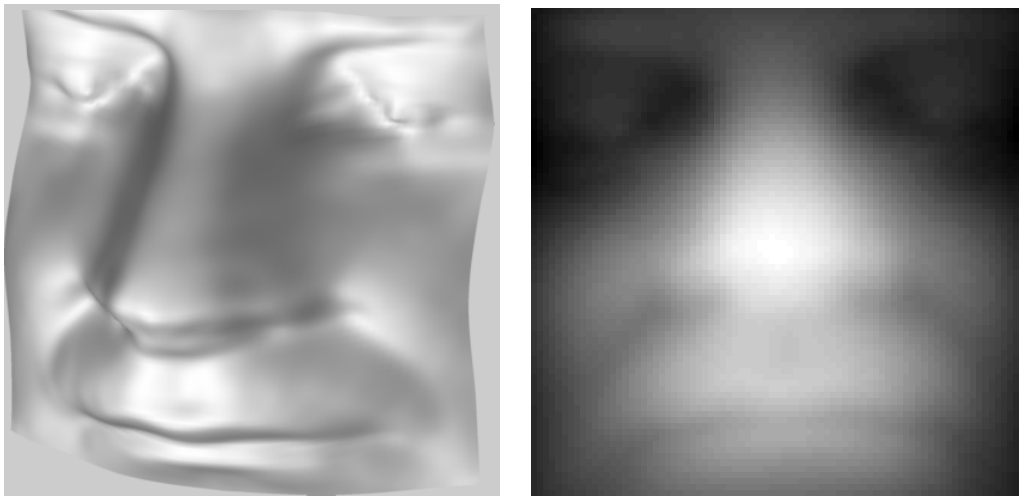
Table 4.3: Recognition performance using different norm-based representations on 1000 frontal expressionless sessions of 61 subjects classified with the Fisherface algorithm. For comparison the results using PMCC are very similar (93.2%, 94.9%, 93.2%, 96.8%).

is generated from integrating across the surface normals, in this case using the Frankot-Chellappa [64] algorithm, an example of which can be seen in Fig. 4.10.

4.5.4 Shape index

The shape index is a 2D representation of the curvature of a 3D surface. The shape index describes regions of a surface ranging from spherical cup (-1) to spherical cap (+1) with saddle at 0. A plane is undefined. A useful property of the shape index representation is that curvature is pose invariant. It is a seemingly under explored representation in the face recognition literature although Gökberk *et al.* [73] show it to outperform PCA and LDA of depth maps while performing worse than point cloud or surface normal representations and Colbry *et al.* [38] use it successfully for fiducial feature detection.

The shape index is calculated from the surface normals by where x, y, z are the components at each pixel location as in Eqn. 4.17.



a b

Figure 4.10: (a) An example of a surface generated by integrating across the normals and (b) the depth map.

$$s = \frac{2}{\pi} \arctan \frac{\left(\frac{\partial z}{\partial x}\right)_x + \left(\frac{\partial z}{\partial y}\right)_y}{\sqrt{\left(\left(\frac{\partial z}{\partial x}\right)_x - \left(\frac{\partial z}{\partial y}\right)_y\right)^2 + 4\left(\frac{\partial n}{\partial x}\right)_y \left(\frac{\partial z}{\partial y}\right)_x}} \tag{4.17}$$

4.5.5 Local Binary Patterns (LBP)

LBP were first used for face recognition by Ahonen *et al.* [7]. The approach seeks to recode a pixel in terms of the relative intensity of surrounding pixels – if a pixel has a higher intensity it is coded with a "1", if it is lower it is coded with a "0". This then gives a binary string of eight digits (one for each surrounding pixel) which can be converted to a decimal value which ranges from 0-255. Alternatively an extra step can be added which classifies the binary pattern as being either uniform or non-uniform. A uniform pattern is classified as one that has less than three bitwise transitions from 0-1 or vice versa. Ojala *et al.* [140] found that 90% of textures were made up of uniform LBP and for 8 bits there are 58 uniform patterns. Patterns with more than two transitions are termed as non-uniform and do not represent a robust feature. Those pixels with a non-uniform pattern are all coded with the same value. A diagram of the process can be seen in Fig. 4.12

Figure 4.11: Examples of shape index representations of four subjects. White is +1 (cap), black is -1 (cup).

and examples can be seen in Fig. 4.13.

There are many extensions to LBP such as the recent Multi-scale Extended LBP implementation which outperformed state-of-the-art approaches [92] and although it is one of the simplest methods, it has shown itself to be a very powerful tool for face recognition and texture classification in general as it is invariant to monotonic illumination changes and computationally efficient.

The most commonly used version is implemented as a benchmark here: classifying each pixel from 0-58 based on which uniform pattern the binary value given by the eight surrounding pixels belongs to, and one bin for all non-uniform patterns. The LBP's are created using the x and y-components of the surface normals.

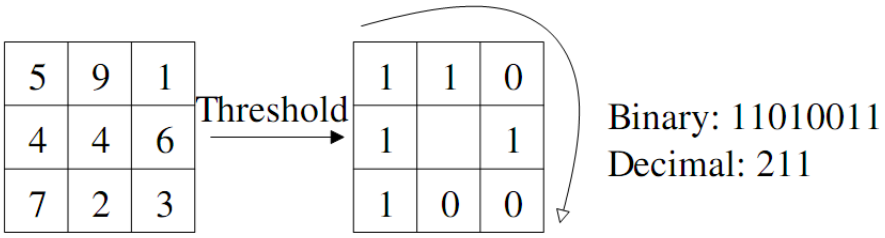


Figure 4.12: An illustration of the LBP process (reproduced from [7]). The pixel values are thresholded against the central pixel, and the resultant pattern is converted to binary, which in turn may then be converted to decimal.

4.5.6 Image Preprocessing

Data is cropped for the benchmark experiments as follows: the median anterior canthi and nose tip across all sessions are used for alignment via linear transforms. The aligned images are then cropped into a square region as shown in Fig. 4.14 to preserve main features of the face (eyes, nose, mouth), and exclude the forehead and chin regions which can frequently be occluded by hair.

In order to remove any artefacts from the Photoface database which are caused by the flashguns having different brightness, the greyscale intensity of the images is normalised. This is achieved by making the mean of each image the same

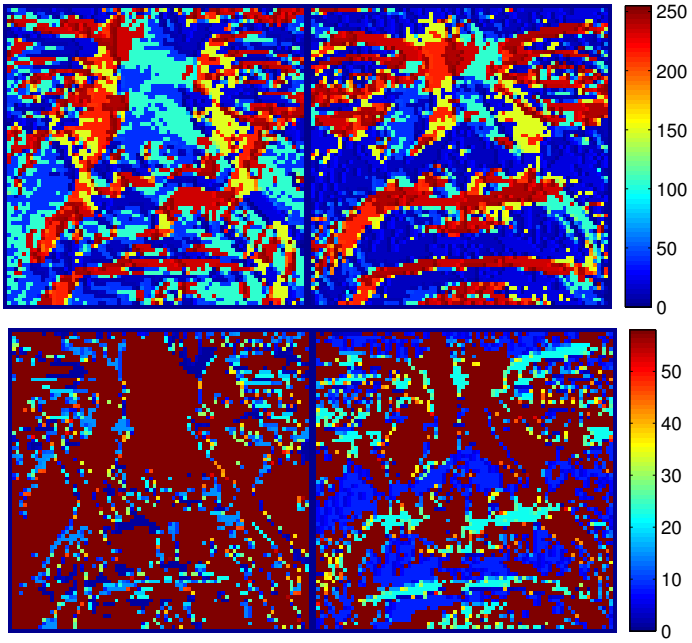


Figure 4.13: The top row shows an example of applying the LBP process to x and y components to give a 255 level LBP image. The bottom row shows the same results but with the commonly used 58 level image which employs the notion of uniform and non-uniform binary patterns.

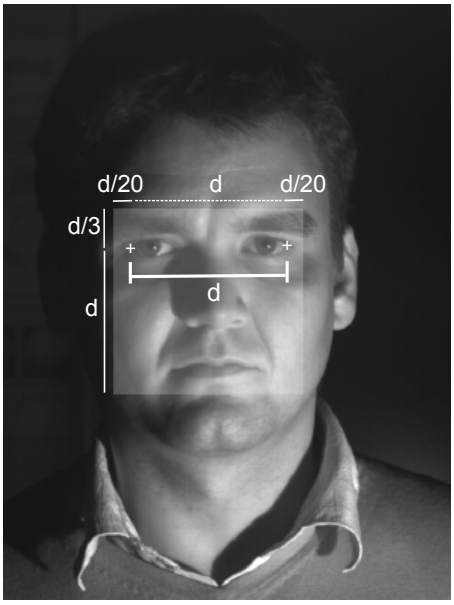


Figure 4.14: The cropped region of a face. The distance between the anterior canthi (d) is used to calculate the cropped region.

as the mean of all session images. Other normalisation techniques such as histogram equalisation, contrast limited adaptive histogram equalisation and linearly increasing the range of intensity values to a maximum 0-255 were investigated in terms of their effect on recognition performance, but no improvement in recognition performance was found.

The 2D images in the Photoface database are the estimated albedo images which are also aligned and cropped in the same way as the 3D data. Due to memory limitations, both the 2D and 3D data are then resized to 80$\times$80 px and are reshaped into a 6400-dimension and a 12800-dimension (x and y components of the surface normals are concatenated) vector respectively.

The data consists of 1000 sessions of 61 frontally facing subjects with neutral expression. These 61 subjects were selected as they had six or more sessions (e.g. sufficient to train on). The number of sessions per subject varies from six to 70 with the distribution shown in Fig. 4.15. A listing of the exact sessions used can be found in Appendix C, and is available electronically from the author. This is a subset of the whole Photoface database whose distribution can be seen in Fig. 4.2.

4.5.7 Experimental Paradigm

The method used to test recognition accuracy is the leave-one-out paradigm. This dictates that every session is used as a probe against a gallery of all other sessions once. There are therefore 999 classifications per session of which the percentage correctly identified is shown. The benefits of using the leave-one-out paradigm are that all sessions are used for probe and gallery images making it a far more efficient paradigm than using a static probe and gallery data partition. It is however far more computationally intensive as a new subspace must be calculated for each trial.

Combining the already computationally intensive Independent Components

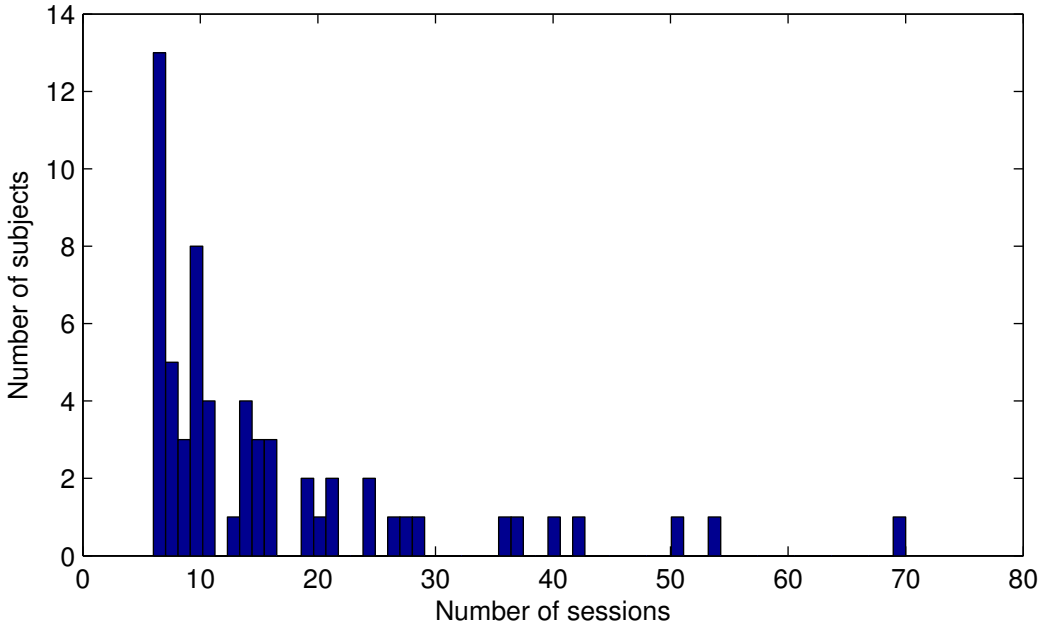


Figure 4.15: Distribution of sessions per subject for the 1000 session dataset - for example the first bar represents the fact that 13 subjects have six sessions.

Analysis with the leave-one-out paradigm would take an excessive length of time and therefore a more conventional data partition paradigm was used, in which the first session of each subject formed the probe partition and the rest were used as the gallery partition. This results in 61 classifications being performed – some reliability of the tests is sacrificed in order to perform them in a reasonable time. The Fisherface algorithm takes approximately twice as long as the eigenface algorithm due to the additional matrix inversion step to solve the general eigenproblem, but a standard desktop computer can process the 999 sessions in just over four hours. Once the face space has been calculated, the difference in classification times is negligible as this only involves calculating the Euclidean distance between vectors.

4.6 Results

Results for benchmarks can be seen in Table 4.4. In common with findings in the literature, it shows better performance of Fisherfaces over eigenfaces on the

Representation \ Algorithm	Eigenface	Fisherface	ICA (Arch1)	ICA (Arch2)	PMCC	Mean
2D (texture)	81.8	97.2	73.77	67.21	94.5	82.9
3D (depth map)	66.6	95.6	50.82	80.33	95.9	77.85
3D (surface normals)	96.7	97.7	93.44	95.08	96.8	95.94
Shape Index	96	91.1	98.36	98.36	97	96.16
LBP	98.4	73.4	95.08	95.08	96.7	91.73
Mean	87.90	91.00	82.29	87.21	96.18	88.92

Table 4.4: Benchmark results using a range of data formats and well known recognition algorithms. The number of components selected for the eigenface algorithm was set to account for 85% variance which meant that it varied for different data representations.

first 3 rows for 2D and 3D data. For texture and depth maps the improvement is large. Why there is a decrease in performance for shape index and LBP is unclear, but it would seem that not all representations benefit from discriminant analysis. The different architectures of ICA do not appear to confer any significant advantage over one another: Architecture I gives better performance for 2D texture, but worse results for depth map and surface normals, and they give the same performance for shape index and LBP. The very best performance is given by ICA on the shape index representation (98.36% for both architectures). However, some caution should be exercised with regards to the ICA results because of the far smaller number of comparisons being performed (61 vs 1000) due to time constraints. Running the ICA tests on shape index take about 100s which includes one calculation of the ICA subspace and then 61 nearest neighbour classifications. In comparison, running the Fisherface algorithm on surface normals takes around 800s to generate the Fisherface subspace and a nearest neighbour classification 1000 times. ICA is obviously a powerful tool, but for the purposes of this research is too slow to be used for dimensionality reduction.

If overall means are compared, using PMCC as a classifier outperforms the common benchmarks and is also the fastest algorithm. This is a surprising result as it is not used commonly in the literature. The reason behind this (as noted previously) is likely due to the fact that it cannot be used with any sort of threshold validation. However as a general measure of how similar two sessions of data are, it provides a very useful tool.

If ICA is taken out of the comparison for the above reasons, then raw surface normals provide a good representation in terms of eigenface and Fisherface performance combined with the fact that no additional processing is required. As a whole, Table 4.4 shows how the choice of data representation and analysis method can impact the recognition performance very clearly and one must be wary of statements that often appear in the literature regarding generalisations on methods e.g. “Fisherfaces are better than eigenfaces”.

4.7 Discussion

This chapter has described two databases which have been produced as part of this project. The Photoface database is one of the largest 3D databases currently available to the global research community. It is unique in that it is the first large scale PS database captured in unconstrained conditions in a realistic environment and it allows 2D, 3D and fusion approaches to be tested. With the bundled Photoface Query Tool application, it is simple to extract certain types of data, thereby making it an attractive addition to the existing available face databases. The smaller 3DE-VISIR database is unique in that it captures visible and NIR images of the subject under a variety of facial expressions. This database is used for experiments in classifying expressions in Chapter 7.

The last chapter demonstrated that the Photoface device was capable of accurate 3D capture and this chapter has shown that the data are suitable for face recognition in a range of representations and algorithms.

These methods are slow and intensive. Are there ways to improve the performance? The next two chapters look at methods that reduce the dimensionality of the data by direct and computationally efficient methods. Both take inspiration from the idiosyncrasies of the human visual system. The first looks at the effect of image resolution based on the fact that humans can recognise familiar faces at low resolutions (7 $\times$10px [167]) and the second takes the idea that caricatures may represent how humans store and/or retrieve faces and implements the idea using a percentile and a variance based approach.

Chapter 5

Human Inspired Low Resolution Face Recognition

Baseline results were presented in the previous chapter with a recognition rate of 97.7% recorded using the Fisherfaces algorithm on 1000 sessions of 61 people using a raw surface normal representation. This chapter presents findings with regards to reducing the amount of data by downscaling the images. The purpose of this to explore how different resolutions affect face recognition performance – if it is possible to reduce the amount of data in a direct way without greatly decreasing the recognition performance then this represents an efficient way to improve processing time and storage needs which in turn decreases cost and increases practicality. Along with computational and storage efficiency, another motivation is that it has been shown that humans can recognise low resolution faces of familiar people with a high degree of accuracy. There is an evolutional advantage to this ability in that it could aid in friend–or–foe type decision making when attempting to classify people in the distance. The further away that this can be judged, the longer the person has to raise an alarm or otherwise prepare. Another reason is that the structure of the face is made up of relatively large structures (nose, eyes and lips) which are configured in such a way that should

be robust to reasonably large reductions in resolution. This is true of faces in general, and the risk in overly reducing the resolution is that the information that allows discrimination between individuals will be lost. This chapter examines the effects of reducing the resolution on 2D and 3D data on recognition performance.

The main contribution of this chapter is to show that the recognition rate found in the previous chapter remains constant when the resolution is reduced to only 10×10 px for 3D surface normal data, and this is proven on both the Photoface database as well as a subset of the FRGC database. 2D data can also withstand considerable reduction in size to 15×15 px without seeing the recognition rate fall.

5.1 Introduction

One obvious method of reducing the amount of data is to downscale the images. A great deal of research has gone into increasing the resolution of poor quality images (super-resolution [11, 195], hallucinating [209]) by combining images or using statistical techniques to reproduce a more accurate representation of a face (e.g. from CCTV footage). By contrast, little research attempts to directly investigate resolution as a function of recognition rates on 3D data. Toderici *et al.* state that there is little to be gained from using high resolution images [178], Boom *et al.* state that the optimum face size is 32×32 px for registration and recognition [27], a view which is reinforced by a more recent study by Lui *et al.* who state that optimum face size lies between 32 and 64 pixels [123] for statistically based techniques, but that modern approaches can benefit from higher resolution. Czyz and Vandendorpe find that there is little effect of reducing resolution down to 16×16 px using LDA [46]. These experiments have used 2D images. Chang *et al.* use both 2D and 3D data and conclude that there is little effect of decreasing resolution down to 25% on 2D data and 50% on 3D of the original size (130×150px) [33] using PCA. The research suggests that relatively low resolutions give optimum

recognition (for the given recognition algorithms). These findings are conducive to the fact that the same appears to be true of human recognition [167], but there is little exploration or exhaustive work on the effect of resolution. With the obvious benefits associated with using less data, this chapter determines an optimum size for cropped faces in terms of face recognition performance. An explanation for the optimum size is then offered in terms of it preserving low spatial frequencies, backed by further experiments.

5.2 Methods and Data

This section details the datasets, preprocessing steps, and methods used in the experiments contained in this and the next chapter. In order to prove the validity of any findings the experiments are performed on the Photoface database as well as the well known FRGCv2.0 dataset. This allows comparison of the algorithms using the benchmark FRGC database and the richer feature set (higher resolution with direct calculation of surface normals) of the new Photoface database.

5.2.1 Data and Image Preprocessing

If techniques are only applied to the Photoface database, there is the danger that any reported results may only be applicable to that data and that any proposed general findings will not apply to other face recognition databases. Therefore the experiments in this and the following chapter, are performed on both Photoface and FRGCv2.0 data. A subset of each database is used which consists of 10 sessions each of 40 subjects. These numbers were chosen as 10 sessions provides a representative sample size, and there are only 40 subjects in the Photoface database with 10 or more expressionless frontal sessions. The FRGCv2.0 database categorises the emotion type of each scan. Those with the emotion type of 'BlankStare' have been selected from the database which is equivalent

to an expressionless face. The first 40 subjects with 10 or more sessions in the FRGCv2.0 database categorised as 'BlankStare' were used to keep the datasets the same size. A listing of the exact sessions used can be found in Appendices D and E, and are available electronically from the author.

The FRGCv2.0 dataset comes in point cloud format which is converted to a mesh via uniform sampling across facets. Spikes are removed by median smoothing (a 12×12 px sized filter gave the best recognition performance) and holes filled by interpolation. Normals are then estimated by differentiating the surface. The depth map images are all normalized to have a minimum value of 0.

Data is cropped for both databases as follows: the median anterior canthi and nose tip across all sessions are used for alignment via linear transforms; the aligned images are then cropped into a square region in the same way as for the benchmarking experiments in Chapter 4 as shown in Fig. 4.14 to preserve main features of the face (eyes, nose, mouth), and exclude the forehead and chin regions which can frequently be occluded by hair.

Our 2D experiments are based on data as follows: the accompanying colour image for each FRGCv2.0 scan is converted to greyscale, aligned and cropped in the same way as the 3D scan. The 2D images in the Photoface database are the estimated albedo images which are also aligned and cropped in the same way as the 3D data. Due to memory limitations, both the 2D and 3D data are then resized to 80×80 px and are reshaped into a 6400-dimension and a 12800-dimension (x and y components of the surface normals are concatenated) vector respectively.

5.2.2 Image Resizing

The effect of different resizing techniques on linear subsampling are investigated in terms of their effect on recognition as a function of resolution. Resizing is performed via the MATLAB® `imresize()` function using the default bicubic kernel type and with antialiasing enabled, as these settings were found to provide the

best performance.

Examples of faces at the different resolutions that are tested can be seen in Fig. 5.1.

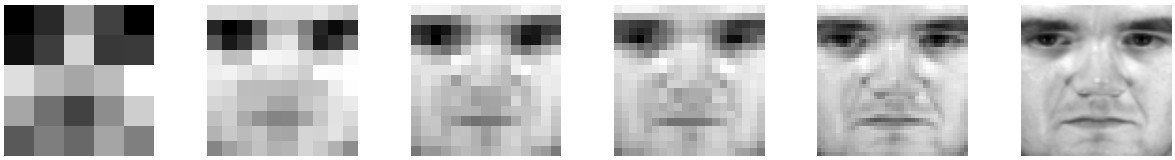


Figure 5.1: Examples of albedo faces at 5, 10, 15, 20, 40 and 80px resolutions

5.2.3 Experimental Paradigm

Again the leave-one-out paradigm is employed. This dictates that every session is used as a probe against a gallery of all other sessions once. There are therefore 399 classifications per condition of which the percentage correctly identified is shown.

As the focus of this research is more concerned with feature extraction efficiency, the actual choice of classifier is not as important. Therefore PMCC is used as a similarity measurement between a probe vector and the gallery vectors. The gallery session with the highest coefficient is regarded as a match. Experimentally, it was found that PMCC gives similar performance on baseline conditions to the Fisherface algorithm but is approximately eight times faster.

5.3 Results

The effect of image resolution on 2D and 3D recognition performance is shown in Fig. 5.2 with accompanying numeric data in Table 5.1. This clearly shows that reducing the resolution down to as little as 10×10 px does not affect the recognition performance on both 3D datasets. The same pattern appears in the 2D databases – the resolution can be reduced to 15×15 px before recognition performance degrades. In either case, these figures are lower than often reported

in the literature and certainly appears contrary to the notion that high resolution imagery is required for good face recognition.

For clarity, it is worth reiterating what was stated in the introductory chapter with regards to what is meant by '3D'. Unless otherwise stated, '3D' in these experiments refers to surface normal vector representations.

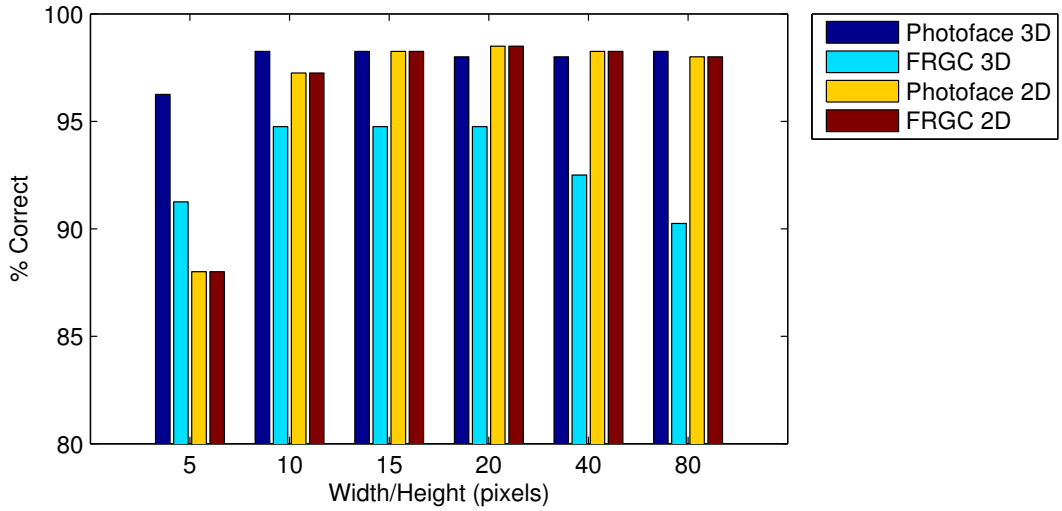


Figure 5.2: Performance of 2D (grey scale images from FRGC/albedo images from Photoface) and 3D (surface normal from both databases) face recognition on Photoface and FRGC data. No degradation in performance is produced even when resolutions are reduced to as little as 10×10 px for 3D data and 15×15 px for 2D data.

Data \ Resolution	5	10	15	20	40	80
Photoface 3D	96.25	98.25	98.25	98	98	98.25
FRGC 3D	91.25	94.75	94.75	94.75	92.5	90.25
Photoface 2D	88	97.25	98.25	98.5	98.25	98
FRGC 2D	88	97.25	98.25	98.5	98.25	98

Table 5.1: Table of results corresponding to Fig. 5.2. 3D results are for surface normal representations and Photoface 2D refers to results for albedo images and FRGC 2D refers to results for grey scale FRGC data.

For completeness, it should be noted that the following methods of resizing were also applied, but led to very little difference in performance compared with the baseline performances in Fig. 5.2 of 91.75% for FRGC and 97.3% for Photoface data:

1. resize depth map and recompute normals for FRGC (83.75%)
2. ensure unit length of normal vectors after resizing operation for Photoface (97.1%)
3. resize gradient fields (tilt and slant – $\frac{\partial z}{\partial x}$, $\frac{\partial z}{\partial y}$) and recompute normals (97%)
4. resize (raw) images and then perform PS to estimate normals (97%)

The trend is clearly shown that low resolution imagery is not detrimental to face recognition. It occurs in both datasets so there is no reason to believe that this is due to something unique in the Photoface database and occurs for both 2D and 3D images. In order to ensure that the effect is not related to using PMCC results for Photoface database are shown in Table 5.2 using Fisherfaces over the optimal range for 3D data – in this instance across the full set of 1000 sessions and 61 subjects. This confirms the previous finding that reducing resolution to around 10×10 px produces no degradation in recognition rate (the best result is 100% using Fisherfaces at 7×7 px).

5.4 Discussion

It would appear that face recognition on surface normal and 2D data can occur at very low resolutions which is consistent with human findings but not with previous research in the sense that a 10×10 px representation is very much smaller than reported in the literature.

By downsampling the images, data is discarded and only the dominant structures will remain. These structures will likely coincide with low frequency data while the high frequency data will be lost due to smoothing effects. It has long been known that low spatial frequencies play a role in human face recognition [206] so this is next investigated to see whether this is the cause of the improved recognition rates by comparing various methods. These methods are (examples of which can be seen in Fig. 5.3):

Classifier \ Resolution (px)	2	3	4	5	6	7	8	9	10	11	12	13	14	15
PMCC	43.6	83	93.7	95.2	95.9	96.9	97.1	97.1	97.4	97.4	97.1	97.2	97.2	97.1
Fisherface	47.8	89.1	98.4	98.9	99.5	100	99.9	99.7	99.9	99.7	99.9	99.9	99.6	99.8

Table 5.2: Recognition rates across the optimal resolution range (2-15 px) using the PMCC and Fisherface algorithm for comparison on surface normal data of the full Photoface frontal expressionless dataset (1000 sessions of 61 subjects).

1. mean filter: block sizes of 1, 3, 5, 10, 15 and 20 pixels

2. low pass filtering: increasing cut-off radii in the Fourier domain to include frequencies of 1, 2, 3, 4, 5 to 40 px radius (0.0125 Cycles Per Pixel (cpp), 0.0250, 0.0375, 0.05, 0.0625 to 0.5 cpp in increments of 0.0625 cpp). This allows us to see the effect of adding increasingly higher frequencies to the images.

3. high-pass filtering: the reverse of the above to see the effect of removing lower frequencies from the image.

4. resizing the low resolution image back to its original size to ensure that the recognition is a real function of spatial frequency rather than image size per se.

The results are shown in Fig. 5.3. The graph of mean filtering indicates an optimal point where high frequencies are removed before the face is over smoothed and performance decreases. Interestingly the low-pass results show that the addition of high frequency data leads to poorer performance, and the high-pass results show that high frequency data on its own, is insufficient for good recognition performance. Fig. 5.4 compares the best results obtained using each method, demonstrating that resizing the image leads to the best performance with the low-pass filter coming second. While all of the proposed comparisons lead to better performance than that of the control, none lead to better performance than the resizing of the image to 10×10 px. This has the benefits of being a direct method compared to better known wavelet approaches/FFT methods of extracting frequency bands and also leads to a smaller representation which may have implications for storage and will certainly improve processing efficiency.

The examples shown in Fig. 5.5 represent the processed y-component data which gives the best recognition performance for the different methods. There is sufficient similarity between the first three to suggest that it is the low spatial

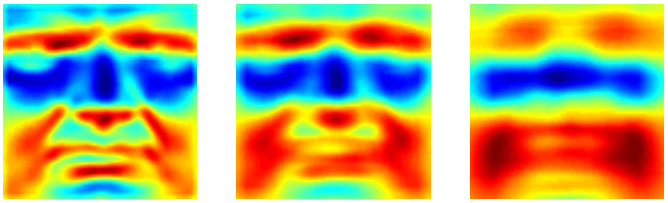
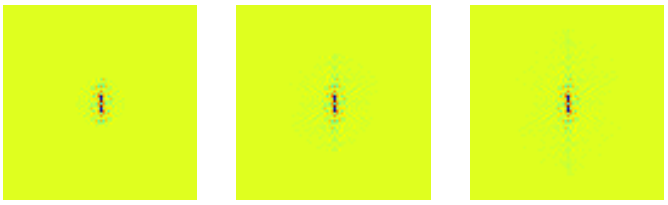
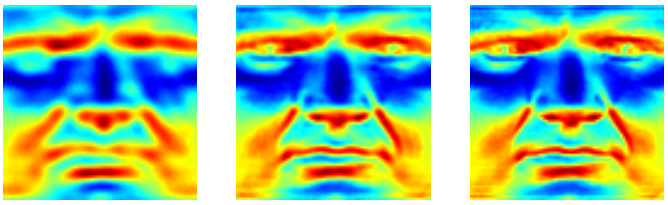
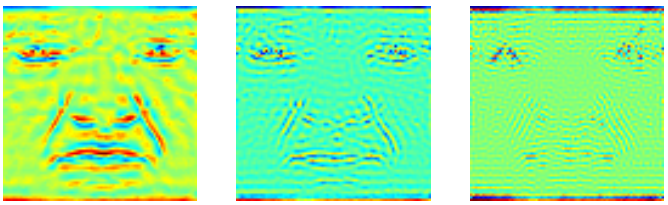
Manipulation		Examples
Mean smoothing		
Low-pass	Fourier space	
	y-component	
High Pass	Fourier space	
	y-component	

Table 5.3: Examples of smoothing using a mean filter of different block sizes at 1, 5, 20 px, low-pass filtering using FFT at 10, 20 and 30 px radius, and high-pass filtering at 10, 20 and 30 px radius on y-components.

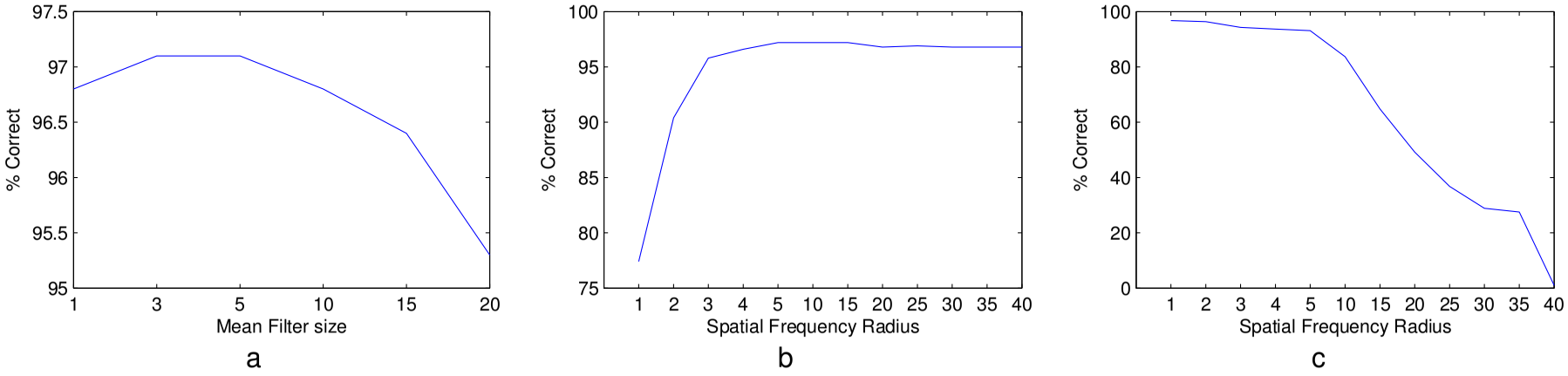


Figure 5.3: Performance effects of smoothing and spatial frequency filtering. Recognition results for (a) mean smoothing, (b)low-pass filtering, and (c) high-pass filtering for a range of values. (a) shows that a mean filter size of between 3 and 5 px produce the best results. (b) shows increasing performance up to around a radius of four after which the addition of higher frequencies has little effect. (c) shows that only using high frequency data (radius four and above) has a negative effect on recognition performance.

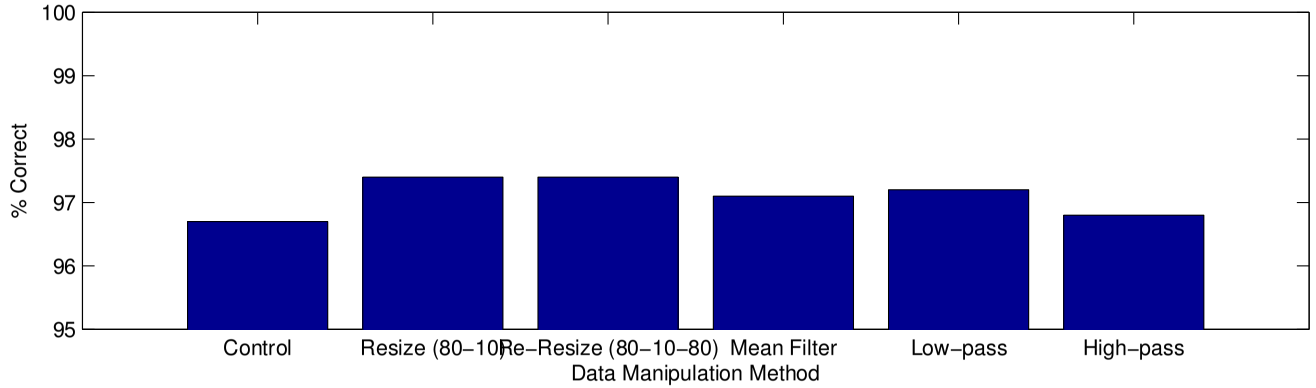


Figure 5.4: The bar plot shows the best result obtained from each method. The 'Control' result is the benchmark result for 80×80 px for comparison on the same dataset using PMCC.

frequency data which leads to improved face recognition performance and that by resizing the image effectively preserves the information contained in these frequencies with the additional benefit of reduced dimensionality. The fourth example (high pass) contains the higher spatial frequencies as well as the lower and as such does not perform as well as the others which have only the lower spatial frequencies.

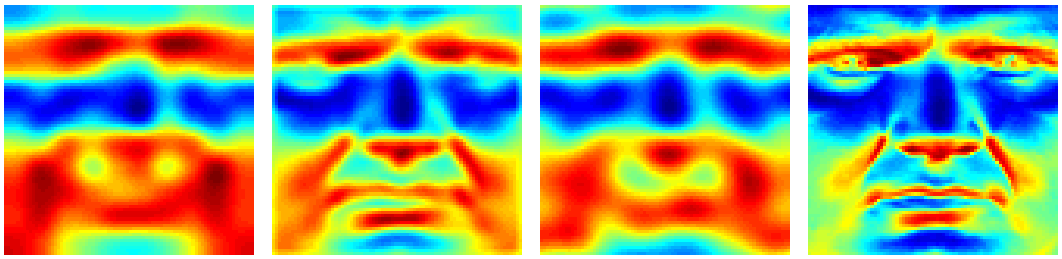


Figure 5.5: An example processed image for each method with the parameter which gave the optimum recognition result. From left to right – resizing the 10×10 image back to 80×80, mean filtering, low-pass filter, high-pass filter. Note that the low-pass filter looks the most similar to the resized image (the third image compared with the first); the low pass filtering also produced the closest recognition performance to image resizing suggesting that the low spatial frequencies play an important role and may provide an explanation as to why image resizing works so well.

Dakin *et al.* [48] present a novel representation of faces based around the concept of a bar code which they suggest explains some of the idiosyncrasies of human face recognition (such as spatial and contrast polarity inversion) by offering a theory of low-level vision. Using Gabor filters in the Fourier domain, they extracted the low-level horizontal features from 2D face photographs. The outputs are similar to those found in this chapter. They then go on to produce a 1D bar code representation through the vertical mid-line of the face which they propose tolerant to certain degradations (scale, distortion and noise) giving comparisons with the HVS. They don't test the method empirically and testing this and an extension incorporating the above work on 3D using surface normals could be an area for further research.

It is interesting that the psychophysical experiments into face recognition have often shown the lower spatial frequencies play a more important role in face

recognition than the higher frequencies (see Section. 2.2), but to date there has been little research into whether this effect can be exploited by automated systems. This chapter has demonstrated the effect on two datasets of 2D and 3D data and is backed by psychophysical findings that humans use the lower spatial frequencies for face recognition. This has implications for all face recognition systems, although it is unclear whether it would be true of very large databases (although this is arguably true of all face recognition systems to date) where it may be necessary to incorporate some slightly higher frequency data to boost the discriminability of the data. Anecdotally this is the sort of information which humans may use to distinguish between identical twins. While the overall geometry of the faces will be close to identical, small scale features amongst the twins will be different e.g. moles, wrinkles, scars and these allow for discrimination.

The earliest psychophysical experiments suggested an optimum resolution of 2.5 cycles per face [86] and a more recent experiment suggested that between 5 and 11 cycles per face give the best human performance [145]. This compares favourably with the frequency of 0.0625 cpp or 5 cycles per face found in this research. Given that the data used in these experiments is tightly cropped compared with the psychophysical stimuli, the experiments performed here fit reasonably well with the findings in humans. Performance for their experiments dropped off after 11 cycles per face which corresponds with the region around 0.1375 cpp or an 11px radius in the high-pass filter results and again matches with findings here with a sharp drop at 10px radius.

If it is possible to automatically recognise low resolution images accurately then this could have applications for CCTV footage. However, along with the low resolution properties common to CCTV footage, there are also commonly other artefacts due to compression and poor illumination which would need addressing. However the common assumption that one needs to build up a high resolution face image in some way would not be necessary. For static PS face

capture devices such as Photoface, this chapter has shown that employing very low resolutions (of the order of 10×10 px) enables optimal recognition rates and as such offers an efficient way to reduce the dimensionality by exploiting low spatial frequency data. The fact that the spatial frequencies involved at these resolutions are in the same range as those used by humans indicates that using findings about the HVS is beneficial. This notion is explored further in the next chapter in which two methods of dimensionality reduction inspired by caricaturing based outlying data values or by selecting pixel locations with high variance are presented.

Chapter 6

Caricature Inspired Variance and Percentile Based Dimensionality Reduction

The previous chapter found a successful method of dimensionality reduction by lowering the resolution to 10×10 px without introducing a corresponding drop in recognition rate. Interestingly this corresponds to a spatial frequency in the optimal range at which it is suggested that humans recognise faces. The motivation for this chapter comes from the observation that humans are able to recognise a caricatured face as well or better than a veridical image [154]. It is suggested that this tells us that humans somehow use a caricatured representation of faces either for retrieval or storage (or both) [128]. If only the caricatured regions of a face (e.g. those that deviate sufficiently from the norm) need to be processed then this has useful implications for data reduction. In this thesis, promising results into using caricaturing initially came from combining low resolution imagery with exaggerating a normal using the method described in [125]. If the 5% of pixels which deviate most from the norm are exaggerated in this way, recognition performance is 96.6% – a 0.2% improvement over baseline. This is a promis-

ing finding which also justifies further exploration of using caricatures to improve automatic face recognition.

This chapter presents two direct methods inspired by caricatures and implemented using a more statistical approach hypothesized by Unnikrishnan [182] using percentile and variance based techniques.

6.1 Introduction

Ultimately, dimension reduction techniques based on a percentile and variance based inclusion principle (to exclude 90% of the data) are compared with a baseline condition containing all pixels. In doing so, a disproportionately small drop in performance is seen. The use of these direct techniques as preliminary step to reduce the search space of potential matches is explored, which can then be analysed using a more rigorous but time consuming algorithm.

In this chapter, the following contributions for both the FRGCv2.0 database [147] and the Photoface database are demonstrated:

1. Empirical support for a pixel-based interpretation of Unnikrishnan's hypothesis that outlying data contains disproportionately more discriminatory information than other data.

2. The exclusive use of just 10% of the data (chosen to be those pixel locations with the greatest variance) is sufficient to maintain recognition rates to within 10% of those rates that include all of the data.

3. By combining the resolution findings from the previous chapter with variance based pixel selection, a recognition accuracy of 96.25% for 40 subjects using only 61 dimensions (pixels) is achieved. This compares to 98% when the full 80$\times$80 px resolution is used on all data.

4. The combination of methods allows for efficient search space reduction as

a preliminary step to before a more intensive and accurate method is used (Elastic Graph Matching in this case).

Caricaturing essentially enhances those facial features that are unusual or deviate sufficiently from the norm. It has been shown that humans are better able to recognise a caricature than they are the veridical image [128, 154]. This finding is interesting as caricaturing is adding distortion or noise to an image. However this noise aids human recognition and this, in turn, provides insights into the storage or retrieval mechanism used by the human brain.

Unnikrishnan [182] conceptualises an approach similar to face caricatures for human recognition. In this approach, only those features which deviate from the norm by more than a threshold are used to uniquely describe a face. Unnikrishnan suggests using those features whose deviations lie below the 5^{th} percentile and above the 95^{th} percentile, thereby discarding 90% of the data. Unnikrishnan provides no empirical evidence in support of his hypothesis, so this theory is experimentally tested. This is done in two ways: the first directly tests his theory, finding the thresholds for each pixel which represent the 5^{th} and 95^{th} percentile values and only including those pixels in each scan which lie outside them (outliers). The second is loosely based on Unnikrishnan's idea, and looks at the variance across the whole database to calculate the pixel locations with the largest variance. Only the pixels at these locations are then used for recognition. It should be noted that the experiments here test a pixel-based interpretation of Unnikrishnan's hypothesis rather than the actual one which he published which relied on higher level features such as nose length or hair colour.

It is envisaged that this approach could be best used as a first step in the recognition process to reduce the search space of possible matches before more intensive and time consuming algorithms are employed. It is obviously important for the proposed algorithm to be sufficiently accurate not to exclude matching sessions; if this is not the case the secondary algorithm will clearly be unable to

find the match. The effectiveness of this approach is demonstrated in combination with a variant of Elastic Graph Matching (EGM) with morphological features [201, 202, 200].

6.2 Methods and Data

The analysis is performed on the same dataset as in the previous chapter in Section 5.2.1. Briefly this consists of 40 subjects with 10 sessions each, with neutral expression captured using Photoface and an equivalent subset from the FRGCv2.0 dataset.

6.2.1 Calculating outliers and variance

The thresholds for each pixel location are calculated which represent the 5^{th} and 95^{th} percentile values. Of interest is the norm across the whole dataset for each pixel location rather than the norm for each image. For the 2D images, percentile values are calculated for the greyscale intensity value for each pixel location. There are 400 sessions, so there are 400 values for each pixel from which the percentile thresholds are calculated. The same process is performed for 3D surface normal data, giving x and y surface normal component thresholds for each pixel. Pixels which have a value between the 5^{th} and 95^{th} percentile are discarded, leaving only the 10% outlying data. This shall be referred to as the "percentile inclusion criterion".

Examples can be seen in Fig. 6.1. An important assumption is made when performing recognition using these representations that very similar (ideally identical) groups of pixels will be selected for the same person across different sessions. This appears to be a reasonable assumption providing the sessions are aligned sufficiently well. The vector size remains unchanged using the ?percentile inclusion criterion? as the vector size for those pixels which are included can be

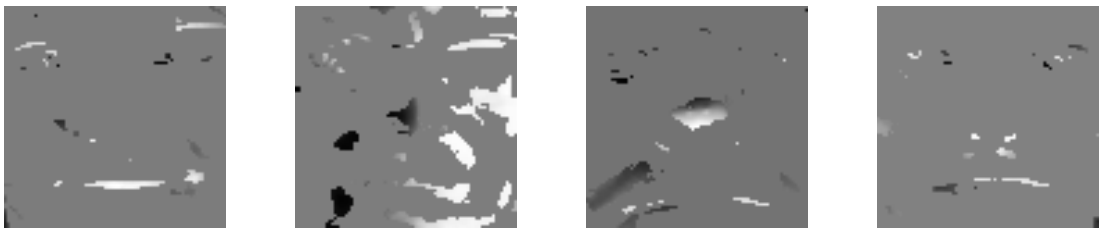


Figure 6.1: Examples of the y-components of the surface normals that have values outside the 5^{th} and 95^{th} percentiles for four subjects. A sparse representation of features is generated.

different for different subjects and even sessions. The recognition algorithms employed here require the vectors to be of equal size and it is therefore not possible to remove them, it is only possible to zero them. This is discussed further and alternatives are presented in Section 6.4.

The above method extracts the least common data from each session and that is what is used for recognition. Alternatively, the variance at each pixel location can be used as a measure of discriminatory power. If a pixel shows a large variance across the dataset, then this might make it useful for recognition (assuming that variance within the class or subject is small). Therefore the standard deviation of each pixel is calculated over all the sessions. Whether or not a particular pixel location is used in recognition depends on whether or not the variance is above a pre-determined threshold. Examples of the use of different thresholds are shown in Fig. 6.2. This is referred to as the "variance inclusion criterion".

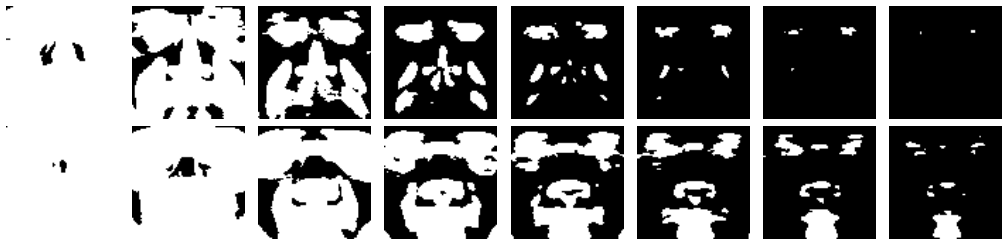


Figure 6.2: Examples of the regions which remain for x (top row) and y-components (bottom row) as the threshold variance is increased from left to right. White regions are retained and black regions are discarded.

6.2.2 Recognition algorithm

Again the same recognition algorithm (PMCC) as outlined in Section 5.2.3 is used for these experiments.

6.3 Results

6.3.1 Dimensionality reduction via the percentile inclusion criterion

Unnikrishnan's theory states that reliable performance should result from using only the data which lies outside the 5^{th} and 95^{th} percentiles [182]. Table. 6.1 shows recognition rates on 2D and 3D data using both all data and the outliers only. Note in particular that, for the 3D surface normal data, the rates drop by under 10% when using outlier data only. This effect seems limited to the surface normal data and is not seen in either the 2D or depth map data. Also included are results from a fusion technique using the Photoface surface normal data combined with the albedo image. There is a small decrease in baseline performance and using only the outlying data leads to a severe decrease of about 34%.

		Baseline	Outliers	Diff
2D	FRGC	90	73.75	16.26
	Photoface	98	64	34
3D	FRGC Surface normals	90.25	84.25	6
	FRGC Depth map	71.5	23.25	48.25
	Photoface	98.25	89.25	9
Fusion	Photoface 2D + 3D	97	63.25	33.75

Table 6.1: Percentile inclusion criterion results. Baseline (all pixels) versus outlier (10% of pixels) performance (% correct). The surface normal results (which show the effect best) are highlighted. Proportionately less performance is lost compared to the number of pixels for the 3D data. The effect is not present for the 2D data.

Fig. 6.3 shows a plot of recognition rate as a function of which percentile range is used for recognition on 3D Photoface and FRGC data. It should be noted that

similar patterns of results were found for all datasets (2D, 3D and FRGC). As predicted, the figure shows that the best recognition performance is obtained using the most outlying percentiles. As expected also, the recognition rate reduces as the percentile ranges used tend toward the inliers. However, for the most inlying data of all (i.e. percentiles 45–55), there is an unexpected increase in performance. Contrary to Unnikrishnan's theory, this implies that there is discriminative data that is useful for face recognition in the most common data as well as the most outlying.

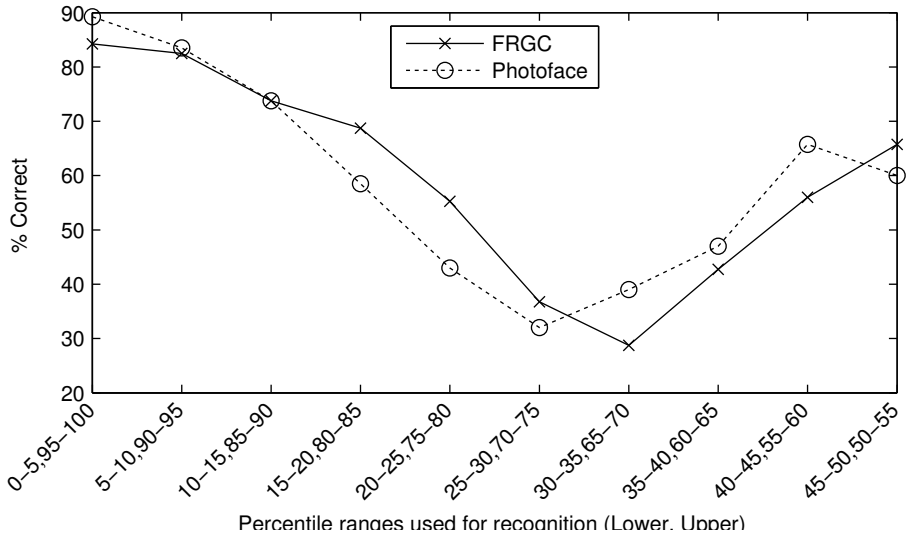


Figure 6.3: Recognition performance using pairs of percentile ranges for 3D data with PMCC. As predicted by Unnikrishnan, the outlying percentiles give the best performance. The rise in performance near the central percentiles however, is unpredicted and unexplained.

In a related experiment, single 5% ranges of data were used for recognition (i.e. $[0^{th} - 5^{th}], [5^{th} - 10^{th}]$ etc.) as shown in Fig. 6.4. It can be seen that the increase in recognition performance for the most inlying data is not replicated implying that it is the combination of the ranges which leads to the increase around the $45^{th} - 55^{th}$ percentiles. This difference indicates that there is an interdependence between the two percentile ranges.

However, performance increases by combining ranges are not always observed. For example, the $25 - 30^{th}$ and $70 - 75^{th}$ percentiles for the FRGCv2.0

data individually give a performance around the 50% mark in Fig. 6.4 , but when combined, the performance drops to around 40% in Fig. 6.3.

The reason for the decrease in performance is unclear, and a fuller comparison of the additive effect of different percentile regions on the Photoface dataset can be seen in Fig. 6.5 which shows the results of a large variety of range combinations. This figure shows a turquoise region with a red border. The turquoise region indicates relatively poor performance and highlights those percentiles which appear not have any useful discriminatory data (or at least data which should not be combined with other percentile data). The red areas show good performance – the top right (or bottom left) corner shows the performance corresponding to outliers as suggested by Unnikrishnan, but what can also be seen is that combining the two percentiles ranges at either end of the scale (eg $0 - 5^{th}$ and $5 - 10^{th}$, or $90 - 95^{th}$ and $95 - 100^{th}$) also results in very good performance. This clearly shows the interdependence between the two ranges of percentiles, although the underlying factor(s) is unknown. It would be useful to explore this relationship further, perhaps by exhaustively combining different length ranges, and perhaps even different numbers of discrete ranges.

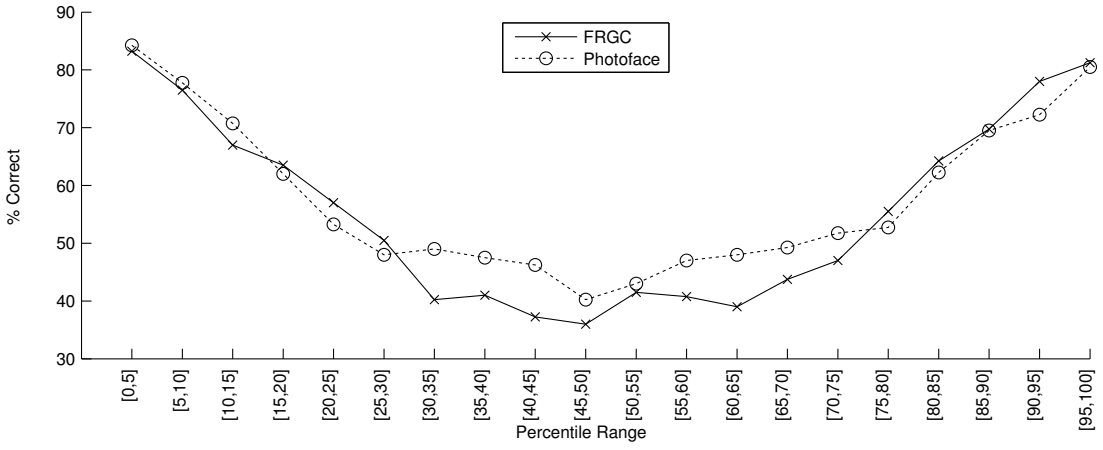


Figure 6.4: As Unnikrishnan's theory would predict, FRGC and Photoface data show a marked symmetry across ranges of percentiles, without the rise in performance across central ranges which occurs with pairs of percentile ranges. Recognition is performed using PMCC.

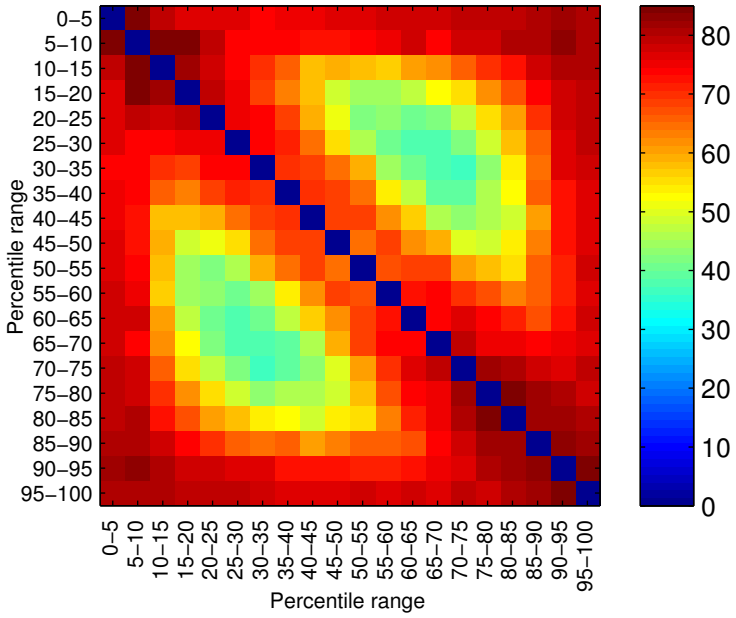


Figure 6.5: Confusion matrix for combinations of all percentile range pairs. The effect of combining two different percentile ranges on recognition performance (% Correct) on Photoface 3D data. Fig. 6.3 corresponds to the line starting from the bottom left corner to the (45-50, 50-55) pixel (near the centre of the plot). The turquoise areas either side of the diagonal show combinations of that have a negative effect on recognition, while the red border areas show combinations with good recognition performance.

6.3.2 Dimensionality reduction via the variance inclusion criterion

One problem with the above method is that the outlying points tend to be scattered across different parts of the images, making inter- and intra-comparisons between individuals somewhat unstructured. For the next method therefore, the same pixel locations are used for all images. Instead of using the percentiles as an inclusion criterion, the variance of a particular pixel across all subjects as explained in Sec. 6.2.1 is used.

Fig. 6.6 shows plots combining the number of pixels which remain as those with least variance (bar plot) are removed against the recognition performance (line plot). It is apparent that close to optimal performance can be achieved while losing a large proportion of the pixels. Approximately 75% of the least varying pixels can be discarded and a corresponding drop of less than 10% in recognition performance is observed on the FRGC data. Indeed, for Photoface data specifically, only lose a few percent are lost.

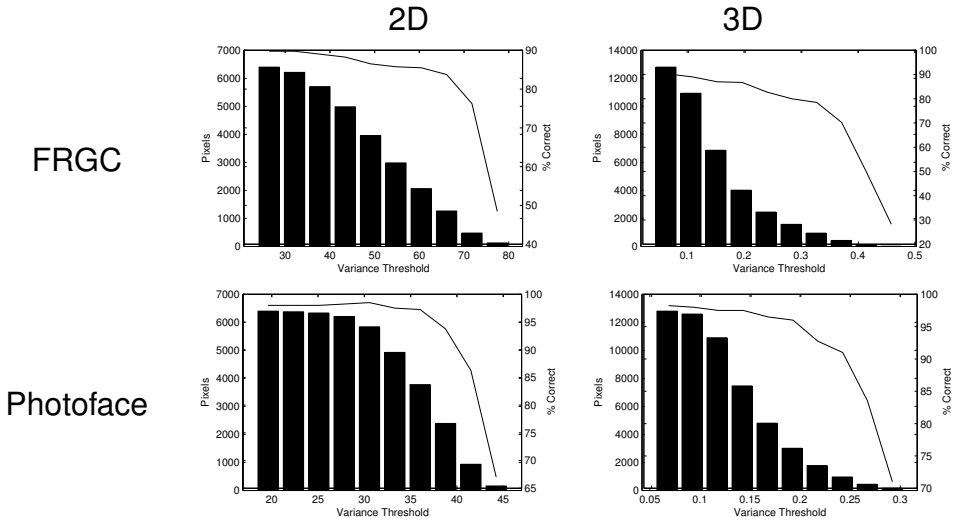


Figure 6.6: Variance inclusion criterion results. Recognition (line plot) as a function of retained pixels (bar chart) using PMCC. The pattern is shown in both sets of data (FRGC on the top row and Photoface on the bottom). 2D (grayscale for FRGC and albedo for Photoface) on the left, and surface normal data is shown on the right. The fact that there is not a linear relationship between the number of pixels and the recognition performance indicates that we are selectively retaining pixels with more discriminatory power.

When only 10% of pixels are retained by discarding the data that varies the least, reasonably high recognition rates are maintained for both the percentile and variance inclusion criteria e.g. 89% and 92% for the Photoface dataset. However the processing time for the percentile based method is far greater: 800s compared to 100s to calculate the outliers/most varying pixels and perform 400 classifications. In a practical setting the processing time is far less as there is no requirement to calculate the outliers/most varying pixels for each classification; they can be calculated offline.

	Percentiles	Variance
FRGC	84.25%	$\approx 79\%$
Photoface	89%	$\approx 92\%$
Processing time	800.64s	180.95s

Table 6.2: A comparison of recognition performance using percentiles and variance methods to select the most discriminatory 10% of the data. The processing time includes the calculation of the outliers/most varying pixels and 400 classifications

The processing time improvement for the variance approach is due to having decreased the vector size by 90%. This compares to 973.09s for the equivalent Fisherface analysis which provides an accuracy of 99.5% so both methods offer considerable time savings at a small cost to accuracy.

By combining the resolution findings of Chapter 5 (using a low resolution of 10$\times$10 px) with the variance method above, comparable recognition performance to an 80$\times$80 px image can be achieved using only 64 pixels for FRGC data and 61 pixels for Photoface data. Recognition rates of 87.75% and 96.25% are recorded (a loss of only 7% and 2% respectively from resizing the images to 10$\times$10 px as shown in Table 5.1). The processing time is also reduced to 10.5s for variance analysis and 400 classifications. The same analysis using the Fisherface algorithm takes 118s and achieves a comparable rate of 89.25% on the FRGC data.

6.3.3 Using Variance Inclusion to reduce search space

The two proposed methods show that data can be selected which contains disproportionately more discriminatory data than that which is discarded. Both give similar performance, but the variance-based approach has the additional benefit of being faster. The use of the variance inclusion criterion as a precursor to a more intensive recognition algorithm (Elastic Graph Matching (EGM)) as a means of reducing the search space is explored next.

EGM is a well known technique in face recognition with proven performance [148] which comes at the expense of complexity. One of the best known implementations is the Elastic Bunch Graph Map [191] where faces are represented as a graph of nodes. Each node is situated on a fiducial feature and consists of a number of Gabor coefficients at different orientations and scales. The Gabor coefficients and the distance vectors between nodes are used as biometrics. Here, a version of EGM that incorporates morphological features instead of Gabor filters is employed[1], full details of which can be found in [201, 202, 200]. First the images are resized to 10×10 px, and the variance inclusion criterion is used to reduce the size to 20% of the original resulting in 18 dimensions. Then PMCC is employed, of which the n closest matches are chosen and are then each compared to the probe using EGM. The method is applied to a subset of the Photoface Database consisting of 1000 frontal sessions of 61 subjects with no expression. The results are shown in Table 6.3.

Each EGM comparison takes about 1.5 seconds so the total time to compare one probe to all 939 gallery sessions is $1.5\times939 = 1409$s or just under 25 minutes. The time saved in reducing the number of comparisons is therefore important and the time is reduced to about 2.5 minutes by only performing EGM comparisons against the top 100 matches.

Alternatively the number of pixels which are retained during the variance anal-

[1] binaries supplied by the author S. Zafeiriou, Imperial College London.

Method	Surface Normals	Depth Map	Approx. Proc. time
Proposed alg. (18px)	85.4%	44.3%	0.12s
Proposed alg. (18px) + EGM ($n = 100$)	95.08%	91.8%	150s
Proposed alg. (18px) + EGM ($n = 10$)	91.8%	73.74%	15s
Proposed alg. (61px) + EGM ($n = 10$)	95.08%	95.08%	15s
EGM	100%	100%	1409s

Table 6.3: Recognition performance on surface normals and depth maps using resizing to 10×10 px and applying the variance inclusion criterion to reduce the search space before using EGM. By using EGM it is possible to dramatically increase the performance of the proposed algorithm while minimizing and increase in processing time as a result of using a more intensive algorithm. The Approx Proc. time provides the time for one probe to be recognized. The dataset was 1000 frontal expressionless Photoface sessions.

ysis can be increased to a similar amount that gave good performance in the previous section (i.e. 61px). By doing this performance when $n = 10$ is at a par with performance when $n = 100$ with 20% of pixels retained.

By examining the rank-n performance plots (Fig. 6.7) it is possible to get an indication of a suitable number of dimensions and the approximate number of candidates to select for the subsequent EGM step. It is clear that 18px is too few as 100% accuracy is not achieved even at rank-200 performance, which would mean that the correct gallery image would not be selected amongst the set used by EGM. However the performance using 61px is 100% at rank-100 which suggests that these would be suitable parameters. The EGM algorithm would only have to run a tenth of the number of comparisons and still achieve 100% accuracy.

6.4 Discussion

The results show that using direct techniques based on using outlying data or the most varying pixel locations are effective at reducing the dimensionality of the data without a corresponding drop in recognition performance. Here, the resolution effects explored in the previous chapter are combined with the variance inclusion criterion.

Computationally efficient methods using variance analysis and image resiz-

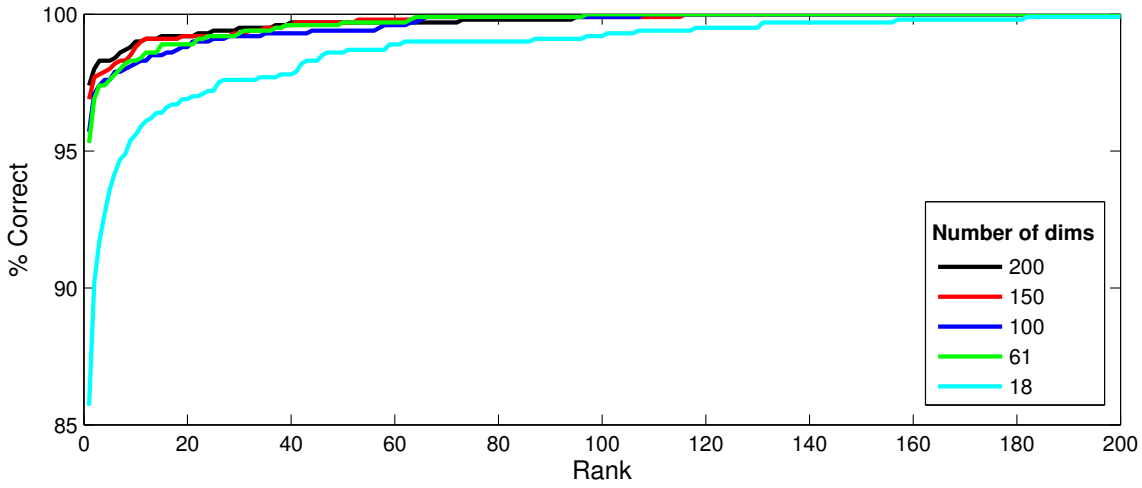


Figure 6.7: Rank-n performance results at different dimensionality. 100% performance is not achieved even at rank-200 using only 18px which suggests that this is too small a number of dimensions. However at 61px, 100% accuracy is achieved at rank-100, suggesting that this might be a suitable number of dimensions to use to effectively reduce the search space needed by the computationally intensive EGM algorithm.

ing have been shown to be powerful means of reducing data but maintaining discriminatory information. Table 6.4 compares the commonly used dimension reduction techniques (eigenfaces and Fisherfaces) with the proposed variance and percentile inclusion criterion techniques at different resolutions in terms of classification accuracy and processing time. For convenience these results have also been plotted (see Fig. 6.8). All experiments were carried out in MATLAB®on a Quad Core 2.5GHz Intel PC with 2GB ram running Windows XP. Only one percentile inclusion criterion result has been included as performance (especially processing time) was not at the same level as other conditions.

The number of components which are used for eigenfaces depends on the specific test as follows: 61 components (61PCA, row 6 of Table 6.4) were chosen for a direct comparison with the 61 variables of the variance inclusion criterion which gave good performance in Fig. 6.6. 15 components (15PCA condition, rows 7, 10 & 14) were chosen arbitrarily as an extra step after the variance inclusion criterion for its low dimensionality and relatively good performance. For other tests using eigenfaces, the number of components are chosen which describe

	Res. (px)	Data Reduction	Classifier	No. Dimensions	% Correct	Proc. time(s)
1.	10x10	None	PMCC	200	98.25	12.02
2.	10x10	VI	PMCC	19 (10%)	82.75	12.52
3.	10x10	VI	PMCC	61	95.75	13.02
4.	10x10	PCA	Euc. dist.	21	94.5	92.47
5.	10x10	PCA	PMCC	21	92.25	97.16
6.	10x10	61PCA	Euc. dist.	61	96.25	102.91
7.	10x10	VI $\rightarrow$ 15PCA	PMCC	61 $\rightarrow$ 15	89.75	128.54
8.	10x10	VI $\rightarrow$ FF	Euc. dist.	19 $\rightarrow$ 19	90.5	129.74
9.	80x80	None	PMCC	12800	98.25	129.86
10.	10x10	VI $\rightarrow$ 15PCA	PMCC	19 (10%) $\rightarrow$ 15	79	132.56
11.	10x10	VI $\rightarrow$ FF	Euc. dist.	61 $\rightarrow$ 39	99	134.69
12.	10x10	FF	Euc. dist.	39	100	144.25
13.	80x80	VI	PMCC	1235 (10%)	92.25	180.95
14.	80x80	VI $\rightarrow$ 15PCA	PMCC	1235 (10%) $\rightarrow$ 15	85.25	331.40
15.	80x80	VI $\rightarrow$ FF	Euc. dist.	1235 (10%) $\rightarrow$ 39	90.75	549.25
16.	80x80	PCA	Euc. dist.	61	96.75	573.52
17.	80x80	PI	PMCC	12800	89	800.64
18.	80x80	FF	Euc. dist.	39	99.5	973.09

Table 6.4: A comparison of the proposed variance (VI) and percentile (PI) inclusion techniques with PCA and Fisherface (FF) algorithms sorted by processing time.

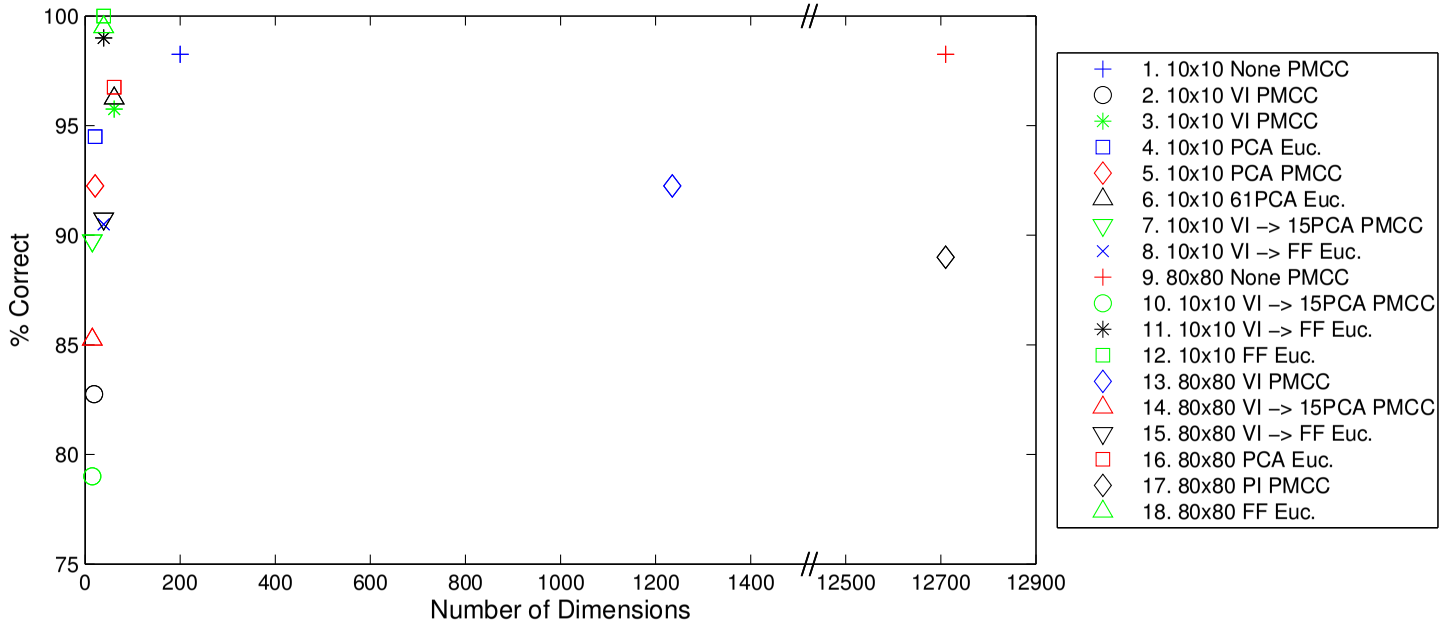


Figure 6.8: Plot of results shown in Table 6.4. Points in the top left corner represent the best performance with the least number of dimensions. The top three are '12. 10x10 FF Euc. dist.', '80x80 FF Euc. dist.' and '11. 10x10 VI $\rightarrow$ FF Euc. dist.'. NB the x-axis values between 1500 and 12500 have been omitted as no data fall in this range.

85% of the variance. Some entries in the "No. Dimensions" column have (10%) shown next to them. This is a reminder that only 10% of the data remains after applying the variance inclusion criterion. Finally some of the rows contain a "$\rightarrow$" symbol representing a combination of processes e.g. Variance Inclusion followed by Fisherface.

Generally resizing the image to 10$\times$10 pixels gives a clear processing time advantage with little or no compromise on accuracy. Without additional dimensionality reduction, a recognition rate of 98.25% (row 1) is achieved. The dimensionality can be reduced by a further 66% and only 2.5% performance is lost by additionally using the variance inclusion criterion to select 61 pixel locations (row 3). This appears to give the best compromise in terms of the number of dimensions, processing time and accuracy. The Fisherface algorithm gives excellent performance (10$\times$10 Fisherface gives 100% accuracy, row 12) but at the cost of increased processing time.

These results only apply to the simplest case in face recognition – the frontal, expressionless face. The variance inclusion algorithm would be unlikely to produce similarly good results if expressions were present in the dataset, as these are likely to produce areas of high variance which will not be discriminatory. Nonetheless these could be used for the purposes of expression analysis instead of recognition or alternatively areas which change greatly with expression could be omitted from the variance inclusion criterion.

It should be noted that although the outliers method reduces the amount of data used for classification, it does so by a sparse representation rather than by discarding the data. This means that the actual image dimensions are not reduced, but methods such as run length encoding could be used for compression, or methods that rely on sparse representations could be employed for recognition. Alternatively, one could provide the x and y coordinates of each non-zero pixel in addition to the pixel value as the feature vector, and discard all zero data. The variance approach does not suffer from this drawback, as the same pixel lo-

cations are used for each image, and so the vector for each image represents the same parts of the face.

It is clear that effective dimensionality reduction can be achieved via more direct, psychologically inspired models in contrast to conventional mathematical tools such as PCA. Processing speed is also drastically increased – if the Fisherface algorithm is used on 80×80 px images, it takes 973.09s. Using 10×10 px images, processing time drops to only 13.02s using the proposed variance inclusion method to extract 61 pixel locations with only a 3.75% drop in performance.

Additionally, the effectiveness of combining the resizing and variance inclusion criterion approach with more computationally intensive and accurate algorithms, namely EGM, is demonstrated. Processing time can be reduced greatly with only a small cost to accuracy.

In summary, a number of important findings have been presented regarding the effects of resolution and the use of different inclusion criteria on face recognition performance. The findings have implications on real-world applications in that they point to computationally attractive means of reducing the dimensionality of the data. Empirical support of Unnikrishnan's hypothesis [182] regarding the use of outlying percentile ranges is provided on both the FRGCv2.0 database as well as the Photoface database although the results suggest that there are complex interactions between different percentile ranges rather than just the outlying 10% per se (Fig. 6.5). One of the most promising results comes from combining the resolution effects explored in the previous chapter from 80×80 pixels to 10×10 pixels and applying the variance based inclusion approach yielding an accuracy of 95.75% using just 61 dimensions and the fact that this heuristic was inspired by the human process of caricaturing. Using this combination of techniques, processing speeds can also be increased tenfold over the conventional Fisherface algorithm. Additionally it has been shown that these methods can be used as a preliminary step to effectively reduce the search space in a large database before employing a more intensive algorithm.

Using a direct variance analysis approach identifies pixel locations which contain disproportionately large amounts of discriminatory information for face recognition. The data used for this analysis had neutral expressions and, as noted earlier in the discussion, expressions will lead to large amounts of variance, potentially confounding the system. In the next chapter the 3DE-VISIR database which contains three expressions for each subject is used to show that it is possible to improve face recognition performance by removing those pixel locations which vary most between certain expressions.

Chapter 7

Classification of Expression using the Variance Inclusion Criterion

The two previous chapters have explored using direct statistical methods to downsample the image or to identify either outliers or pixels which show high variation as a means of dimensionality reduction. Performance using a variance based approach paired with image downsampling is comparable to more intensive methods for the purposes of face recognition. This chapter concerns itself with two aspects of expression – its classification and its removal to improve face recognition performance.

7.1 Introduction

In this chapter, ways of incorporating the variance based approach into expression processing are explored. A commonly cited problem in face recognition is expression, and using the variance approach will almost certainly fail on faces with different expressions to those on which the variance analysis is performed. This is because the variance inclusion criterion relies on the assumption that high variance at certain pixel locations is caused by differences between subjects and

not by within-subject variations. Expressions cause large scale within-subject variations and this chapter looks at the differences in performing variance analyses on different face expressions and how these might affect face recognition.

Also, this chapter extends the use of variance information to simple expression classification. The applications for expression recognition are not so immediately apparent as for face recognition. It is nonetheless a fast growing area of research interest, the outcomes of which will likely impact future systems. Suggested uses are systems that are able to modify the feedback depending on a subject's expression which might be useful in video tutoring, or detecting deceit during interviews/interrogations and in human-machine interaction. With an increasingly aged population, robots to perform care duties could also benefit from being able to classify expressions in order to gauge an expected response better. Current research is also looking at expression recognition for helping those with developmental disorders such as autism to improve their expressions through an interactive game called *SmileMaze* [36]. This system uses the Facial Action Coding System (FACS) which is an attempt to quantify aspects of an expression. The system is able to automatically score a face in terms of FACS and then use this to estimate the expression. If the smile generates a sufficiently high action unit score, an obstacle is removed from the maze and the subject progresses.

A different possible use for an expression classification system would be as a pre-capture step in order to screen the probe subject and ensure that their facial expression matches whatever the system expects. Beveridge *et al.* found that, contrary to most experimental data, smiling faces lead to better recognition than neutral [21], so it may be useful to be able to recognise that a subject is smiling before a probe image is used for recognition.

Similar to face recognition, expression recognition research broadly falls into the same two categories, feature-based and statistical approaches. In the same way, feature based methods analyse fiducial features, the locations and ratios of landmarks to classify expressions (the above automated FACS coding system

is an example of this), and statistical techniques such as PCA, FLD and ICA are used to model the way in which the global appearance of a face varies with expressions.

Again, in common with face recognition research, one of the most commonly used and reliable feature representations is to use Gabor filters [203, 54] and in turn this has motivated research into using the functionally similar but more efficient Local Binary Patterns (LBP) to improve expression classification [161]. There are also temporal approaches which measure the onset, apex and offset phases used by FACS of an expression in order to perform classification using the optimal fiducial feature displacement for the expression [183]. Optical flow has also been used to quantify temporal differences between frames in face expression recognition [60].

Expression classification is a research area in its infancy compared with face recognition. The Cohn-Kanade Action Unit Coded Facial Expression Database [101] is perhaps the most commonly used dataset (recently updated to CK+ [122]) due to its rich metadata which codes all the action units of each face. However, as stated by Valstar *et al.* , researchers often use different parts of the database with different training and evaluation protocols. This year Valstar *et al.* ran the first Facial Expression Recognition and Analysis Challenge (FERA) which uses its own database and standardises an evaluation methodology [184]. It will be interesting to see whether this challenge is adopted as successfully and whether the technology advances as quickly in this area as it did with the FRGCv2.0 challenge.

Little research exists into using 3D data for expression recognition, probably due to the lack of availability of 3D and 4D databases until recently (BU-3DFE [197] and BU-4DFE [196]). The first comparison of 2D versus 3D data for expression classification showed that 3D provided better results for detecting action units for classification [157] except for some action units in the upper face. They also state that a neutral expression is useful for improving classification to provide a

difference image for 3D data. Kakadiaris *et al.* report extremely good results (over 95% accuracy) on the FRGCv2.0 subset of expression data using an Annotated Face Model (AFM) [99]. A variant on this approach also scored very well on a new dataset involving identical twins [186] .

As with face recognition approaches, the methods used are often complex and computationally intensive. This chapter explores whether the variance and resolution based approaches explored in the previous chapters can be employed for expression classification and whether this in turn can be used to improve recognition rates. The main contributions of this chapter are a demonstration of the 3DE-VISIR database as an effective database for expression analysis (with 90% accuracy in classifying a postive expression from others), evidence that high variance pixel locations on the face encode more expression information than identity recognition information and that by removing these pixels, recognition rates can be increased by just under 5%.

7.2 Methods and Data

The data used in this chapter comes from the 3DE-VISIR database and consists of 80 subjects and 644 sessions in total captured using visible light and NIR light source PS with three expressions (positive, negative and neutral – see Sec. 4.2 for more details). While this is a far smaller dataset than the Photoface database it contains controlled capture of different expressions. The faces are cropped tightly for reasons outlined in 5.2.1 and examples can be seen in Fig. 7.1.

Three expressions are captured for each subject and classification accuracy is assessed. For this classification, the Fisherface algorithm is chosen which treats the different expressions as different classes. Whether visible, NIR or combined light source data provides best performance is investigated. It might be expected for algorithms to perform the best on the combined dataset as it doubles the number of samples per condition, which has been shown to improve the performance

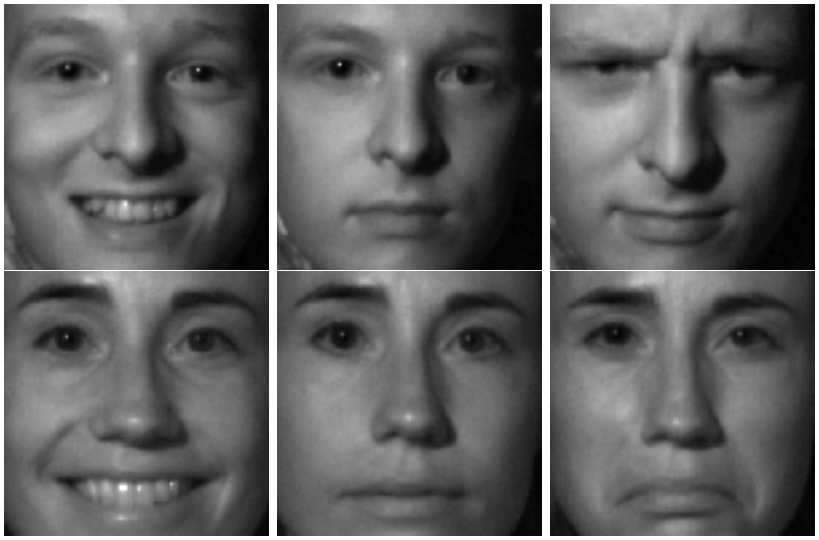


Figure 7.1: Examples of two subjects from the 3DE-VISIR showing typical expressions of the cropped data. Left to right, *Positive*, *Neutral* and *Negative*.

Figure 7.2: Example showing the face symmetry altering with the *Negative* expression. The jaw has moved the left.

of the Fisherface algorithm. By chance alone a 33.3% accuracy would be expected. A number of pairwise tests are then performed to explore whether any expression is more easily identifiable than the others. Again the leave-one-out paradigm is used to maximise the number of trials.

As expression is likely to account for a great deal of variance in pixel values, can this be used as useful information for expression classification and inversely, to aid identity recognition? Experiments performed here test this hypothesis by using the variance inclusion criterion.

Finally the resolution effects that were found to aid face recognition at low resolution rates are explored to see whether they can aid expression recognition in a similar way as is suggested by research into the HVS.

7.3 Results

Table 7.1 contains results for expression classification using combined light source data and visible and NIR data. As expected, combining the datasets provides the best results which is likely due to Fisherfaces being better able to discriminate given more data. The combined results show correct classification of the three expressions 63.75% of the time. The best performances are distinguishing between a *Positive* and a *Neutral* face (88.03%) and a *Positive* face from a *Negative* face at (87.5%). It also shows that it is not possible to reliably distinguish between negative and neutral faces (50.75% is only just above chance of 50%). This is likely due to the frown (the commonest *Negative* expression) being far more similar to the *Neutral* face than the *Positive* expression. Therefore the *Negative* and *Neutral* faces are combined into the *Other* class and results show that the *Positive* faces can successfully be separated from these *Other* expressions reliably (90.51%).

Light \ Expression	+ive,-ive,Neu	+ive,-ive	+ive,Neu	+ive,-ive	+ive,Other
Vis+NIR	63.75	87.5	88.03	53.75	90.51
Vis	61.02	83.60	87.32	56.91	87.12
NIR	57.97	85.71	82.16	47.87	86.44

Table 7.1: Fisherface classification results for *Positive* (+ive), *Negative* (-ive) and *Neutral* (Neu) faces for Visible, NIR and combined light source data. By chance alone, 33% accuracy for the first column and 50% for the other columns would be expected. *Positive* expressions are most distinguishable. The expression category *Other* combines *Negative* and *Neutral* data and gives the highest classification performance when distinguishing from *Happy* expressions.

These levels of correct classification (except for the *Neutral* and *Negative* condition) show that there is reliable discriminatory data. If it is assumed that this comes from areas of highest variance, can the variance inclusion principle be used (which is successful for subject recognition), in the same way on expression classification? For example, is it true to say that pixel locations with the highest variance are the most discriminative in expression classification? Furthermore,

if it is possible to identify regions/pixels which vary greatly with expression these can be removed and then the remaining (expression invariant) pixels can be used for identification.

Fig. 7.3 shows the standard deviations of the pixel locations for different expressions and the dataset as a whole (i.e. all expressions) together with thresholded variance masks. These are generated in the same way as described in Section 6.2.1 and shown in Fig. 6.2. Visually comparing the figures, of most interest is the row containing the results for the *Positive* data (containing images with the red border). The standard deviation map for the x-components (first column, top row) highlights the cheek areas while the y-components (first column, bottom row) clearly show an upwardly bowed mouth area (this is also present in the corresponding *Negative* expression map, but is less pronounced). These areas correspond with the typical changes that occur with a smiling mouth – the cheeks bunch as the lips are extended laterally and raised at the edges.

7.3.1 Expression Classification using Variance Inclusion Criterion

In order to test whether the variance inclusion principle can be used successfully for expression classification it is necessary to identify which pixel locations with a high variance are caused by different expressions rather than different identities. As the *Positive* expressions gave the most reliable performance, the variance inclusion principle is employed in an attempt to distinguish between a *Positive* expression and the *Neutral* subset. To do this, the pixel locations with the highest variance must be found in the *Positive* and *Neutral* subsets and then use the difference between these as the mask. The standard deviation σ of each dataset is used to provide a map of variance d as shown in Eqn. 7.1.

$$d = ||\sigma_{positive} - \sigma_{all}| - |\sigma_{neutral} - \sigma_{all}|| \tag{7.1}$$

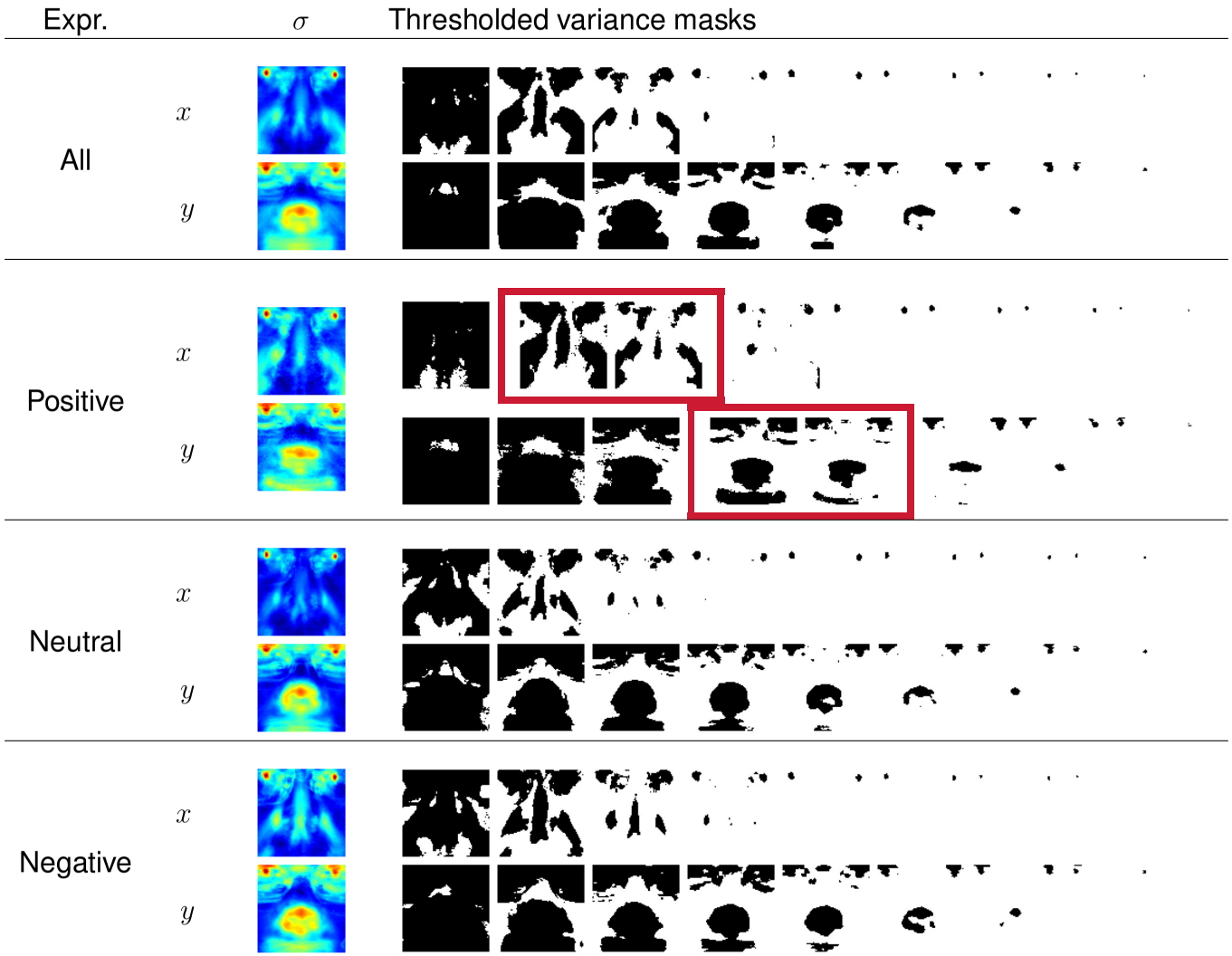
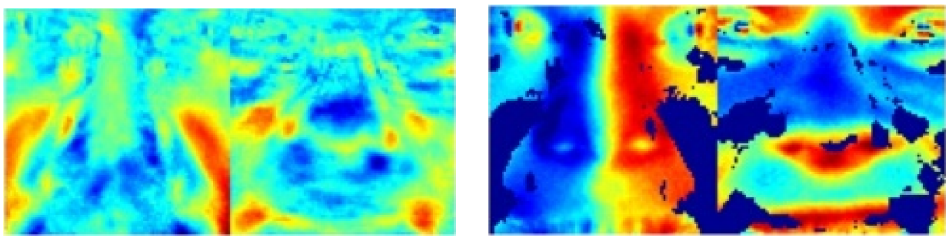


Figure 7.3: Standard deviation (σ) maps and thresholded variance masks across four different expressions for x-component (top row of each expression) and y-component (bottom row of each expressions). The first (coloured) image in each row shows the standard deviation for each pixel, the next eight images show the results of thresholding at equal intervals between the minimum and maximum standard deviation. Those images in red boxes indicate the recognizable aspects of a smiling face – the cheeks bunch (shown in the y-component) as the lips are extended laterally and raised at the edges (shown mostly in the x-component).

An example of the map can be see in Fig 7.4(a).



(a) (b)

Figure 7.4: a) The difference in variance between *Positive* and *Neutral* faces for x-components (left) and y-components (right). b) The results of applying a mask to surface normals. The darkest blue regions are removed from the faces before comparison. The mask is produced by thresholding (a) to produce something similar to those masks shown in Fig. 7.3.

Applying a threshold to the difference images provides a mask which is applied to the probe and gallery images to remove the pixels which contain the most expression variant information. An example of the mask applied to a face can be seen in Fig 7.4(b) where the dark blue areas correspond to the largest difference in variance between the *Positive* and *Neutral* faces and are removed from the faces before comparison.

Expr. Px/Rec. Px	Expression Correct	Recognition Correct
12800/12800	88.03	61.74
3200/9600	84.51	63.62
6400/6400	86.85	65.26
9600/3200	87.32	64.32

Table 7.2: Percentage accuracy at classifying between *Positive* and *Neutral* expressions using the variance inclusion criterion to determine the most useful pixel locations together with subject recognition based on the remaining pixels. There are 12800 pixels in the raw image data. The reader may be wondering why the recognition rates are so low – likely causes include small sample size and facial expressions.

Table 7.2 compares three threshold values with the baseline *Positive/Neutral* expression classification and subject recognition. The first column provides the number of pixels on which classification/recognition occurs, and the second and third columns respectively provide the percentage of expressions and subjects correctly classified. It can be seen that for the expression classification the base-

line performance is not improved upon (88.03%), but that the number of pixels used in the analysis can be reduced with only a relatively small reduction in performance. This is likely an indication that the selected pixel locations encode expression data to a larger extent than the discarded pixels. Instead of using the pixels that code expression for expression classification, the inverse of the mask can be used for face recognition. The assumption is that the pixels which remain after the expression variant pixels have been removed encode the identity of the person. The third column shows the results of this, and there is almost a 4% increase in performance when those half of the pixel locations that show the highest variance are removed.

In summary by using the pixel locations which vary the most between *Positive* and *Neutral* expressions, just over half of the pixels can be removed and only a couple of percent in expression recognition accuracy is lost, and that by using the inverse of these pixel locations, the subject recognition accuracy is improved by just under 4%.

7.3.2 Resolution Effects on Expression Classification

Reducing the resolution down to 10×10 px was found to increase the recognition accuracy in Chapter 5. Here, the effects of resolution on expression recognition accuracy are tested. Research into the HVS has suggested that lower spatial frequencies code the expression of a face as well as the identity [158] so it might be expected that lower resolutions are also beneficial for automated expression recognition.

As shown in Fig. 7.5, the size which gives the best performance in terms of recognition rate on the 3DE-VISIR database is 38×38 px (50.85%) but because the graph peak is small, the recognition rate remains steady down to resolutions as low as 20×20 where there is only 1% decrease. This is still larger than that for the FRGCv2.0 or Photoface database datasets. The reason for this is unknown,

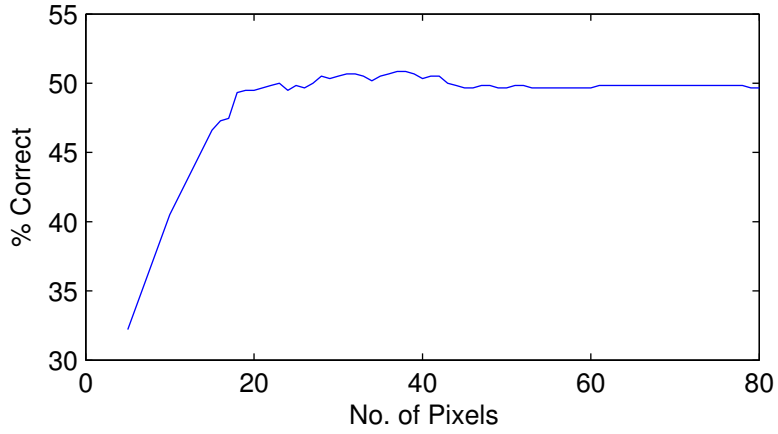


Figure 7.5: Recognition performance on the 3DE-VISIR dataset at different resolutions. Again recognition performance is not affected by low resolutions.

but is likely due in some way to the variety in expressions. Next, the optimum size for expression classification for the same categories as in Table 7.1 is explored, the results of which can be seen in Table 7.3 and Fig. 7.6. These show that the best (or close to) expression classification rates occur at low resolutions of 10×10 pixels. An unexpected finding is that the rate drops to a minimum at about 15×15 pixels before rising again to near optimal performance at 40×40 pixels. These are interesting findings, but need validation on a larger dataset before any firm conclusions can be drawn from them.

Expression \ Resolution	5	10	15	20	40	80
+ive,-ive,Neu	68.31	69.66	58.98	58.47	68.31	70.85
+ive,-ive	88.89	87.30	70.90	86.24	88.36	89.15
+ive,Neu	87.09	86.85	65.49	81.92	87.56	88.03
-ive,Neu	64.36	61.45	60.11	62.50	65.43	64.63
+ive,Others	88.47	91.19	81.70	76.78	89.15	90.51

Table 7.3: Percentage accuracy at classifying expressions at different resolution. These are the data for Fig. 7.6.

7.4 Discussion

This chapter has explored whether direct methods (resizing and variance based selection) which were found to improve recognition rates in previous chapters,

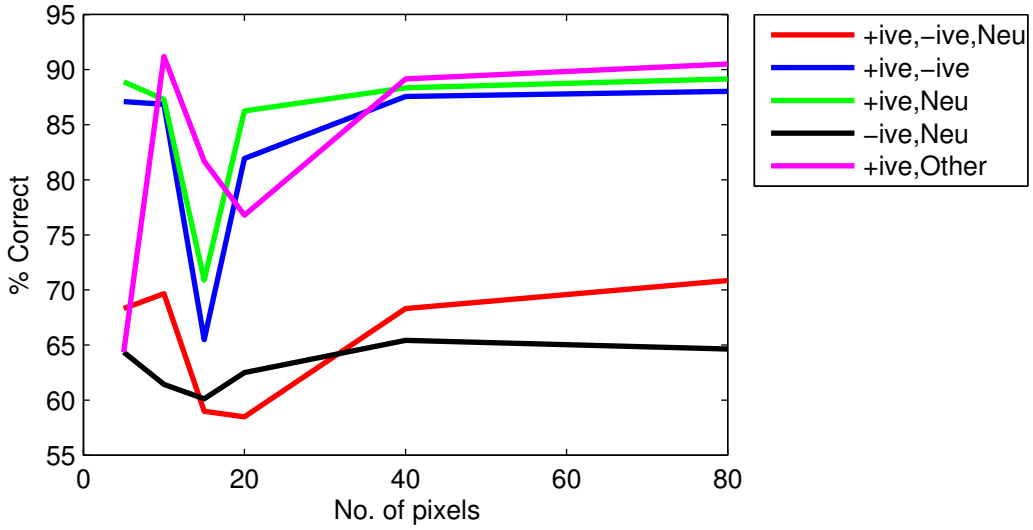


Figure 7.6: Expression classification performance on the 3DE-VISIR dataset at different resolutions. Again in line with human findings, low resolutions do not hinder expression classification. The reason for the consistent drop in performance around the 15×15 px resolution is unknown.

could also be used to classify expressions and furthermore to provide some level of expression invariance.

The ability to classify an expression as *Positive* versus any other expression is by far the most reliable (around 90%). This is interesting and a potentially useful finding if a recognition system requires smiling rather than neutral faces.

The results also provide some evidence that the pixel locations whose values vary the most between expressions are those that code the expression information, and that the inverse pixel locations code identity, and that an improvement in recognition rate shows some expression invariance of these pixel locations. By removing the pixels which vary the most with expression the subject recognition rate can be improved slightly. Instead of removing the pixels, it might be interesting to replace them with the mean value of that pixel across all captured sessions - in this way we might expect the expression to become more neutral.

Agreeing with findings in Chapter 5, expression classification appears to not be adversely affected by very low resolutions (typically only about a percent in performance is lost in resizing from 80×80 px to 10×10 px). However, there is

a dip in performance across all conditions (except the Neutral versus Negative expressions) at 15×15 pixels. The reason for this unclear. Again it is interesting that these findings are similar to those found in psychological studies [158] – that lower spatial frequencies code expression.

There are limitations to the experiments here in that they use a database which contains only three expressions, and those three are captured under constrained conditions meaning that the expressions are likely to be forced. It would be interesting to apply the technique to a larger and more unconstrained database of 3D surface normal data, but to date none exists so this would be interesting future work.

The final chapter looks at how the findings here and in previous chapters can be combined into a fully developed face recognition system.

Chapter 8

A Proposed Fully Automatic Face Recognition System

The previous chapters have tested how accurate PS is for face capture, developed methods of dimensionality reduction which are inspired by the HVS and attempted to use these findings to improve robustness to expression. These are all important findings which contribute toward a fully automated face recognition system. Even though taken together, these findings represent a sizeable amount of work toward a fully automated system, it seems natural to identify the areas which are still missing and make some preliminary investigations into them. It is suggested that automated feature detection to allow for automated alignment and defining a threshold to allow unknown subjects to be rejected need to be found. Here, all the work of previous chapters are brought together into a fully implemented system and some of the wider implications to face recognition are discussed. Fig. 8.1 shows the steps required to capture and recognise a face. Those steps in green represent two areas which are missing from the system as discussed so far, and this chapter goes some way to address them.

For the most part the pipeline is self-explanatory, although a few points need some expansion. For example, in the second and third processes, expression classification and roll correction are mentioned. The purpose of the expression

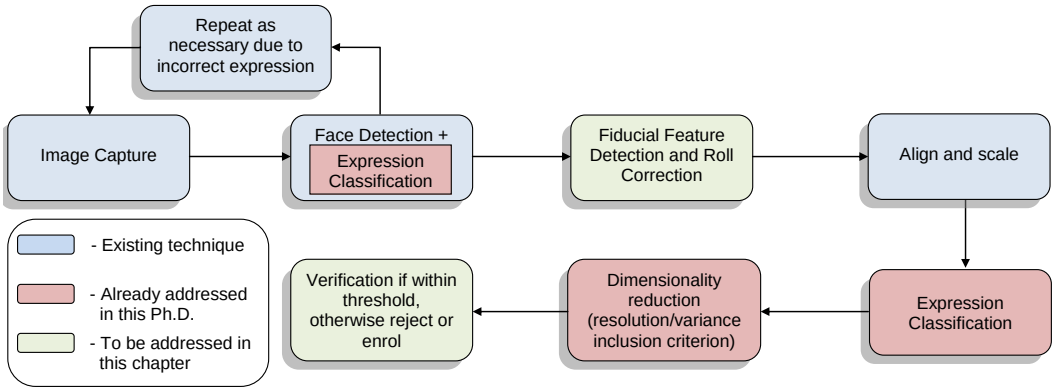


Figure 8.1: A pipeline of a developed system. Those in green represent the two areas addressed in this chapter, while red boxes highlight contributions made in this Ph.D..

classification at this point is to reject and retake the capture if they do not have the required expression. There is currently no pose estimation system, but one would be required in the same way to ensure that the subject is frontal with minimal pitch. Because the face is captured in 3D there is the opportunity to correct for minor pose deviations (as long as no part of the face is obscured by another). However to do so would require surface generation, rotation correction and then conversion back into surface normals – additional processing which is undesirable. An alternative to this would be to employ a rejection mechanism whereby instead of the computer making the required corrections, the user can simply be prompted to make them before walking through the device again. The fifth box, “Expression Classification” would come into play if such a rejection mechanism was not employed in order to remove the most expression variant pixels to boost the recognition performance as shown in Chapter 7.

This thesis has focussed on methods of reducing dimensionality through direct methods which in some ways have been inspired by the HVS – reducing the resolution and selecting pixel location with high variance or outlying values. Throughout the experiments, a simple true positive (the number of correctly identified individuals) recognition rate has been used to assess the efficacy of the proposed methods. However it tells us nothing about the number of individu-

als who are incorrectly classified as another i.e. “Bob” is recognised as “Alice”. This is arguably the worst case scenario for an automated recognition system – granting access to an individual who should be denied. Conversely, incorrectly denying access to an individual would pose more of an inconvenience than a security issue, as secondary methods such as keys or passwords could be used. In order for a face recognition system to be able to perform such functions it is necessary to define a threshold distance which will reject a probe image if it is not within a certain threshold of a gallery image. If this threshold is too lenient then the false positive rate will be unacceptably high (and unauthorised individuals could be granted access), if it is too strict then the true positive rate will be too low (and authorised individuals will be rejected). Such a threshold is empirically determined later in the chapter.

The tightly cropped face images used in the experiments throughout the thesis provided high recognition rates and it would seem advantageous to keep to this format. In order to crop the images, certain fiducial features are used to scale and align the images (for more information see Section 5.2.1). The fiducial features are all manually labelled on the data which is necessary for experimental purposes but is impractical for a deployed system. We therefore need some way of automating this process. There are many landmark detectors in the literature e.g. eyes, nose, nasion etc. but here we look at using an extension of Haar-like features for detection of landmarks on surface normal representations. First we explore using the standard Haar-like features that Viola and Jones used in their seminal paper [187] to detect the lateral canthus and then look to extend a Haar-like feature for direct detection of spherical caps and cups through surface normal analysis in order to detect the nose tip, nasion and interior canthi areas. As a consequence of this process, we show that a clear horizontal pattern of the face presents itself which has the potential to correct roll pose.

8.1 Introduction

Viola and Jones [187] presented a novel face detection method which worked via a boosted cascade of Haar-like features. These features are shown in Fig. 8.2. Each of these features are positioned exhaustively over the image and at each position, the sum of pixel intensities within the white areas are subtracted from the sum of the pixel intensities in the black areas. The features are then resized and reapplied to the image to build up a large set of features (about 45,000 which is in contrast to the 180,000 stated in the paper [1]. To speed the calculation Viola and Jones used what they termed the *integral image* in which each pixel's value is the sum of all those to the left and above its location. Therefore instead of calculating the sum of all the individual pixels in each area of the Haar-like feature, any rectangular area can be computed in four array references, and the difference between two rectangular areas can be calculated in just eight. They then use the Adaboost algorithm for selecting the best features and to train the classifier over a large training set of positive and negative examples (about 10,000 of each at 24$\times$24 pixels). This is very time consuming but needs to only be performed once. Empirically they then determine which combinations of these features provide the best performance in their cascade. The benefit of the cascade is that the descending classifiers eliminate many negative examples without the need of further processing at each stage. They report excellent results on their 38-level cascade of classifiers which is made up of 6061 features which is capable of processing a 384$\times$288 pixel image in just 0.067 seconds.

Due to their reported performance on 2D photographs, the Haar-like features would seem an ideal candidate to extend for use in this system. Additionally the Haar-like features are likely to give similar, albeit simplified, results to more complex Gabor filter type processing which have been shown to be very effective for automated systems but also model the HVS. Vertical and horizontal edges, verti-

[1] In personal correspondence, one of the authors (Mike Jones) responded that this was due to them only analysing every other pixel.

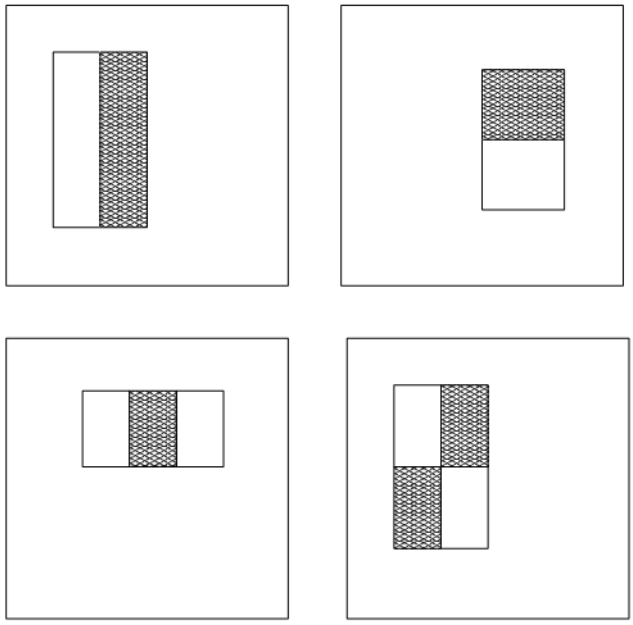


Figure 8.2: Representative examples of the Haar-like features used by Viola and Jones, reproduced from [187]. These are windowed over an image, and the sum difference between the pixels under the shaded and the unshaded regions are calculated for different scales of the features to produce weak classifiers for a face. Adaboost is then used to select the best performing weak classifiers and these are combined into a cascade which enables rapid real-time face detection.

cal lines and diagonal lines will be detected by the features shown in Fig. 8.2. The resizing of the features means that different scales are detected. Firstly, detection of the lateral canthi in the surface normals is attempted using the same features as those used by Viola and Jones with promising results, although computational time was found to be prohibitive. Then a new feature (inspired by the Haar-like features but tuned to work on surface normal data), containing three square areas (Fig. 8.4) is used in order to detect spherical caps and cups e.g. the nose and interior canthi regions.

Once suitable fiducial features have been identified it is possible to align and scale the image as required and for the recognition algorithm to be run. Regardless of the algorithm run, some sort of comparison between probe and gallery images must occur which results in a similarity score. In previous chapters the highest correlation score using PMCC or the lowest Euclidean distance using the Fisherface algorithm have been used to return what is known as the Rank-1 result

– the top result is the correct one. It is common in the literature to report other rank scores e.g. Rank-3, Rank-5, Rank-10 in which the correct subject is returned in the top three, five and ten matches.

A stricter measure which would be more suited to a real-world deployment would be the use of thresholding on the distance measures to accept or reject probes. For example, the FRVT2006 tests set performance goals of 0.1% FAR (which equates to one person falsely being identified in 1000 trials) and a verification rate of 98%. A simple and commonly used method to compare performance is to use the Equal Error Rate (EER) which is the point at which False Accept Rate (FAR) and False reject rate (FRR) are equal – the lower the EER the better the performance of the system. In order to inform the decision on a suitable threshold the FRR or True Accept Rate (TAR) versus FAR can be plotted to generate a ROC. The compromise between a system that is overly selective and falsely rejects probes and one which misidentifies probes can then be visualised and a threshold which provides the desired level of performance can be chosen.

8.2 Methods and Data

The initial experiments described here look at detecting the left exterior lateral canthus using Haar-like features. Essentially, this applies the Viola and Jones [187] method to surface normal components. The features were applied separately to the x and y-components of 300 images of the left exterior lateral canthus (positive) and 600 images of parts of the face other than the left exterior lateral canthus (negative) training examples. The training images were 24$\times$24px in size. Haar-like features were moved across the image in turn and the resultant score is calculated and stored for each. In-keeping with the original work, every other pixel was skipped in order to increase efficiency. After this had been performed with all of the Haar-like features, at all scales, at all locations the features are ordered by

score. Without calculating the cascade the 100 top performing were run against the surface normal components of an unseen image to give the result shown in Fig. 8.3. For comparison the features were also run against a grey scale albedo image having been trained on the same numbers of positive and negative albedo examples.

Figure 8.3: Landmark feature detection using Haar-like features from the method of Viola and Jones. Highest scores for the albedo image (left) and y-component (right) using the 100 top performing features. The top score for the albedo is located very close to the actual left lateral canthus (red square), but there are many false hits for the surface normal data (the blue square indicates the closest match).

While this appears to work well for the albedo image, and highlights the correct region in the image of y-components, there are also a great deal of false positives. This may be caused by too small a number of training examples. While this is an indication of the possible usefulness of using the technique developed by Viola and Jones, the Haar-like features are designed for use with intensity images / photographs rather than surface normals. Therefore, experiments follow that look at creating a new Haar-like feature which extends those in Fig. 8.2 for detecting spherical caps and cups.

The nose tip is the feature targeted for extracting as it is a useful marker given

its central location in the face. The nose tip can broadly be thought of as a spherical cap shape, which lends itself well to detection via a two-stage Haar-like detector – one on the x-components and one on the y-components of the surface normals. Due to the symmetrical nature of the spherical dome of the nose tip, it is also rotationally invariant. Fig. 8.4 shows the nose tip as an enlarged region – what can clearly be seen is the ridge structure travelling vertically and horizontally in x and y-components respectively (the green pixels). Therefore an extension of the Haar-like feature shown on the far right of Fig. 8.4 is applied, which is designed to respond optimally to this ridge. By rotating it 90 degrees it should then respond well to the y-components. Each of the differently coloured blocks contain a matrix of either -1, 0 or +1 as indicated on the diagram of 10×10 px. This feature is moved across the surface normal components, the difference between the -1 and +1 blocks calculated and then the results for the x and y-components are multiplied together. The results of this procedure can be seen in Fig. 8.5 of the results section.

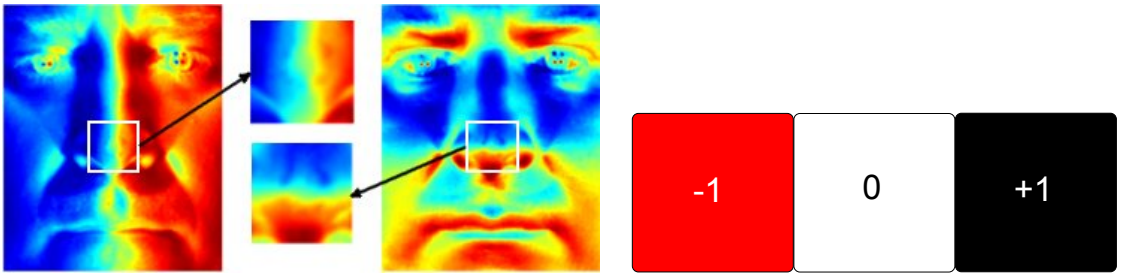


Figure 8.4: (a) illustrates the enlarged nose tip showing the x and y-components respectively, and (b) is the proposed basic filter for surface normals inspired by the Haar-like features used by Viola and Jones.

With regards to selecting a suitable threshold the next section presents ROC curves comparing the performance using PMCC and Fisherface on 10×10 px images.

8.3 Results

This section presents preliminary work into using surface normals for landmark detection and empirically deciding on a suitable threshold for the given database.

8.3.1 Feature Detection

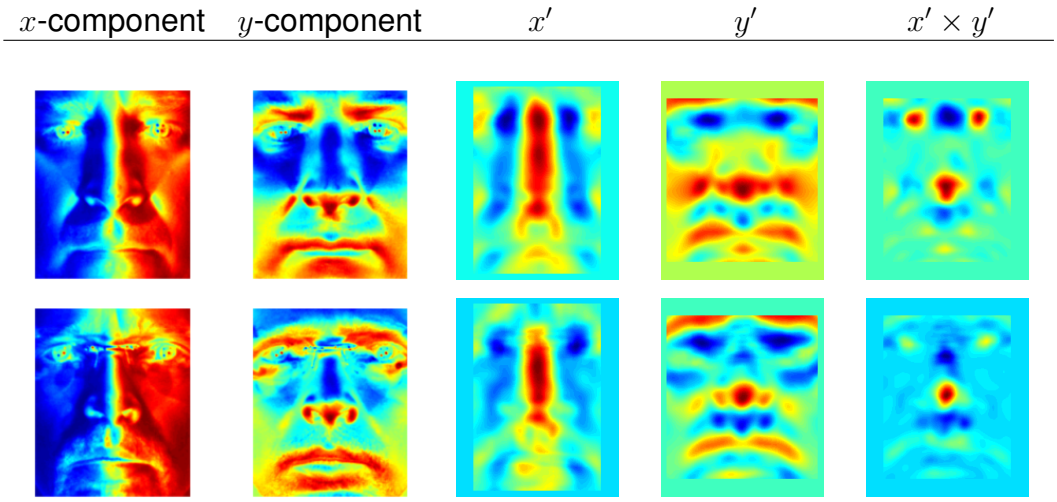


Figure 8.5: Results of applying the extended Haar-like feature to surface normals of two subjects. The first two columns show the raw surface normals, the next two columns show the result of applying the filter, and the final column shows the result of multiplying the third and fourth columns.

It is clear that for these two subjects the nose tip area is highlighted well (red regions). The interior canthi are also well highlighted (the tearduct region) and the nasion is lowlighted clearly for the first subject and a little less so for the second subject even though he is wearing glasses (blue regions).

The nose tip region has been effectively localized using this simple filter on the surface normals. This region can then be used to mask the z-component from which the maximum value can be extracted which then corresponds to the nose tip (assuming the face is frontal). See Fig. 8.6. If the face is not entirely frontal, then instead of using the z-component, the shape index for that region could be calculated and the value closest to 1 (spherical cap) could be used. Additionally, the nasion appears to be have a large negative value, and by simply

choosing the lowest value, an estimation of its position is also provided. Using the approximate position of the nasion, it is then possible to make an informed estimate of the interior canthi position by excluding the region below the nasion region from the search space assuming that the canthi correspond to the maxima either side of the canthi.

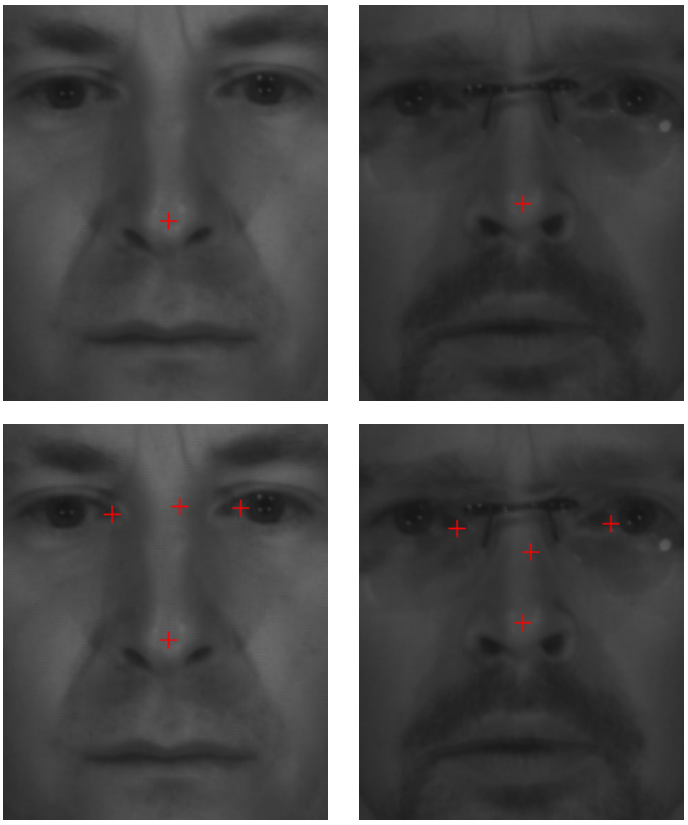


Figure 8.6: Estimation of the nose tip from combining the results from filtering to localize a region from which the maximum z-component value is chosen (top), and the less reliable nasion and interior lateral canthi estimation (bottom).

The ability to highlight the nose tip region has been attempted on over 300 raw captures from the Photoface device. The majority are frontal facing although there are many that are looking slightly to the side. Additionally in certain cases the Viola and Jones algorithm which detects the face before cropping the raw image (part of the Photoface capture system) has failed leading to an image consisting mainly of background. Even under these circumstances the proposed algorithm performs well. It is robust to roll and pitch but less so to extremes of yaw.

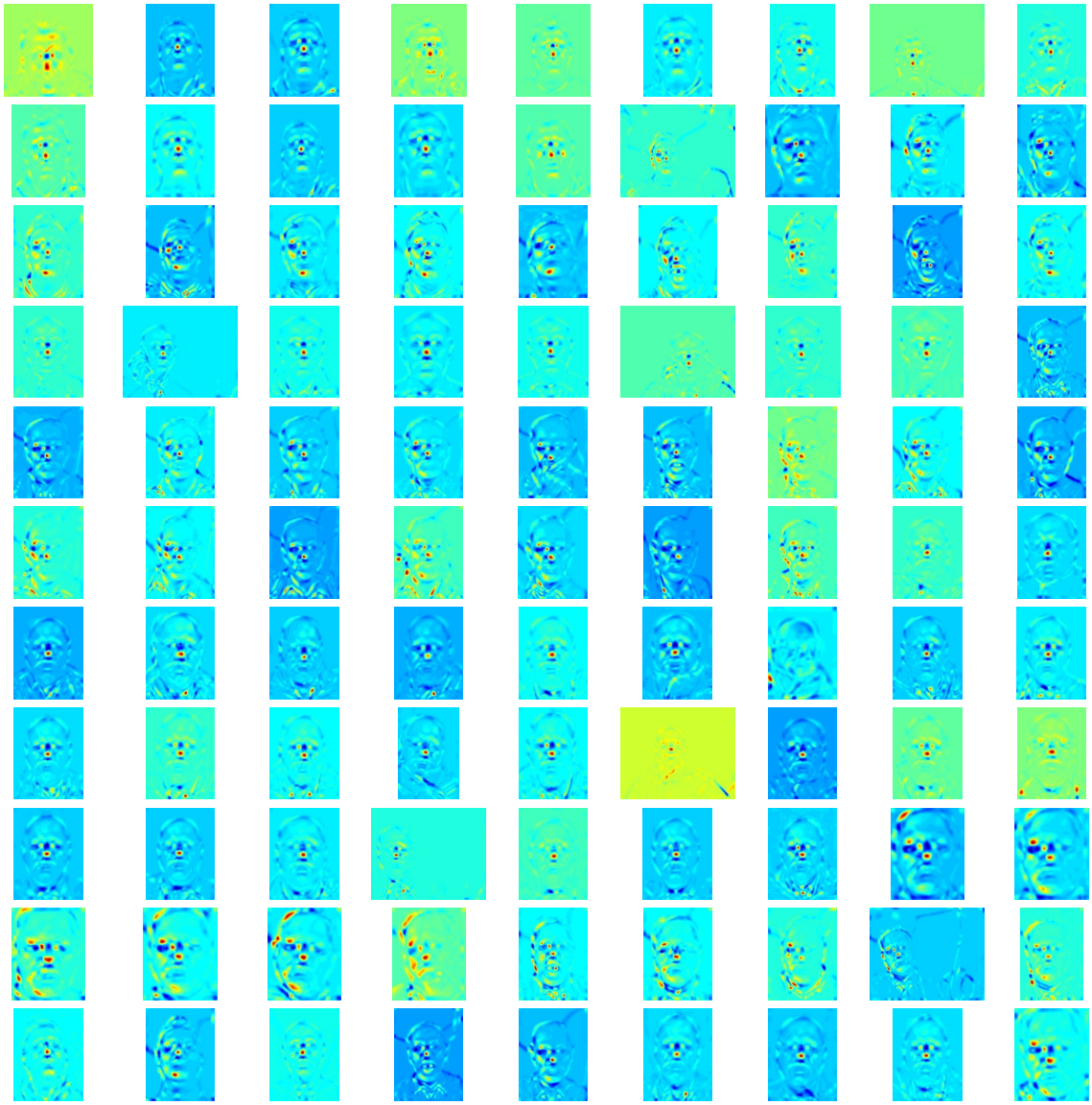
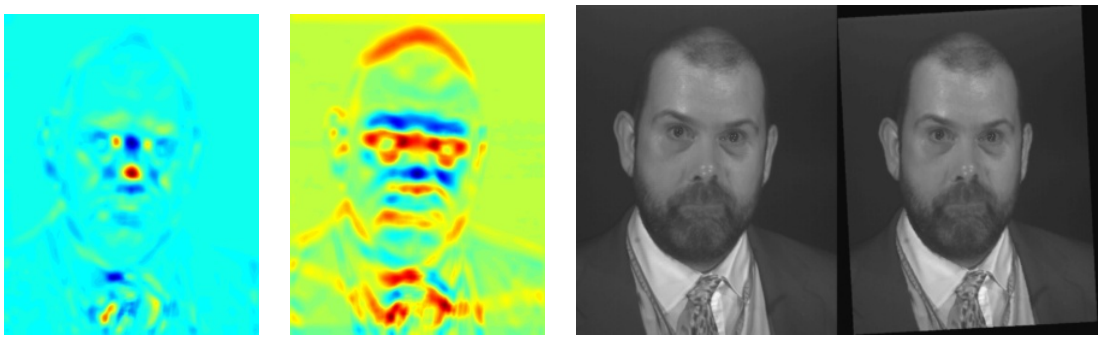


Figure 8.7: 100 examples of how the algorithm highlights the nose tip region.

An interesting outcome of applying the filter to the y-components in Fig. 8.5 is that the horizontal structure of the face is accentuated into something somewhat resembling a bar-code. Bar-codes were found to be a useful representation by Dakin & Watt [48]. The claims made in their paper are not explored further here, but not only does this seem a simpler method for bar-code generation, but also what seems obvious is that the bars should lend themselves well to roll pose correction by looking for symmetry. The results of a naive approach are shown in Fig. 8.8 demonstrating the efficacy of this simple approach. This could be extended to use more advanced techniques such as a radon transform which

would give the angle along which the maximum energy occurs, the inverse of which would correct the image.



a b c d

Figure 8.8: The nose tip region is highlighted (a) and used to mark the middle of the face along which the symmetry across the vertical axis of (b) is tested at different rotations. The rotation with the least amount of difference between the two sides is then applied to the original image (c) resulting in the roll corrected result shown in (d).

A benefit of this approach to roll correction is its robustness compared to using a feature based method such as the axis between the eyes or along nose which will give unexpected results if the subject has eyes which are not horizontally aligned or if the nose is crooked. Using the symmetry across the whole face makes it unlikely to fail as long as the roughly horizontal structure of the face is preserved. It is also likely to be robust to rotational variance as the lateral banding will be preserved if the head rotates in any direction. Results of applying this method to the bar-code like representation can be seen in Fig. 8.9 which shows that the lowest asymmetry occurs at about 4°, which when the opposite rotation is applied to the original image (Fig. 8.8.c) the resulting image is corrected (Fig. 8.8.d). Performing the same algorithm on the albedo image did not work – the most symmetrical rotation was reported to be that of the original image as can be seen in which also shows that this method does not work on the raw normals themselves.

Although inspired by the work of Viola and Jones on Haar-like filters, the actual filter investigated here is similar to the features extracted by Microsoft's Kinect system for pose estimation [164]. Whereas the Kinect's features are pixel based,

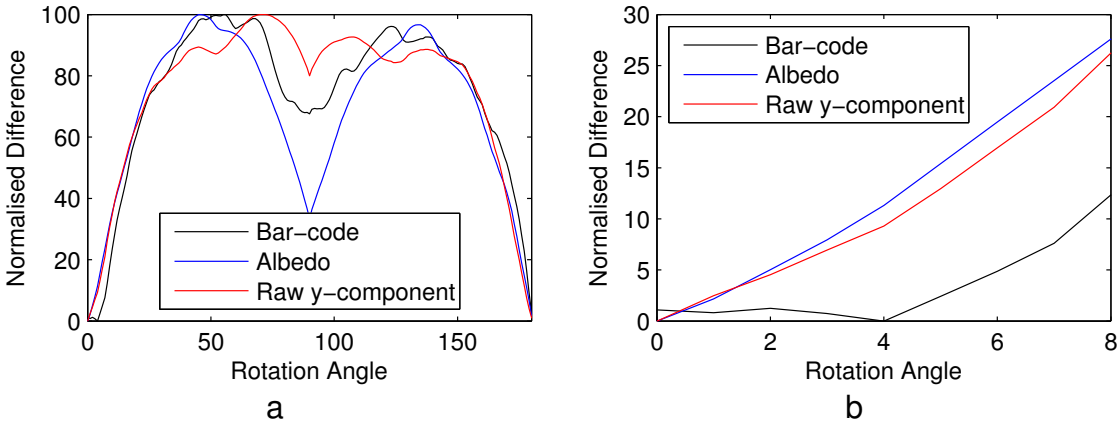


Figure 8.9: Plot (a) shows asymmetry (measured as the difference between the left hand side and right hand side of the face about the nose) as a function of rotation angle for the bar-code representation (as shown in Fig. 8.8.b), the raw y-component and albedo image for comparison. Plot (b) shows an enlarged range of 0-8 degrees of (a) and clearly shows that the bar-code format is the only representation which gives the correct pose correction angle – a normalized difference of 0 at 4°. A lower normalized difference corresponds to low asymmetry (which is expected when a face is upright).

the filter presented here is region based and combines some knowledge of the feature it is optimised for (e.g. a ridge). The Kinect uses training to organise the features into the most discriminatory, and indeed this may be a useful approach in future work for discovering new Haar-like features.

8.3.2 Threshold Selection

The two areas identified that have not been covered in this thesis were feature detection and threshold selection. The previous section covered feature detection, and this section covers threshold selection.

Two ROC curves can be seen in Fig. 8.10, one showing the results for adjusting the threshold using Fisherfaces, the other adjusting the threshold for PMCC. From the shape of the graphs it an be concluded that PMCC is unsuitable for use in a automated recognition system as there is no suitable compromise when the TAR is high and the FAR is low. Although it gives Rank-1 performance that is approximately equal to Fisherface, it is not suitable for applications where a verifica-

tion is required as the FAR rises at approximately the same rate as the TAR meaning that there is no acceptable compromise at which a threshold can be used. The Fisherface results on the other hand show very good results: a 98.65% recognition rate at FAR=0.01 at a threshold distance of 0.3896 (EER=0.0117, threshold of 0.3866). The FAR=0.01 rate means that for every 100 verifications only one false positive will occur and 98.65% of positives will be correctly identified. At a stricter level of FAR=0.001 (or one false verification in every 1000 comparisons) the recognition rate is 95.04%.

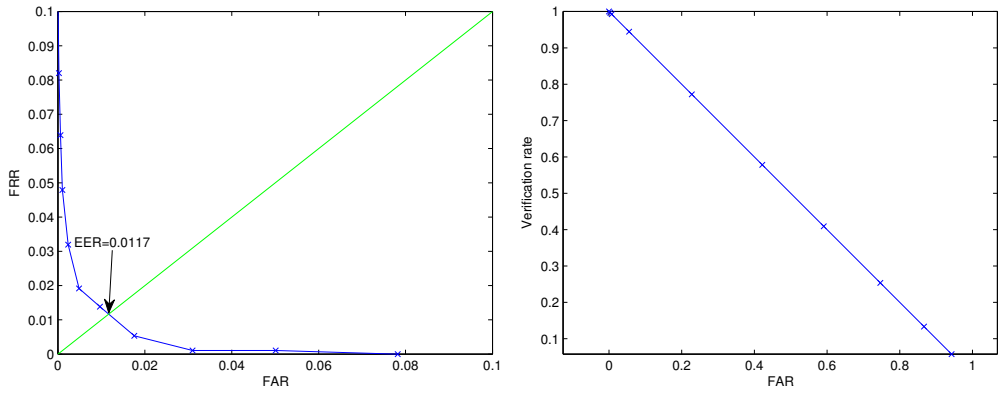


Figure 8.10: ROC curves for Fisherfaces (left) and PMCC (right) for 1000 sessions at 10×10 px. The linear nature of PMCC makes it unsuitable for deployment in any real world system, but its performance makes it a suitable for choice for algorithm performance testing.

8.4 Discussion

Presented in this chapter are two preliminary investigations into two of the final major parts of an operational 3D photometric stereo recognition system. The possibilities of using extended Haar-like features for both landmark detection and roll correction are demonstrated, as well as empirically defining a threshold for Fisherface recognition. The threshold is probably only applicable to this particular dataset but the calculation for a different dataset would be the same.

Outside the scope of this project and concerning the Photoface device itself

are issues involved with its use in the environment. For the majority of users the device will function well as long as they pass through in alignment with the camera. If they do not, then the light source angles will alter and lead to poor surface normal estimations. Therefore an improvement could be made which involves detecting the coordinates of the face and adjusting the light source vectors accordingly to improve the accuracy of the PS reconstruction. In the current system the light source unit vectors are calculated from a point at the centre of the camera's field of view and this is used for all reconstructions regardless of where the face is actually located. For this reason, the light source unit vectors are less accurate if the person walking through the device does not locate their face near the centre of the camera's field of view. The exact error caused by this inaccuracy is unknown, but amending the light source angles on a per person basis will improve the surface normal estimates. It is unfortunate that this shortcoming only came to light after the collection of the Photoface database, as it would have been straightforward to have stored the position of the clipped face in the whole image. Future work can learn from this and include the data.

Another issue concerning the device is how it deals with extremely tall or short people, or wheelchair bound persons which would probably trigger the device correctly, but the location of the face could be outside of the field of view of the camera. Two possible solutions for this are (1) to use two cameras and trigger sensors at different heights or (2) to increase the field of view of the camera. The first solution would work by using the most suitable camera depending on which sensor had been triggered e.g. if only the lower sensor was triggered, then the lower camera would be used. While this is a straightforward solution it would increase the cost of the equipment considerably as the camera is the most expensive piece of apparatus. Increasing the field of view is also straightforward and would provide an adequate solution so long as wide-angle lens distortions did not become evident and that the face remains large enough on the image to provide discriminating information for the later recognition process.

Currently the system is only designed to cope with one face in the captured image. If there are two or more, then only the first that is found by the face detection algorithm will be processed. If there are multiple heads it is very likely that they will obstruct the light sources and introduce shadowing which will have an adverse effect on PS. Therefore an additional step should be introduced to ensure that only one head is present in the field of view.

The last process in the pipeline diagram shown in Fig. 8.1 is for the verification step. It states that if the distance between the probe and gallery is within an acceptable threshold then the identity can be verified, and if it is not then the probe must be rejected or the option given to enrol the probe subject onto the system. When a subject is enrolled for the first time, a number of sessions must be captured if the Fisherface algorithm is to be used (interestingly, through experimentation on the Photoface surface normal data it has been found that the small sample size problem is not so apparent as it is when using conventional 2D data), so five captures would be sufficient. It is important that the expression of the subject in each captured session used for enrolment is the same. As it is possible to correctly identify a positive expression with reasonably high accuracy, and the fact that there is some research supporting its use for enrolment [21], a smile is probably a suitable choice. Once the captures are completed, the face space will need to be recalculated and the coefficients for each session stored in the database. There is some risk that with a sufficient number of enrolments, the threshold value will shift such that an unacceptable number of false positives occur. It would therefore be necessary to run a suite of tests to build up the ROC and check that the threshold falls within a certain distance of the existing one, or indeed replace it.

If the expression classification system is employed to reject captures that do not match the expected expression, then people with facial paralysis might be excluded from the system. For example, if the system expects a smiling face to enrol a subject, a person suffering from facial paralysis may not be able to

generate the symmetrical smile that the system has been trained to respond to. Obviously this relates to the training data employed, so either there would need to be examples of expressions from those with facial paralysis or an alternative mode of enrolment could be devised for those who are unable to produce expressions close enough to those expected by the system. Indeed any norm-based system (of which Fisherface is one) would likely fail on abnormal faces caused by injuries or syndromes, and it is unknown how the Viola and Jones face detection cascade would function, but one would expect it to also fail in extreme cases.

There are other operational caveats which would also need to be addressed. People's appearance alters with time, either gradually though ageing, or through sudden changes such as facial hair, sunglasses, spectacles, wounds, bruising and swelling or hairstyles. In the case of ageing, re-enrolment could occur once a year, or alternatively on a successful verification the captured image could be used to update that subject's sessions. In either case, the face space would need to be recalculated and all enrolled subjects' coefficients updated. For the other cases, there is no readily available solution; these are truly difficult problems. Attempts to deal with the occlusion type problems include using patch or region based approaches and have proven to be effective [193], but to date little or perhaps no work has focused on dealing with injuries such as bruising or swelling. Facial hair such as moustaches are seen by humans as being distinguishing features, but are treated as quite the opposite by machine vision systems, and facial hair invariant representations are sought. Other distinguishing features which are useful for humans are scars and moles on the face, which fit in well with Unnikrishnan's [182] idea. Due to the rarity of such features though (which is also their defining feature!) research into these is considerably more difficult. Burton *et al.* demonstrated that a more accurate representation of a subject may improve recognition accuracy by using an image made by averaging many different photographs of the subject [97]. This would then allow some robustness to different lighting. What they also suggest is using PCA to model the intra-subject variance

along with this central tendency in order to represent the dimensions along which a subject can look different on separate occasions.

There are also some issues concerning the level of cooperation required by the subject which would affect how well an automated face recognition system works. While the overall design of the Photoface device means that the level of cooperation required is minimal i.e. walking through the archway while looking at the camera, a higher level of cooperation is required for the actual algorithm to function correctly. If the subject has an incentive to pass the verification (for example to gain entry to their home, computer or smart-phone) then they are likely to be as cooperative as possible and all should work as intended. However, in situations where being recognised is a disadvantage, minor changes to expression, pose and appearance via glasses or make-up would prevent verification to the benefit of the individual. Indeed automated systems are unlikely to be able to cope with all the varieties of variance for the foreseeable future and so are particularly unsuited to situations where subject cooperation is likely to be limited in any way. However for those that have an incentive to be authenticated, then such systems should work well with 3D data providing an additional level of security over 2D which can be tricked into authenticating a photograph of an enrolled individual.

Being able to incorporate some sort of label for faces that provides a description e.g. spikey face, looks a bit like celebrity 'X' etc. would also prove very useful for police work. Kumar *et al.* [109] have recently researched this and developed a system that can automatically add metadata to images such as gender, age, jaw shape and nose size which can then be used at least to narrow down a search or even uniquely identify an individual in a way which is arguably more intuitively human. This sort of metadata could be employed by Photoface in deployments where recognition is required using far larger databases in exactly this sort of way – as a preliminary step to reduce the search space.

This chapter has looked at automatic feature detection using Haar-like fea-

tures on surface normals and has empirically determined a threshold which can achieve a high verification rate with a low false positive rate. It has also proposed a method for roll correction that uses features extracted from the y-component of the surface normals based on the horizontal structure and symmetrical nature of the face. Proposals to overcome some limitations in the Photoface device are given and some wider aspects of face recognition are discussed; of which more consideration will be given in the next chapter. It brings to an end the experimental chapters of this thesis, and in the next and final chapter the major findings and implications of the thesis as a whole are summarised.

Chapter 9

Conclusions, Limitations and Future Work

This thesis has examined face recognition using surface normals captured via PS and focussed on methods of dimensionality reduction which are inspired by the HVS. As part of this, two databases with metadata have been created, one of which has been made available to the research community. This thesis has shown that PS lends itself well to face recognition and a future device would benefit from using NIR light sources as these provide a more accurate reconstruction. Recognition rates are not decreased when resolution is decreased to as little as 10×10 px for the raw surface normal data. The reason for this is likely that the spatial frequencies which code the major structural features are preserved and some evidence exists that these frequencies are important in human face recognition. The low resolution is beneficial both for storage and processing requirements. Unusual, or outlying pixel values contain disproportionately more discriminatory information than more common values as do pixel locations in the face which have the highest variance. This lends itself well to the concept of caricaturing a face. Problems with this approach arise when the data contains expressions, as the expressions themselves (rather than the subject identity) are responsible for the largest variance. By using the variance approach to locate and exclude

the pixels which vary the most for a given expression in the database population, recognition rates can be improved. The original Haar-like features as used by Viola & Jones are unlikely to be optimal for surface normal data, but it may be possible to design new ones which work well such as the one presented to locate the nose tip. Additionally a straightforward method of roll correction is suggested which used the symmetry of the output of the new feature.

For reference these are summarised in Section 9.

Summary of Contributions

This section briefly outlines the major novel findings and contributions of this thesis by chapter.

- Chapter 3 showed that the Photoface device captures accurate surface normal data of faces by comparing captured sessions with a ground truth model. It was also shown that using NIR light sources is slightly more accurate than using visible light sources and this it was suggested that this finding can explained in terms increased sub-surface scatter of NIR.

- Chapter 4 introduces two novel databases created as part of thesis, one of which is now publicly available to the research community. The Photoface database is the largest publicly available PS database of faces and is unique in that it was captured under relatively unconstrained conditions. Benchmarking experiments are then performed on a subset of frontal expressionless sessions both to prove that the data is suitable for face recognition and to provide performance against which algorithms in later chapters can be compared.

- Humans are excellent at recognising familiar faces and areas of the human brain appear to be dedicated to processing faces. Chapters 5 and 6 investigate two idiosyncrasies of human face recognition in an attempt to

incorporate them into an automated system. It is shown that recognition performance is robust to reducing the resolution of the images to as little as 10×10 px on both the Photoface and FRGCv2.0 (the most commonly used in the literature) 3D databases. Taking caricaturing as inspiration, Chapter 6 uses a percentile based and a variance based method to select pixel locations which are shown to contain proportionately more discriminatory information. Removing up to 90% of the data leads to recognition performance being affected by less than 10%. These findings have important implications for automated systems as they offer direct methods of dimensionality reduction which means improved efficiency through reduced computation and storage needs. In turn this likely means reduced cost.

- Expression is commonly cited as a confounding factor to 3D face recognition. Chapter 7 explores expression classification on the novel 3DE-VISIR database and finds that *Happy* expressions are the most reliable to classify (up to 90% accuracy), and that this performance remains when resolution is lowered to 10×10 px. Being able to classify expressions is useful in Human Computer Interaction (HCI) applications, but it has also been suggested to be useful for recognition as smiling faces have been demonstrated to be easier to recognise. Additionally this chapter showed that by applying the variance based approach in the previous chapter, pixel locations which encode expression can be identified and, in turn removed in order to (modestly) boost subject recognition performance.

- Chapter 8 brings the findings of the previous chapters together and identifies two areas which would need to be developed as part of fully automated PS face recognition system: automatic face alignment and defining a threshold so that imposters can be rejected. It builds upon the work of Viola and Jones into using a rapid cascade of Haar-like features to propose and test an extension of these features to one specifically designed for surface nor-

mals to detect caps/cups so that the nose-tip and nasion can be localised. This in turn was found to generate a bar-code representation which lends itself well explaining a number of findings relating to human face recognition as well as providing a simple means of roll correction. Finally through the use of ROC curves a threshold is determined which allows the system to perform at over 98% accuracy at a FAR=0.01.

Limitations and Future Work

The findings of this thesis are useful contributions to face recognition. The Photoface device is a convenient and relatively cheap 3D capture device and the recognition efficiency can be improved by simply resizing the surface normal estimates. Biometrics are likely to be used increasingly as methods become more reliable, affordable and perhaps most importantly, public acceptance grows as the convenience offered becomes apparent. Although 3D face recognition overcomes many problems with 2D (or can theoretically do so) it is likely not the panacea to biometric security. While it is the most convenient amongst the commonest methods (iris, retina, palmprint, fingerprint) it is also undoubtedly the easiest to break. While facial appearance for the most part does not change quickly, a disguise can be produced to prevent the system from recognising an individual either temporarily through cosmetics or permanently through plastic surgery, arguably making the system unsuitable for watchlist checking scenarios. For this reason, face recognition is better suited to a verification scenario (e.g. “I say I am ’X’. Am I”?) as opposed to a recognition or surveillance scenario. Perhaps if combined with other modalities a more robust system could be developed.

A major problem for most 2D systems is that they can be fooled by a photograph of an enrolled individual. One solution to overcome this would be to require a video sequence of the face in motion (perhaps speaking a phrase) but even then it could get fooled by a video on a screen. This isn’t likely to be such a

problem with 3D face recognition and if it were to additionally include some sort of motion capture it would represent a very unique feature of 3D which makes it almost unbeatable in theory – fake 3D models of realistically moving (elastic) faces are not readily available. Arguably the best currently available are those created by Hollywood or Hiroshi Ishiguro's Gemini robot. Going to those lengths to compromise a system is very unlikely; there will almost certainly be a simpler alternative.

When capturing the face, Photoface currently relies on visible light sources and the Lambertian assumption. As part of this thesis it was shown that skin reflectance is more Lambertian under NIR. Future work will therefore focus on converting the Photoface device to work with NIR lamps and is in fact the focus of the EPSRC funded Photoskin project which additionally aims to overcome the simplifying assumption that skin is Lambertian by generating a "per session" reflectance model which will improve reconstruction accuracy further.

Preprocessing the faces crops and aligns them with a baseline set of coordinates via an affine transform. The reason for this to ensure that the major features of the face are in approximately the same place for each capture so that any following statistical analysis will not be drawn into irrelevant areas of variance (e.g. if the landmarks were not aligned then this would be a major source of variance, and if the faces were not closely cropped, background regions would come into play). While the effect of these is not shown by the work in this thesis (and the cropped and aligned data is shown to give good results), it is likely that some useful, discriminatory data especially for unusual features e.g. a long nose, or close-knit eyes will be lost. Future work should concentrate on testing the effects of different alignment algorithms in order to see how large these effects are.

Although 3D capture is often stated to overcome the problems of pose with 2D face recognition, it is not so straightforward with PS capture. The fact that there is only one camera means that a full 3D rendering of the object as whole is not created, only the surface(s) that face the camera (and therefore as stated in

the introductory chapter it is technically not 3D but 2.5D). This limits the amount of pose correction available but for reasonable small deviations from frontal this should be acceptable. However for wider deviations, occlusions and foreshortening will come into effect meaning that there will be far less data from which to reconstruct the surface before attempting pose correction. For example if a face is presented at a 3/4 or portrait pose, the nose will occlude a region of cheek and little of that side of the face will be visible. If a face is presented with a high pitch angle (e.g. the nose is pointing upwards) then there will be very little of the face surface available to reconstruct the surface with. With a full 3D system, neither of these situations would cause so much of a problem because the scan would capture the entire object (although pose alignment is still not entirely trivial even in full 3D). In order to overcome the problem with a 2.5D system, one could either employ a pre-capture step to ensure that the subject is in a suitable pose, or in the case of a face being presented in 3/4 pose, perhaps exploit the symmetry of the face to hallucinate any missing or occluded data.

The previous chapter presented some preliminary work into automated feature detection. The nose tip detector appears to work well but further research is required to locate other features which are also required for alignment or pose correction such as the lateral canthi, nasion etc. Perhaps if NIR light sources were used for PS, the raw images could be used for eye-detection owing to the *bright-eye* effect where the pupil is highly reflective to NIR.

One of the very first steps in the current pipeline employs the face detection cascade of Viola and Jones. If this fails, then no face recognition will occur. A Masters student recently showed that by combining cosmetics and hair styling to disrupt key areas corresponding to features used by the cascade, he was able to prevent a face being detected [87]. While not inconspicuous as they currently stand and therefore likely to attract greater attention, they do highlight a simple method to defeat a complex technology.

Two of the contributions of this Ph.D. are reducing dimensionality using direct

methods of resolution reduction and variance analysis. Both of these methods individually show the importance of a holistic representation for recognition (as opposed to parts/features in isolation). Further work to find whether there are better pixel locations than those suggested by the variance or percentile based approaches could be employed e.g. by using a simulated annealing approach to stochastically locate such features. It would be satisfying to find a better way of combining the two approaches but the reduced resolution produces such an effective reduction that further processing with the variance inclusion criterion are far less powerful. Nonetheless, both approaches deliver real effects, indicating that following a human approach in this instance pays off. It would be interesting to see whether the effects are limited to faces or whether they generalise to other object categories.

The use of low resolution images is useful for both storage and processing efficiency and the fact that raw surface normals are employed and shown to be the best representation for recognition means that no additional processing is required for surface estimation. Unlike most 2D systems, Photoface cannot be fooled by a photograph or make-up unless prosthetics are employed to change the actual 3D shape. Future work could look at ways to ensure that the presented surface is that of a living person, perhaps using thermal infrared.

The system has been shown to work well under near ideal situations in a real-world environment (neutral expressions and front facing) for 61 individuals. Importantly it has demonstrated that the system works on data that has been captured while the subject uses it realistically and "casually". This number is in the employee range of many SMEs. The limits of face recognition algorithms are generally unknown but this is an area which would be very beneficial to industry and governments. Data collection on this sort of scale would be logistically difficult so it may be better to create synthetic faces using modelling software (e.g. FaceGen Modeller [4], or MakeHuman [5]) to increase the gallery size. Alternatively, if one has the finances and inclination, the data could be manually collected. India has

started enrolling every one of its 1.2 billion population into the Unique Identification Authority of India (UIDAI) program which assigns each individual a unique number and records their irises, ten fingerprints and a photograph of their face. They expect to have 600 million people enrolled by 2014 [6].

This thesis has almost exclusively relied on statistical methods for face recognition. Sirovich & Kirby's realisation that the N$\times$N problem could be presented as M$\times$M made possible the multitude of work that followed in this area. However, if the numbers of subjects enrolled is sufficiently large so that M $>>$ N, then the original N$\times$N problem becomes a M$\times$M problem, and the likelihood of the matrix being non-invertible is large. As they stand, their use must be restricted to databases where M $<<$ N.

Biometrics and face recognition in particular have many benefits and we are likely to see them being used more and more as the convenience they offer become apparent to the public, and used more in office settings as the cost decreases. This thesis has investigated 3D face recognition from PS, presenting two novel methods of dimensionality reduction, a novel method of improving robustness to expression variance and making available the largest PS annotated face database as well as showing that PS using NIR is more accurate for face reconstruction than visible light. In the well studied area of face recognition, these represent useful contributions that have a direct impact with regards to implementations and suggest many areas for future research.

References

[1] VIVID 910-Non-contact 3D Digitizer | KONICA MINOLTA. http://www.konicaminolta.com/instruments/products/3d/non-contact/vivid910/ [accessed Dec. 2011].

[2] 3D Face Webportal. http://www.3dface.org/ [accessed: Dec. 2011], December 2011.

[3] 3dMDface System. http://www.3dmd.com/3dmdface.html [accessed: Dec. 2011], December 2011.

[4] FaceGen Modeller. http://www.facegen.com/modeller.htm [accessed: Dec. 2011], December 2011.

[5] MakeHuman. http://www.makehuman.org [accessed: Dec. 2011], December 2011.

[6] Unique Identification Authority of India. https://en.wikipedia.org/wiki/Unique_Identification_Authority_of_India [accessed: Jan 2012], 2012.

[7] T. Ahonen, A. Hadid, and M. Pietikäinen. Face recognition with local binary patterns. In *Proc. European Conference on Computer Vision*, pp. 469–481, 2004.

[8] T. Ahonen, A. Hadid, and M. Pietikäinen. Face Description with Local Binary Patterns: Application to Face Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 2037–2041, 2006.

[9] R. R. Anderson and J. A. Parrish. The Optics of Human Skin. *Journal of Investigative Dermatology*, 77(1):13–19, 1981.

[10] G. A Atkinson and M. L. Smith. Facial Feature Extraction and Change Analysis Using Photometric Stereo. In *Proc. Pattern Recognition and Image Analysis*, pp. 96–103, 2009.

[11] S. Baker and T. Kanade. Limits on Super-Resolution and How to Break Them. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(9):1167–1183, 2002.

[12] D. J. Bardsley and L. Bai. 3D Surface Reconstruction and Recognition. In *Proc. Biometric Technology for Human Identification*, vol. 6539, 2007.

[13] S. Barsky and M. Petrou. The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(10):1239–1252, 2003.

[14] M. S. Bartlett. *Face image analysis by unsupervised learning*. Kluwer Academic Publishers, 2001.

[15] M.S. Bartlett, J.R. Movellan, and T.J. Sejnowski. Face recognition by independent component analysis. *IEEE Transactions on Neural Networks*, 13(6):1450–1464, 2002.

[16] P. N Belhumeur, D. J Kriegman, and A. L Yuille. The bas-relief ambiguity. *International Journal of Computer Vision*, 35(1):33–44, 1999.

[17] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.

[18] P. J Benson and D. I Perrett. Perception and Recognition of Photographic Quality Facial Caricatures: Implications for the Recognition of Natural Images. *European Journal of Cognitive Psychology*, 3(1):105–135, 1991.

[19] P. J. Besl and H. D. McKay. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992.

[20] C. Beumier. 3D Face Recognition. In *Proc. IEEE International Conference on Industrial Technology*, pp. 369–374, 2006.

[21] J. R. Beveridge, G. H. Givens, P. J. Phillips, and B. A. Draper. Factors that influence algorithm performance in the Face Recognition Grand Challenge. *Computer Vision and Image Understanding*, 113(6):750–762, 2009.

[22] D. J Beymer. Face recognition under varying pose. *AI Lab Technical Report 1461, MIT*, 1993.

[23] V. Blanz and T. Vetter. A morphable model for the synthesis of 3D faces. In *Proc. Computer graphics and interactive techniques*, pp. 187–194, 1999.

[24] V. Blanz and T. Vetter. Face Recognition Based on Fitting a 3D Morphable Model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9):1063–1074, 2003.

[25] W. W. Bledsoe. Man-machine facial recognition. *Panoramic Research Inc., Palo Alto, CA*, 1966.

[26] C. Boehnen and P. Flynn. Accuracy of 3D scanning technologies in a face scanning scenario. In *Proc. 3-D Digital Imaging and Modeling*, pp. 310–317, 2005.

[27] B. J. Boom, G. M. Beumer, L. J. Spreeuwers, and R. N. J. Veldhuis. The effect of image resolution on the performance of a face recognition system.

In *Proc. International Conference on Automation, Robotics, Control and Vision*, pp. 1–6, 2006.

[28] A. M. Bronstein, M. M. Bronstein, and R. Kimmel. Three-Dimensional Face Recognition. *International Journal of Computer Vision*, 64(1):5–30, 2005.

[29] V. Bruce, Z. Henderson, K. Greenwood, P. J. B. Hancock, A. M. Burton, and P. Miller. Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied*, 5(4):339–360, 1999.

[30] V. Bruce and A. Young. Understanding face recognition. *British journal of psychology*, 77(3):305–327, 1986.

[31] R. Brunelli and T. Poggio. Face recognition: features versus templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(10):1042–1052, 1993.

[32] A. M Burton, V. Bruce, and P. J. B. Hancock. From pixels to people: A model of familiar face recognition. *Cognitive Science: A Multidisciplinary Journal*, 23(1):1–31, 1999.

[33] K. Chang, K. Bowyer, and P. Flynn. Face recognition using 2D and 3D facial data. In *ACM Workshop on Multimodal User Authentication*, pp. 25–32, 2003.

[34] C. F. Chen, Y. Tseng, and C. Chen. Combination of PCA and Wavelet Transforms for Face Recognition on 2.5D Images. In *Proc. Image and Vision Computing*, pp. 343–347, 2003.

[35] S. Chopra, R. Hadsell, and Y. LeCun. Learning a Similarity Metric Discriminatively, with Application to Face Verification. In *Proc. Computer Vision and Pattern Recognition*, 2005.

[36] J. Cockburn, M. Bartlett, J. Tanaka, J. Movellan, M. Pierce, and R. Schultz. SmileMaze: A Tutoring System in Real-Time Facial Expression Perception

and Production in Children with Autism Spectrum Disorder. In *Proc. International Conference on Automatic Face and Gesture Recognition, Workshop on Facial and Bodily expression for Control and Adaptation of Games*, 2008.

[37] D. Colbry and G. Stockman. The 3DID face alignment system for verifying identity. *Image and Vision Computing*, 27(8):1121–1133, 2009.

[38] D. Colbry, G. Stockman, and A. Jain. Detection of Anchor Points for 3D Face Verification. In *Proc. Computer Vision and Pattern Recognition*, vol. 93, pp. 94, 2005.

[39] E. N Coleman and R. Jain. Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry. *Computer Graphics and Image Processing*, 18(4):309–328, 1982.

[40] C. Conde, Á. Serrano, and E. Cabello. Multimodal 2D, 2.5 D & 3D Face Verification. In *Proc. International Conference on Image Processing*, pp. 2061–2064, 2006.

[41] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001.

[42] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active Shape Models - Their Training and Application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[43] N.P. Costen, D.M. Parker, and I. Craw. Effects of high-pass and low-pass spatial filtering on face identification. *Attention, Perception, & Psychophysics*, 58(4):602–612, 1996.

[44] D. Cox and N. Pinto. Beyond simple features: A large-scale feature search approach to unconstrained face recognition. In *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 8–15, 2011.

[45] K. Crookes and E. McKone. Early maturity of face recognition: no childhood development of holistic processing, novel face encoding, or face-space. *Cognition*, 111(2):219–247, 2009.

[46] J. Czyz and L. Vandendorpe. Evaluation of LDA-based face verification with respect to available computational resources. In *Proc. International Workshop on Pattern Recognition in Information Systems*, 2002.

[47] G.W Dailey, M. N Cottrell, and T. A Busey. Facial Memory is Kernel Density Estimation (Almost). In *Proc. Advances in neural information processing systems*, vol. 11, pp. 24, 1999.

[48] Steven C. Dakin and Roger J. Watt. Biological “bar codes” in human faces. *Journal of Vision*, 9(4), 2009.

[49] J. Daugman. How Iris Recognition Works. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):21–30, January 2004.

[50] John G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A*, 2(7):1160–1169, 1985.

[51] G. M. Davies, H. D. Ellis, and J. W. Shepherd. Face recognition accuracy as a function of mode of representation. *Journal of Applied Psychology*, 63(2):180–187, 1978.

[52] K. A Deffenbacher, T. Vetter, J. Johanson, and A. J O’Toole. Facial aging, attractiveness, and distinctiveness. *Perception*, 27:1233–1244, 1998.

[53] W. Deng, J. Hu, J. Guo, W. Cai, and D. Feng. Emulating biological strategies for uncontrolled face recognition. *Pattern Recognition*, 43(6):2210–2223, 2010.

[54] G. Donato, M. S Bartlett, J. C Hager, P. Ekman, and T. J Sejnowski. Classifying facial actions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(10):974–989, 1999.

[55] C. Donner and H. W. Jensen. Light diffusion in multi-layered translucent materials. *ACM Transactions on Graphics*, 24(3):1032–1039, 2005.

[56] J. Dowdall, I. Pavlidis, and G. Bebis. Face Detection in the Near-IR Spectrum. *Image and Vision Computing*, 21(7):565–578, 2003.

[57] B. A. Draper, K. Baek, M. S. Bartlett, and J. R. Beveridge. Recognizing faces with PCA and ICA. *Computer Vision and Image Understanding*, 91(1-2):115–137, 2003.

[58] R. O Duda, P. E Hart, and D. G Stork. *Pattern classification*. Citeseer, 2nd edition, 2001.

[59] G. J. Edwards, T. F. Cootes, and C. J. Taylor. Face recognition using active appearance models. *Proc. European Conference on Computer Vision*, pp. 581–595, 1998.

[60] I. A Essa and A. P Pentland. Facial expression recognition using a dynamic model and motion energy. In *Proc. International Conference on Computer Vision*, pp. 360, 1995.

[61] M. J Farah, K. D Wilson, H. Maxwell Drain, and J. R Tanaka. The inverted face inversion effect in prosopagnosia: Evidence for mandatory, face-specific perceptual mechanisms. *Vision Research*, 35(14):2089–2093, 1995.

[62] R. A Fisher. The use of multiple measures in taxonomic problems. *Annals of Eugenics*, 7:179–188, 1936.

[63] D. A. Forsyth and J. Ponce. *Computer Vision: A modern approach*. Prentice Hall Professional Technical Reference, 2002.

[64] R. T Frankot and R. Chellappa. A method for enforcing integrability in shape from shading algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(4):439–451, 1988.

[65] C. Fredembach, N. Barbusica, and S. Süsstrunk. Combining visible and near-infrared images for realistic skin smoothing. In *Proc. IS&T/SID 17th Color Imaging Conference*, 2009.

[66] C. D Frowd, V. Bruce, D. Ross, A. H McIntyre, and P. J.B Hancock. An application of caricature: how to improve the recognition of facial composites. *Visual cognition*, 15:954–984, 2007.

[67] N. Furl, P. J. Phillips, and A. J. O'Toole. Face recognition algorithms and the other-race effect: computational mechanisms for a developmental contact hypothesis. *Cognitive Science*, 26(6):797–815, 2002.

[68] R. E Galper. Recognition of faces in photographic negative. *Psychonomic Science*, 19(4):207—208, 1970.

[69] A. S. Georghiades. Recovering 3-D shape and reflectance from a small number of photographs. In *Proc. Eurographics workshop on Rendering*, pp. 230–240, 2003.

[70] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman. From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):643–660, 2001.

[71] A. Ghosh, T. Hawkins, P. Peers, S. Frederiksen, and P. Debevec. Practical modeling and acquisition of layered facial reflectance. *ACM Transactions on Graphics*, 27(5):1–10, 2008.

[72] V. Goffaux, J. Peters, J. Haubrechts, C. Schiltz, B. Jansma, and R. Goebel. From Coarse to Fine? Spatial and Temporal Dynamics of Cortical Face Processing. *Cerebral Cortex*, 21(2):467 –476, 2011.

[73] B. Gökberk, M. O. İrfanoğlu, and L. Akarun. 3D shape-based face representation and feature extraction for face recognition. *Image and Vision Computing*, 24(8):857–869, 2006.

[74] B. Gökberk, A. Salah, and L. Akarun. Rank-based decision fusion for 3D shape-based face recognition. In *Proc. Audio-and Video-Based Biometric Person Authentication*, pp. 119–130, 2005.

[75] G. G. Gordon. Face recognition from frontal and profile views. *Proc. International Workshop on Automatic Face- and Gesture-Recognition*, pp. 47—52, 1995.

[76] G.G. Gordon. Face recognition based on depth and curvature features. In *Proc. Computer Vision and Pattern Recognition*, pp. 808—810, 1992.

[77] C.C. Goren, M. Sarty, and P.Y.K. Wu. Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics*, 56(4):544, 1975.

[78] Graphics Lab, Beijing University of Technology. The BJUT-3D Large-Scale Chinese Face Database. Technical report, Technical report, 2005.

[79] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-PIE. *Image and Vision Computing*, 28(5):807–813, 2010.

[80] R. Gross, J. Shi, and J. Cohn. Quo vadis face recognition. *In Third Workshop on Empirical Evaluation Methods in Computer Vision*, pp. 119—132, 2001.

[81] S. Gupta, J.K. Aggarwal, M.K. Markey, and A.C. Bovik. 3D Face Recognition Founded on the Structural Diversity of Human Faces. In *Proc. Computer Vision and Pattern Recognition*, pp. 1–7, 2007.

[82] P. W Hallinan, G. G Gordon, A. L Yuille, P. Giblin, and D. Mumford. *Two-and three-dimensional patterns of the face*. AK Peters, Ltd., 1999.

[83] P. J.B Hancock, V. Bruce, and A. M Burton. Recognition of unfamiliar faces. *Trends in Cognitive Sciences*, 4(9):330–337, 2000.

[84] P. J.B Hancock, A. M Burton, and V. Bruce. Face processing: Human perception and principal components analysis. *Memory & Cognition*, 24(1):26–40, 1996.

[85] M. F. Hansen, G. A. Atkinson, L. N. Smith, and M. L. Smith. 3D face reconstructions from photometric stereo using near infrared and visible light. *Computer Vision and Image Understanding*, 114(8):942–951, August 2010.

[86] L.D. Harmon. The recognition of faces. *Scientific American*, (229):70–82, 1973.

[87] A. Harvey. CV Dazzle: Open-Source Camouflage From Computer Vision by Adam Harvey. http://cvdazzle.com/ [accessed Nov. 2011], 2011.

[88] C. Hernández, G. Vogiatzis, and R. Cipolla. Shadows in three-source photometric stereo. In *Proc. European Conference on Computer Vision*, pp. 290–303, 2008.

[89] T. Heseltine. The UOY 3D Face Database. http://www-users.cs.york.ac.uk/˜nep/research/3Dface/tomh/3DFaceDatabase.html [accessed: Dec. 2011].

[90] B. K. P. Horn. Shape from shading: A method for obtaining the shape of a smooth opaque object from one view. *Ph.D Thesis, MIT*, 1970.

[91] B. K. P. Horn. Extended gaussian images. *Proceedings of the IEEE*, 72(12):1671–1686, 1984.

[92] D. Huang, M. Ardabilian, Y. Wang, and L. Chen. A novel geometric facial representation based on multi-scale extended local binary patterns. In *Proc. IEEE International Conference on Automatic Face & Gesture Recognition*, pp. 1–7, 2011.

[93] G. B Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. *University of Massachusetts, Amherst, Technical Report 07*, 49:1, 2007.

[94] D. H Hubel and T. N Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160(1):106, 1962.

[95] D. H. Hubel and T. N. Wiesel. Sequence regularity and geometry of orientation columns in the monkey striate cortex. *The Journal of Comparative Neurology*, 158(3):267–293, 1974.

[96] M. Hüsken, M. Brauckmann, S. Gehlen, and C. Von der Malsburg. Strategies and benefits of fusion of 2D and 3D face recognition. In *IEEE workshop on face recognition grand challenge experiments*, pp. 174, 2005.

[97] Rob Jenkins and A. Mike Burton. Stable face representations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1571):1671 –1683, June 2011.

[98] A. Johnston, H. Hill, and N. Carman. Recognising faces: effects of lighting direction, inversion, and brightness reversal. *Perception*, 21(3):365–375, 1992.

[99] I. A Kakadiaris, G. Passalis, G. Toderici, M. N Murtuza, Y. Lu, N. Karampatziakis, and T. Theoharis. Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):640—649, 2007.

[100] T. Kanade. *Computer recognition of human faces*. Birkhauser, Basel, Switzerland and Stuttgart, Germany, 1973.

[101] T. Kanade, J. F Cohn, and Y. Tian. Comprehensive database for facial expression analysis. In *Proc. IEEE International Conference on Automatic Face & Gesture Recognition*, pp. 46–53, 2000.

[102] N. Kanwisher, J. McDermott, and M.M. Chun. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience*, 17(11):4302, 1997.

[103] M. Kawulok, J. Wu, and E. R Hancock. Supervised relevance maps for increasing the distinctiveness of facial images. *Pattern Recognition*, 44(9):929—939, 2011.

[104] S. C. Kee, K. M. Lee, and S. U. Lee. Illumination invariant face recognition using photometric stereo. *IEICE Transactions on Information And Systems E Series D*, 83(7):1466–1474, 2000.

[105] M. D. Kelly. *Visual identification of people by computer. Techical Report AI-130*. Stanford AI Project, Stanford, CA, 1970.

[106] A. Kolb, E. Barth, R. Koch, and R. Larsen. Time-of-Flight Cameras in Computer Graphics. In *Computer Graphics Forum*, vol. 29, pp. 141–159, 2010.

[107] S. G. Kong, J. Heo, B. R. Abidi, J. Paik, and M. A. Abidi. Recent advances in visual and infrared face recognition—a review. *Computer Vision and Image Understanding*, 97(1):103–135, 2005.

[108] S Krishna. Open CV Viola-Jones Face Detection in Matlab. http://www.mathworks.com/matlabcentral/fileexchange/19912 [accessed: Dec. 2011].

[109] N. Kumar, A. Berg, P. N Belhumeur, and S. Nayar. Describable Visual Attributes for Face Verification and Image Search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(10):1962–1977, 2011.

[110] M. Lades, J. C. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R. P. Wurtz, and W. Konen. Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers*, 42(3):300–311, 1993.

[111] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[112] H. Lee and D. Kim. Expression-invariant face recognition by facial expression transformations. *Pattern Recognition Letters*, 29(13):1797–1805, 2008.

[113] J. C. Lee and E. Milios. Matching range images of human faces. In *Proc. International Conference on Computer Vision*, pp. 722–726, 1990.

[114] K. Lee, G. Byatt, and G. Rhodes. Caricature effects, distinctiveness, and identification: Testing the face-space framework. *Psychological Science*, 11(5):379, 2000.

[115] K. C Lee, J. Ho, and D. J Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(5):684—698, 2005.

[116] Y. Lee and D. Marshall. Curvature based normalized 3D component facial image recognition using fuzzy integral. *Applied Mathematics and Computation*, 205(2):815–823, 2008.

[117] L. Li and C.S. Ng. Rendering human skin using a multi-layer reflection model. *International Journal of Mathematics and Computers in Simulation*, 3:44–53, 2009.

[118] S. Z. Li, R. F. Chu, S. C. Liao, and L. Zhang. Illumination Invariant Face Recognition Using Near-Infrared Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):627–639, 2007.

[119] S. Z Li and A. K Jain. *Handbook of face recognition*. Springer Verlag, 2005.

[120] S. Z. Li, L. Zhang, S. C. Liao, X. X. Zhu, R. F. Chu, M. Ao, and R. He. A Near-Infrared Image Based Face Recognition System. In *Proc. IEEE International Conference on Automatic Face & Gesture Recognition*, pp. 455–460, 2006.

[121] R. Lienhart and J. Maydt. An Extended Set of Haar-like Features for Rapid Object Detection. In *IEEE International Conference on Image Processing*, vol. 1, pp. 900 – 903, 2002.

[122] P. Lucey, J. F Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews. The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp. 94–101, 2010.

[123] Y. M Lui, D. Bolme, B. A Draper, J. R Beveridge, G. Givens, and P. J Phillips. A meta-analysis of face recognition covariates. In *Proc. IEEE International Conference on Biometrics: Theory, applications and systems*, pp. 139–146, 2009.

[124] R. S. Malpass and J. Kravitz. Recognition for faces of own and other race. *Journal of Personality and Social Psychology*, 13(4):330–334, 1969.

[125] T. Malzbender, B. Wilburn, D. Gelb, and B. Ambrisco. Surface enhancement using real-time photometric stereo and reflectance transformation. In *Proc. Eurographics Symposium on Rendering*, 2006.

[126] B. S. Manjunath, R. Chellappa, and C. von der Malsburg. A feature based approach to face recognition. In *Proc. Computer Vision and Pattern Recognition*, pp. 373–378, 1992.

[127] A. M. Martínez and A. C. Kak. PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2):228—233, 2001.

[128] R. Mauro and M. Kubovy. Caricature and face recognition. *Memory & Cognition*, 20(4):433–440, 1992.

[129] K. Messer, J. Kittler, M. Sadeghi, M. Hamouz, A. Kostin, F. Cardinaux, S. Marcel, S. Bengio, C. Sanderson, and N. Poh. Face authentication test on the BANCA database. In *Proc. International Conference on Pattern Recognition*, pp. 523–532, 2004.

[130] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre. XM2VTSDB: The extended M2VTS database. In *International conference on Audio and Video-based Biometric Person Authentication*, vol. 964, pp. 965–966, 1999.

[131] A. S. Mian, M. Bennamoun, and R. Owens. An Efficient Multimodal 2D-3D Hybrid Approach to Automatic Face Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(11):1927–1943, 2007.

[132] E. Mingolla and J. T. Todd. Perception of solid shape from shading. *Biological Cybernetics*, 53(3):137—151, 1986.

[133] N. Miura, A. Nagasaka, and T. Miyatake. Feature extraction of finger-vein patterns based on repeated line tracking and its application to personal identification. *Machine Vision and Applications*, 15(4):194–203, 2004.

[134] A.A. Mohammed, R. Minhas, Q.M. Jonathan Wu, and M.A. Sid-Ahmed. Human face recognition based on multidimensional PCA and extreme learning machine. *Pattern Recognition*, 44(10-11):2588–2597, 2011.

[135] A. B. Moreno and A. Sanchez. Gavabdb: A 3d face database. In *Proc. Workshop on Biometrics on the Internet*, pp. 75–80, 2004.

[136] A. B. Moreno, A. Sanchez, J. F. Velez, and F. J. Diaz. Face recognition using 3D surface-extracted descriptors. In *Irish Machine Vision and Image Processing Conference*, vol. 2, 2003.

[137] C. H. Morimoto, D. Koons, A. Amir, and M. Flickner. Pupil detection and tracking using multiple light sources. *Image and Vision Computing*, 18(4):331–335, 2000.

[138] M. Moscovitch, G. Winocur, and M. Behrmann. What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. *Journal of Cognitive Neuroscience*, 9(5):555–604, 1997.

[139] H. Nejati and T. Sim. A study on recognizing non-artistic face sketches. In *IEEE Workshop on Applications of Computer Vision*, pp. 240–247, 2011.

[140] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.

[141] B. A Olshausen and D. J Fteld. Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision research*, 37(23):3311–3325, 1997.

[142] M. Oren and S.K. Nayar. Generalization of the Lambertian model and implications for machine vision. *International Journal of Computer Vision*, 14(3):227–251, 1995.

[143] A. J. O'Toole, T. Vetter, H. Volz, and E. M. Salter. Three-dimensional caricatures of human heads: distinctiveness and the perception of facial age. *Perception*, 26(6):719 – 732, 1997.

[144] G. Pan, S. Han, Z. Wu, and Y. Wang. 3D Face Recognition using Mapped Depth Images. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, pp. 175, 2005.

[145] D.M. Parker and N.P. Costen. One extreme or the other or perhaps the golden mean? Issues of spatial resolution in face processing. *Current Psychology*, 18(1):118–127, 1999.

[146] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine Series 6*, 2(11):559, 1901.

[147] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *Proc. Computer Vision and Pattern Recognition*, 2005.

[148] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The FERET Evaluation Methodology for Face-Recognition Algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, 2000.

[149] P. J. Phillips, W. T. Scruggs, A. J. O'Toole, P. J. Flynn, K. W. Bowyer, C. L. Schott, and M. Sharpe. FRVT 2006 and ICE 2006 Large-Scale Results. *National Institute of Standards and Technology*, 7408, 2006.

[150] P. J Phillips, H. Wechsler, J. Huang, and P. J Rauss. The FERET database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing*, 16(5):295–306, 1998.

[151] P.J. Phillips, J.R. Beveridge, B.A. Draper, G. Givens, A.J. O'Toole, D.S. Bolme, J. Dunlop, Y.M. Lui, H. Sahibzada, and S. Weimer. An introduction to the good, the bad, & the ugly face recognition challenge problem. In *Proc. IEEE International Conference on Automatic Face & Gesture Recognition and Workshops*, pp. 346–353, 2011.

[152] P.J. Phillips, P. Grother, R. Micheals, D.M. Blackburn, E. Tabassi, and M. Bone. Face recognition vendor test 2002. In *Proc. IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, pp. 44, 2003.

[153] J. T. Reason. *Human error*. Cambridge University Press, 1990.

[154] G. Rhodes, S. Brennan, and S. Carey. Identification and ratings of caricatures: Implications for mental representations of faces. *Cognitive Psychology*, 19(4):473–497, 1987.

[155] R. Sala Llonch, E. Kokiopoulou, I. Tosic, and P. Frossard. 3D face recognition with sparse spherical representations. *Pattern Recognition*, 43(3):824–834, 2010.

[156] A. Savran, O. Celiktutan, A. Akyol, J. Trojanová, H. Dibeklioglu, S. Esenlik, N. Bozkurt, C. Demirkır, E. Akagündüz, K. Calıskan, et al. 3D face recognition performance under adversarial conditions. In *Proc. eNTERFACE07 Workshop on Multimodal Interfaces*, 2007.

[157] A. Savran, B. Sankur, and M. Taha Bilge. Comparative evaluation of 3D vs. 2D modality for automatic detection of facial action units. *Pattern Recognition*, 45(2):767–782, 2012.

[158] P. G. Schyns and A. Oliva. Dr. Angry and Mr. Smile: when categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition*, 69(3):243–265, 1999.

[159] J. Sergent. Microgenesis of face perception. *Aspects of face processing*, pp. 17–33, 1986.

[160] T. Serre, A. Oliva, and T. Poggio. A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences*, 104(15):6424, 2007.

[161] C. Shan, S. Gong, and P. W McOwan. Facial expression recognition based on Local Binary Patterns: A comprehensive study. *Image and Vision Computing*, 27(6):803–816, 2009.

[162] M. Sharkas and M. A Elenien. Eigenfaces vs. fisherfaces vs. ICA for face recognition; a comparative study. In *Proc. IEEE International Conference on Signal Processing*, pp. 914–919, 2008.

[163] L. Shen and L. Bai. A review on Gabor wavelets for face recognition. *Pattern Analysis & Applications*, 9(2):273—292, 2006.

[164] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-time human pose recognition in parts from single depth images. In *Proc. Computer Vision and Pattern Recognition*, 2011.

[165] T. Sim, S. Baker, and M. Bsat. The CMU pose, illumination, and expression database. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1615–1618, 2003.

[166] S. Singh, A. Gyaourova, G. Bebis, and I. Pavlidis. Infrared and visible image fusion for face recognition. In *Proc. SPIE Defense and Security Symposium*

(Biometric Technology for Human Identification), vol. 5404, pp. 585–596, 2004.

[167] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell. Face Recognition by Humans: Nineteen Results All Computer Vision Researchers Should Know About. *Proceedings of the IEEE*, 94(11):1948–1962, 2006.

[168] L. Sirovich and M. Kirby. Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America A*, 4(3):519–524, 1987.

[169] M. L. Smith and L. N. Smith. Dynamic photometric stereo—a new technique for moving surface analysis. *Image and Vision Computing*, 23(9):841–852, 2005.

[170] W. A.P Smith and E. R Hancock. Facial Shape-from-shading and Recognition Using Principal Geodesic Analysis and Robust Statistics. *International Journal of Computer Vision*, 76(1):71–91, 2008.

[171] D. A. Socolinsky, A. Selinger, and J. D. Neuheisel. Face recognition with visible and thermal infrared imagery. *Computer Vision and Image Understanding*, 91(1-2):72–114, 2003.

[172] F. Solomon and K. Ikeuchi. Extracting the shape and roughness of specular lobe objects usingfour light photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(4):449–454, 1996.

[173] S. Subramaniam and I. Biederman. Does contrast reversal affect object identification. *Investigative Ophthalmology and Visual Science*, 38(4):998, 1997.

[174] J. Sun, M. L. Smith, L. N. Smith, S. Midha, and J. Bamber. Object surface recovery using a multi-light photometric stereo technique for non-

Lambertian surfaces subject to shadows and specularities. *Image and Vision Computing*, 25(7):1050–1057, 2007.

[175] H.T. Tanaka, M. Ikeda, and H. Chiaki. Curvature-based face surface recognition using spherical correlation. Principal directions for curved object recognition. In *Proc. IEEE International Conference on Automatic Face & Gesture Recognition*, pp. 372–377, 1998.

[176] J. W. Tanaka and M. J. Farah. Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology Section A*, 46(2):225–245, May 1993.

[177] P. Thompson. Margaret Thatcher: A new illusion. *Perception*, 9(4):483—484, 1980.

[178] G. Toderici, S. O'Malley, G. Passalis, T. Theoharis, and I. Kakadiaris. Ethnicity- and Gender-based Subject Retrieval Using 3-D Face-Recognition Techniques. *International Journal of Computer Vision*, 89(2):382–391, 2010.

[179] K. E Torrance and E. M Sparrow. Theory for off-specular reflection from roughened surfaces. *Journal of the Optical Society of America A*, 57(9):1105–1112, 1967.

[180] M. Turk. A Random Walk through Eigenspace. *IEICE Transactions on Information And Systems E Series D*, 84(12):1586–1595, 2001.

[181] M. Turk and A. Pentland. Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[182] M.K. Unnikrishnan. How is the individuality of a face recognized? *Journal of Theoretical Biology*, 261(3):469–474, 2009.

[183] M. Valstar and M. Pantic. Fully automatic facial action unit detection and temporal analysis. In *Proc. Computer Vision and Pattern Recognition Workshop*, pp. 149—149, 2006.

[184] M. F Valstar, B. Jiang, M. Mehu, M. Pantic, and K. Scherer. The first facial expression recognition and analysis challenge. In *Proc. IEEE International Conference on Automatic Face & Gesture Recognition*, pp. 921–926, 2011.

[185] M. Varma and A. Zisserman. Texture classification: Are filter banks necessary? In *Proc. Computer Vision and Pattern Recognition*, vol. 2, pp. 691—698, 2003.

[186] V. Vijayan, K. Bowyer, P. Flynn, D. S Huang, L. Chen, M. Hansen, O. Ocegueda, S. Shah, and I. Kakadiaris. Twins 3D Face Recognition Challenge. In *International Joint Conference on Biometrics*, 2011.

[187] P. Viola and M. Jones. Rapid Object Detection Using a Boosted Cascade of Simple Features. In *Proc. Computer Vision and Pattern Recognition*, 2001.

[188] X. Wang and X. Tang. Face photo-sketch synthesis and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(11):1955—1967, 2008.

[189] Wang, K. Award#1131883 - US-German Collaboration: Towards a Neural Theory of 3D Shape Perception. http://www.nsf.gov/awardsearch/showAward.do?AwardNumber=113188, 2011.

[190] P. Welti, J. Suchier, and C. Busch. Improving Border Control with 3D Face Recognition. In *Proc. of the Workshop Biometric Border Control & Federated Identity Management*, pp. 123—134, 2008.

[191] L. Wiskott, J. M. Fellous, N. Krüger, and C. von der Malsburg. Face Recognition by Elastic Bunch Graph Matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):775—779, 1997.

[192] R. J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):139—144, 1980.

[193] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust Face Recognition via Sparse Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):210–227, 2009.

[194] C. Xu, S. Li, T. Tan, and L. Quan. Automatic 3D face recognition from depth and intensity gabor features. *Pattern Recognition*, 42(9):1895—1905, 2009.

[195] J. Yang, J. Wright, T. S Huang, and Y. Ma. Image super-resolution via sparse representation. *IEEE Transactions on Image Processing*, 19(11):2861–2873, 2010.

[196] L. Yin, X. Chen, Y. Sun, T. Worm, and M. Reale. A high-resolution 3D dynamic facial expression database. In *Proc. IEEE International Conference on Automatic Face & Gesture Recognition*, pp. 1–6, 2008.

[197] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J Rosato. A 3D facial expression database for facial behavior research. In *Proc. IEEE International Conference on Automatic Face & Gesture Recognition*, pp. 211—216, 2006.

[198] A. W. Young, D. Hellawell, and D. C. Hay. Configurational information in face perception. *Perception*, 16(6):747–759, 1987.

[199] S. Zafeiriou, M. F. Hansen, G. A. Atkinson, V. Argyriou, M. Petrou, M.L. Smith, and L.N. Smith. The PhotoFace Database. In *Proc. Computer Vision and Pattern Recognition, Biometrics Workshop*, pp. 161—168, 2011.

[200] S. Zafeiriou, A. Tefas, and I. Pitas. Elastic graph matching versus linear subspace methods for frontal face verification. In *International Workshop on Nonlinear Signal and Image Processing*, 2005.

[201] S. Zafeiriou, A. Tefas, and I. Pitas. Exploiting discriminant information in elastic graph matching. In *IEEE International Conference on Image Processing*, vol. 3, pp. 768–71. IEEE, 2005.

[202] S. Zafeiriou, A. Tefas, and I. Pitas. The discriminant elastic graph matching algorithm applied to frontal face verification. *Pattern Recognition*, 40(10):2798–2810, 2007.

[203] Z. Zhang. Feature-based facial expression recognition: Sensitivity analysis and experiments with a multilayer perceptron. *International Journal of Pattern Recognition and Artificial Intelligence*, 13(6):893–911, 1999.

[204] S. Zhao and R. Grigat. An Automatic Face Recognition System in the Near Infrared Spectrum. *Machine Learning and Data Mining in Pattern Recognition*, pp. 437–444, 2005.

[205] W. Zhao and R. Chellappa. *Face processing: advanced modeling and methods*. Academic Press, 2006.

[206] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):399—458, 2003.

[207] S. K. Zhou, G. Aggarwal, R. Chellappa, and D. W. Jacobs. Appearance characterization of linear lambertian objects, generalized photometric stereo, and illumination-invariant face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(2):230–245, 2007.

[208] S. K. Zhou, R. Chellappa, and D. W. Jacobs. Characterization of human faces under illumination variations using rank, integrability, and symmetry

constraints. *Proc. European Conference on Computer Vision*, pp. 588—601, 2004.

[209] Y. Zhuang, J. Zhang, and F. Wu. Hallucinating faces: LPH super-resolution and neighbor reconstruction for residue compensation. *Pattern Recognition*, 40(11):3178–3194, 2007.

[210] J. Zivanov, P. Paysan, and T. Vetter. Facial Normal Map Capture using Four Lights - An Effective and Inexpensive Method of Capturing the Fine Scale Detail of Human Faces using Four Point Lights. In *Proc. International Conference on Computer Graphics Theory and Applications*, pp. 13–20, 2009.

[211] X. Zou, J. Kittler, and K. Messer. Face Recognition Using Active Near-IR Illumination. In *Proc. British Machine Vision Conference*, pp. 209–219, 2005.

[212] X. Zou, J. Kittler, and J. Tena. A morphing system for effective human face recognition. In *Proc. Visual Information Engineering*, pp. 215–220, 2008.

Glossary

3DE-VISIR database

A set of 363 faces belonging to 115 individuals captured using near-infrared and visible light photometric stereo at UWE with 3 captures within each session of positive, neutral and negative emotions.

3D Morphable Model

A 3D model of the face created by Blanz and Vetter [23] which is parameterizes the variations of faces. This allows for synthesis of realistic 3D face models from 2D photographs.

Active Appearance Model

An extension to the ASM which also models and parameterizes changes in an object's texture.

Active Shape Model

A statistical model which parameterizes shape variations so that the model can accurately be fitted to a novel image to extract the object boundaries.

Annotated Face Model

A method by Kakadiaris *et al.* [99] which aligns and deforms a 3D model onto an annotated 3D face to aid segmentation as a preprocessing step.

Discrete Cosine Transform

The discrete cosine transform is a means of representing a given signal as the sum of cosine functions at different frequencies. It is commonly used

for compression of media (e.g. jpeg for images) and is similar to the Fourier Transform.

Dynamic Link Architecture

An extension to neural networks that used the results of convolving Gabor wavelets over images as inputs before using Elastic Graph Matching for recognition..

Eigenface

The “Eigenface” approach was invented by Turk and Pentland [181] and refers to the PCA components describing the most variant dimensions, which when reshaped into a 2D image, resemble a face.

Elastic Graph Matching

One of the most successful techniques used in face recognition that deforms a grid so that the nodes fall into correspondence with facial landmarks, the coefficients of which can be used as features for recognition.

Equal Error Rate

The position on a ROC curve where the FAR and FRR are the same. Ideally this should be a low value, and it is useful for comparing algorithm performance.

Extended Gaussian Image

A data representation invented by Horn [91] which maps surface normals on a unit sphere, the area that is taken up on the sphere’s surface is proportional to the number of normals pointing in that direction.

Face Recognition Grand Challenge

A series of biometric grand challenges that followed the FRVT challenge running from 2002-2006 and superseded by the Multiple Biometrics Evaluation. Funded by the American National Institute of Science and Technology,

it provides what is arguably the de-facto database in 3D face recognition..

Face Recognition Vendor's Test

Preceded the FRGC and provided a large scale objective comparison of face recognition technologies against 2D faces, running from 2000-2006.

Facial Action Coding System

A system that objectively quantifies facial expressions.

False Accept Rate

The number of trials in which an unknown probe is incorrectly identified expressed as a percentage or ratio of the total number of trials..

False Reject Rate

The number of trials in which a known probe is not recognised expressed as a percentage or ratio of the total number of trials. This is the same as 1-TAR.

Facial Recognition Technology

A program setup by the US Department of Defense counterdrug Technology program with an accompanying database of faces which are used for evaluating the effectiveness of algorithms..

Fisher's Linear Discriminant

Fisher's Linear Discriminant is an example of LDA with the assumption that the covariance in each and every group is uniform..

Fisherface

The Fisherface approach to face recognition was first performed by Belheumeur *et al.* [17] and combines FLD with PCA in order to reduce the dimensionality before solving the generalized eigenproblem..

Fusiform face area

An area in the temporal lobe of the brain which fires preferentially for face

stimuli and damage to which causes Prosopagnosia (face blindness).

Human Computer Interaction

An area of research that aims to improve the interface(s) between humans and machines..

Human Visual System

A term to describe all aspects of human vision from the low level functions within the retina to higher level feature understanding in the cortex..

Independent Components Analysis

Similar to PCA but not restricted to describing variation along orthogonal dimensions.

Iterative Closest Point

An algorithm for aligning two roughly aligned 3D surfaces such that the RMS between them is minimized..

Linear Discriminants Analysis

Linear Discriminants Analysis (of which Fisher's Linear Discriminant is a commonly used example) is a statistical tool which is used to maximise the separation of different groups of labelled data while minimising the separation with each group. This is primarily used improve classification by making more distinct clusters of groups..

Local Binary Patterns

A simple convolution that is sensitive to certain primitives such as edges and lines and has been found to be an effective representation for face recognition..

ℓ_p-norm

A distance measure in p-space. The familiar euclidean distance is the ℓ_2-

norm, the Manhattan distance (so called because of the block-by-block taxi-cab type constraint) is the ℓ_2-norm. The Chebyshev distance or ℓ_{inf} norm is the greatest distance between points along any dimension.

Multi-Dimensional Scaling

Is primarily a data mapping tool commonly used in psychological experiments for assessing subject judgements. It preserves the ratio of distances of the differences between observations so that the layout of the data remains the same after transformation.

Near Infrared

A wavelength of light just outside the visible range above the red part of the visible spectrum. More specifically it refers to a wavelength in the region of 850nm for the purposes of this thesis.

Pearson's product-moment correlation coefficient

A measure of correlation between two variables ranging between -1 and +1, where -1 indicates a negative correlation, 0 indicates no correlation, and +1 indicates a positive correlation. In this thesis it is used as a Rank-1 classifier with classification performance similar to the Fisherface algorithm, but about 10 times faster. However it is found to be unsuitable for use in a face recognition system due to an almost linear ROC graph..

Photoface database

A publicly available database of 3187 faces belonging to 453 individuals captured using photometric stereo in an unconstrained office corridor..

Photometric stereo

Invented by Woodham [192], photometric stereo is a method of estimating the surface orientation using multiple illuminations of an object from a single viewpoint..

Principal Components Analysis

Principal Components Analysis is a statistical tool commonly used for dimension reduction in computer vision applications. It redefines the data in terms of accounted variance along orthogonal principle components or dimensions and is the basis of the popular eigenfaces approach for face recognition..

Receiver Operating Characteristic

A plot of FAR against FRR which can be used to calculate the EER and the TAR as a certain FAR.

Root Mean Squared

Is a measure of the magnitude of variation between two quantities. The RMS error is used to quantify the difference between surface reconstructions in this thesis.

Shape-from-Shading

Refers to the technique of estimating the 3D shape of an object from a 2D image. Photometric stereo is a type of shape from shading where certain associated ambiguities are removed (e.g. by separating texture from gradient) through multiple illumination sources.

Support Vector Machine

The support vector machine is a supervised learning technique used to construct the hyperplane which maximally separates two classes of data. It has been extended to non-linear via kernel functions and multi-class to perform classifications of more than two classes..

True Accept Rate

Also known as the verification rate, this is the number of trials in which a probe is correctly identified expressed as a percentage or ratio of the total number of trials..

Acronyms

3DE-VISIR

3D Expression-VISible and near-InfraRed database.

3DMM

3D Morphable Model.

AAM

Active Appearance Model.

AFM

Annotated Face Model.

ASM

Active Shape Model.

DCT

Discrete Cosine Transform.

DLA

Dynamic Link Architecture.

EER

Equal Error Rate.

EGI

Extended Gaussian Image.

EGM

Elastic Graph Matching.

FACS

Facial Action Coding System.

FAR

False Accept Rate.

FERET

Facial Recognition Technology.

FFA

Fusiform face area.

FLD

Fisher's Linear Discriminant.

FRGC

Face Recognition Grand Challenge.

FRR

False Reject Rate.

FRVT

Face Recognition Vendor's Test.

HCI

Human Computer Interaction.

HVS

Human Visual System.

ICA

Independent Components Analysis.

ICP

Iterative Closest Point.

LBP

Local Binary Patterns.

LDA

Linear Discriminants Analysis.

MDS

Multi-Dimensional Scaling.

NIR

Near Infrared.

PCA

Principal Components Analysis.

PMCC

Pearson's product-moment correlation coefficient.

PS

Photometric stereo.

RMS

Root Mean Squared.

ROC

Receiver Operating Characteristic.

SFS

Shape-from-Shading.

SVM

Support Vector Machine.

TAR

True Accept Rate.

Glossary of Symbols

A

Matrix of the difference between each image and the mean image (**X**-**u**).

B_i

Number of samples in class i.

c

Number of classes or labels.

C

Covariance matrix of image data.

D

Mixing matrix in ICA.

e

Eigenvector.

ε

Euclidean distance.

$\mathbf{I}_{1...3}$

Three images taken under different illumination.

K_i

Specifies a particular class K_i.

λ

Eigenvalue.

M

Number of images.

N

Number of pixels/dimensions per image.

n_x, n_y, n_z

Surface normal components.

$\mathbf{\Omega}$

Face space projection.

$\boldsymbol{\omega}$

Face space projection.

P

Number of eigenvectors used to represent face space.

ρ

Surface albedo.

$\mathbf{S}_{1...3}$

Three light source vectors.

s

Independently separable variables in ICA.

$\mathbf{S}_B$

Between/inter-class scatter matrix.

$\mathbf{S}_T$

Total scatter matrix.

$\mathbf{S}_W$

Within/intra-class scatter matrix.

u

Mean intensity value for each pixel across all images ($N \times M$).

v

Estimate of actual **s** signals.

V

Separating matrix (inverse of **D**).

W

Linear transformation matrix.

X

Matrix of image data ($N \times M$).

x

A matrix of pixel intensities (a photograph/image).

Appendix A

Selected publications

The appendix contains copies of selected journal and conference papers which have been produced as part of this Ph.D.. Each is prefaced with a short introductory paragraph regarding its significance.

A.1 3D face reconstructions from photometric stereo using near infrared and visible light

The first appendix was published in Computer Vision and Image Understanding in August 2010. It provides details of the Photoface 3D capture device which was developed prior to the start of this Ph.D. This author's contribution was writing the software to perform the necessary image processing and analysis for assessing the accuracy of the device compared with 3dMD reconstructions and performing reflectance analysis.

Computer Vision and Image Understanding 114 (2010) 942–951

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

3D face reconstructions from photometric stereo using near infrared and visible light

Mark F. Hansen *, Gary A. Atkinson, Lyndon N. Smith, Melvyn L. Smith

Machine Vision Laboratory, University of the West of England, Bristol BS16 1QY, UK

ARTICLE INFO

Article history:
Received 3 August 2009
Accepted 1 March 2010
Available online 7 April 2010

Keywords:
3D reconstruction
Near infrared
Photometric stereo
Skin reflectance
3D face recognition

ABSTRACT

This paper seeks to advance the state-of-the-art in 3D face capture and processing via novel Photometric Stereo (PS) hardware and algorithms. The first contribution is a new high-speed 3D data capture system, which is capable of acquiring four raw images in approximately 20 ms. The results presented in this paper demonstrate the feasibility of deploying the device in commercial settings. We show how the device can operate with either visible light or near infrared (NIR) light. The NIR light sources offer the advantages of being less intrusive and more covert than most existing face recognition methods allow. Furthermore, our experiments show that the accuracy of the reconstructions is also better using NIR light. The paper also presents a modified four-source PS algorithm which enhances the surface normal estimates by assigning a likelihood measure for each pixel being in a shadowed region. This likelihood measure is determined by the discrepancies between measured pixel brightnesses and expected values. Where the likelihood of shadow is high, then one light source is omitted from the computation for that pixel, otherwise a weighted combination of pixels is used to determine the surface normal. This means that the precise shadow boundary is not required by our method. The results section of the paper provides a detailed analysis of the methods presented and a comparison to ground truth. We also analyse the reflectance properties of a small number of skin samples to test the validity of the Lambertian model and point towards potential improvements to our method using the Oren–Nayar model.

1. Introduction

Face recognition is now one of the most active areas of computer vision research. A wide range of different approaches have been proposed for the detection, processing, analysis and recognition of faces within images [1]. A recent trend has been to incorporate 3D information to aid recognition [2]. Unlike for 2D methods, the process of data capture is a complex procedure for 3D methods and may involve expensive and bulky hardware with computationally intensive algorithms.

In this paper, we make significant contributions to 3D face capture and processing by presenting a novel Photometric Stereo (PS) hardware device, a new PS algorithm for mitigating the effects of shadows within images, and a detailed set of experiments to assess the accuracy and practicality of the device. The new variation of PS estimates a field of surface normals by selecting the optimal combination of light sources to reduce the effects of shadow without requiring knowledge of the exact shadow boundaries. This is demonstrated on a novel high-speed practical 3D facial geometry capture device. We have also been successful at facial PS using near infrared (NIR) light. This offers several benefits to existing methods including exploiting skin phenomenology, creating a more covert capture system and making the system less intrusive. Extensive experimental results of these proposed advances are presented, including an analysis of skin reflectance qualities under NIR and visible light in terms of the Lambertian assumption.

In summary, the contributions of this paper are fourfold:

1. The development of 3D data capture hardware suitable for practical face recognition environments.
2. The development of a new algorithm for choosing the optimal light source configuration for each pixel in order to reduce the effects of shadows.
3. Detailed experiments to test the accuracy of the device on a variety of faces under visible and NIR light sources in terms of ground truth reconstructions and the Lambertian assumption.
4. Detailed experiments to assess the validity of the Lambertian assumption and a test to determine any possible improvements that may be possible using the Oren–Nayar reflectance model [3].

The remainder of this section provides an overview of related work and outlines the contributions to the state-of-the-art. Section 2 presents details of the hardware arrangement and image

* Corresponding author.
E-mail addresses: mark.hansen@uwe.ac.uk (M.F. Hansen), gary.atkinson@uwe.ac.uk (G.A. Atkinson), lyndon.smith@uwe.ac.uk (L.N. Smith), melvyn.smith@uwe.ac.uk (M.L. Smith).

1077-3142/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.cviu.2010.03.001

acquisition process and our method to mitigate the effects of shadows. Detailed experimental results are then provided in Section 3. The implications and potential limitations of the work are discussed in Section 4.

1.1. Related work

The use of 3D information for face recognition has been attracting increasing attention in recent years [2,4,5]. This is due to the ability to overcome certain limitations associated with 2D recognition, e.g. problems of illumination and pose variance. Methods using 3D specific information also allow for representations which offer robustness to facial expression. Such methods include the 3D morphable model of Blanz and Vetter [6] and the geodesic representations of Bronstein et al. [7] and Mpiperis et al. [8]. Frequently, research which directly compares 2D and 3D recognition reports improved success rates for 3D recognition and that the best results occur when 2D and 3D information is fused [2]. As demand for practical face recognition systems is likely to increase, it is important that the most accurate methods are used and that the acquisition devices are both practical and affordable. There are a number of existing ways to capture and reconstruct 3D face information and the benefits and limitations of the most common approaches will now be discussed with the aim of putting our device into context.

Structured light scanning is perhaps the best known approach to generating 3D models of faces. This was used for generating the morphable head model in [6] and also for all the 3D faces used in the FRGC2.0 dataset [9], currently the largest publicly available 3D face database. For face capture, this technique works by scanning the object with a horizontal plane of laser light, capturing the line of light on a sensor and then calculating the location of each point via triangulation. The technique provides potentially very accurate scans; the Minolta Vivid 910 device [10] used for the FRGC2.0 dataset has a quoted accuracy of ±0.10 mm. However, these devices take about 2.5 s to capture the data, during which time the subject could move, thus distorting the reconstruction. It therefore requires a great deal of cooperation from the subject. They are also sensitive to high levels of ambient illumination. For these reasons, and the fact that they are financially costly, laser scanners are currently not suitable for many practical applications. The speed of acquisition can be improved by using striped patterns projected across the whole surface instead of using a scanning line. Distortions in this pattern can then used to calculate the 3D geometry of the surface [11], however accuracy is likely to be compromised and calibration can be time consuming.

The commercially available 3dMD system [12], which is used in this paper to acquire ground truth models, is an example of a projected pattern range finder. This device uses a number cameras to take images of an object from different positions. It uses a projected pattern to solve the correspondence problem between the images. The benefits of this system are its high accuracy (reported as <0.2 mm) and the speed of image acquisition (1.5 ms). However, the processing time is approximately 90 s for a face. This type of system is also expensive and requires a time consuming calibration procedure.

Shape-from-shading (SFS) is a technique for estimating 3D geometry from a single image [13]. Gradients of the surface are estimated from the patterns of intensity changes in an image. However the problem is ill-posed, meaning that there is no guarantee of a unique solution for a given image [14]. The main advantage of SFS is that it does not require any specialist capture apparatus such as laser scanners or projected pattern devices; merely a single ordinary camera. For this reason, finding solutions for the SFS problem are attractive to researchers. One way of overcoming the ill posed problem of SFS is to photograph the object multiple times under different illumination. This technique is known as Photometric Stereo (PS) and was first devised by Woodham [15] who showed that for any Lambertian surface, three differently illuminated images are sufficient to remove the ambiguity associated with a single image. Further details of this method will be given in Section 2.3.

A unique surface normal can be estimated by using three PS images provided that none of the light sources cast a shadow and the surface is Lambertian. In the case of a human face, shadows are frequently cast by features such as the nose. Indeed overcoming the detrimental effects of shadow on PS has been the subject of a number of papers. Smith and Hancock [16] use a statistical model to recover geometry in the presence of shadows. Hernández et al. [17] use two images where shadow is not present to estimate the value of the third where the shadow is present via integration. Coleman and Jain [18] use four light sources to over-determine the surface orientation. If a shadow is present in one image, it can simply be discarded. Similarly, Solomon and Ikeuchi [19] use four sources, but determine shadows and specularities by considerations of anomalies in albedo estimates that cannot be statistically attributed to camera noise. Barsky and Petrou [20] suggest a similar alternative solution to highlights and shadows by using a four source coloured light PS technique.

Georghiades extended PS beyond Lambertian surfaces to incorporate the Torrance and Sparrow model of reflectance [21] and created very accurate reconstructions [22]. However, a large number of images were required for the reconstruction which significantly increases surface computation and image acquisition time. Sun et al. [23] use five lights to handle shadows and specularities on non-Lambertian surfaces and show that a minimum of six lights are required in order fully realise any convex surface using photometric stereo. Using 20 images of a face, Ghosh et al. [24] build up a very detailed model of the skin's reflectance taking into account specular reflection and single, shallow and deep scattering. However, the images are captured over "a few seconds" which makes this approach unsuitable for our needs (i.e. practical applications). Also, their method would add a large amount of complexity for relatively little gain as skin follows Lambert's Law reasonably well (as shown in this paper for example).

Of the vast amount of research into automatic face recognition during the last two decades [1], relatively little work has involved PS. Kee et al. investigate the use of 3-source PS under dark room conditions [25]. They were able to determine the optimal light source arrangement and demonstrate a working recognition system. Zhou, Chellappa and Jacobs apply rank, integrability and symmetry constraints to adapt PS to face-specific applications [26]. Zhou et al. extended a PS approach to unknown light sources [27]. Georghiades et al. show how reconstructions from PS can be used to form a generative model to synthesise images under novel pose and illumination [28].

Comparing point clouds, the shape index, depth maps, profiles and surface normals in terms of face recognition performance, Gökberk et al. [5] concluded that surface normals provide the best features for face recognition. It is surprising therefore, that so few applications to date utilise PS, which inherently generates surface normals. The reason for this is likely to be that the availability and affordability of cameras with high enough frame rates, sensitivity and synchronisation capabilities for PS have only recently reached the market. Such cameras are necessary in commercial and industrial applications to effectively freeze the motion of the person while they may be moving by capturing several images in a short burst.

The majority of past work on PS has been conducted using visible illumination. As explained above, we also consider NIR light in this paper. Studies into the optical properties of skin have shown it to be increasingly reflective in the NIR light band up to wave-

lengths of about 1.1 μm [29]. This suggests that NIR, which is more covert and less intrusive, is a viable alternative to visible light. Furthermore, NIR can be used as a replacement for visible light because its proximity to the visual spectrum means that it is likely to behave in a similar manner on skin. We might expect some fine surface detail to be lost due to sub-surface scattering as reported by Zivanov et al. [30], but this is unlikely to affect overall face shape estimation. In addition to our work, infrared light has been used previously in 2D face recognition to mitigate the negative impact of ambient illumination [31,32] and to aid eye detection algorithms using the "bright eye" effect [33].

1.2. Contributions

The contributions of this paper to the state-of-the-art in 3D face capture and processing are via a novel system of hardware and algorithms. The new PS-based 3D face shape capture device is suitable for practical recognition environments and consists of four illumination sources placed evenly around a high-speed camera, as shown in Fig. 1. Individuals walk through the archway towards the camera located on the back panel and exit through the side. Compared to existing technologies, our device is cheap to build and involves exceptionally short image capture and processing times. The device is also able to operate at high resolution, is robust to ambient illumination and requires only minimal calibration. All images are captured in approximately 20 ms, resulting in only very small misalignment between frames. This allows subjects to be imaged as they casually walk through the archway.

We have tested our device using both visible and NIR illumination sources and found the latter to yield more accurate reconstructions when compared with ground truth. To the best of our knowledge, no published research has looked at using NIR light sources in PS for the purpose of face recognition. These considerations make our method attractive for use in many commercial and industrial settings such as at entrances to high security areas, airport check-in and border control.

For the main algorithmic contribution of this paper, we show how the effects of shadow can be mitigated using a related approach to that of Solomon and Ikeuchi [19] and Barsky and Petrou [20]. The method relies on estimating the likelihood of each pixel being in shadow and weights the contributions of the light sources in the PS computation accordingly. One advantage of our method is that neither the exact shadow boundary nor the camera noise parameters are required.

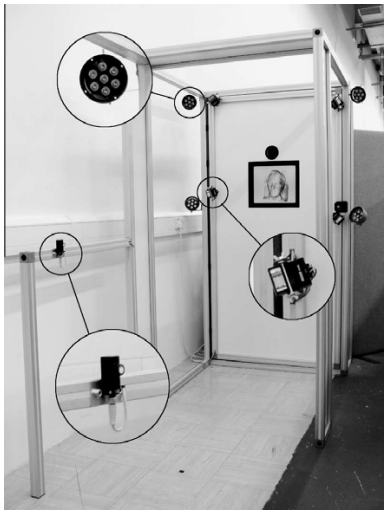


Fig. 1. The geometry capture device. Enlarged areas from top to bottom: a NIR light source, a visible light source and an ultrasound trigger. The camera can be seen on the back panel.

The final contribution of the paper is a detailed analysis of the quality of reconstructions and the nature of the skin reflectance properties. The device is tested on a variety of subjects and the RMS height errors and ℓ_2-norm errors are presented. Ground truth data is supplied by a 3dMD scanner. A quantitative analysis on the validity of the Lambertian assumption on skin reflectance is then presented. The extent of the discrepancies between the measured skin reflectance and Lambert's Law are demonstrated graphically and shown to be relatively minor for non-grazing angles. Lastly, we show that skin is more Lambertian under NIR illumination, solidifying our earlier claims about the feasibility of NIR as an alternative to visible light. This reflectance analysis also demonstrates the possibilities of improving the reconstructions by incorporating the Oren–Nayar reflectance model into the method.

2. Method

This section first outlines the overall PS image acquisition hardware, before moving on to describe the reconstruction process. We also discuss the differences between our use of visible and NIR light sources. The problem of shadowing is then addressed, by presenting a new PS method to automatically select the optimal light source configuration.

2.1. Hardware

This section details the acquisition device hardware. The device, shown in Fig. 1, is designed for practical 3D face geometry capture and recognition. The presence of an individual is detected by an ultrasound proximity sensor placed before the archway. This can be seen in Fig. 1 on the horizontal beam towards the left-hand side of the photograph. The sensor triggers a sequence of high speed synchronised frame grabbing and light source switching.

The aim is then to capture five images at a high frame rate: four images illuminated by the main light sources in sequence and an additional control image with only ambient illumination. Either one image per visible light is captured, or one image per NIR source. Note that the ambient lighting is uncontrolled (for the experiments presented in this paper, overhead fluorescent lights are present). The four visible light sources are low-cost Jessops M100 flashguns (colour temperature 5600 K), while the NIR lights are stripped down X-vision VIS080IR lensed 7-LED clusters, which emit light at ≈850 nm.

It was found experimentally that for people walking through the device, a minimum frame rate of approximately 150 fps was necessary to avoid significant movement between frames. The device currently operates at 200 fps, and it should be noted that it is only operating for the period required to capture the five images. That is, the device is left idle until it is triggered. A monitor is included on the back panel to show the reconstructed face or to display other information.

For visible light, the following sequence of events takes place to capture the five images as an individual passes through the device.

1. Await signal from ultrasound sensor.
2. Send trigger to camera.
3. Await integration enabled signal from camera.
4. Discharge first flashgun.
5. Await end of integration enabled signal.
6. Repeat from step 2 for the remaining light sources.
7. Capture control image with ambient lighting only.

All interfacing code is written in NI LabVIEW. The ultrasound sensor is a highly directional Baumer proximity switch. When its beam is broken within a distance of 70 cm, it transmits a signal to an NI PCI-7811 DIO card fitted to a computer. When this signal is received, a trigger is sent to the camera. This is a Basler 504 kc camera with a 55 mm, f5.6 Sigma lens, placed 2 m from the subject. As with many silicon-based sensors, the Basler chip is responsive to both visible and NIR irradiance. The trigger is transmitted to the camera from a frame grabber via Camera Link®. The frame grabber is an NI PCIe-1429, which communicates with the DIO card via a RTSI bus for triggering purposes.

To ensure that the signal has reached the camera, and that the camera has commenced frame capture (i.e. is integrating), a second connection from the camera to the DIO card as added. This connection is TTL-high while the camera is integrating. When the computer receives this signal, the first light source is to be immediately illuminated. A flashgun is discharged by making a short circuit between its input pins. This is achieved here by sending a short pulse from the DIO card to the input pins via a phototransistor opto-isolator IC. This electrically isolates the sensitive DIO card from the high voltages of the flashgun terminals. Finally, the DIO card awaits the falling edge of the camera integration enabled signal before moving on to the next light source.

For NIR light, a slightly different procedure is adopted whereby synchronous TTL signals are sent to the camera and LEDs. This is because the LEDs can be illuminated for the duration of the camera exposure, while the flashguns only last for a small fraction of the exposure. The NIR LEDs are powered independently from the DIO card and interfaced via a simple transistor circuit. As the LEDs are illuminated for only 5 ms, it is possible to overpower them, in order to increase their brightness without causing damage. We therefore apply 20 V across the LEDs, compared to the recommended 12 V.

2.2. Visible and NIR comparison

One possibly negative aspect of the visible light set-up is that the firing of flashguns is obvious to the subject and possibly intrusive to any surrounding people. A possible advantage of NIR is that there may be additional subcutaneous or vascular structures present in the raw images taken under NIR light which may be used to aid recognition. Unfortunately, we found that such features were not visible in the wavelength band considered in this paper, but we aim to study this further in future work. NIR light is also more covert for a face recognition environment and subjects are less inclined to "pose" for the camera, meaning that more neutral expressions are likely. Finally, it is worth noting the advantage that many camera sensors are inherently more sensitive to NIR light.

One disadvantage of NIR illumination is the relative difficulty in obtaining the necessary brightness for the required short exposure times. While the flashguns were easily bright enough with an exposure time of 1 ms, an exposure of 5 ms was needed for the NIR LEDs (i.e. the maximum possible exposure for the given frame rate). Although this was adequate for our experiments, we had to use LED lenses that provided a narrow divergence angle, meaning that the face had to be more precisely positioned to obtain full illumination. For the visible light sources, the images were bright enough even for large diversion angles, removing the need for accurate positioning of apparatus and allowing subjects to pass through the archway without having to consider their exact location with respect to the camera.

To account for ambient illumination, the control image is subtracted from the other four images. These images are then normalised in terms of intensity before reconstruction takes place. This was done by linearly scaling the greylevels of each image so that the mean intensity was equal for each image. A detailed comparison of the resulting reconstructions is presented in Section 3.2.

2.3. Photometric stereo

Fig. 2 shows an example of four raw images of an individual using our prototype operating with the visible light sources. The person was slowly (≈1 m/s) but casually walking through the device. Each image has pixel dimensions of 500 × 400 and there are typically just a few pixel lengths misalignment between the first and last images. The face detection method of Lienhart and Maydt [34] is used to extract the face from the background of the image.

The four intensity images are processed using a MATLAB implementation of a standard PS method [35, Section 5.4]. Denote the general operation of PS by

$$\{\mathbf{n}_i\} = \mathbb{P}(\{I_{1,i}\}, \{I_{2,i}\}, \ldots, \mathbf{L}_1, \mathbf{L}_2, \ldots) \tag{1}$$

where $\{\mathbf{n}_i; i = 1,\ldots,N\}$ is the resulting set of surface normals, N is the total number of pixels, $\{I_{k,i}; i = 1,\ldots,N\}$ is the set of intensities for image k, and $\mathbf{L}_j$ is the jth light source vector. For the bulk of this paper, we use four light sources, resulting in set of surface normals

$$\{\mathbf{n}_i\} = \mathbb{P}(\{I_{1,i}\}, \{I_{2,i}\}, \{I_{3,i}\}, \{I_{4,i}\}, \mathbf{L}_1, \mathbf{L}_2, \mathbf{L}_3, \mathbf{L}_4) \tag{2}$$

The general equation for PS using four sources for pixel i is

$$\begin{bmatrix} I_{1,i} \\ I_{2,i} \\ I_{3,i} \\ I_{4,i} \end{bmatrix} = \rho_i \begin{bmatrix} \mathbf{L}_1^T \\ \mathbf{L}_2^T \\ \mathbf{L}_3^T \\ \mathbf{L}_4^T \end{bmatrix} \mathbf{n}_i \tag{3}$$

where ρ_i is the reflectance albedo. The intensity values and light source positions are known, and from these the albedo and surface normal components can be calculated by solving (3). The resultant dense field of surface normals are then integrated to form height maps using the well-known Frankot and Chellappa method [36]. Fig. 2 shows the resultant reconstruction.

2.4. Optimising light sources

In many cases of PS usage, it is desirable to use all available light sources in the reconstruction in order to maximise robustness. However, where one or more sources do not illuminate the entire surface due to a self/cast shadow, it becomes disadvantageous to

Fig. 2. Left: four raw images. Right: reconstructions using standard PS (Eq. (3)), and the optimal source algorithm (Section 2.4).

use all the sources. In the case of a face, this is most likely to happen around the nose and outer edges of the cheeks, as shown in Fig. 2. In cases where the light source zenith angles (the angles between the viewpoint vector and the light source vectors) are small and the face is looking directly towards the camera, shadows tend not to be too problematic. When this is not the case, then more of the face is in shadow and the reconstructions become distorted. In such cases, it becomes beneficial to omit one or more sources from the PS computation at certain pixels. Assume for now that the faces are frontal, but that the light sources are repositioned to the archway (see Fig. 1) to increase the size of the shadows.

For this paper, we assume that no points on the face are shadowed by more than one source. In other words, each pixel is visible to at least three sources. In practice, a few areas to the sides of the nose and the extreme edges of the cheeks are sometimes shadowed by two sources, but we found that these areas get smoothed over by the subsequent surface integration and therefore have little bearing on the overall reconstructions. Ideally, we would like to determine which points are visible to which sources and only use these lights in the PS computation. However, performing this task precisely is difficult [20] and the resulting fields of surface normals tend to exhibit discontinuities around the estimated shadow boundaries. We therefore propose to adopt the following surface normal for an arbitrary pixel i:

$$\mathbf{n}_{\text{opt},i}(\varepsilon_i) = \varepsilon_i \mathbf{n}_{3,i} + (1-\varepsilon_i)\mathbf{n}_i \tag{4}$$

where ε is a measure of the likelihood of a pixel being in shadow and $\mathbf{n}_3$ is the surface normal estimated from the optimal three light sources. Where $\varepsilon = 0$, the pixel is definitely not in shadow and all four light sources are used. Where $\varepsilon = 1$, the pixel definitely in shadow and only three sources are used. For intermediate values of ε, a mixture of $\mathbf{n}$ and $\mathbf{n}_3$ are used. This has the dual effect that the shadow boundary does not need to be known precisely and that the discontinuities mentioned above become smoothed out.

Methods are therefore needed to determine $\mathbf{n}_3$ (i.e. which are the best three light sources to use) and ε for each pixel. For the former of these, it is adequate to simply use the three light sources that cause the brightest pixel for each point.

Each pixel has four measured intensities, one for each light source. Let us call the brightest intensity I_a, with corresponding light source vector $\mathbf{L}_a$. We shall call the second brightest pixel I_b and so on. We can therefore write our estimates of $\mathbf{n}_3$ as

$$\{\mathbf{n}_3\} = \mathbb{P}(\{I_a\},\{I_b\},\{I_c\},\mathbf{L}_a,\mathbf{L}_b,\mathbf{L}_c) \tag{5}$$

where we are omitting the index suffix, i, for the sake of simplifying the notation.

We know that for a pixel which is in a self shadow, the angle between $\mathbf{L}_d$ and the surface normal is greater than 90°. We can therefore define a self shadow condition as follows:

$$\arccos(\mathbf{L}_d \cdot \mathbf{n}_3) \geqslant 90^\circ \tag{6}$$

where this condition is satisfied, we know that the fourth light source is of no use and so we set $\varepsilon = 1$. Where the condition is not met, then *cast* shadows may or may not be present.

To obtain a suitable value of ε for pixels that do not satisfy condition (6), we compare the value of I_d to the value that which we would expect to measure in the absence of a cast shadow. Call this value I_{ex}. It is possible to estimate this quantity using a combination of $\mathbf{n}_3$, $\mathbf{L}_d$, the albedo found from the brightest three pixels (call this ρ_3), and Lambert's Law:

$$I_{\text{ex}} = \rho_3 \mathbf{L}_d \cdot \mathbf{n}_3 \tag{7}$$

For areas potentially in cast shadow from one source, we use the ratio between I_d and I_{ex} to determine ε. For areas deep in cast shadow, the expectation is that $I_d \ll I_{\text{ex}}$. For these regions we would like $\varepsilon \approx 1$, while for pixels away from shadow, we have $I_d \approx I_{\text{ex}}$, so we require $\varepsilon \approx 0$. Combining this logic with condition (6) we arrive at our final definition of ε:

$$\varepsilon = \begin{cases} 1 & \arccos(\mathbf{L}_d \cdot \mathbf{n}_3) \geqslant 90^\circ \\ \max\left(1 - \frac{I_d}{I_{\text{ex}}}, 0\right) & \text{otherwise} \end{cases} \tag{8}$$

where the "max" is required to deal with points that are not in shadow, but where I_d happens to be slightly greater than I_{ex}.

Fig. 2 shows the surface reconstructions resulting from both standard PS and using the method presented here. The new method was able to more accurately recover the regions of the face that are in shadow. This is especially true for areas of large surface zenith angle, such as the sides of the nose and outer edges of the face. The method proposed here is also able to restore a greater definition in the fine details of the face. Note however, that the method breaks down slightly near the far edges of the cheek, where the region is shadowed by two light sources. The result is that discontinuities in the reconstructed height appear in such regions. Further analysis of the method will be presented in Section 3.2, which also includes a comparison with the Solomon and Ikeuchi method [19].

3. Results

3.1. Basic reconstructions

Fig. 3 shows a series of reconstructions from the method described in Section 2 using visible light. The device was placed at the entrance to a workplace to ensure casual (and thus realistic) usage. The general 3D structures of the faces have clearly been well estimated. Note however, that the spectacles of one of the subjects have been "blended" into the face. This is a combined consequence of the rim of the spectacles being highly specular and the surface being non-integrable for this region of the image [36]. Although, we would ideally be able to estimate the shape of the spectacles accurately, the blending effect can potentially be beneficial to face recognition algorithms because it means that such details have a lesser impact on the overall reconstruction. A set of images and reconstructions using both visible and NIR light sources can be seen in Fig. 4. It is clear that NIR is also capable of providing good estimates of the 3D geometry of the face.

We now compare the accuracy of the face reconstructions against ground truth data. To do this, we scanned eight different faces using a commercial 3dMD projected pattern range finder

Fig. 3. Estimated geometry of three different subjects using visible light sources.

Fig. 4. Example raw images and reconstructions using visible (top) and NIR light sources for four subjects. For these experiments only, the subjects were asked to rest their chin on a support in order to ensure that all subjects are compared to each other in fair conditions.

[12]. The 3dMD models were rescaled so that the distance between tear ducts was the same as in the visible PS reconstruction. All reconstructions were then cropped to 160×200 px regions centred on the nose tip that encompass the eyebrows and mouth. Part of the forehead is omitted by this choice of cropping region as it is frequently occluded by hair and is therefore deemed unreliable for face recognition. An example of the face regions used for comparison can be seen in Fig. 5, which also shows a ground truth reconstruction acquired using a 3dMD scanner. The face regions from visible and NIR light sources are then aligned to ground truth using the Iterative Closest Point (ICP) algorithm [37].

Individual RMS and ℓ_2-norm error results between the reconstructions and ground truth are displayed in Fig. 6. The eight subjects consist of 6 males and 2 females and a mixture of Caucasian and Asian ethnicities. The variations in residual errors and ℓ_2-norm distances between visible and NIR reconstructions are significant according to paired t-tests ($p = 0.05$). This demonstrates that PS using NIR as a light source is a perfectly valid approach and leads to more accurate reconstructions.

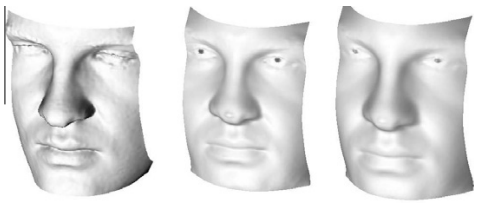


Fig. 5. 3D reconstructions for one subject from a 3dMD scanner (left) which is used as ground truth, PS using visible light sources (middle), and PS using NIR sources (right).

In order to get an indication of the regions where the greatest differences occur between ground truth and PS reconstructions, the residuals and ℓ_2-norm errors at each pixel are plotted in Fig. 7. Typically, the largest variations occur in regions with the highest curvatures, such as eye sockets, nose tips and the sides of the nose.

In attempting to produce the most accurate reconstructions possible via PS, it was found that the estimated surface normals could be enhanced by using normals acquired by re-differentiating the reconstructed height map estimate. It is unclear as to why this should be the case but preliminary analysis indicates that the reason may be due to the imposition of integrability constraints and the fitting of limited basis functions in the Fourier domain [36], as required by our adopted integration method. These factors may cause errant normals to be "smoothed out" leading to a more accurate reconstruction. However, if this method of improving reconstructions is used, a second integration step would be needed thus removing one of the benefits of PS for face recognition: that the surface normals (and hence distinctive differential surface features) are recovered directly. More research is required into this area in order to confirm that the improvements result from the imposed integrability constraints.

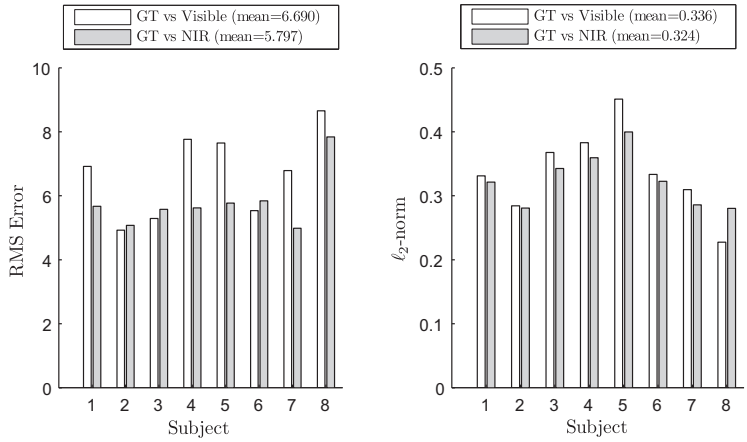


Fig. 6. RMS (left) and ℓ_2-norm (right) errors between Ground Truth (GT) and visible PS and NIR PS for each subject. *NB*: The order of subjects is arbitrary, i.e. there is no significance to the pattern that can be inferred from the ℓ_2-norm errors figure.

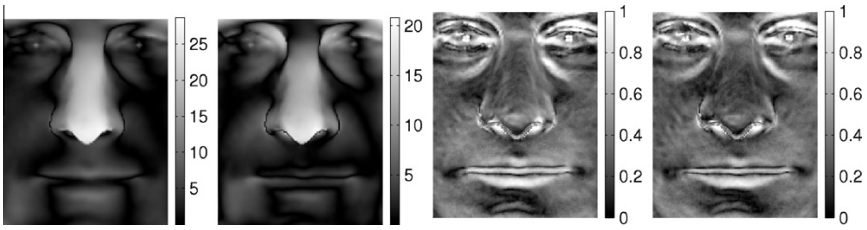


Fig. 7. Representative examples of the residuals and the ℓ_2-norm errors at each pixel. Left to right: residuals for visible and NIR respectively, ℓ_2-norm errors for visible and NIR respectively. Lighter areas represent larger errors.

3.2. Mitigating shadow effects

To demonstrate the improvements in surface normal estimates using the novel technique described in Section 2.4, we have manually selected a square region of the image that contains a strong shadow. Fig. 8 illustrates the surface normals of this region estimated by various methods. Two improvements of the proposed method compared to standard PS are noteworthy. Firstly, slightly finer details of the face are estimated using the new method. Secondly, the area immediately to the left of the nose is badly corrupt in the standard PS estimate, whereas the shadow is barely noticeable for the new method. To quantify the improvement, the ℓ_2-norm error was calculated between the two estimates and ground truth. Using standard PS for the region in Fig. 8, the ℓ_2-norm error was 0.32, while for the new method, the error dropped to 0.30. The difference in error between methods for the entire face is negligible as most regions are not in strong shadows.

Although the difference in ℓ_2-norm error is very small, this could be significant in certain applications as the nose and surrounding regions of the face offer useful biometrics. This area is seldom occluded by headgear/spectacles, varies considerably between individuals [38] and is relatively invariant to expression. Interestingly also, psychological research has shown that the nose is a preferred fixation point for humans attempting face recognition [39]. We should point out however, that although the surface normals and depth are improved through our method, the discontinuities in surface orientation at two-source shadow boundaries may cause reduced Sobolev-norm errors in some cases. In future work, we hope to reduce Sobolev-norm errors by treating two-source shadowed regions in a different manner to one-source shadowed regions.

For comparison, we have also implemented the Solomon and Ikeuchi PS method [19]. Their method is similar to ours in that combinations of three light sources are used to address shadowing/specularity issues. Where the albedo estimates from each combination of sources differs by an amount related to the standard deviation of camera noise at each pixel, σ_i, it is assumed that a shadow or specularity is present. For simplicity here, we assumed that σ_i is constant for all i and estimated the camera noise from 100 images of a planar white surface at close range. The ℓ_2-norm error for the region in Fig. 8 was 0.29 using the calculated value of $\sigma_i = 0.54\,\forall\, i$. Therefore, our method is comparable to the Solomon and Ikeuchi method in terms of accuracy. However, our method does not require camera noise information in order to attain optimum quality.

The method described in this section can enhance PS shape estimates for images that contain shadows. This means that PS can be used for facial reconstruction with somewhat arbitrarily positioned light sources. For the sake of simplicity, we will assume that the sources are positioned as in Fig. 1 for the rest of the paper and conduct the remainder of our analysis work using standard PS.

3.3. Reflectance analysis

To determine whether Lambert's law is obeyed more strictly under NIR light than visible, we have plotted graphs of I/ρ against θ, the angle between the light source and the normal vector. For a purely Lambertian surface, the relationship between the two should follow a cosine law. The results can be seen in Fig. 9. To generate the graph, values of I, ρ and θ were estimated for each pixel of each image for each of eight faces. The angle θ is calculated for each point of the face from the 3dMD scan data and the known light source vectors. The average values of I/ρ are used for each 1° increment in θ. The line at $\theta = 60°$ indicates a reasonable cut-off point after which data points become too sparse to be significant. The RMS difference between the measured curves and the cosine curve in the range of $0 \leqslant \theta \leqslant 60$ is 0.04 (s.d. 0.11) for NIR light and 0.06 (s.d. 0.12) for visible. For completeness, the RMS difference across the whole curve is 0.11 (s.d. = 0.13) for NIR light and 0.17 (s.d. = 0.12) for visible. The figure demonstrates that skin under NIR light is marginally more Lambertian than under visible light.

Although the data suffers from significant noise levels (as indicated by a standard deviation exceeding 10% of the range for both conditions), the NIR condition has a lower RMS error and is therefore closer to the Lambertian curve than for visible light. This difference is significant given the large numbers of pixels and subjects used in the trials. This represents an average pixel intensity error of 10 greylevels for NIR and 15 for visible light across the image, assuming a maximum of 256 grey level intensities. This supports the hypothesis that skin is more Lambertian under NIR

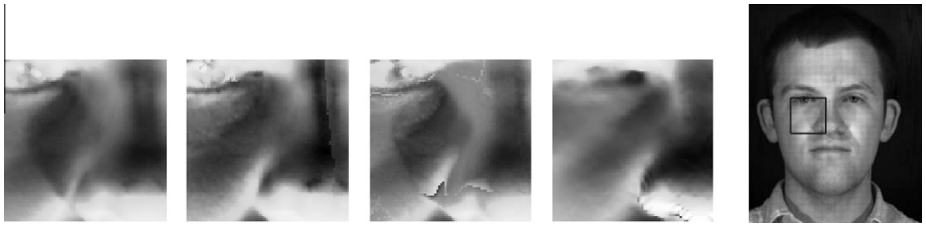


Fig. 8. Image of the vertical component of the surface normal (lighter areas indicate more downward pointing normals) for a region of a face in shadow. From left: estimates using standard four-source PS, using the proposed new algorithm, using the Solomon and Ikeuchi method, using the 3dMD scanner.

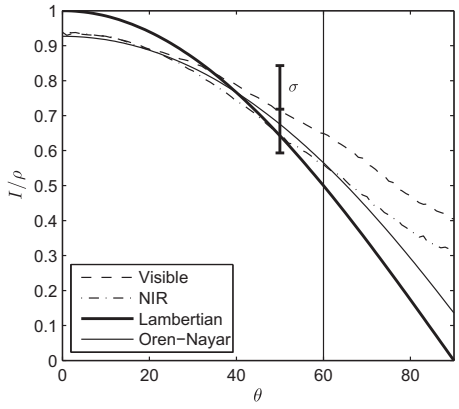


Fig. 9. Mean I/ρ values averaged over eight subjects against θ. To the right of the vertical line at $\theta = 60°$, data were too sparse to be of significance. For reference *one* standard deviation is shown to give an indication of spread.

illumination. We believe that this result is related to the fact that NIR light penetrates more deeply into the skin than visible light [40], which facilitates a more uniform scattering than surface reflection. Note however, that neither the Lambertian model nor the Oren–Nayar model (see below) take account of internal scattering or Fresnel effects. The results in Section 3.1 demonstrate that the more Lambertian behaviour associated with NIR light also leads to more accurate reconstructions.

A more detailed analysis for two individual subjects is shown in Fig. 10 and Table 1. What can be noted immediately is the similarity across the plots. There are small differences in I/ρ caused by different light sources but this appears to have little negative impact on the reconstructions and is likely to be due to environmental effects. The figure suggests that PS using both visible and NIR is robust to different skin types and light intensities. A more thorough analysis of the effects of gender and race on reflectance properties will be the subject of future work.

3.3.1. Comparison to the Oren–Nayar model

We have also compared our reflection measurements to the Oren–Nayar reflectance model [3], as shown in Fig. 9. The Oren–Nayar model represents the reflecting surface as an array of V-shaped groves of random orientation, commonly called "microfacets". The distribution of microfacet orientations is characterised by a roughness parameter and each facet is assumed to act as perfect Lambertian reflector. This model is able to account for the

Table 1
The RMS collective error across all eight reconstructions and for the first two reconstructions shown in Fig. 4 separately. The standard deviations are shown in brackets.

	Visible		NIR	
	RMS, $\theta \leqslant 60°$	RMS, overall	RMS, $\theta \leqslant 60°$	RMS, overall
All faces	0.06 ($\sigma = 0.11$)	0.16 ($\sigma = 0.12$)	0.04 ($\sigma = 0.12$)	0.11 ($\sigma = 0.13$)
Subject 1	0.07 ($\sigma = 0.09$)	0.16 ($\sigma = 0.18$)	0.05 ($\sigma = 0.12$)	0.10 ($\sigma = 0.22$)
Subject 2	0.07 ($\sigma = 0.10$)	0.17 ($\sigma = 0.18$)	0.04 ($\sigma = 0.13$)	0.12 ($\sigma = 0.21$)

common feature of limb-brightening and is itself based on the earlier Torrance–Sparrow model [41] where each microfacet is assumed to be mirror-like.

We have chosen to use the Oren–Nayar model as skin is not a smooth surface (especially on older people) and the model has been shown previously to be successful on a range of materials of varying degrees of roughness [3]. We do not believe that the microscopic structure of skin closely matches the Oren–Nayar model, but are merely demonstrating how alternate methods for reflection may improve our framework in future work. Investigating the various degrees of freedom of the BRDFs is also reserved for future work. Furthermore, there are additional models for skin reflectance which take account of a huge range of physical phenomena [42,43], but these are out of the scope of this paper.

The Oren–Nayar curve in Fig. 9 represents an example intensity profile for reference with a roughness parameter of 0.2. Clearly, this model fits the measured reflectance data significantly more accurately than the Lambertian curve, suggesting that the model could be incorporated into the method in the future. This will however, add significant complexity and computation time to the algorithm. This is because a minimisation method must be implemented in order to recover all the model parameters and to accommodate the increased number of angular degrees of freedom in the model.

4. Discussion

The results presented in Section 3 demonstrate that PS is an effective method for producing 3D facial reconstructions in terms of quality. Our method also requires a relatively short computation time. Using the device with standard PS, LabVIEW interfacing, MatLab processing and a typical modern PC, the time between device trigger and the reconstructed height map was approximately 4 s. Use of the optimised light source algorithm adds a further 3 s of computation time. The construction of the hardware also lends itself well to relatively unobtrusive data capture with a minimum

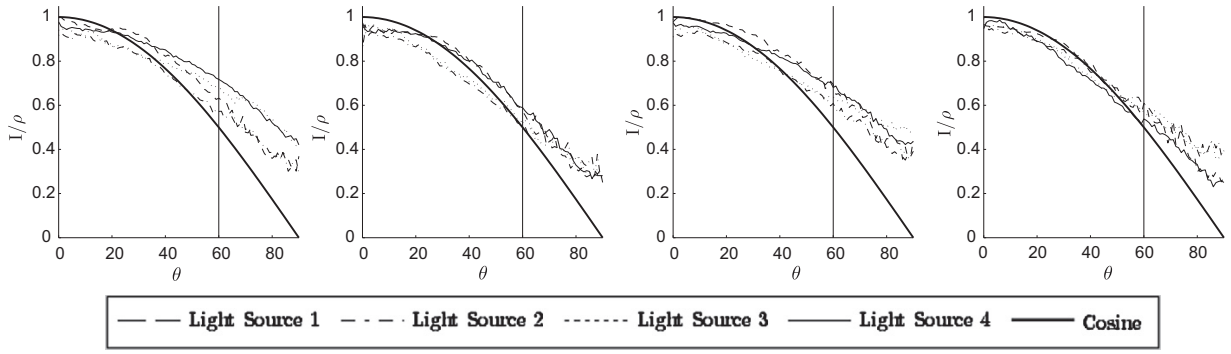


Fig. 10. I/ρ values from individual light sources plotted against θ for the first two reconstructions shown in Fig. 4. Left to right: Subject 1 under visible, Subject 1 under NIR, Subject 2 under visible, Subject 2 under NIR. The light sources are labelled clockwise from the bottom-left in Fig. 1.

amount of effort from the subject. Of particular interest are the following points:

1. The PS technique offers a valid alternative to existing, more expensive and processor intensive, 3D face capture methods.
2. The PS technique is robust to common facial features such as spectacles, makeup and facial hair (see also [44]).
3. NIR light sources produce reconstructions that are more accurate than visible light sources.
4. The optimised light source method described in Section 2.4 permits arbitrary light source arrangements and the presence of shadows.

Our system offers several benefits over commonly used existing laser triangulation and projected pattern 3D shape capture devices:

1. It is significantly cheaper to construct.
2. Acquisition time is shorter than laser triangulation systems.
3. Data processing time is shorter than projected pattern systems.
4. The method is robust to typical ambient illumination conditions.
5. It is very robust against accidental collisions (because it is tolerant to errors in the light source vectors).
6. Very fine details of the face can be reconstructed.
7. Calibration is very quick and simple and only needs to be performed after the initial light source positioning.
8. Due to the optimised light source method, the light sources can be positioned conveniently for different physical environments.
9. Although our system cannot reconstruct hair with high levels of accuracy, it can at least provide some details of its overall shape (see Fig. 3, for example). In contrast, laser triangulation and projected pattern systems usually fail completely with hair.

At present, the 3D reconstructions are not yet as accurate as those from projected pattern range finders. The reconstructions tend to be flatter than their real-world counterparts, with most protrusions understated. They do however provide extremely fine detail of a face such as wrinkles and pores. Even though the reconstructions suffer from a flattening of the features, they would still appear to be viable for recognition purposes (each reconstruction is clearly of a distinct identity) and the additional fine details could potentially be used as supplementary information to aid recognition.

The reconstructions under NIR were shown to be more accurate than those under visible light, but provided no additional 2D texture information. They also diminish the need for flashing lights, making the system less intrusive compared to visible light.

Zivanov et al. [30] offer an alternative argument to ours, stating that shorter wavelength light gives better results. Their justification is that shorter wavelengths undergo less internal scattering and thus provide a crisper, more defined reconstruction. It would appear therefore that a compromise must be reached in deciding between fine detail (using Zivanov's short wavelength suggestion) and overall geometry and covertness (using our NIR method).

4.1. Limitations and future research

One current limitation of the hardware described in this paper is that it does not cope with large deviations of peoples' height. Extremely tall or short people, or wheelchair bound persons would probably trigger the device correctly, but the location of the face could be outside of the field of view of the camera. Two possible solutions for this are (1) to use two cameras and trigger sensors at different heights or (2) to increase the field of view of the camera. The first solution would work by using the most suitable camera depending on which sensor had been triggered. While this is a straightforward solution it would increase the cost of the equipment considerably as the camera is the most expensive piece of apparatus. Increasing the field of view is also straightforward and would provide an adequate solution so long as wide-angle lens distortions did not become evident and that the face remains large enough on the image to provide discriminating information for the later recognition process.

Another improvement which could be made involves detecting the coordinates of the face and adjusting the light source vectors accordingly to improve the accuracy of the PS reconstruction. In the current system the light source unit vectors are calculated from a point at the centre of the camera's field of view and this is used for all reconstructions regardless of where the face is actually located. For this reason, the light source unit vectors are less accurate if the person walking through the device does not locate their face near the centre of the camera's field of view. The exact error caused by this inaccuracy is unknown, but amending the light source angles on a per person basis will improve the surface normal estimates.

5. Conclusion

This paper has brought together a number of advances in state-of-the-art 3D capture and processing technology. We have presented an algorithm for selecting the optimal light sources used for PS reconstruction which has the advantage over similar algorithms of not requiring knowledge of the exact shadow boundary.

The novel 3D facial geometry capture device has proved to be capable of reconstructing 3D models of faces under realistic workplace conditions using both visible and NIR light sources. It is cheaper, more robust and requires less calibration than alternative 3D acquisition devices. Although its reconstructions are less accurate than those of the state-of-the-art commercial 3dMD system, they are suitable for face recognition, which will be the focus of further study. The paper has also shown how human skin is more Lambertian under NIR light which is offered as an explanation for the associated improved accuracy of the reconstructions. A detailed error analysis for these results was also presented.

Acknowledgments

We would like to thank the EPSRC for funding this research and General Dynamics UK Ltd. for assistance in testing and data collection. We would also like to thank Imperial College, London and the Home Office Scientific Development Branch for their ongoing collaboration on this project.

References

[1] W. Zhao, R. Chellappa, Face Processing: Advanced Modeling and Methods, Academic Press, 2006.
[2] K.W. Bowyer, K. Chang, P. Flynn, A survey of approaches and challenges in 3D and multi-modal 3D+ 2D face recognition, Comput. Vis. Image Und. 101 (1) (2006) 1–15.
[3] M. Oren, S.K. Nayar, Generalization of the Lambertian model and implications for machine vision, Int. J. Comput. Vis. 14 (1995) 227–251.
[4] W. Zhao, R. Chellappa, P.J. Phillips, A. Rosenfeld, Face recognition: a literature survey, ACM Comput. Surv. (2003) 399–458.
[5] B. Gökberk, M.O. İrfanoğlu, L. Akarun, 3D shape-based face representation and feature extraction for face recognition, Image Vis. Comput. 24 (8) (2006) 857–869.
[6] V. Blanz, T. Vetter, Face recognition based on fitting a 3D morphable model, IEEE Trans. Pattern Anal. Mach. Intell. (2003) 1063–1074.
[7] A.M. Bronstein, M.M. Bronstein, R. Kimmel, Three-dimensional face recognition, Int. J. Comput. Vis. (2005) 5–30.
[8] I. Mpiperis, S. Malassiotis, M.G. Strintzis, 3-D face recognition with the geodesic polar representation, IEEE Trans. Inform. Forensic Security (2007) 537–547.
[9] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, Overview of the face recognition grand challenge, in: Proceedings of the CVPR, vol. 1, 2005.

[10] <www.konicaminolta.com/sensingusa/products/3d/non-contact/vivid910> (accessed 31.03.10).
[11] F. Chen, G.M. Brown, M. Song, Overview of three-dimensional shape measurement using optical methods, Opt. Eng. 39 (2000) 10.
[12] <www.3dmd.com/3dmdface.html> (accessed 31.03.10).
[13] B.K.P. Horn, Shape from shading: a method for obtaining the shape of a smooth opaque object from one view, Ph.D. thesis, MIT, 1970.
[14] P.N. Belhumeur, D.J. Kriegman, A.L. Yuille, The bas-relief ambiguity, Int. J. Comput. Vis. 35 (1999) 33–44.
[15] R.J. Woodham, Photometric method for determining surface orientation from multiple images, Opt. Eng. 19 (1) (1980) 139–144.
[16] W. Smith, E. Hancock, Facial shape-from-shading and recognition using principal geodesic analysis and robust statistics, Int. J. Comput. Vis. 76 (2008) 71–91.
[17] C. Hernández, G. Vogiatzis, R. Cipolla, Shadows in three-source photometric stereo, in: Proceedings of the ECCV, 2008, pp. 290–303.
[18] E.N. Coleman, R. Jain, Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry, Comput. Vis. Image Process. (1982) 309–328.
[19] F. Solomon, K. Ikeuchi, Extracting the shape and roughness of specular lobe objects using four light photometric stereo, IEEE Trans. Pattern Anal. Mach. Intell. 18 (1996) 449–454.
[20] S. Barsky, M. Petrou, The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows, IEEE Trans. Pattern Anal. Mach. Intell. 25 (2003) 1239–1252.
[21] K.E. Torrance, E.M. Sparrow, Theory for off-specular reflection from roughened surfaces, J. Opt. Soc. Am. A 57 (9) (1967) 1105–1112.
[22] A.S. Georghiades, Recovering 3-D shape and reflectance from a small number of photographs, in: Proceedings on Eurographics Workshop on Rendering, Eurographics Association, Leuven, Belgium, 2003, pp. 230–240.
[23] J. Sun, M. Smith, L. Smith, S. Midha, J. Bamber, Object surface recovery using a multi-light photometric stereo technique for non-Lambertian surfaces subject to shadows and specularities, Image Vis. Comput. 25 (7) (2007) 1050–1057.
[24] A. Ghosh, T. Hawkins, P. Peers, S. Frederiksen, P. Debevec, Practical modeling and acquisition of layered facial reflectance, in: International Conference on Computer Graphics and Interactive Techniques, 2008.
[25] S.C. Kee, K.M. Lee, S.U. Lee, Illumination invariant face recognition using photometric stereo, IEICE Trans. Inf. Syst. E Ser. D 83 (7) (2000) 1466–1474.
[26] S.K. Zhou, R. Chellappa, D.W. Jacobs, Characterization of human faces under illumination variations using rank, integrability, and symmetry constraints, Proc. ECCV (2004) 588–601.
[27] S.K. Zhou, G. Aggarwal, R. Chellappa, D.W. Jacobs, Appearance characterization of linear Lambertian objects, generalized photometric stereo, and illumination-invariant face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 29 (2) (2007) 230–245.
[28] A.S. Georghiades, P.N. Belhumeur, D.J. Kriegman, From few to many: illumination cone models for face recognition under variable lighting and pose, IEEE Trans. Pattern Anal. Mach. Intell. (2001) 643–660.
[29] R.R. Anderson, J.A. Parrish, The optics of human skin, J. Invest. Dermatol. 77 (1) (1981) 13–19.
[30] J. Zivanov, P. Paysan, T. Vetter, Facial normal map capture using four lights – an effective and inexpensive method of capturing the fine scale detail of human faces using four point lights, in: GRAPP, 2009, pp. 13–20.
[31] S.Z. Li, R.F. Chu, S.C. Liao, L. Zhang, Illumination invariant face recognition using near-infrared images, IEEE Trans. Pattern Anal. Mach. Intell. (2007) 627–639.
[32] S.G. Kong, J. Heo, B.R. Abidi, J. Paik, M.A. Abidi, Recent advances in visual and infrared face recognition – a review, Comput. Vis. Image Und. 97 (1) (2005) 103–135.
[33] C.H. Morimoto, D. Koons, A. Amir, M. Flickner, Pupil detection and tracking using multiple light sources, Image Vis. Comput. 18 (4) (2000) 331–335.
[34] R. Lienhart, J. Maydt, An extended set of haar-like features for rapid object detection, in: IEEE ICIP, vol. 1, 2002, pp. 900–903.
[35] D.A. Forsyth, J. Ponce, Computer Vision: A Modern Approach, Prentice Hall Professional Technical Reference, 2002.
[36] R.T. Frankot, R. Chellappa, A method for enforcing integrability in shape from shading algorithms, IEEE Trans. Pattern Anal. Mach. Intell. 10 (4) (1988) 439–451.
[37] P.J. Besl, H.D. McKay, A method for registration of 3-D shapes, IEEE Trans. Pattern Anal. Mach. Intell. 14 (2) (1992) 239–256.
[38] A. Moorhouse, A.N. Evans, G.A. Atkinson, J. Sun, M.L. Smith, The nose on your face may not be so plain: using the nose as a biometric, in: International Conference on Imaging for Crime Detection and Prevention, 2009.
[39] J.H.W. Hsaio, G. Cottrell, Two fixations suffice in face recognition, Psychol. Sci. 19 (2008) 998–1006.
[40] C. Fredembach, N. Barbuscia, S. Süsstrunk, Combining visible and near-infrared images for realistic skin smoothing, in: Proceedings of the IS&T/SID Color Imaging Conference, 2009.
[41] K. Torrance, M. Sparrow, Theory for off-specular reflection from roughened surfaces, J. Opt. Soc. Am. 57 (1967) 1105–1114.
[42] C. Donner, H.W. Jensen, Light diffusion in multi-layered translucent materials, ACM Trans. Graph. 24 (2005) 1032–1039.
[43] L. Li, C.S. Ng, Rendering human skin using a multi-layer reflection model, Int. J. Math. Comput. Simul. 3 (2009) 44–53.
[44] G. Atkinson, M. Smith, Facial feature extraction and change analysis using photometric stereo, in: Proceedings of the IbPRIA, 2009, pp. 96–103.

A.2 Computationally efficient bio-inspired dimension reduction for 3D faces

This poster was presented after being selected as one of ten candidates for the Doctorial Consortium of the Automated Face and Gesture Recognition conference held in Santa Barbara, USA in March 2011. This competitive process attracted funding of $1500 to cover travel and accommodation. The British Machine Vision Association also provided additional funding through a travel bursary available to postgraduate students. This was a phenomenal experience whereby members of the Doctorial Consortium were assigned mentors during their stay. These were world class researchers who were able to provide critical review of one's work, general career advice, and a means of introduction to other leaders in the field. In this author's case, Prof. Kevin Bowyer from University of Notre Dame, USA was assigned. Kevin Boywer is perhaps best known for his contributions to the ubiquitous FERET and FRGC face databases, and has been in contact numerous times since leading to collaborative work on his latest project on recognising identical twins, which can be seen in the next section.

Computationally efficient bio-inspired dimensionality reduction for 3D faces

Mark Hansen, Machine Vision Laboratory
PhD Student. Supervisory Team: Dr. Gary Atkinson, Prof. Melvyn Smith, Dr. Lyndon Smith

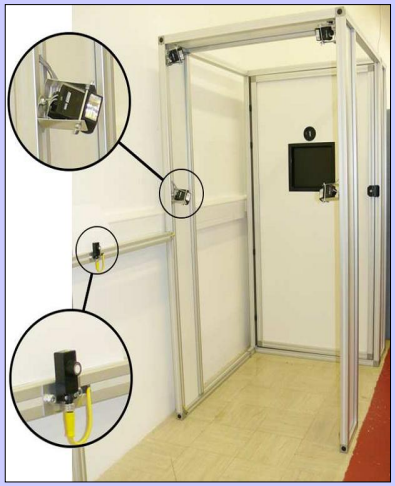
EPSRC Engineering and Physical Sciences Research Council

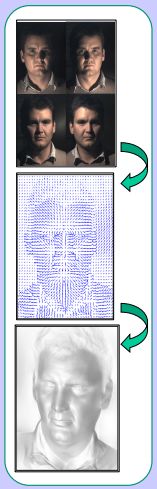
University of the West of England, BRISTOL

Aims and Objectives

- Images of faces are of high dimensionality
- This is compounded in 3D which has some advantages over 2D
- Can we use high level features of human vision to guide extraction of the "useful" data?

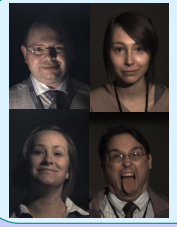
Image Acquisition & Hardware





- 4-source photometric stereo Photoface portal as seen on the Discovery Channel.
- Ultrasound triggers a capture time of approximately 15ms.
- Acquisition is unobtrusive and suited to work environments
- EPSRC, Home Office, Imperial College and General Dynamics

The Photoface database



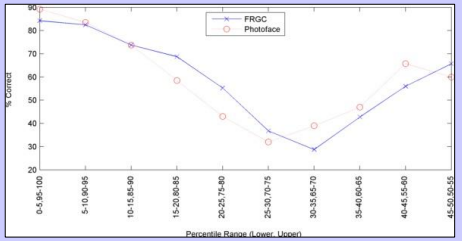
- Captured over 6 months in an unconstrained office environment
- Expressions, with/without facial hair, glasses etc
- Over 3000 sessions of 457 individuals

Unnikrishnan's Hypothesis

- Humans can recognise a caricature as quickly as a photograph
- Unnikrishnan (2009) suggests that it is uncommon features (e.g. a short or long nose) that provide useful discriminatory information.
- We test this by using only data which lies below the 5th and above the 95th percentile (outliers)
- 40 subjects, 10 sessions each, Pearson's correlation coefficient used for classification

	Baseline	Outliers
2D	98%	64%
3D	98.25%	89.25%

The outlier results on the left for the 3D data show that although we have discarded 90% of the data we have lost under 10% recognition performance. This is not shown by the 2D data.

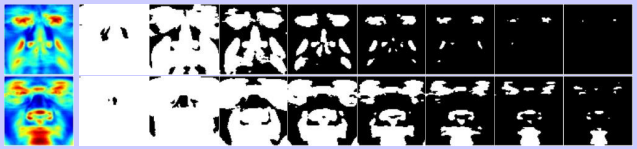


The most discriminating data can be found in the outliers, but that there is also useful information around the median. This finding is replicated in the FRGCv2.0 data.

1. M. K. Unnikrishnan, "How is the individuality of a face recognized?," *Journal of theoretical biology*, 2009.

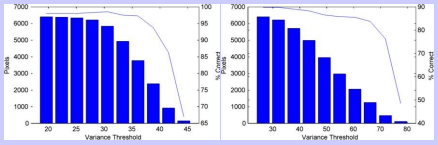
Variance Analysis

- Instead of using percentiles can we identify useful pixels based on their variance.

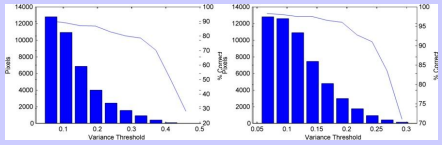


Masks of increasing variance thresholds for x (top) and y (bottom) components (white pixels are retained for recognition). Standard deviation across pixels is shown on the left.

- 75% of the least varying pixels can be removed and less than 10% of performance is lost for 3D data. (e.g. 92.75% accuracy after reducing 12800px to 1783px)



Photoface (left) and FRGC (right) 2D recognition performance (line) as a function of number of pixels retained (bar).



Photoface (left) and FRGC (right) 3D recognition performance (line) as a function of number of pixels retained (bar).

Implications

- Dimensionality can be greatly reduced using simple methods inspired by high level vision functions.
- Outlying pixels and those with high variance contain a disproportionate amount of discriminatory data
- This is useful for face storage and processing for recognition applications
- Future work will address invariance or robustness to expressions.

A.3 Twins 3D Face Recognition Challenge

This paper was presented at the International Joint Conference on Biometrics held in Washington, USA in 2011 which is a special combination of two major biometrics research conference traditions, the International Conference on Biometrics (ICB) and the Biometrics Theory, Application and Systems (BTAS) conference. This author was invited to collaborate after discussing work with Prof. Kevin Bowyer at the earlier Automated Face and Gesture Recognition conference. The paper was well received, and there is likely to be further collaboration opportunities.

Twins 3D Face Recognition Challenge

Vipin Vijayan [1], Kevin Bowyer [1], Patrick Flynn [1], Di Huang [2], Liming Chen [2], Mark Hansen [3], Omar Ocegueda [4], Shishir Shah [4], Ioannis Kakadiaris [4]

Abstract

The best performance in 3D face recognition algorithms has become high enough in large datasets like FRGC v2 [1] that it is difficult to achieve further significant increases in recognition performance. The 3D TEC dataset is a more challenging dataset which consists of 3D scans of 107 pairs of twins taken in a single session, with each subject having a scan of a neutral expression and a smiling expression. The combination of factors related to the facial similarity of identical twins and the variation in facial expression makes this a challenging dataset. We conduct four experiments using state of the art face recognition algorithms and present the results. Our results indicate that 3D face recognition of identical twins in the presence of varying facial expressions is far from a solved problem.

1. Introduction

Distinguishing between faces of identical twins is generally considered to be a difficult problem. In this paper, we conduct a study on the performance of state of the art 3D face recognition algorithms on a large set of identical twins using the 3D Twins Expression Challenge ("3D TEC") dataset. The dataset contains 107 pairs of identical twins and is the largest dataset of 3D scans of twins known to the authors.

Recently, there have been some twins studies in biometrics research. Phillips et al. [2] assessed the performance of three of the top algorithms submitted to the Multiple Biometric Evaluation (MBE) 2010 Still Face Track [3] on a dataset of twins acquired at Twins Days [4] in 2009 and 2010. They examined the performance using images acquired on the same day, and also using images acquired a year apart (i.e., where the face images acquired in the first year are gallery images and the face images acquired in the second year are probe images). They also examined the performance with varying illumination conditions and expressions. They found that results ranged from approximately 2.5% Equal Error Rate (EER) for images acquired on the same day with controlled lighting and neutral expressions, to approximately 21% EER for gallery and probe images acquired in different years and with different lighting conditions.

Sun et al. [5] conducted a study on multiple biometric traits of twins. They found no significant difference in performance when using non-twins compared to using twins for their iris biometric system. For their fingerprint biometric system, they observed that performance when using non-twins was slightly better than using twins. In addition, their face biometric system could distinguish non-twins much better than twins.

It is already known that iris texture is different enough between identical twins that it can be used to distinguish between them. Hollingsworth et al. [6] examined whether iris textures from a pair of identical twins are similar enough that they can be classified by humans as being from twins. They conducted a human classification study and found that people can classify two irises as being from the same pair of twins with 81% accuracy when only the ring of iris texture was shown to them.

Jain et al. [7] conducted a twins study using fingerprints. They found that identical twins tend to share the same fingerprint class (fingerprints are classified into whorls, right/left loops, arches, etc.) but their fingerprint minutiae were different. They concluded that identical twins can be distinguished using a minutiae based automatic fingerprint system with slightly lower performance compared to distinguishing random persons. They also hypothesized that the difference in performance is because the minutiae based system may depend on fingerprint class.

To date, there have been no studies conducted in 3D face recognition that focused mainly on twins. Bronstein et al. [8] tested the performance of their 3D face recognition algorithm on a dataset of 93 adults and 64 children which contained one pair of twins, and stated that "our methods make no mistakes in distinguishing between Alex and Mike".

2. The Dataset

The Twins Days 2010 dataset was acquired at the Twins Days Festival in Twinsburg, Ohio [4]. Phillips et al. [2] provides more details about the overall dataset. It contains 266 subject sessions, with the 3D scans in the dataset containing two scans taken using a range scanner: one with a neutral

[1]Deparment of Computer Science and Engineering, University of Notre Dame. 384 Fitzpatrick Hall, Notre Dame, IN 46556, USA.
{vvijayan, kwb, flynn}@nd.edu

[2]Université de Lyon, CNRS, Ecole Centrale Lyon, LIRIS UMR 5205, 69134, Ecully, France.

[3]Machine Vision Lab, DuPont Building, Bristol Institute of Technology, University of the West of England, Frenchay Campus, Coldharbour Lane, Bristol BS16 1QY, UK.

[4]Department of Computer Science, University of Houston. 4800 Calhoun Road, Houston, TX 77004, USA.

Figure 1: Images of two twins taken in a single session. The top row shows the images obtained from one twin and the bottom row, the other twin. The left two images on each row shows the 3D and texture images taken of the neutral expression and the right two shows the smiling. (The texture images were brightened to increase visibility in this figure.)

expression and another with a smiling expression. There were 106 sets of identical twins, one set of triplets, and the rest were non-twins. Three pairs of twins came in for two recording sessions and the other twins only had a single session. The twins in the database only declared themselves to be identical twins. No tests were done to prove this.

The experiments in this paper use the "3D TEC" subset of the Twins Days dataset, which consists of 3D face scans of 107 pairs of twins (two of the triplets were included as the 107th set of twins), and where only the first session for each person was used. To our knowledge, this is the only dataset of 3D face scans in existence that has more than a single pair of twins. For information on obtaining the 3D TEC dataset, see [9].

The scans were taken using a Minolta VIVID 910 3D scanner [10] in a controlled light setting, with the subjects posing in front of a black background. For each pair of twins, their neutral and smile images were taken in a 5 to 10 minute window of time.

The Minolta scanner takes a texture image and a range image of 480×640 resolution. The telephoto lens of the Minolta scanner was used since it gives a more detailed scan. The distance of the scanner from the subject was approximately 1.2 m. A scan using the telephoto lens contains 70,000 to 195,000 points for the Twins 2010 dataset, with an average of 135,000 points.

3. Algorithms

We describe the four algorithms employed in this study. Table 1 shows the performance of these algorithms on the FRGC v2 dataset.

3.1. Algorithm 1

Algorithm 1, described in McKeon [11] has a number of optimizations over Faltemier et al. which include (i) the symmetric use of the two point clouds and doing score fusion on the results, (ii) score normalization of the match scores, and (iii) weighting the scores for the regions. Faltemier et al. [12] performed Iterative Closest Point (ICP) using an ensemble of 38 spherical regions and fused the match scores of each region to calculate the final match score.

Algorithm 1 is not an exact replica of the algorithm in McKeon [11]. The major difference is the preprocessing step where first roughly align the face using the rough symmetry plane estimation method described in Spreeuwers [13], and then finely align the image to a reference face using ICP.

We have two point clouds, a probe image p and a gallery image g. Each region in the ensemble is created by selecting a point in p at a certain offset from the origin, and then intersecting a sphere around the selected point with the face. For Algorithm 1, the nose-tip is set as the origin.

The match scores are calculated by applying ICP to each region in order to calculate the region's final alignment error from the gallery image. The scores are then fused by weighting and then adding each of the regions scores. A modification of Sequential Forward Search (SFS) [14], called SFS Weighted (SFSW) is used to find the region ensemble with the best performance. SFSW is a greedy algorithm and works as follows. Begin with a sequence $S = \{(\beta_1, R_1), (\beta_2, R_2), ...(\beta_n, R_n)\}$ where R_i are the match scores for the region, and β_i are integer weights, initialized to 0, associated with each region. The fused match

score from S would be $E_{SFSW} = \sum_{i=1}^{n} \beta_i R_i$. For each iteration, find the region R_k that when its weight, β_k, is incremented gives the maximum TAR at 0.1% FAR using E_{SFSW} as the match score. Then S is modified by incrementing the weight, β_k. This is repeated n times.

Let $E(p_1, p_2) = E_{SFSW}(p_1, p_2)$, be the match score of point clouds p_1 and p_2. The ICP algorithm is not symmetric, which means that $E(p_1, p_2) \neq E(p_2, p_1)$ for almost all cases. The two scores are fused using the minimum rule: $E_{min}(p_1, p_2) = \min(E(p_1, p_2), E(p_2, p_1))$, which provides better performance.

The match scores are then normalized in two ways. First, the match scores are normalized such that the normalized score is

$$E_{pkn}(p, g_k) = \frac{E_{min}(p, g_k)}{\sum_{j=1, j\neq k}^{N} \frac{E_{min}(g_j, g_k)}{N-1}} \quad (1)$$

where p is a probe image, g_k are the gallery images, and N is the number of gallery images.

Then we perform min-max normalization over the resulting match score from the first normalization, E_{pkn}, so that the final match score is

$$E_{minmax}(p, g_k) = \frac{E_{pkn}(p, g_k) - \min(V_p)}{\max(V_p) - \min(V_p)} \quad (2)$$

where $V_p = [E_{pkn}(p, g_1), E_{pkn}(p, g_2), ..., E_{pkn}(p, g_N)]$.

Algorithm 1 uses E_{min} along with score normalization and SFSW (E_{SFSW}) with only the regions calculated by setting the nose-tip as the origin.

3.2. Algorithm 2

Algorithm 2 consists of two main steps: intermediate facial representation and Scale Invariant Feature Transform (SIFT) based local matching.

Currently, local feature extraction mostly operates directly on smooth facial range images, and it hence leads to limited number of local features or ones with low discriminative power. To solve this problem, intermediate facial representation is used to highlight local shape changes of 3D facial surfaces in order to improve their distinctiveness. In this paper, we evaluated three types of intermediate facial maps: Shape Index [15], extended Local Binary Patterns [16] as well as Perceived Facial Images [17]. Figure 2 shows examples of these facial maps. The three types of facial maps are described below.

Shape Index (SI) [15] was first proposed to describe shape attributes. For each vertex p of a 3D facial surface, its SI value can be calculated using

$$S(p) = \frac{1}{2} - \frac{1}{\pi} \arctan \frac{k_1(p) + k_2(p)}{k_1(p) - k_2(p)} \quad (3)$$

where k_1 and k_2 are the maximum and minimum principal curvatures respectively. Based on the SI values of all the vertices, we can produce the SI map of a given facial surface.

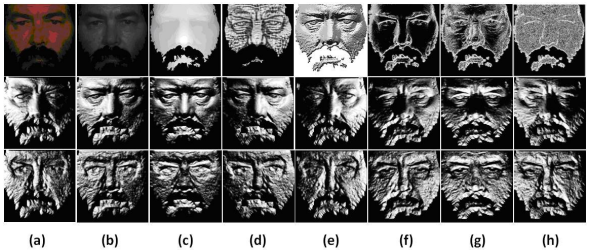


Figure 2: Some examples of intermediate facial representation. The first row contains (a) original RGB image; (b) grayscale texture image; (c) original range image; (d) SI map; (e)-(h) eLBP maps of different layers. The second row contains eight PFIs of quantized orientations of facial range image. The third row contains eight PFIs of quantized orientations of facial texture image.

In the Extended Local Binary Pattern (eLBP) [16] approach, a set of multi-scale eLBP maps are generated to represent a given facial range image. eLBP maps consist of four layers. Layer 1 is LBP, which encodes the gray value differences between neighboring pixels into a binary pattern. eLBP also considers their exact value differences and encodes this information into Layers 2 to 4. The eLBP maps are generated by regarding the eLBP codes of each pixel as intensity values. As the neighborhood size of the given pixel changes, multi-scale eLBP maps are formed.

Perceived Facial Image (PFI) [17] aims at simulating the complex neuron response using a convolution of gradients in various orientations within a pre-defined circular neighborhood. Given an input facial image I, a certain number of gradient maps L_1, L_2,..., L_o, one for each quantized direction o, are first computed. Each gradient map describes gradient norms of the original image in an orientation o at every pixel. We then simulate the response of complex neurons by convolving its gradient maps with a Gaussian kernel G, and the standard deviation of G is proportional to the radius value of the given neighborhood area R.

$$\rho_o^R = G_R * L_o \quad (4)$$

The purpose of the Gaussian convolution is to allow the gradients to shift in a neighborhood without abrupt changes. At a certain pixel location (x, y), we collect all the values of the convolved gradient maps at that location and form the vector $\rho^R(x, y)$, which a response value of complex neurons for each orientation o.

$$\rho^R(x, y) = \left[\rho_1^R(x, y), \cdots, \rho_O^R(x, y)\right]^t \quad (5)$$

The vector $\rho^R(x, y)$ is further normalized to a unit norm vector $\underline{\rho}^R(x, y)$, which is called response vector. This, a new Perceived Facial Image (PFI), J_o, is achieved using

complex neurons for each orientation o defined as

$$J_o(x, y) = \underline{\rho}_o^R(x, y) \quad (6)$$

After the three types of intermediate facial representations are computed, a SIFT-based matching process [18] is used to find robust keypoints from the facial representations. It is expected to associate many more keypoints between facial maps of the same subject than those of different subjects. Furthermore, since SIFT has good tolerance to moderate pose variations and all the data in the 3D TEC dataset are nearly frontal scans, we did not perform any registration in preprocessing. All parameter settings of intermediate facial representations are presented in detail in [15, 16, 17].

In addition, SI maps and eLBP maps are mainly proposed for 3D facial range images, while PFIs can be either applied to facial range or texture images as done in Huang et al. [17] for 3D face recognition using shape and texture. Therefore, in this paper, we also tested the performance based on 2D PFIs with SIFT matching for comparison.

3.3. Algorithm 3

Algorithm 3 converts the 3D image to a surface normal representation, then discards data with less discriminatory power and resizes the image. It then matches the images using Euclidean distance of the variance of the remaning surface normals.

Surface normals have been shown to lend themselves well to face recognition tasks [19]. We convert the depth maps of 3D images to surface normal representations, applying median smoothing and hole filling to reduce noise.

Unnikrishnan [20] conceptualised an approach similar to face caricatures for human recognition. In this approach, only those features which deviate from the norm by more than a threshold are used to uniquely describe a face. Unnikrishnan suggested using features whose deviations lie below the 5th percentile and above the 95th percentile, thereby discarding 90% of the data. In a similar vein, the algorithm that we present here is based on what we call the "Variance Inclusion Criterion". We can use the surface normal variance at each pixel location as a measure of discriminatory power. If a pixel shows a large variance across the dataset, then it can be used for recognition (assuming that variance within the class or subject is small). Therefore, the standard deviation of each pixel is calculated over all the images in the gallery. Whether or not a particular pixel location is used in recognition depends on whether or not the variance is above a pre-determined threshold.

Another key step of this algorithm is resizing the image. Sinha et al. [21] summarised a number of findings indicating that humans can recognise familiar faces from very low resolution images. We resize the surface normal maps to 10×10 pixels before applying the Variance Inclusion Criterion to get the number of pixels used for recognition down to just over 60 pixels. The reason for choosing this value is due to experimentation on frontal and neutral expression subsets of the FRGC v2 and Photoface [22] datasets. In these experiments it was found that when retaining only 64 pixels for FRGC v2 data and 61 pixels for Photoface data, rank-one recognition rates of 87.75% and 96.25% were achieved respectively (a loss of only 7% and 2% from the baseline). This is taken as an indication that the high variance pixel locations contain disproportionately more discriminatory information than low variance pixel locations.

Because of the two expressions used between gallery and probe images in the 3D TEC dataset, it was felt that the most variance would occur around the mouth region and bottom half of the face. Therefore we only performed the variance analysis on the top half of the face.

Additional pre-processing is performed by aligning all the images to the median left and right lateral canthus and nose tip coordinates for the dataset. A tight crop around the facial features is then applied to remove areas in a straightforward way that can be occluded by hair. Euclidean distance is used for classification.

It is envisaged that this algorithm be used as a means of pruning the search space due to its computational efficiency before applying more rigorous algorithms.

3.4. Algorithm 4

The UR3D algorithm proposed by Kakadiaris et al. [23] consists of three main steps: (i) the 3D facial meshes are aligned to a common reference Annotated Face Model (AFM), (ii) the AFM is deformed to fit the aligned data, and (iii) the 3D fitted mesh is represented as a three-channel image using the global UV-parametrization of the AFM. The benefit of representing the 3D mesh as a multi-channel image is that standard image processing techniques can be applied directly to the images. In this approach, the full Walsh wavelet packet decomposition is extracted from each band of the geometry and normal images and a subset of the wavelet coefficients are selected as the signature of the mesh. The signature can be compared directly using a weighted L^1 norm. Recently, Ocegueda et al. [24] presented an extension to UR3D that consists of a feature selection step that reduces the number of wavelet coefficients retained for recognition, followed by a projection of the signatures to a subspace generated using Linear Discriminant Analysis (LDA). The feature selection step was necessary because the high dimensionality of the standard UR3D signature made it infeasible to apply standard algorithms for LDA. However, by using the algorithm proposed by Yu and Yang [25], we can directly apply LDA to the original UR3D metric. We found that applying LDA to the original signature yields slightly better results. We will use this variation of the UR3D algorithm in our experiments. We used

the frontal, non-occluded facial meshes from the Bosphorus dataset developed by Savran et al. [26] as training set for LDA.

Algorithm	Rank 1 RR	VR (ROC III)
Alg. 1	98.0%	98.8%
Alg. 2 (SI)	91.8%	85.8%
Alg. 2 (eLBP)	97.2%	95.0%
Alg. 2 (Range PFI)	95.5%	90.4%
Alg. 2 (Text. PFI)	95.9%	
Alg. 3	97.0%	97.0%
Alg. 4	87.8%	

Table 1: Rank-one recognition rates and verification rates (TAR at 0.1% FAR) of the algorithms on the FRGC v2 dataset. For recognition, the earliest image acquired of each subject is in the gallery set and the rest of the images are probes. For the ROC III verification experiment, the gallery set contains the images acquired in the first semester and the probe set contains the images in the second semester.

4. Experimental Design

We arbitrarily label one person in each pair of twins as Twin A and the other as Twin B and perform four experiments with the different gallery and probe sets shown in Table 2.

No.	Gallery		Probe	
I	A Smile,	B Smile	A Neutral,	B Neutral
II	A Neutral,	B Neutral	A Smile,	B Smile
III	A Smile,	B Neutral	A Neutral,	B Smile
IV	A Neutral,	B Smile	A Smile,	B Neutral

Table 2: List of experiments performed.

Experiment I has all of the images with a smiling expression in the gallery and the images with a neutral expression as the probe. Experiment II reverses these roles. This models a scenario where the gallery has one expression and the probe has another expression. In the verification scenario, both the match and non-match pairs of gallery and probe images will have different expressions. In the identification scenario, theoretically the main challenge would be to distinguish between the probe's image in the gallery and his/her twin's image in the gallery since they look similar.

Experiment III has Twin A smiling and Twin B neutral in the gallery with Twin A neutral and Twin B smiling as the probe. Experiment IV reverses these roles. This models a worst case scenario in which the system does not control for the expressions of the subject in a gallery set of twins. In the verification scenario, the match pairs would have opposite expressions like in Experiments I and II but the non-match pairs which are twins would have the same expression. In the identification scenario, theoretically the main challenge would be to distinguish between the probe's image and his/her twin's image in the gallery. This is more difficult than Experiments I and II since the probe's expression is different from his/her image in the gallery but is the same as his/her twin's image in the gallery.

5. Results and Discussion

We evaluate performance using the following characteristics: True Accept Rate at 0.1% False Accept Rate (TAR at 0.1% FAR), Equal Error Rate, and Rank 1 Recognition Rate. Receiver Operating Characteristic (ROC) curves is also used for some of the algorithms to show their performance.

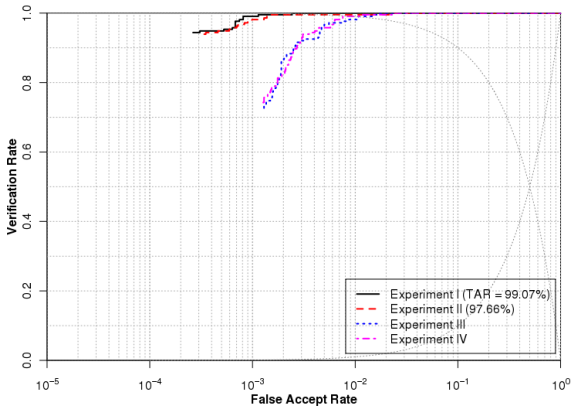


Figure 3: ROC curves of the four experiments for Algorithm 1. TAR at 0.1% FAR could not be calculated because of the normalization techniques used in the algorithm.

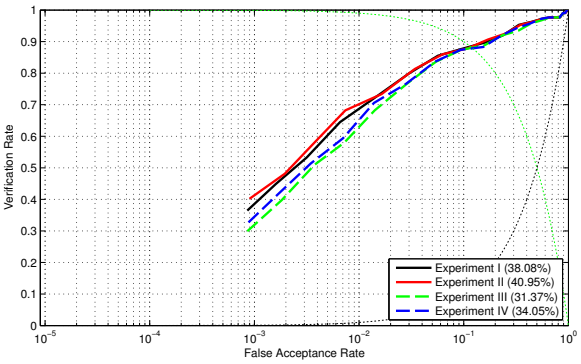


Figure 4: Verification performance of Algorithm 3.

In the first two of our four experiments, all subjects are enrolled with a 3D face scan that has one expression, and all recognition attempts are made with the other expression.

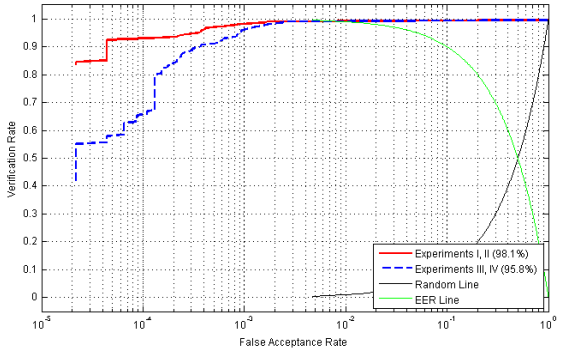


Figure 5: Verification performance of the extended UR3D algorithm.

Algorithm	True Accept Rate			
	I	II	III	IV
Alg.1	99.1%	97.7%		
Alg.2 (SI)	91.1%	89.7%	83.2%	81.8%
Alg.2 (eLBP)	94.4%	95.3%	79.0%	78.0%
Alg.2 (Range PFI)	93.5%	94.4%	68.7%	69.2%
Alg.2 (Text. PFI)	96.7%	96.7%	93.0%	93.5%
Alg.3	38.1%	41.0%	31.4%	34.1%
Alg.4	98.1%	98.1%	95.8%	95.8%

Table 3: TAR at 0.1% FAR of the algorithms.

Algorithm	Equal Error Rate			
	I	II	III	IV
Alg.1	0.2%	0.5%	1.1%	0.9%
Alg.2 (SI)	2.7%	3.7%	4.2%	4.5%
Alg.2 (eLBP)	3.7%	3.3%	4.2%	4.2%
Alg. 2 (Range PFI)	4.1%	2.8%	4.7%	4.6%
Alg.2 (Text. PFI)	2.7%	2.8%	3.3%	2.8%
Alg.3	11.6%	11.8%	12.0%	12.2%
Alg.4	0.8%	0.8%	0.8%	0.8%

Table 4: Equal Error Rate of the different algorithms.

Algorithm	Rank 1 Recognition Rate			
	I	II	III	IV
Alg.1	94.4%	93.5%	72.4%	72.9%
Alg.2 (SI)	92.1%	93.0%	83.2%	83.2%
Alg.2 (eLBP)	91.1%	93.5%	77.1%	78.5%
Alg.2 (Range PFI)	91.6%	93.9%	68.7%	71.0%
Alg.2 (Text. PFI)	95.8%	96.3%	91.6%	92.1%
Alg.3	62.6%	63.6%	54.2%	59.4%
Alg.4	98.1%	98.1%	91.6%	93.5%

Table 5: Rank-one recognition rates.

Thus, the difference in expression between enrollment and recognition is the same for all subjects. In these two experiments, we find that 3D face recognition accuracy for twins exceeds 90% for most of the algorithms. In the last two of the four experiments, the facial expression differs between the twins' enrollment images and also between their images for recognition. In these experiments, 3D face recognition accuracy ranges from the upper 60% to the lower 80%, except for Algorithm 2 (Texture PFI) which makes use of the texture information, and Algorithm 4. An exception is Algorithm 3, which showed reasonable performance on the FRGC v2 and Photoface [22] datasets but vastly degrades in performance on the 3D TEC dataset.

Why do some algorithms perform very well on this dataset while others don't? Algorithm 3, for example, discards a large amount of data by resizing and uses thresholded Euclidean distance which is a fairly simple classification method. Algorithm 1 on the other hand discards almost no data; it matches using the original point cloud that was scanned after some standard processing. The results also show a stark difference in the performances in Experiments I and II compared to III and IV for some of the algorithms. This demonstrates how good the algorithm is when dealing with different expressions.

The 3D TEC dataset contains only "same session" data, meaning that there is essentially no time lapse between the image used for enrollment and the image used for recognition. Therefore, any performance estimates from this data are biased to exceed those that can be expected in any practical application.

6. Conclusion

3D face recognition continues to be an active research area. We have presented results of different state of the art algorithms on a dataset representing 107 pairs of identical twins with varying facial expressions, the 3D Twins Expression Challenge ("3D TEC") dataset. These algorithms have previously been reported to achieve good performance on the FRGC v2 dataset, which has become a de facto standard dataset for evaluating 3D face recognition. However, we observe lower performance on the 3D TEC dataset. The combination of factors related to the facial similarity of identical twins and the variation in facial expression makes for an extremely challenging problem.

The 3D TEC Challenge is smaller and therefore computationally simpler than the FRGC v2 Challenge. It combines a focus on fine discrimination between faces and handling varying expression. There have been claims in the literature of 3D face recognition algorithms that can distinguish between identical twins. To our knowledge, this is the first time that experimental results have been reported for 3D face recognition involving more than a single pair of identical twins. The results demonstrate that 3D face recognition of identical twins in the presence of varying facial expressions remains an open problem.

7. Acknowledgements

Acquisition of the dataset used in this work was supported by the Federal Bureau of Investigation under US Army contract W91CRB-08-C-0093. This work was funded in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), through the Army Research Laboratory (ARL). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing official policies, either expressed or implied, of IARPA, the ODNI, the Army Research Laboratory, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

The contribution by D. Huang and L. Chen in this paper was supported in part by the French National Research Agency (ANR) through the FAR 3D project under Grant ANR-07-SESU-004-03.

References

[1] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the Face Recognition Grand Challenge," in *Proc. CVPR*, San Diego, CA, USA, Jun. 2005, pp. 947–954.

[2] P. J. Phillips, P. J. Flynn, K. W. Bowyer, R. W. Vorder Bruegge, P. J. Grother, G. W. Quinn, and M. Pruitt, "Distinguishing identical twins by face recognition," in *Proc. FG*, Santa Barbara, CA, USA, Mar. 2011.

[3] P. J. Grother, G. W. Quinn, and P. J. Phillips, "Report on the evaluation of 2D still-image face recognition algorithms," in *NIST Interagency/Internal Report (NISTIR) - 7709*, 2010.

[4] "Twins days." [Online]. Available: http://www.twinsdays.org/

[5] Z. Sun, A. A. Paulino, J. Feng, Z. Chai, T. Tan, and A. K. Jain, "A study of multibiometric traits of identical twins," in *Proc. SPIE, Biometric Technology for Human Identification VII*, vol. 7667, Orlando, FL, USA, Apr. 2010, pp. 1–12.

[6] K. Hollingsworth, K. Bowyer, and P. Flynn, "Similarity of iris texture between identical twins," in *Proc. CVPR Workshop on Biometrics*, San Francisco, CA, USA, Jun. 2010, pp. 22–29.

[7] A. Jain, S. Prabhakar, and S. Pankanti, "On the similarity of identical twin fingerprints," *Pattern Recognition*, vol. 35, no. 11, pp. 2653–2663, Nov. 2002.

[8] A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Expression-invariant 3D face recognition," in *Audio- and Video-based Biometric Person Authentication*, ser. LNCS, no. 2688, 2003, pp. 62–70.

[9] "CVRL data sets." [Online]. Available: http://www.nd.edu/~cvrl/CVRL/Data_Sets.html

[10] "Konica minolta catalogue." [Online]. Available: http://www.konicaminolta.com/instruments/products/3d/non-contact/vivid910/

[11] R. McKeon, "Three-dimensional face imaging and recognition: A sensor design and comparative study," Ph.D. dissertation, University of Notre Dame, 2010.

[12] T. C. Faltemier, K. W. Bowyer, and P. J. Flynn, "A region ensemble for 3D face recognition," *IEEE Trans. on Information Forensics and Security*, vol. 3, no. 1, pp. 62–73, 2008.

[13] L. Spreeuwers, "Fast and accurate 3d face recognition," *Int. J. Comput. Vision*, vol. 93, pp. 389–414, Jul. 2011.

[14] S. D. Stearns, "On selecting features for pattern classifiers," in *Proc. ICPR*, Coronado, CA, USA, 1976, pp. 71–75.

[15] D. Huang, G. Zhang, M. Ardabilian, Y. Wang, and L. Chen, "3D face recognition using distinctiveness enhanced facial representations and local feature hybrid matching," in *Proc. BTAS*, Washington D.C., USA, Sep. 2010, p. 1.

[16] D. Huang, M. Ardabilian, Y. Wang, and L. Chen, "A novel geometric facial representation based on multi-scale extended local binary patterns," in *Proc. FG*, Santa Barbara, CA, USA, Mar. 2011.

[17] D. Huang, W. Ben Soltana, M. Ardabilian, Y. Wang, and L. Chen, "Textured 3D face recognition using biological vision-based facial representation and optimized weighted sum fusion," in *Proc. CVPR Workshop on Biometrics*, Colorado Springs, CO, USA, Jun. 2011.

[18] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–110, Nov. 2004.

[19] B. Gökberk, M. O. Irfanoglu, and L. Akarun, "3D shape-based face representation and feature extraction for face recognition," *Image and Vision Computing*, vol. 24, no. 8, pp. 857–869, 2006.

[20] M. K. Unnikrishnan, "How is the individuality of a face recognized?" *Journal of Theoretical Biology*, vol. 261, no. 3, pp. 469–474, 2009.

[21] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell, "Face recognition by humans: Nineteen results all computer vision researchers should know about," *Proceedings of the IEEE*, vol. 94, no. 11, pp. 1948–1962, 2006.

[22] S. Zafeiriou, M. Hansen, G. Atkinson, V. Argyriou, M. Petrou, M. Smith, and L. Smith, "The photoface database," in *In review for CVPR*, 2011.

[23] I. A. Kakadiaris, G. Passalis, G. Toderici, M. N. Murtuza, Y. Lu, N. Karampatziakis, and T. Theoharis, "Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach," *IEEE Trans. PAMI*, vol. 29, no. 4, pp. 640–649, Apr. 2007.

[24] O. Ocegueda, S. K. Shah, and I. A. Kakadiaris, "Which parts of the face give out your identity?" in *Proc. CVPR*, Colorado Springs, CO, USA, Jun. 2011, pp. 641–648.

[25] H. Yu and J. Yang, "A direct LDA algorithm for high-dimensional data with application to face recognition," *Pattern Recognition*, vol. 34, pp. 2067–2070, Oct. 2001.

[26] A. Savran, N. Alyuz, H. Dibeklioglu, O. Celiktutan, B. Gokberk, B. Sankur, and L. Akarun, "Bosphorus database for 3D face analysis," in *Proc. First COST 2101 Workshop on Biometrics and Identity Management*, Roskilde, Denmark, May 2008, pp. 47–56.

A.4 Psychologically inspired dimensionality reduction for 2D and 3D face recognition

This was presented as part of the British Machine Vision Conference at the Student Workshop, Dundee, UK held in September 2011. The conference and workshop was a useful opportunity to refresh acquaintances made at previous conferences and to network with other researchers at a similar stage of their career.

Psychologically inspired dimensionality reduction for 2D and 3D Face Recognition

Mark F. Hansen
mark.hansen@uwe.ac.uk
Gary A. Atkinson
gary.atkinson@uwe.ac.uk
Melvyn L. Smith
melvyn.smith@uwe.ac.uk
Lyndon N. Smith
lyndon.smith@uwe.ac.uk

The Machine Vision Laboratory
University of the West Of England
Bristol, UK

Abstract

We present a number of related novel methods for reducing the dimensionality of data for the purposes of 2D and 3D face recognition. Results from psychology show that humans are capable of very good recognition of low resolution images and caricatures. These findings have inspired our experiments into methods of effective dimension reduction. For experimentation we use a subset of the benchmark FRGCv2.0 database as well as our own photometric stereo "Photoface" database. Our approaches look at the effects of image resizing, and inclusion of pixels based on percentiles and variance. Via the best combination of these techniques we represent a 3D image using only 61 variables and achieve 95.75% recognition performance (only a 2.25% decrease from using all pixels). These variables are extracted using computationally efficient techniques instead of more intensive methods employed by Eigenface and Fisherface techniques and can additionally reduce processing time tenfold.

1 Introduction

Automatic face recognition has been an active area of research for over four decades and a key part of this research is understanding how different data representations affect recognition rates and efficiency. Digital images of faces have a very high data dimensionality: a 200×200px image defines a point in a 40000-dimensional space, making computation a slow and resource hungry process. This is compounded when faces images are extended into 3D models. Reducing the dimensionality of the data without discarding the discriminatory information is the aim of this research. If a face can effectively be reduced down from many thousands of dimensions of raw data to a few tens of dimensions as in this paper, then storage needs become far less and processing becomes far faster. This has obvious applications for industrial and commercial implementations.

In this paper, we prove the following contributions for both the FRGCv2.0 database [11] and our own photometric stereo database [22] captured using the "Photoface" device [6]:

1. Optimal recognition results for close-cropped faces are obtained when the resolution is reduced to a mere 10x10 pixels.

2. The exclusive use of just 10% of the data (chosen to be those pixel locations with the greatest variance) is sufficient to maintain recognition rates to within 10% of those rates that include all of the data.

3. When combining the above two contributions we perform recognition at an accuracy of 96.25% for 40 subjects using only 61 dimensions (pixels). This compares to 98% when the full 80x80 resolution is used on all data.

Ultimately we aim to compare dimension reduction techniques based on a percentile and variance based inclusion principle (to exclude 90% of the data) with a baseline condition containing all pixels.

Our own database, Photoface, provides over 3000 sessions of 457 individuals, and scans are captured using photometric stereo [17] which results in estimated surface normals at each pixel. Full details of the actual device used can be found in [6] and an example of a scan can be seen in Fig. 1. The FRGCv2.0 database, which we also use in this paper, does not provide the surface normals. They can be calculated by numerically differentiating the point cloud data. We also include experiments on the depth map images to rule out any errors introduced by differentiation.

Figure 1: Examples of FRGCv2.0 (left) and Photoface (right) 3D scans. NB They are not of the same person.

Using 3D data for face recognition allows for pose and illumination correction which are two commonly cited problems with conventional 2D images. Better recognition rates have also been reported using 3D over 2D data [4], although this is not always replicated [7]. One reason for this may be the representation of the 3D data used in the analysis. Gökberk *et al*. [5] performed recognition experiments using numerous 3D representations. They concluded that '*... surface normals are better descriptors than the 3D coordinates of the facial points.*' This is at odds with most research which uses the 3D point coordinates as a starting point. Surface normals are used in the experiments performed in this paper for this reason.

There are many mathematical techniques for dimensionality reduction, and in particular the Eigenface [15] and Fisherface [2] (based on Principle Components Analysis (PCA) and Fisher's Linear Discriminant (FLD) respectively) techniques are commonly used in face recognition. With an added dimension, 3D face models potentially compound the problem for large data storage. Recent techniques such as sparse representation (such as non-negative matrix factorization) and manifold learning (such as local linear embedding [8]) show that effective methods of dimension reduction are a key topic. Methods that can reduce the amount of data without discarding discriminatory information are essential for faster processing and optimal solutions. There have been many attempts in the literature to extend and generalise PCA, FLD and other methods [19, 20, 21] in order to improve robustness to pose, illumination, etc, typically at the expense of computational efficiency. The main contribution of this paper by contrast, is to show that for the constrained case of frontal 2.5D data, then the

efficiency can be improved even compared to PCA by using more direct analysis without the need to project into a new subspace.

Caricaturing essentially enhances those facial features that are unusual or deviate sufficiently from the norm. It has been shown that humans are better able to recognise a caricature than they are the veridical image [10, 12]. This finding is interesting as caricaturing is simply distorting or adding noise to an image. However this noise aids human recognition and this, in turn, provides insights into the storage or retrieval mechanism used by the human brain.

Unnikrishnan [16] conceptualises an approach similar to face caricatures for human recognition. In this approach, only those features which deviate from the norm by more than a threshold are used to uniquely describe a face. Unnikrishnan suggests using those features whose deviations lie below the 5^{th} percentile and above the 95^{th} percentile, thereby discarding 90% of the data. Unnikrishnan provides no empirical evidence in support of his hypothesis, so an aim of this paper is to test the theory experimentally. We do this in two ways: the first directly tests his theory, finding the thresholds for each pixel which represent the 5^{th} and 95^{th} percentile values and only including those pixels in each scan which lie outside them (outliers). The second is loosely based on Unnikrishnan's idea, and looks at the variance across the whole database to calculate the pixel locations with the largest variance. Only the pixels at these locations are then used for recognition.

An obvious method of reducing the amount of data is to downscale the images. A great deal of research has gone into increasing the resolution of poor quality images (super-resolution [1, 18], hallucinating [23]) by combining images or using statistical techniques to reproduce a more accurate representation of a face (*e.g.* from CCTV footage). By contrast, little research attempts to directly investigate resolution as a function of recognition rates on 3D data. Toderici *et al.* state that there is little to be gained from using high resolution images [14], Boom *et al.* state that the optimum face size is 32×32 px for registration and recognition [3], a view which is reinforced by a more recent study by Lui *et al.* who state that optimum face size lies between 32 and 64 pixels [9]. These experiments have used 2D images. Chang *et al.* use both 2D and 3D data and conclude that there is little effect of decreasing resolution up to 25% on 2D data and 50% on 3D [4] using PCA. In summary, the research suggests that relatively low resolutions give optimum recognition (for the given recognition algorithms). These findings are conducive to the fact that the same appears to be true of human recognition [13].

2 Methods and data

This section details the datasets, preprocessing steps, and the methods used in the experiments.

2.1 Data and preprocessing

Experiments were performed on 10 sessions of 40 subjects facing frontally without expression on the FRGCv2.0 and our own photometric stereo database. 2D and 3D data are used in separate experiments.

The FRGCv2.0 dataset comes in point cloud format which is converted to a mesh via uniform sampling across facets. Noise is removed by median smoothing and holes filled by interpolation. Normals are then estimated by differentiating the surface. The depth map images are all normalized to have a minimum value of 0.

Figure 2: The cropped region of a face. The distance between the anterior canthi (d) is used to calculate the cropped region.

Data is cropped for both databases as follows: the median anterior canthi and nose tip across all sessions are used for alignment via linear transforms. The aligned images are then cropped into a square region as shown in Fig. 2 to preserve main features of the face (eyes, nose, mouth), and exclude the forehead and chin regions which can frequently be occluded by hair.

Our 2D experiments are based on data as follows: the accompanying colour image for each FRGCv2.0 scan is converted to greyscale, aligned and cropped in the same way as the 3D scan. The 2D images in the Photoface database are the estimated albedo images which are also aligned and cropped in the same way as the 3D data. Due to memory limitations, both the 2D and 3D data are then resized to 80×80 px and are reshaped into a 6400-dimension and a 12800-dimension (x and y components of the surface normals are concatenated) vector respectively.

2.2 Calculating outliers and variance

The thresholds for each pixel are calculated which represent the 5^{th} and 95^{th} percentile values. We are interested in the norm across the whole dataset for each pixel location rather than the norm for each image. For the 2D images, percentile values are calculated for the greyscale intensity value for each pixel location. There are 400 sessions, so there are 400 values for each pixel from which we calculate the percentile thresholds. The same process is performed for 3D surface normal data, giving x and y surface normal component thresholds for each pixel. Pixels which have a value between the 5^{th} and 95^{th} percentile are discarded, leaving only the 10% outlying data. We shall refer to this as the "percentile inclusion criterion". Examples can be seen in Fig. 3.

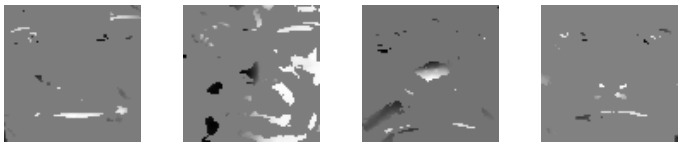


Figure 3: Examples of the y-components of the surface normals that have values outside the 5^{th} and 95^{th} percentiles for four subjects which are used for recognition.

The above method extracts the least common data from each session and that is what is used for recognition. Alternately, we can use the greyscale variance at each pixel location as a measure of discriminatory power. If a pixel shows a large variance across the dataset, then this might make it useful for recognition (assuming that variance within the class or subject is small). Therefore the standard deviation of each pixel is calculated over all the sessions. Whether or not a particular pixel location is used in recognition depends on whether or not the variance is above a pre-determined threshold. Examples of the use of different thresholds are shown in Fig. 4. We refer to this as the "variance inclusion criterion".

2.3 Image resizing

The effect of different resizing techniques on linear subsampling are investigated in terms of their effect on recognition as a function of resolution. Resizing is performed via the Matlab

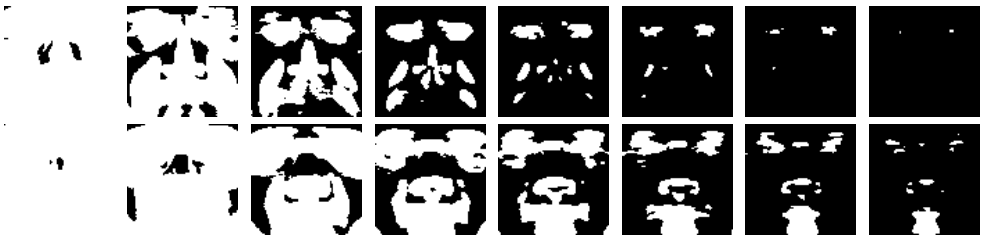


Figure 4: Examples of the regions which remain for x (top row) and y-components (bottom row) as the threshold variance is increased from left to right. White regions are retained and black regions are discarded.

`imresize()` function using the deafult bicubic kernal type and with antialiasing on, as these settings were found to provide the best performance.

2.4 Recognition algorithm

Our experiments used to test recognition accuracy employ the leave-one-out paradigm. This dictates that every session is used as a probe against a gallery of all other sessions once. There are therefore 400 classifications per condition of which the percentage correctly identified is shown.

As the purpose of this research is feature extraction efficiency, the actual choice of classifier is not so important. We therefore implement Pearson product-moment correlation coefficient (PMCC) as a similarity measurement between a probe vector and the gallery vectors. The gallery session with the highest coefficient is regarded as a match. Experimentally, we found that PMCC gives similar performance on baseline conditions to the Fisherface algorithm but is approximately eight times faster.

3 Results

3.1 Dimensionality reduction via the percentile inclusion criterion

Unnikrishnan's theory states that we should expect reliable performance using only the data which lies outside the 5^{th} and 95^{th} percentiles [16]. Table. 1 shows recognition rates on 2D and 3D data using both all data and the outliers only. Note in particular that, for the 3D surface normal data, the rates drop by under 10% when using outlier data only. This effect seems limited to the surface normal data and is not seen in either the 2D or depth map data. We have included results from a fusion technique using the Photoface surface normal data combined with the albedo image. There is a small decrease in baseline performance and using only the outlying data leads to a severe decrease of about 34%.

		Baseline (All pixels)	Outliers (10% of pixels)
2D	FRGC	90	73.75
	Photoface	98	64
3D	FRGC Surface normals	90.25	84.25
	FRGC Depth map	71.5	23.25
	Photoface	98.25	89.25
Fusion	Photoface 2D + 3D	97	63.25

Table 1: Baseline versus outlier performance (% correct).

Fig. 5 shows a plot of recognition rate as a function of which percentile range is used for recognition on 3D Photoface data. It should be noted that similar patterns of results were found for all datasets (2D, 3D and FRGC). As predicted, the figure shows that the best recognition performance is obtained using the most outlying percentiles. As expected also, the recognition rate reduces as the percentile ranges used tend toward the inliers. However, for the most inlying data of all (i.e. percentiles 45–55) we find a significant increase in performance. Contrary to Unnikrishnan's theory, this implies that there is discriminative data that is useful for face recognition in the most common data as well as the most outlying.

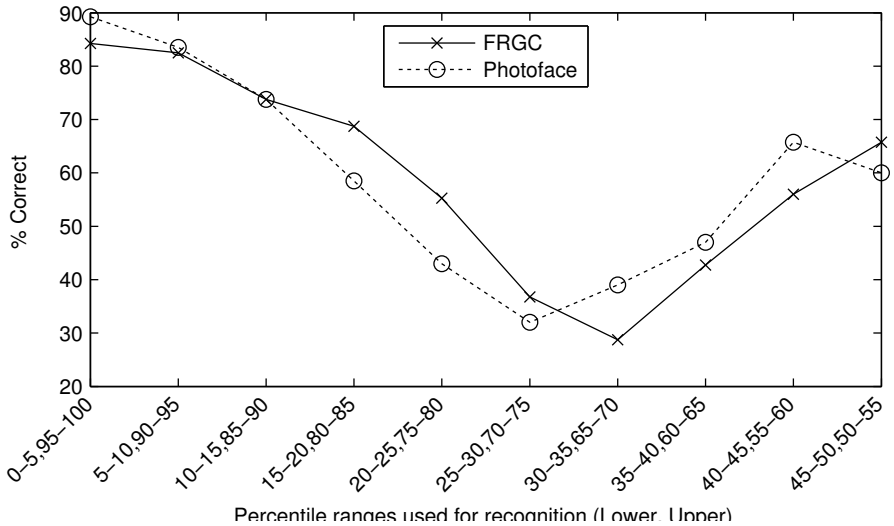


Figure 5: Recognition performance using pairs of percentile ranges for 3D data.

In a related experiment, we used single 5% ranges of data for recognition (i.e. $[0^{th} - 5^{th}], [5^{th} - 10^{th}]$ etc.) as shown in Fig. 6. Note that the increase in recognition performance for the most inlying data is not replicated. The slightly lower performance compared with Fig. 5 is because only 5% of the data is used instead of 10%.

Performance increases by combining ranges are not always observed. Consider, for example, the $25 - 30^{th}$ and $70 - 75^{th}$ percentiles for the FRGCv2.0 data. Individually the two percentiles give a performance around the 50% mark in Fig. 6, but when combined, the performance drops to around 40% in Fig. 5.

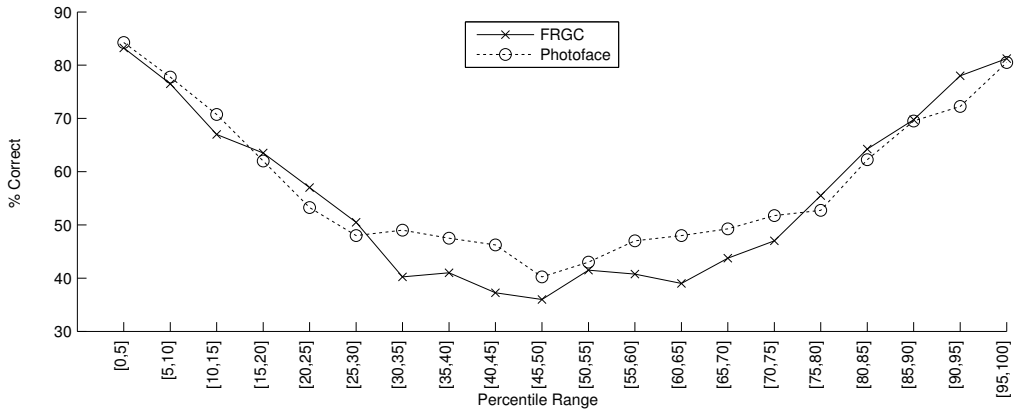


Figure 6: FRGC and Photoface data show a marked symmetry across ranges of percentiles.

3.2 Dimensionality reduction via the variance inclusion criterion

One problem with the above method is that the outlying points tend to be scattered across different parts of the images, making inter- and intra-comparisons between individuals somewhat unstructured. For the next method therefore, we use the same pixel locations in our recognition test for all images. Instead of using the percentiles defined within a single image as an inclusion criterion, we use the variance of a particular pixel across all subjects as explained in Sec. 2.2.

Fig. 7 shows plots combining the number of pixels which remain as we remove those with least variance (bar plot) against the recognition performance (line plot). It is apparent that we can achieve close to optimal performance while losing a large proportion of the pixels. We can discard approximately 75% of the least varying pixels and observe a corresponding drop of less than 10% in recognition performance on the FRGC data. Indeed, for Photoface data specifically, we only lose a few percent.

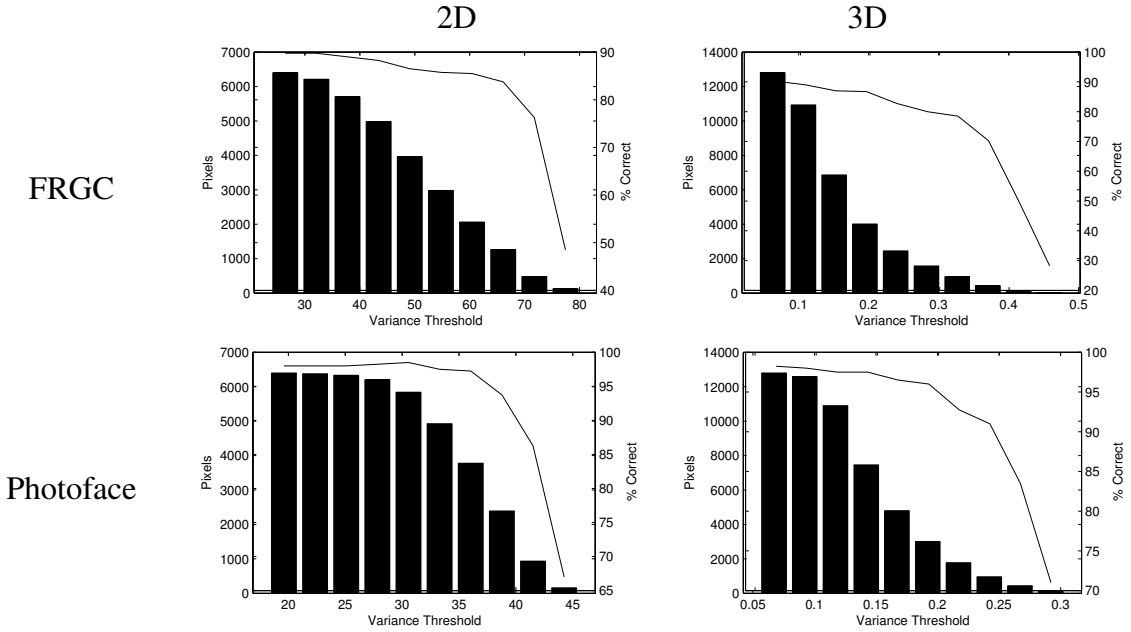


Figure 7: Recognition (line) as a function of retained pixels (bar chart). The pattern is shown in both sets of data (FRGC on the top row and Photoface on the bottom). 2D (grayscale for FRGC and albedo for Photoface) on the left, and surface normal data is shown on the right.

Table 2 shows a performance comparison of the two types of inclusion criteria when only 10% of pixels are retained. It is clear that by discarding the data that varies the least, we can maintain reasonably high recognition rates.

	Percentiles	Variance
FRGC	84.25%	$\approx 79\%$
Photoface	89%	$\approx 92\%$
Processing time	800.64s	180.95s

Table 2: A comparison of recognition performance using percentiles and variance methods to select the most discriminatory 10% of the data. The processing time includes the calculation of the outliers/most varying pixels and 400 classifications

The processing time improvement for the variance approach is due to having decreased the vector size by 90 %. This compares to 973.09s for the equivalent Fisherface analysis

which provides an accuracy of 99.5% so both methods offer considerable time savings at a small cost to accuracy.

3.3 Optimisation of Image resolution

Finally the effect of image resolution on 3D recognition performance is shown in Fig. 8. This clearly shows that a resolution of 10×10 px provides optimal or close to optimal recognition performance (the result for 40×40 px is 0.25% higher for FRGC) on both 3D datasets. The same pattern appears in the 2D Photoface database, but there is a small decrease of just under 3% for the 2D FRGC data. Nonetheless, if we take the 10×10 px as an optiumum size, this figure is lower than often reported in the literature. This may be because the data used in these experiments is already highly cropped, and other research may be using other metrics such as the distance across the uncropped head. Although not shown in the figure, not antialiasing the resampled images led to poorer performance in all cases.

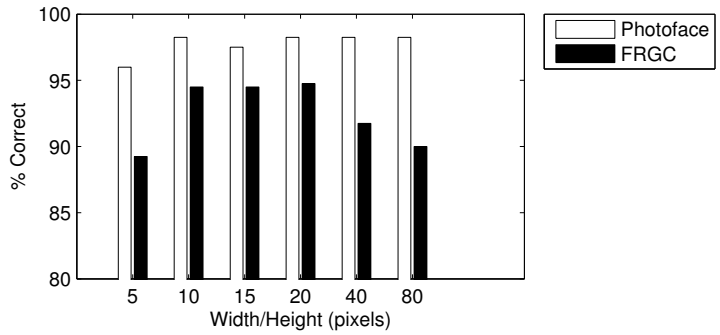


Figure 8: The effect of resolution on 3D recognition performance. Recognition rates for 10×10 px are 94.75% for FRGC data and 98.25% for Photoface data.

Combining the optimal resolution of 10×10 px with the variance method above we can achieve virtually the same recognition performance as an 80×80 px image but using only 64 pixels for FRGC data and 61 pixels for Photoface data. Recognition rates of 87.75% and 96.25% are recorded (a loss of only 7% and 2% respectively). The processing time is also reduced to 10.5s for variance analysis and 400 classifications. The same analysis using the Fisherface algorithm takes 118s and achieves a comparable rate of 89.25%.

4 Discussion

This paper describes methods to effectively reduce data dimensions while maintaining recognition performance. Computationally efficient methods using variance analysis and image resizing have been shown to be powerful means of reducing data but maintaining discriminatory information. Table 3 compares commonly used dimension reduction techniques of PCA and Fisherface with our variance and percentile inclusion criterion techniques at different resolutions in terms of classification accuracy and processing time. All experiments were carried out in Matlab on a Quad Core 2.5GHz Intel PC with 2GB ram running Windows XP. Only one percentile inclusion criterion result has been included as performance (especially processing time) was not at the same level as other conditions.

The number of components which are used for PCA depends on the specific test as follows: 61 components (61PCA, row 6 of Table 3) were chosen for a direct comparison

	Res. (px)	Data Reduction	Classifier	No. Dimensions	% Correct	Proc. time(s)
1.	10x10	None	PMCC	200	98.25	12.02
2.	10x10	VI	PMCC	19 (10%)	82.75	12.52
3.	10x10	VI	PMCC	61	95.75	13.02
4.	10x10	PCA	Euc. dist.	21	94.5	92.47
5.	10x10	PCA	PMCC	21	92.25	97.16
6.	10x10	61PCA	Euc. dist.	61	96.25%	102.91
7.	10x10	VI $\rightarrow$ 15PCA	PMCC	61 $\rightarrow$ 15	89.75	128.54
8.	10x10	VI $\rightarrow$ FF	Euc. dist.	19 $\rightarrow$ 19	90.5	129.74
9.	80x80	None	PMCC	12800	98.25	129.86
10.	10x10	VI $\rightarrow$ 15PCA	PMCC	19 (10%) $\rightarrow$ 15	79	132.56
11.	10x10	VI $\rightarrow$ FF	Euc. dist.	61 $\rightarrow$ 39	99	134.69
12.	10x10	FF	Euc. dist.	39	100	144.25
13.	80x80	VI	PMCC	1235 (10%)	92.25	180.95
14.	80x80	VI $\rightarrow$ 15PCA	PMCC	1235 (10%) $\rightarrow$ 15	85.25	331.40
15.	80x80	VI $\rightarrow$ FF	Euc. dist.	1235 (10%) $\rightarrow$ 39	90.75	549.25
16.	80x80	PCA	Euc. dist.	61	96.75	573.52
17.	80x80	PI	PMCC	12800	89	800.64
18.	80x80	FF	Euc. dist.	39	99.5	973.09

Table 3: A comparison of our variance (VI) and percentile (PI) inclusion techniques with PCA and Fisherface (FF) algorithms sorted by processing time.

with the 61 variables of the variance inclusion criterion which gave good performance in Fig. 7. 15 components (15PCA condition, rows 7, 10 & 14) were chosen arbitrarily as an extra step after the variance inclusion criterion for its low dimensionality and relatively good performance. For other tests using PCA, the number of components are chosen which describe 85% of the variance. Some entries in the "No. Dimensions" column have (10%) shown next to them. This is a reminder that only 10% of the data remains after applying the variance inclusion criterion. Finally some of the rows contain a "$\rightarrow$" symbol representing a combination of processes eg Variance Inclusion followed by Fisherface.

Generally resizing the image to 10x10 pixels gives a clear processing time advantage with little or no compromise on accuracy. Without additional dimensionality reduction we achieve a recognition rate of 98.25% (row 1). We are able to reduce the dimensionality by a further $\frac{2}{3}$ and only lose 2.5% performance by additionally using the variance inclusion criterion to select 61 pixel locations (row 3). This appears to give the best compromise in terms of the number of dimensions, processing time and accuracy . The Fisherface algorithm gives excellent performance (10x10 Fisherface gives 100% accuracy, row 12) but at the cost of processing time.

These results only apply to the simplest case in face recognition – the frontal, expressionless face. The variance inclusion algorithm would be unlikely to produce similarly good results if expressions were present in the dataset, as these are likely to produce areas of high variance which will not be discriminatory. Nonetheless these could be used for the purposes of expression analysis instead of recognition or alternatively areas which change greatly with expression could be omitted from the variance inclusion criterion.

It is clear that effective dimensionality reduction can be achieved via more direct, psychologically inspired models in contrast to conventional mathematical tools such as PCA. Processing speed is also drastically increased – if we perform recognition by the Fisherface algorithm on 80×80 pixel images, it takes 973.09s. Using 10×10 pixel images, processing

time drops to only 13.02s using our proposed variance inclusion method to extract 61 pixel locations with only a 3.75% drop in performance.

5 Conclusion

We have presented a number of important findings that affect face recognition performance regarding the effects of optimum image size and the use of different variance measures to select discriminatory data. The findings have implications on real-world applications in that they point to computationally attractive means of reducing the dimensionality of the data. Empirical support of Unnikrishnan's hypothesis [16] regarding the use of outlying percentile ranges is provided on both the FRGCv2.0 database as well as our own photometric stereo face database. One of the most promising results comes from resizing the original 3D data from 80x80 pixels to 10x10 pixels and applying the variance based inclusion approach yielding an accuracy of 95.75% using just 61 dimensions and the fact that this heuristic was inspired by the human process of caricaturing. Using this combination of techniques, processing speeds can be also be increased tenfold over the conventional Fisherface algorithm.

References

[1] Simon Baker and Takeo Kanade. Limits on Super-Resolution and how to break them. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(9):1167–1183, 2002. ISSN 0162-8828. doi: http://doi.ieeecomputersociety.org/10.1109/TPAMI.2002.1033210.

[2] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997. ISSN 0162-8828.

[3] B. J. Boom, G. M. Beumer, L. J. Spreeuwers, and R. N. J. Veldhuis. The effect of image resolution on the performance of a face recognition system. In *9th International Conference on Control, Automation, Robotics and Vision, 2006. ICARCV'06*, page 1–6, 2006.

[4] K. Chang, K. Bowyer, and P. Flynn. Face recognition using 2D and 3D facial data. In *ACM Workshop on Multimodal User Authentication*, pages 25–32, 2003.

[5] B. Gökberk, M. O. İrfanoğlu, and L. Akarun. 3D shape-based face representation and feature extraction for face recognition. *Image and Vision Computing*, 24(8):857–869, 2006.

[6] Mark F. Hansen, Gary A. Atkinson, Lyndon N. Smith, and Melvyn L. Smith. 3D face reconstructions from photometric stereo using near infrared and visible light. *Computer Vision and Image Understanding*, 114(8):942–951, August 2010. ISSN 1077-3142.

[7] M. Hüsken, M. Brauckmann, S. Gehlen, and C. Von der Malsburg. Strategies and benefits of fusion of 2D and 3D face recognition. In *IEEE workshop on face recognition grand challenge experiments*, page 174, 2005.

[8] B. Li, C. H Zheng, and D. S Huang. Locally linear discriminant embedding: An efficient method for face recognition. *Pattern Recognition*, 41(12):3813–3821, 2008. ISSN 0031-3203.

[9] Y. M Lui, D. Bolme, B. A Draper, J. R Beveridge, G. Givens, and P. J Phillips. A meta-analysis of face recognition covariates. In *Proceedings of the 3rd IEEE international conference on Biometrics: Theory, applications and systems*, page 139–146, 2009.

[10] R. Mauro and M. Kubovy. Caricature and face recognition. *Memory & Cognition*, 20 (4):433–440, 1992.

[11] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *Proc. CVPR*, volume 1, 2005.

[12] G. Rhodes, S. Brennan, and S. Carey. Identification and ratings of caricatures: Implications for mental representations of faces. *Cognitive Psychology*, 19(4):473–497, 1987.

[13] Pawan Sinha, Benjamin Balas, Yuri Ostrovsky, and Richard Russell. Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proceedings of the IEEE*, 94(11):1948–1962, November 2006. ISSN 0018-9219.

[14] George Toderici, Sean O'Malley, George Passalis, Theoharis Theoharis, and Ioannis Kakadiaris. Ethnicity- and gender-based subject retrieval using 3-D Face-Recognition techniques. *International Journal of Computer Vision*, 89(2):382–391, 2010. doi: 10.1007/s11263-009-0300-7.

[15] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[16] M.K. Unnikrishnan. How is the individuality of a face recognized? *Journal of Theoretical Biology*, 261(3):469–474, December 2009. ISSN 0022-5193. doi: 16/j.jtbi.2009.08.011.

[17] R. J. Woodham. Photometric method for determining surface orientation from multiple images. *Opt. Eng.*, 19(1):139–144, 1980.

[18] J. Yang, J. Wright, T. S Huang, and Y. Ma. Image super-resolution via sparse representation. *Image Processing, IEEE Transactions on*, 19(11):2861–2873, 2010. ISSN 1057-7149.

[19] Jian Yang, David Zhang, Alejandro F. Frangi, and Jing-yu Yang. Two-Dimensional PCA: a new approach to Appearance-Based face representation and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(1):131–137, 2004. ISSN 0162-8828. doi: http://doi.ieeecomputersociety.org/10.1109/TPAMI.2004.10004.

[20] M. H Yang. Kernel eigenfaces vs. kernel fisherfaces: Face recognition using kernel methods. In *fgr*, page 0215, 2002.

[21] J. Ye, R. Janardan, Q. Li, et al. Two-dimensional linear discriminant analysis. *Advances in Neural Information Processing Systems*, 17:1569–1576, 2004.

[22] Stefanos Zafeiriou, Mark F. Hansen, Gary A. Atkinson, V. Argyriou, Maria Petrou, M.L. Smith, and L.N. Smith. The PhotoFace database. In *Proc. Biometrics Workshop of Computer Vision and Pattern Recognition*, pages 161–168, Colorado Springs, USA, 2011. IEEE.

[23] Y. Zhuang, J. Zhang, and F. Wu. Hallucinating faces: LPH super-resolution and neighbor reconstruction for residue compensation. *Pattern Recognition*, 40(11):3178–3194, 2007. ISSN 0031-3203.

A.5 The Photoface Database

This paper was presented at the Computer Vision and Pattern Recognition Biometrics Workshop, Colorado, USA in June 2011. It presents a preliminary version of the database together with some recognition analysis. This conference was selected to announce the public availability of the database as it is a well respected conference with a large international attendence.

The Photoface Database

Stefanos Zafeiriou[†], Mark Hansen[*], Gary Atkinson[*], Vasileios Argyriou[⋆],
Maria Petrou[‡,•], Melvyn Smith[*] and Lyndon Smith[*]
† Department of Computing, Imperial College London,
180 Queen's Gate, London SW7 2AZ UK.
* Machine Vision Lab, Faculty of Environment and Technology, University of the West of England,
Frenchay Campus, Bristol BS16 1QY UK.
⋆ Faculty of Computing, Information Systems and Mathematics, Kingston University London,
River House, 53-57 High Street, Kingston upon Thames, Surrey KT1 1LQ UK.
‡ Department of Electrical and Electronic Engineering, Imperial College London,
Exhibition Road, South Kensington Campus, London SW7 2AZ UK.
• Informatics and Telematics Institute, Centre of Research & Technology - Hellas
6th km Xarilaou - Thermi, Thessaloniki 57001 Greece
{s.zafeiriou,maria.petrou}@imperial.ac.uk, vasileios.argyriou@kinston.ac.uk
{mark.hansen,gary.atkinson,melvyn.smith,lyndon.smith}@uwe.ac.uk. *

Abstract

In this paper we present a new database suitable for both 2D and 3D face recognition based on photometric stereo, the so-called Photoface database. The Photoface database was collected using a custom-made four-source photometric stereo device that could be easily deployed in commercial settings. Unlike other publicly available databases the level of cooperation between subjects and the capture mechanism was minimal. The proposed device may also be used, to capture 3D expressive faces. Apart from the description of the device and the Photoface database, we present experiments from baseline face recognition and verification algorithms using albedo, normals and the recovered depth maps. Finally, we have conducted experiments in order to demonstrate how different methods in the pipeline of photometric stereo (i.e. normal field computation and depth map reconstruction methods) affect recognition/verification performance.

1. Introduction

Face recognition researchers have been collecting databases of face images for several decades now [15, Chapter 13]. While some databases can be regarded as superior to others, each of them is designed to test different aspects of recognition and have their own strengths and weaknesses. One of the largest databases available is the FERET database [17]. This has a total of 1199 subjects with up to 20 poses, two expressions and two light source directions. The FERET database was originally acquired using a 35mm camera. Others on the other hand, for example the widely used CMU PIE database [19] or the Harvard RL database [11], concentrate more on varying the capture conditions such as pose and illumination.

The PIE database is one of the most extensively researched. This is due to the fact that the faces are captured under highly controlled conditions involving 13 cameras and 21 light sources. The Yale B database [8] offers similar advantages to the PIE database except with an even larger number of lighting conditions (64), but just using ten subjects. Nine poses were considered per subject. The original Yale database [4] was designed to consider facial expressions, with six types being imaged for 15 subjects. Finally, the extended Yale B database was published. It contains 28 subjects with 9 different poses and 64 illumination conditions [14].

Even though the PIE [19], Yale [8] and extended Yale [14] databases provide facial samples taken under different illumination directions, they contain very few persons. More recently, the CMU Multi-PIE database [10] has been constructed with the aim of extending the image sets to in-

*This work was supported by the EPSRC project EP/E028659/1 Face Recognition using Photometric Stereo. This work has been conducted while Stefanos Zafeiriou was with Department of Electrical and Electronic Engineering, Imperial College London

clude a larger number of subjects (337) and to capture faces taken in four different recording sessions. This database was recorded under controlled laboratory conditions, as with the others mentioned above.

Recent trends in face recognition research to incorporate three-dimensional information into the recognition process lead to the collection of databases with 3D facial samples. This trend is due to the fact that a large number of viewing conditions adversely affect the 2D appearance of a face image but not the 3D appearance. This was the motivation for the FRGC2.0 database [16], which consists of a multi-partition 2D and 3D database including a validation set of 4007 scans of 466 subjects. A Minolta Vivid 900/910 series laser range finder [24] was used for data capture.

This paper describes the construction of a new type of database of faces to aid research into face recognition. As explained, there is a growing number of related databases available for public research use. Each of these is designed to test different aspects of recognition, such as expression or illumination invariance. These databases have been built under meticulously designed and calibrated physical arrangements. The purpose of the new database described in this paper, however, is to capture a large number of faces in a more industrial setting. Most existing 3D capture devices (e.g. [23, 24]) are both financially and computationally expensive which can be highly inhibiting for commercial application. By contrast, we use a four-source high-speed capture arrangement, which permits the use of photometric stereo methods [22] to recover the 3D information with minimal computational expense. Furthermore, the device is significantly financially cheaper than most other 3D capture mechanisms.

Two important features of the database are that (1) the level of cooperation between subjects and the capture mechanism is minimal, and that (2) the capture process was carried out in a realistic commercial setting. For this reason, we placed the device near the entrance of a busy workplace and gave all of the volunteer subjects the sole instruction to "walk through the archway". This arrangement accurately simulates one of the ultimate goals for access-control face recognition, where there is no interaction required between the subjects and the technology. The database therefore offers an ideal testbed for face recognition algorithms designed for real world applications. As photometric stereo can be applied to the four images to calculate the 3D structure of the face, the database also allows for both 2D, 3D and hybrid algorithms to be evaluated.

In addition to describing the device and the database, we also present baseline experiments on the Photoface database applying baseline face recognition/verification techniques on albedo, depth and normal images. The focus of the conducted experiments is neither to compare various 2D/3D face recognition and verification methods nor to demonstrate that fusion of information of 3D and 2D data increase the recognition performance [5],[9]. The aim of the experiments conducted in this paper is: 1) to demonstrate how different methods in the pipeline of photometric stereo (i.e. normal field computation and depth map reconstruction methods) affect recognition/verification performance, and 2) to verify that a similar conclusion to [5] can be drawn for the modalities derived from photometric stereo methods. We applied three different photometric stereo methods in order to compute the normal field and the albedo image and five different integration methods that compute the height map from the normal field. To the best of the authors' knowledge this is the first experiment on a real-world photometric stereo database which also explores the effect of the use of different methods in the processing pipeline. In summary, the contributions of this paper are:

- The presentation of the first realistic commercial acquisition arrangement for the collection of 2D/3D facial samples using Photometric Stereo (PS).
- The presentation of the first facial image database collected under such a setting.
- The demonstration of how different methods in the pipeline of PS affect recognition/verification performance via a detailed set of recognition/verification experiments using a range of algorithms.

2. Capturing Device and Database Collection

The Photoface database was collected using a custom-made four-source PS device. Unlike previous constructions, our aim was to capture the data using a hardware that could be easily deployed in commercial settings. The setup is as follows: individuals walk through the archway towards the camera located on the back panel and exit through the side (Fig. 1). This arrangement makes the device suitable for usage at building entrances, high security areas, airports etc. The presence of an individual is detected by an ultrasound proximity sensor placed before the archway. This can be seen in Fig. 1 on the horizontal beam towards the left-hand side of it.

The hardware equipment used to create the entire system was the following:

- Camera: Basler 504kc with Camera Link interface operating at 200fps, 1ms exposure time, placed approximately at a distance of 2m from the head of the subject.
- Lens: 55mm, $f5.6$ Sigma lens.
- Light sources: low cost Jessops M100 flashguns, approximately at a distance of 75cm from the head of the subject.

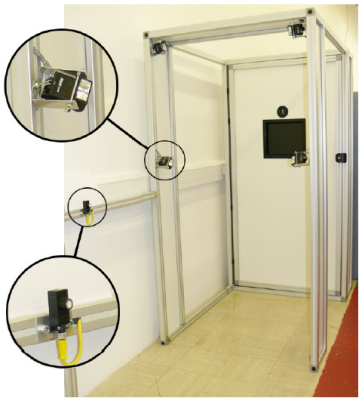


Figure 1. The image capturing device. One of the light sources and an ultrasound trigger are shown on the left. The camera is located at the back panel.

- Device trigger: Baumer highly directional ultrasound proximity switch. Range 70cm.
- Hardware IO card (for receiving and distributing triggers): NI PCI-7811 DIO.
- Frame grabber: NI PCIe-1429.
- Interfacing code: NI LabVIEW (the reconstruction and recognition algorithms were written in MATLAB).

The device also contains a monitor (as can be seen in Fig. 1) that provides instructions and indicates whether or not an individual was recognised, in the case of a recognition scenario, or whether an identity claim was accepted or rejected, in the case of a verification scenario.

The device captures one image of the face for each light source in a total of approximately 20ms. This time, also chosen for our experiments, was regarded as an adequately short period of time in which the inter-frame motion is no greater than a few pixels. The only case in which the performance of the system is expected to deteriorate significantly, is when a person runs passing through the device, due to the large inter-frame motion observed. For each person passing through the device, the following sequence of events takes place to capture the four images:

1. Await signal from ultrasound sensor.
2. Send trigger to camera.
3. Await integration enabled signal from camera.
4. Discharge first flashgun.
5. Await end of integration enabled signal.
6. Repeat from step 2 for the remaining light sources.

Figure 2. Four raw input images.

Figure 2 shows an example of four raw images of an individual (the resolution of the captured images were 1280×1024).

The capturing device was placed at the entrance of a busy workplace for a period of four months. Volunteer employees casually passed through the booth at regular intervals throughout this period. No instructions were given, other than to instructing them to walk through the archway looking at the camera or monitor. Thus, the volunteers typically passed through the device on their way in and out of the building. This arrangement is of great importance as:

1. It means that the capturing conditions were realistic for a real-world example. This is in contrast with existing face databases such as the widely used CMU-PIE database [10] or the FRGC database [16].
2. The whole setup was non-invasive, thus being suitable for any recognition algorithms developed for immediate commercial use.

3. Statistics of the Database

The Photoface database was collected in a period of four months (February 2008 to June 2008). It consists of a total of 1,839 sessions of 261 subjects and a total of 7,356 images. Some individuals used the device only once, while some others walked through it more than 20 times. The majority of people in the database are men (227 men over 34 women). The vast majority of the individuals are Caucasians (257 persons). Since there was no supervision, most of the captured faces in the database display an expression (for example more than 600 smiles and more than 200 surprises, open mouth, scream like expression etc. were recorded).

98 people walked through the device only once. For 126 of the 163 subjects that used the device more than once, the sessions were collected over a period of more than a week's

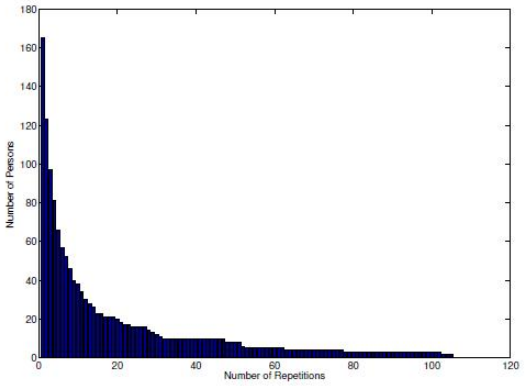


Figure 3. Four raw input images.

interval. For the majority of those (90 people), this interval was greater than one month. A histogram corresponding to the number of subject recordings by the device is depicted in Figure 3.

4. Photometric Stereo (PS) and Surface Reconstruction Methods

In this Section we summarise the standard PS method [6, §5.4], which we implemented using both three and four light sources. For these experiments we used an implementation of the standard PS method of Woodham [22]. We have mainly concentrated on a four-source version of the technique, although we have also compared our results with methods using three light sources. For the latter, we omitted the upper-right source in Fig. 1 from the computation. In order to do so, we examined a few reconstructions using various combinations of sources, resulting in the conclusion that deleting this specific light source was the most safe choice, as the performance was neither enhanced nor decreased, in comparison with the removal of one of the remaining three sources.

The standard PS method we used, assumes three or more greyscale images of a Lambertian object and constructs the following matrix equation from Lambert's Law for each pixel $\mathbf{x} = [x, y]$:

$$[\mathbf{I}_1(\mathbf{x})\ \mathbf{I}_2(\mathbf{x}) \cdots \mathbf{I}_N(\mathbf{x})]^T = \rho(\mathbf{x})\ [\mathbf{l}_1^T\ \mathbf{l}_2^T \cdots \mathbf{l}_N^T]^T\ \mathbf{n}(\mathbf{x}) \tag{1}$$

where $\mathbf{I}_m(\mathbf{x})$ is the mth measured pixel brightness, $\mathbf{l}_m$ is the mth light source vector, N is the number of light sources, $\rho(\mathbf{x})$ is the reflectance albedo and $\mathbf{n}(\mathbf{x})$ is the surface unit normal. Examples of the raw images under 4-lights can be seen in Figure 2. The intensity values and light source positions are known a-priori and from these the albedo and surface normal components can be calculated by solving (1). (example of the computed albedo and surface using the 4-lights PS are shown in Figures 4 In our experiments, apart from the above mentioned PS method, we also applied the PS method proposed in [3].

In the following, we briefly review the problem of reconstructing a surface from the surface normals. In order to compute the shape of the surface, we need to obtain the depth map. This suggests representing the surface as $(\mathbf{x}, f(\mathbf{x}))$, so that the normal is a function of $\mathbf{x}$:

$$\tilde{\mathbf{n}}(\mathbf{x}) = \frac{1}{\sqrt{1 + \frac{\partial f}{\partial x}^2 + \frac{\partial f}{\partial y}^2}} \left(-\frac{\partial f}{\partial x}, -\frac{\partial f}{\partial y}, 1 \right)^T \tag{2}$$

To recover the depth map, we need to determine $f(\mathbf{x})$ from the computed values of the unit normal.

Let us assume that the computed value of the unit normal at some point $\mathbf{x}$ is $\mathbf{n}(\mathbf{x}) = [a(\mathbf{x}), b(\mathbf{x}), c(\mathbf{x})]$, as calculated by (1). Then

$$\frac{\partial f}{\partial x} = \frac{a(\mathbf{x})}{c(\mathbf{x})} \quad \frac{\partial f}{\partial y} = \frac{b(\mathbf{x})}{c(\mathbf{x})}. \tag{3}$$

Here, we also perform another check on the data set. Let the images $\mathbf{P}(\mathbf{x}) = [\frac{a(\mathbf{x})}{c(\mathbf{x})}]$ and $\mathbf{Q}(\mathbf{x}) = [\frac{b(\mathbf{x})}{c(\mathbf{x})}]$. Because

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x} \tag{4}$$

we expect

$$A(\mathbf{x}) \equiv \frac{\partial\left(\mathbf{P}(\mathbf{x})\right)}{\partial y} - \frac{\partial\left(\mathbf{Q}(\mathbf{x})\right)}{\partial x} \tag{5}$$

to be small (close to zero) at each point $\mathbf{x}$.

Assuming that the partial derivatives satisfy the above condition, we can reconstruct the surface up to some constant error in depth. The partial derivatives give the change in surface height with a small step in either the x or the y direction. This means that we can get the surface by summing these changes in height along some path. In particular, we have

$$f(\mathbf{x}) = \oint_C \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right) \cdot \vec{dl} + c \tag{6}$$

where C is a curve starting at some fixed point and ending at $\mathbf{x}$, $\vec{dl}$ is the infinitesimal element along the curve and c is a constant of integration, which represents the unknown height of the surface at the starting point. All methods proposed for surface reconstruction solve the above problem with similar results. In our experiments we applied the surface reconstructions described in [7, 20, 8, 1, 2]. A recent discussion regarding the accuracy of different algorithms for face reconstruction from normals can be found [12].

Figure 4. The reconstructed surface with the computed albedo.

5. Face recognition/verification using Albedo and Depth Images

In this Section we outline the baseline methods used for feature extraction from albedo and depth images for face recognition and verification. The family of methods that we applied extract features using linear projections (also referred to as subspace methods). This family includes Principal Component Analysis (PCA) (the so-called Eigenfaces), Nonnegative Matrix Factorization (NMF) etc. In our experiments NMF [13] produced the best recognition and verification results. In subspace methods, like NMF, the facial images are lexicographically scanned in order to form vectors Let M be the number of samples in the image database $\mathcal{U} = \{\mathbf{u}_1, \mathbf{u}_2, .., \mathbf{u}_M\}$ where $\mathbf{u}_i \in \Re^n$ is a database's image. A linear transformation of the original n-dimensional space onto a subspace with m-dimensions ($m \ll n$) is a matrix $\mathbf{W}^T \in \Re^{m \times n}$. The new feature vectors $\mathbf{y}_k \in \Re^m$ are given by:

$$\mathbf{y}_k = \mathbf{W}^T(\mathbf{u}_k - \bar{\mathbf{u}}), \quad k \in \{1, 2, \ldots, M\} \tag{7}$$

where $\bar{\mathbf{u}} \in \Re^n$ is the mean image of all samples. Classification is performed using a simple distance measure and a nearest neighbour classifier using the normalized correlation.

5.1. Face Recognition using Normalfaces

In this paper we use a face recognition method based on the orientation of the normals. The baseline method is a very easy to implement method that is based on a novel representation of faces, the so-called Normalfaces. For an image $\mathbf{I}$ using the computed $\mathbf{P}$ and $\mathbf{Q}$ from PS we compute:

$$\mathbf{\Phi}(\mathbf{x}) = \text{atan}\frac{\mathbf{Q}(\mathbf{x})}{\mathbf{P}(\mathbf{x})} \tag{8}$$

which is an image that contains the normal orientations. We measure the orientations in the interval $\in [-\frac{\pi}{2}, \frac{\pi}{2}]$. For two images $\mathbf{\Phi}_1(\mathbf{x})$ and $\mathbf{\Phi}_2(\mathbf{x})$ we use the following dissimilarity measure:

$$d(\mathbf{\Phi}_1(\mathbf{x}), \mathbf{\Phi}_2(\mathbf{x})) = 1 - \frac{1}{NM\pi} \sum_{i=1}^{N \times M} |\mathbf{\Phi}_1(\mathbf{x}_i) - \mathbf{\Phi}_2(\mathbf{x}_i)|. \tag{9}$$

The above dissimilarity measure is then used in order to extract features using metric multidimensional scaling. Classification is performed using the normalized correlation in the new space.

6. Baseline Experiments

6.1. Recognition Experiments

We used the subset of images taken with more than a week's interval (126 people). For the majority of them (90 people) the interval was greater than one month. We assessed the recognition performance of all three modalities (i.e., albedo image, normals and height maps). Moreover, we experimented using fusing strategies. For the experiments presented here we tested using two setups:

- In the first one, a very challenging experimental procedure was followed, exploiting only one grayscale albedo image, the surface normals derived from the application of PS, and the depth image derived from the integration of the normal field. Similarly, one grayscale albedo image, one set of normals and a height map was used for testing. Most of the training and testing images display a different facial expression. One sample face recognition is among the most challenging face recognition scenarios with various applications [21]. In [17], a face recognition scenario was designed based only on one sample per person for training. Similar recognition/verification experiments were also described for the FRGC database [16]. A similar scenario was tested in [5].
- In the second, two samples for training and one for testing were used. In our database we have 96 persons with three or more samples per person. The testing image for all 96 subjects was the same one used in the one sample experimental setup. This realization was implemented in order to test whether or not recognition using two samples of the same modality is better than fusing information across different modalities.

6.1.1 Face recognition from Albedo Images

Four source, three source and ray trace-based PS methods were employed for albedo computation. These methods are abbreviated as 4L-PS, 3L-PS and RAY-PS, respectively. The recognition rates using one albedo image for training and one for testing for all the tested PS methods are depicted in Figure 5 (a). As it can be seen the recognition rate

is affected by the PS method applied and noticeably better recognition performance is achieved by PS methods that use all four illuminants. The best recognition rate was equal to 78%.

For the case of the two samples experiment, we used a decision fusion strategy similar to [5]. That is, we combined the matching scores for each person across the two samples of 2D albedo images and ranked the subjects based on the combined scores. Scores from each modality are linearly normalized to the range of $[0, 100]$ before combining. We explored various confidence-weighted versions of the sum, product and minimum rules. Among the fusion rules that we tested, the sum rule provided the best performance overall. The recognition rates for the two samples experiments is depicted in Figure 5 (b). As it can be seen, the use of more than two samples increases the recognition performance. Moreover, the methods which use all four illuminants achieved better recognition rates than those using only three. The best recognition rate was equal to 85%.

It is worth noting here that when using only the 96 persons of the second experiment in the first experiment the recognition rate was also about 78%, as well.

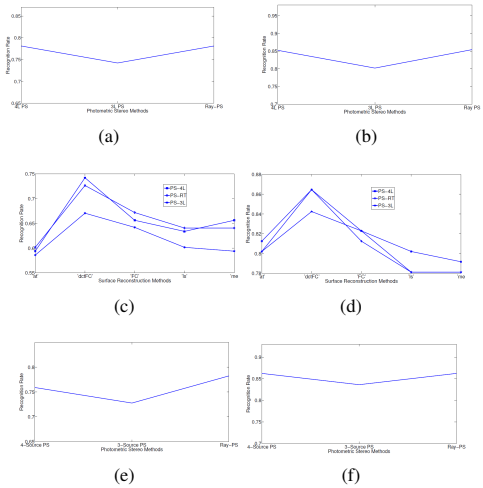


Figure 5. Experiments using, (a) one albedo image for training and one for testing; (b) two albedo images for training and one for testing; (c) one depth image for training and one for testing; (d) two depth images for training and one for testing; (e) one Normalface for training and one for testing; (f) two Normalfaces for training and one for testing .

6.1.2 Face recognition from Depth and Normalface Images

We applied five different methods for surface reconstruction from the normal field. For the reconstruction methods we use the following abbreviations: 1) 'at' for the method in [1] 2) 'dctFC' for the DCT Frankot-Chellappa method [8], 3) 'FC' for the original Frank-Chellappa method [7], 4) 'ls' for the least square solution of the poison equation [20] and 5) 'me' for the reconstruction based on M-estimator [2]. The recognition rates for the one sample experiment and for all reconstruction and PS methods are plotted in Figure 5 (c). The best recognition result it was equal to 74%. As can be seen, PS and reconstruction methods greatly affect the recognition performance. More precisely, four source PS methods always achieve better recognition results. Moreover, the depth maps that were produced by dctFC constantly outperformed the performance of the depth maps produced by all other reconstruction methods.

Experiments using two samples for training and one sample for testing were conducted in a similar manner as the ones for the albedo images. These results are depicted in Figures 5 (d). The best recognition result was equal to 86%.

The experiments using NormalFace for all tested PS methods are depicted in Figures 5 (e) and 5 (f) for one sample and two samples recognition, respectively.

6.1.3 Fusion 2D and 3D

Multimodal decision fusion is performed by combining the match scores for each person across the modalities of 2D albedo and depth image and ranking the subjects based on the combined scores in a similar manner as in the two samples experiments. The sum rule provided the best performance. We performed fusion only on depth images derived from the DCT-FC method. Fusion of intensity and geometry information was conducted only on the subset of persons that have more than 2 samples available in order to be directly comparable with the single modality two samples experiments. The recognition results from multimodal fusion using various PS methods are summarized in Figure 6 (a). The best recognition result was equal to 85%.

A summary of the best recognition results for the single modalities and multimodal fusion is given in Tables 1 and 2.

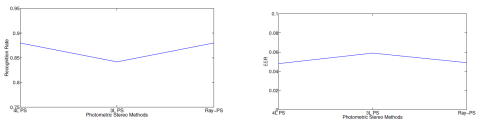


Figure 6. Multimodal fusion results (a) for recognition, (b) for verification.

6.2. Verification Experiments

A person verification system should decide whether an identity claim is valid or not. The performance of face verification systems is measured in terms of the False Rejection Rate (FRR) achieved at a fixed False Acceptance Rate (FAR) [26, 25]. There is a trade-off between FAR and FRR. This trade-off between the FAR and FRR can create a Receiver Operating Characteristic (ROC) curve, where FRR is plotted as a function of FAR. The performance of a verification system is often quoted by a particular operating point of the ROC curve where FAR=FRR. This operating point is called Equal Error Rate (EER). The performance of the algorithms is quoted for the Equal Error Rate (EER) which is the scalar figure of merit that is often used to judge the performance of a verification algorithm. Verification experiments were conducted in the same database as well. The verification protocol was similar to the one defined in the FERET verification protocol in [18]. The probe (or client set) was defined by the 126 persons as in the recognition experiments. The first image is used for training while the second is used for testing client claims. The remaining 135 people in the database, with one image per person, are considered to be impostors.

In the second experiment we used two images from the 96 subjects for training while the third is used for testing client claims. The other 135 persons were used for impostor claims.

6.2.1 Face Verification using Albedo, Depth and Normalface Images

The EERs for various PS methods for the one sample experiment are depicted in Figure 7 (a). The verification results for the two samples experiment and for various PS methods are depicted in Figure 7 (b).

The EERs for various PS and surface reconstruction methods for the one sample experiment are depicted in Figure 7 (c). The verification results for the two samples experiment and for various PS and surface reconstruction methods are depicted in Figure 7 (d).

The EERs for various PS methods for the one sample experiment are depicted in Figure 7 (e). The verification results for the two samples experiment and for various PS methods are depicted in Figure 7 (f).

6.2.2 Multimodal Fusion

Multimodal decision fusion was performed exactly as in the recognition experiments case, by combining the match scores for each person across the modalities of the 2D albedo image and depth map and ranking the subjects based on the combined scores. The fusion results for verification using various PS methods are summarized in Figure 6 (b).

Table 1. A summary of the best percentage of recognition (PR) for all the conducted experiments across different modalities.

Perc. of Recognition (PR%)					
One Sample			Two Samples		
Albedo	Depth	Normal	Albedo	Depth	Normal
78	74	78	85	86	86

Table 2. A summary of the best PR for fusion across different samples and modalities.

Perc. of Recognition (PR%)			
Sample Fusion			Modality Fusion
Albedo	Depth	Normal	Albedo + Depth
85	86	86	85

Table 3. A summary of the best percentage of EER for all the conducted experiments across different modalities.

Verification (EER%)					
One sample			Two samples		
Albedo	Depth	Normal	Albedo	Depth	Normal
7.4	10.5	9.1	5.2	5.7	5.2

A summary of the best verification results for the single modalitities and multimodal fusion is given in Tables 3 and 4.

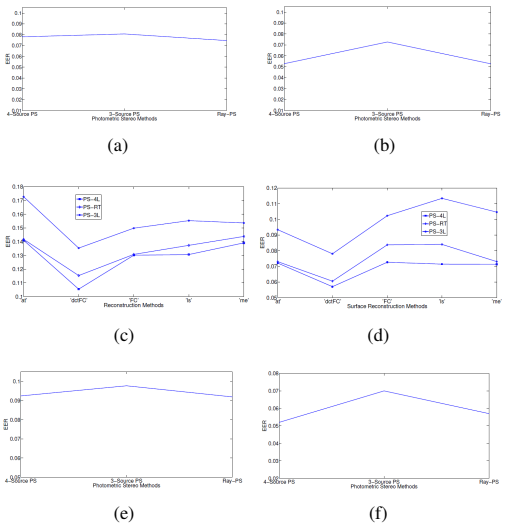


Figure 7. Verification experiments (a) using one albedo image for training; (b) using two albedo images for training; (c) using one depth image for training; (d) using two depth images for training; (e) using one Normalface image for training; (f) using two Normalfaces for training.

Table 4. A summary of the best percentage of EER for fusion across different samples and modalities.

Verification (EER%)			
Sample Fusion			Modality Fusion
Albedo	Depth	Normal	Albedo + Depth
5.2	5.7	5.2	5.2

7. Discussion and Concluding Remarks

In this paper, we presented a new database collected using a real life commercial setting based on photometric stereo. We presented the first experiments which demonstrate how different methods in the pipeline of photometric stereo affect the recognition performance and concluded the following: (1) Four source photometric stereo methods produce facial samples (albedo, normals) that achieve constantly better recognition and verification performance regardless of the reconstruction method applied. (2) The reconstruction methods greatly affect the recognition and verification performance. The method which constantly produces the best recognition/verification performance proved to be the one proposed in [8].

Moreover, we have verified most of the findings of [5]: (1) In most cases the best recognition and verification results of recovered albedo, normals and the reconstructed depth maps achieve approximately the same results, in some cases the recovered albedo produces better results. (2) Fusion of albedo and reconstructed surfaces produce significantly better results than using only the albedo or the depth images. (3) Fusion of two albedo images in the same way that we fused the results of albedo and depth map gave approximately the same recognition and verification results. Details on how the database can be provided to researchers are provided in http://Photoface.iti.gr/ or http://www.uwe.ac.uk/research/Photoface.

References

[1] A. Agrawal, R. Chellappa, and R. Raskar. An algebraic approach to surface reconstruction from gradient fields. In *ICCV*, pages 174–181, 2005. 172, 174

[2] A. Agrawal, R. Raskar, and R. Chellappa. What is the range of surface reconstructions from a gradient field? *ECCV*, 3951:578, 2006. 172, 174

[3] V. Argyriou and M. Petrou. Recursive photometric stereo when multiple shadows and highlights are present. In *CVPR*, pages 1–6, 2008. 172

[4] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE T-PAMI*, 19:711–720, 1997. 169

[5] K. Chang, K. Bowyer, and P. Flynn. An evaluation of multimodal 2D+ 3D face biometrics. *IEEE T-PAMI*, pages 619–624, 2005. 170, 173, 174, 176

[6] D. A. Forsyth and J. Ponce. *Computer Vision, A Modern Approach*. Prentice Hall, Upper Saddle River, NJ, 2003. 172

[7] R. T. Frankot and R. Chellappa. A method for enforcing integrability in shape from shading algorithms. *IEEE T-PAMI*, 10:439–451, 1988. 172, 174

[8] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman. From few to many: illumination cone models for face recognition under variable lighting and pose. *IEEE T-PAMI*, 23:643–660, 2001. 169, 172, 174, 176

[9] B. Gokberk, H. Dutagaci, A. Ulas, L. Akarun, and B. Sankur. Representation plurality and fusion for 3-D face recognition. *IEEE T- SMCB*, 38(1):155–173, 2008. 170

[10] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-pie. *IVC*, 28:807–813, 2010. 169, 171

[11] P. W. Hallinan, G. G. Gordon, A. L. Yuille, P. Giblin, and D. Mumford. *Two- and Three- Dimensional Patterns of the Face*. A K Peters Ltd., 1999. 169

[12] M. Hansen, G. Atkinson, L. Smith, and M. Smith. 3D face reconstructions from photometric stereo using near infrared and visible light. *CVIU*, 114(8):942–951, 2010. 172

[13] D. Lee and H. Seung. Algorithms for non-negative matrix factorization. In *NIPS*, pages 556–562, 2000. 173

[14] K. Lee, J. Ho, and D. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE T-PAMI*, 27(5):684–698, 2005. 169

[15] S. Li and A. Jain. *Handbook of Face Recognition*. Springer-Verlag, 2005. 169

[16] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *In Proc. CVPR*, pages 947–954, 2005. 170, 171, 173

[17] P. J. Phillips, H. Wechsler, and P. Rauss. The feret database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing*, 16:295–306, 1998. 169, 173

[18] S. Rizvi, P. Phillips, and H. Moon. The FERET verification testing protocol for face recognition algorithms. *FG98*, pages 48–53, 1998. 175

[19] T. Sim, S. Baker, and M. Bsat. The cmu pose, illumination, and expression database. *IEEE T-PAMI*, 25:1615–1618, 2003. 169

[20] T. Simchony, R. Chellappa, and M. Shao. Direct analytical methods for solving Poisson equations in computer vision problems. *IEEE T-PAMI*, 12(5):435–446, 1990. 172, 174

[21] X. Tan, S. Chen, Z. Zhou, and F. Zhang. Face recognition from a single image per person: A survey. *Pat. Rec.*, 39(9):1725–1745, 2006. 173

[22] R. J. Woodham. Photometric method for determining surface orientation from multiple images. *Opt. Eng.*, 19:139–144, 1980. 170, 172

[23] www.3dmd.com/3dmdface.html. [Accessed 13 March 2010]. 170

[24] www.konicaminolta.com/sensingusa/products/3d/non contact/vivid910. [Accessed 15 Dec 2009]. 170

[25] S. Zafeiriou, A. Tefas, and I. Pitas. Exploiting discriminant information in elastic graph matching. In *ICIP*, volume 3, pages III–768, 2005. 175

[26] S. Zafeiriou, A. Tefas, and I. Pitas. Learning discriminant person-specific facial models using expandable graphs. *IEEE T-IFS*, 2(1):55–68, 2007. 175

Appendix B

Additional publications

Three conference papers produced as part of this thesis are included on the following pages.

Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 00 (2010) 1–9
Procedia Computer Science

ICEBT 2010

Biologically inspired 3D face recognition from surface normals

Mark F. Hansen, Gary A. Atkinson

Machine Vision Laboratory, University of the West of England, Bristol. BS16 1QY.

Abstract

A major consideration in state-of-the-art face recognition systems is the amount of data that is required to represent a face. Even a small (64×64) photograph of a face has 2^{12} dimensions in which a face may sit. When large (> 1MB) photographs of faces are used, this represents a very large (and practically intractable) space and ways of reducing dimensionality without losing discriminatory information are needed for storing data for recognition. The eigenface technique, which is based upon Principal Components Analysis (PCA), is a well established dimension reduction method in face recognition research but does not have any biological basis. Humans excel at familiar face recognition and this paper attempts to show that modelling a biologically plausible process is a valid alternative approach to using eigenfaces for dimension reduction. Using a biologically inspired method to extract the certain facial discriminatory information which mirrors some of the idiosyncrasies of the human visual system, we show that recognition rates remain high despite 90% of the raw data being discarded.

Keywords: 3D face recognition, surface normals, biometrics, dimension reduction, caricature

1. Introduction

Face recognition has been an area of intense research for over forty years and, although significant progress has been made, a number of major challenges remain. Much of the research focuses on face recognition using 2D images which has highlighted some universal problems that affect recognition accuracy. Two of these problems, pose and illumination variance, can be compensated for using 3D models rather than 2D photographs. Because of this, and the increased availability of 3D capture devices, 3D face recognition has become an active research area over the past decade.

A primary goal of automatic face recognition is to reproduce the phenomenal ability of human face discrimination. Certain approaches have modelled features of the Human Visual System (HVS) with great effect e.g. the use of Gabor filters in [1], but most approaches tend to use more traditional pattern recognition and classification techniques. The reason for this is arguably two-fold: 1) the processes underlying human face recognition are still poorly understood and 2) good results are achieved using classical pattern recognition approaches.

The motivation for this work therefore comes from attempting to improve aspects of automatic face recognition by incorporating features of the HVS. In particular we look at dimension reduction and present a method based upon the idea of caricaturing that was theorised by Unnikrishnan [2]. By only using facial data which falls outside the 5^{th} and 95^{th} percentiles for a given face database (i.e. 90% is discarded) we show that recognition rates only show a proportionally small decrease thus lending support to Unnikrishnan's hypothesis.

Email address: mark.hansen@uwe.ac.uk (Mark F. Hansen)

1.1. Related work

Early research into automatic face recognition focused on describing a face in terms of absolute or ratios of distances between features [3, 4, 5]. Information theory inspired a new statistical approach termed eigenfaces by which Principal Components Analysis (PCA) is used to describe a face in terms of a linear combination of coefficients [6]. Recognition is is then performed using the smallest Euclidean distance between the coefficients of a probe image and the mean coefficients for each identity within the gallery. This approach has the advantage of not needing to mark and measure fiducial features on the faces as was necessary with the earlier approaches. The Fisherface technique [7] incorporates class information (in this case the identities of the photographs) in order to find a better dimensional representation which maximises the clustering of the classes, making discrimination easier. Both eigenfaces and Fisherfaces are commonly used in state-of-the-art research as they represent acknowledged benchmarks, with Fisherfaces providing better recognition performance as long as there are sufficient training examples [8]. For this reason, the Fisherfaces technique is adopted for use in this paper.

A different and biologically motivated approach comes from using Gabor filters [1, 9, 10]. The Gabor filter [11] is fundamentally a sine wave windowed by a Gaussian. By varying the orientation and frequency of these waves, filter banks which mimic functionality of an area in the primary visual cortex (area V1) are created [12, 13]. In the approach used by Wiskott et al. [10], it is not necessary to mark out fiducial features, as an elastic bunch graph map (EBGM) finds the features most similar to those in its database automatically. Testament to the benefits of using biologically inspired Gabor filters comes from the FERET [14] evaluation and FVC2004 [15] face recognition tests, in which the top performing algorithms used Gabor filters for feature extraction.

The main drawback of implementing Gabor filters is that they are computationally intensive. More efficient alternatives are Local Binary Patterns (LBP) which approximate the Gabor function. This approach is most commonly associated with face detection e.g. [16] but it has also been used successfully for face [17] and even expression recognition [18].

The approximation of area V1 functionality by Gabor filters represents the reproduction of a low-level process. While face recognition undoubtably relies on this, it is not something uniquely associated with it. A number of high-level features which are directly involved with human face recognition can be found in [19] including caricaturing. Caricaturing can be defined as the exaggeration of features away from the average e.g. if someone has a larger than average nose, the caricature would exaggerate the nose to make it even larger. Caricaturing essentially enhances those facial features that are unusual or deviate sufficiently from the norm. It has been shown that humans are better able to recognise a caricature than they are the veridical image [20, 21]. This finding is interesting as caricaturing is simply distorting or adding noise to an image, but this noise aids human recognition and this, in turn, provides insights into the storage or retrieval mechanism used by the human brain.

Unnikrishnan [2] conceptualises an approach similar to face caricatures, whereby only those features which deviate from the norm by more than a threshold are used to uniquely describe face. Unnikrishnan suggests using those metrics whose deviations lie below the 5^{th} percentile and above the 95^{th} percentile, thereby discarding 90% of the data. Apart from dimension reduction, an interesting feature of this approach is that because it is norm-based, faces from under-represented groups (in our case ethnicity and gender) will possess features not present in the average population. These features are distinguishing to that group leading to a clustering of minority groups making discrimination for difficult. This is analagous to a well documented feature in human face recognition known as the *own-race effect* [22] by which discrimination of faces from races other than the subject's own is diminished. No empirical support for Unnikrishnan's hypothesis is given in [2], so the aim of this paper is to test the presented theory.

Most face recognition experiments in the research literature are carried out using 2D photographs, but it has been shown that 3D models lead to improved recognition rates because illumination and pose can be compensated for [23], although this finding is not always replicated [24]. The database used for the experiments in this paper consists of surface normal data captured using the PhotoFace device (Fig. 1). PhotoFace is a 3D photometric stereo capture system which was placed in a workplace corridor for six months and left to capture unconstrained images of employees walking through the device (for more details, the interested reader is referred to [25]). Photometric stereo is a technique of illuminating an object from multiple directions and using the known positions of illuminants and pixel intensity to estimate surface orientation [26]. Surface normal data is particularly well suited to face recognition as shown by Gökberk in his meta-analysis [27] on the effect of different data representations for face recognition. He concluded that "*... surface normals are better descriptors than the 3D coordinates of the facial points.*".

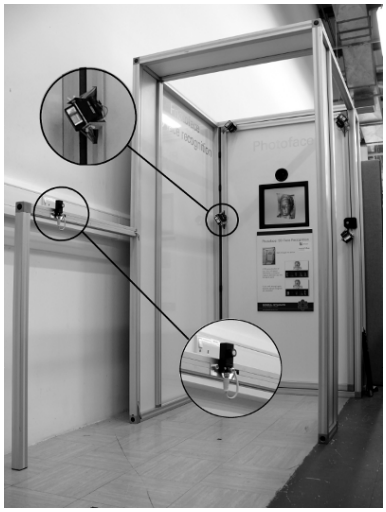


Figure 1: The PhotoFace capture device. The insets show a flashgun light source and the ultrasound trigger, which detects the presence of a person using the device.

If Unnikrishnan's hypothesis is correct (and also applicable to surface normal data) we can expect to see little effect on recognition results when only the outlying 10% of data is used. Also, if Unnikrishnan's hypothesis reflects a real process in the HVS, then we should expect it to exhibit some of the idiosyncracies of human face recognition abilities. As mentioned above, one such phenomenon is the *own-race effect*, and we test whether recognition is worse for a subject from a minority race than the norm. This is also extended to what we term the *own-sex effect* by which we might expect worse recognition on the gender which is under-represented. (NB There is no evidence for the own-sex effect in human recognition, probably because exposure to one sex over another to the same levels as to generate the own-race effect is not feasible).

1.2. Contributions

The contributions of this paper are four-fold:

- We show that discarding 90% of facial data by only keeping the outlying 10% just leads to a 24% drop in recognition performance on 3D surface normal data.
- The drop in performance using 2D data, on the other hand, is much greater (a drop of 43%).
- This provides empirical support for Unnikrishnan's hypothesis concerning the important discriminatory properties of outliers.
- We find no evidence to support Unnikrishnan's assumption that using outlying data reflects a process in the HVS in terms of own-race/sex effect (although further experiments are required).
- We show that 3D surface normal data gives better recognition performance than 2D photographs on a database of images captured in an unconstrained "real world" environment.

2. Data and Methodology

The data used for our experiments consists of 61 subjects with at least six sessions each (that is six sets of photometric stereo images per subject). All images were taken in a frontal pose with neutral expression. The maximum number of sessions per subject is 70, the mean number of sessions per subject is 16 with a mode of 7. Of the 61

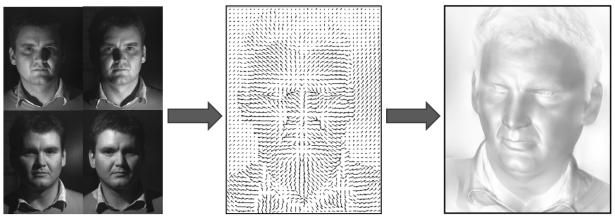


Figure 2: Four differently illuminated images, the needle map of surface normals and the integrated surface

subjects, only two are female, and only one is not caucasian - these are the subjects whose sessions are used for exploring the own-race/sex effect. There are a total of 1000 sessions. Four images are captured per session with different illuminants in ≈20ms. This effectively freezes the subject's motion. For these experiments, visible light flashguns are used (colour temperature ≈5600K). A standard photometric stereo technique [28, Section 5.4] is then used to estimate the surface normals at each pixel. Although not used in this paper, the normals can be integrated to form a surface via, for example, the well known Frankot-Chellappa method [29]. An example set of images can be seen in Fig. 2.

The centre of the eyes and nasion are manually labelled on each image. The images are then scaled and aligned to one another. Fig. 3 shows how the face region is cropped based around the distance between the centres of the eyes. This results in a close crop around the eyes nose and mouth, and excludes areas such as the chin and forehead which can frequently be covered with hair and are therefore unreliable features for recognition. Due to memory limitations the images are then scaled down to 80×80px.

In order to remove any artefacts which are caused by the flashguns having different brightness, the greyscale intensity of the images is normalised. This is achieved by making the mean of each image the same as the mean of all session images. Other normalisation techniques such as histogram equalisation, contrast limited adaptive histogram equalisation and increasing the range of intensity values to a maximum 0-255 were investigated in terms of their effect on recognition performance, but none offered any improvement.

The images that we use for 2D recognition are generated by taking the mean of each pixel of the four differently lit images. This reduces any confounding influence of illumination variance that may be present if only one lighting condition were used e.g. extreme lighting and cast shadows. Each mean image is reshaped into a vector and these vectors are added into a matrix such that columns represent sessions and rows represent greyscale intensities at a particular pixel. As each mean image is 80×80px, the dimension of the matrix used for percentile calculation and subsequent recognition is 6400×1000.

For the 3D surface normal data, only the x and y components of the normals are used, as there is redundancy in the z component. When calculating percentiles and performing recognition using the Fisherface technique, the x and y components of each session are reshaped and then concatenated into a single vector. In the same way as for the 2D mean images, these vectors are added into a matrix such that columns represent sessions and rows represent x and y components at a particular pixel. As there are 80×80 values for both the x and y component, each session is therefore represented by a vector $6400 \times 2 = 12800$ in length. The dimensions of the matrix are therefore 12800×1000.

In order to work out which data in each image falls in the outlying 10% of the data, we first need to calculate the thresholds for each pixel which represent the 5^{th} and 95^{th} percentile values. This is a norm-based approach, and we are interested in the norm across the whole dataset for each pixel rather than the norm for each image. For the 2D photographs, percentile values are calculated for the greyscale intensity value for each pixel. There are 1000 sessions, so there are 1000 values for each pixel from which we calculate the 5^{th} and 95^{th} percentile values. Once reshaped into the original dimensions, this results in two 80×80 matrices (one for the 5^{th} and one for 95^{th} percentile), examples of which can be seen in Fig. 4. In the same way, for 3D surface normal data, percentile values are calculated for x and y surface normal component values for each pixel. Once these thresholds have been calculated, all pixels which have a value between the 5^{th} and 95^{th} percentile are discarded, leaving only the 10% outlying data.

The method used to test recognition accuracy is the leave-one-out paradigm. This dictates that every session is used as a probe against a gallery of all other sessions once. There are therefore 1000 classifications per condition of which the percentage correctly identified is shown.

The Fisherface technique [7] is used for subspace representation and simple pairwise Euclidean distance compar-

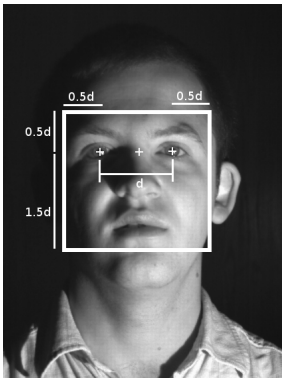


Figure 3: Cropping the face images based on the inter-eye distance. The distance between the eye centres is denoted by d.

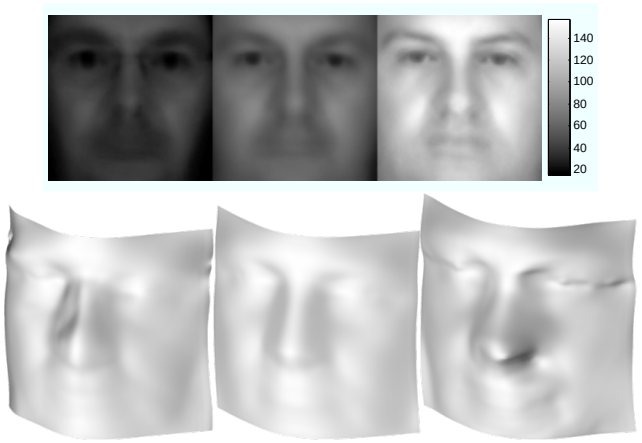


Figure 4: Images of 5th, 50th and 95th percentile values (respectively) for 2D photographs (top row) and 3D surface normals (bottom row). The normals have been integrated here via the Frankot-Chellapa method to form a surface (for illustrative purposes only). In the 2D images, there is a general trend from dark to light as would be expected, but both sets of images also show different identities for each percentile which do not match any individual in the database.

ison between class means and the probe image is used for recognition. This particular method has been chosen as it is well known in the literature, is proven to be effective, it is a linear technique and it is computationally efficient. While there are better algorithms available, absolute performance is not what we require for these experiments; we need to measure relative performance between conditions – a task which the Fisherface technique is well suited to.

3. Results

Example data for two subjects can be seen in Fig. 5. 2D examples are shown in the eight images on the left, and 3D examples are shown on the right. The 3D examples only show y-component data to simplify visualisation – it should be noted that the experiments are also performed on the x-components. Each row represents data for one subject. The first two images on each row of the groups show examples of aligned and cropped greyscale intensity images (2D photographs) and raw y-component surface normals. The next two images show the corresponding outlying data of the first two images (i.e. those pixels with a value whose deviation lies below the 5^{th} or above the 95^{th} percentile). There is visibly more consistency between the outlying 3D data than the 2D data, especially for the first subject.

	Base rate	Outliers
2D photographs	91.2	30.2
3D surface normals	97.5	73.5

Table 1: Recognition rates (%) on 2D photographs and 3D surface normal data. The base rate column shows recognition rates for the raw data, and the outliers column shows recognition rates on the outlying 10% of data (data whose deviation lies below the 5^{th} percentile and above the 95^{th} percentile).

	Base rate	Outliers
2D own-race	100	50
3D own-race	100	75
2D own-sex	97.62	50
3D own-sex	97.61	88.1

Table 2: Recognition rates (%) for subjects using a single race/sex subset of the data.

Table 1 shows the baseline recognition rates for 2D and 3D data, as well as the rates using only the most outlying 10% of data. The table can be summarised as follows:

1. 3D surface normal data gives better recognition rates than 2D photographs (97.5% vs 91.2%)
2. Far better recognition is seen on the 3D outlying data than the 2D outlying data (73.5% vs 30.2%).
3. The decrease in performance when only the outlying 10% of data is used is only 24% on the 3D data which is disproportional to the 90% of data which has been discarded.

Table 2 is designed to investigate the own-race and own-sex effect. It is clear from the table however, that neither the own-race nor the own-sex effect are being exhibited as the performance drop of the outlying data is less than that across the whole group (as seen in Table 1). Caution should be exercised in any interpretation of these results as the number of sessions available for ethnic minority/female subjects is very small (one subjects with 16 sessions and two subjects with 42 sessions respectively). These results are discussed further in Section 4.

It is possible that Unnikrishnan's assumption that the most outlying data provides the most discriminatory information is inaccurate as no empirical evidence is offered. However, these experiments do suggest that more information is generally contained in the outlying data than the rest of the data. It may be that there are other bands of percentiles which provide better recognition. This was investigated by measuring the recognition rate using different bands of percentiles e.g. [10-15, 90-95], [15-20, 85-90] etc. which account for 10% of the data. Fig. 6 shows a plot of the recognition rate against these bands and provides support for Unnikrishnan in that the most outlying 10% of the data gives the best recognition performance. Interestingly, after a decrease in performance, there is a rise as we near the 50th percentile. The reason for this pattern is unknown, but will be explored in further research.

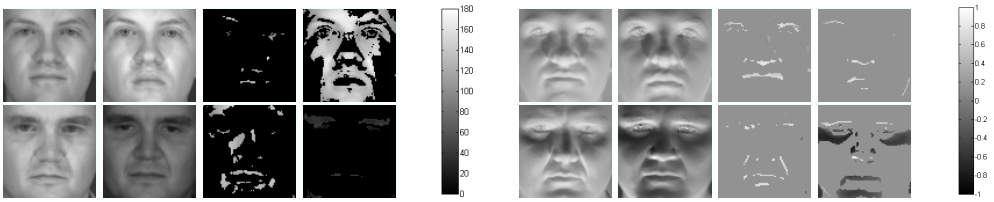


Figure 5: Examples of data from two sessions of two subjects. 2D data is shown on the left and one component (the y-component) of the 3D data on the right. Within each group the first two columns show examples of the baseline condition (all data), and the last two columns show the outlying data which falls outside the 5^{th} and 95^{th} percentile values.

4. Discussion

The results show that recognition rates of over 90% are achieved on the frontal, neutral expression data from the PhotoFace database, with surface normal data providing the highest level of discrimination (97.5%). By applying

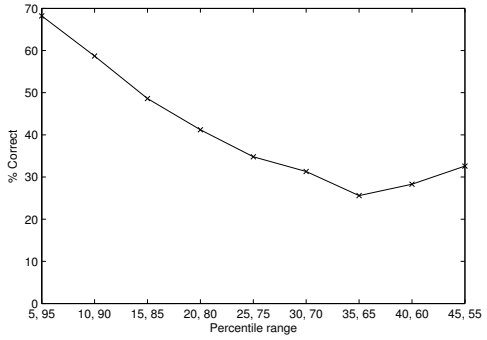


Figure 6: Recognition accuracy as a function of percentile band. Each tick on the x-axis shows the (upper, lower) limits of the 10% percentile band e.g. (0-5, 95-100), (5-10, 90-95) etc. This shows that the best recognition performance is given by the most outlying data that is less than the 5^{th} percentile and greater than the 95^{th} percentile.

Unnikrishnan's theory that most discriminating data can be found in the outlying 5% percentile ranges, we have tested recognition rates after discarding 90% of the data. There is a decrease in recognition performance but it is not proportional to the amount of data that has been discarded e.g. 90% of data has been removed without an accompanying 90% decrease in recognition performance. In the case of the surface normal tests this is a 24% decrease in performance and for the 2D data, a 61% decrease. What we can infer is that there is more reliable discriminatory information in the 3D outliers than in the mass of the data. By looking at the examples of outlying data in Fig. 5 however, it seems unlikely that this discriminatory information is the same as that used to aid human recognition. Although features do indeed appear to be picked out (e.g. the broad nose in the first subject), there is no obviously discernable pattern in the images which one could liken to a caricature (which Unnikrishnan likens his approach to), and for the second subject there is little similarity between the 3D outlying data images. Arguably, one could say that the subject shown on the bottom has distinctive eyebrows and that this is highlighted in the second 2D outlier image, but no such feature is highlighted on the first 2D outlier image.

While using the outlying data does not lead to improved recognition rates, it does offer a relatively simple way of reducing the amount of data without losing the same amount of discrimination. It could also provide a rough (but relatively accurate) metric as to where the face may lie in face-space, which can then be searched more exhaustively to provide an accurate match. This has implications in real world recognition systems, where the numbers of identities which may be stored in a central database could be in the millions (or even billions). If we have a quick, low computation key by which we can reduce the search space then we can use more thorough recognition algorithms on this subset only. Attempting to use complex recognition algorithms across a very large database would cripple even the most advanced systems currently available.

We did not see any evidence of the own-race or own-sex effect. Where we might have expected a far greater performance decrease in the outlying data condition according to Unnikrishnan's hypothesis, we actually have a far smaller one. This implies that these under-represented samples are actually more readily discriminated between. In the case of the own-race test, a problem arises in that there is only one subject available to test against. This means that instead of not being able to tell subjects from the same race apart it actually becomes easier as one can say this person is not of the majority race, and therefore it is that one particular person. However as there are three subjects ($\approx$ 5% of the sample population) for the own-sex effect there is likely be a different reason for the improved performance on the outlying data compared with the whole dataset results. One possible reason could be that they are sufficiently different from the rest of the sample population. This would mean that they form a discrete subspace within the total subspace away from the general population and still provide sufficient between-class scatter amongst themselves to accurately enable recognition. An analysis of the Fisherface subspace would provide evidence for this and will likely be the subject of further work. As mentioned previously, caution must be exercised in drawing any conclusions from this data due to the very small number of samples. Future work will attempt to verify these results using a larger number of samples.

Limitations and future research

- The images were reduced to 80x80 pixels in order to be able to run the experiments on a standard desktop computer (Quadcore 2.5GHz, 2GB RAM, Windows XP SP3). Although good recognition rates are achieved at this resolution, the full size images are likely to offer better data.
- Currently, the images are aligned manually by selecting three points on the face. This task is time consuming and requires vigilance. It is likely that some data will not be aligned perfectly with the rest due to small human errors. This process would be ideally automated using feature detection techniques such as Gabor filters. Ideally any alignment algorithm would also need to take into account 3D rotations.
- Future work will look into whether humans group similar looking faces together in face space. It would be interesting to code the data by hand to group individuals who look similar to one another and see whether these groupings are represented by the outlier face space. It would then be possible to see whether humans group similar looking people together based on their most unusual features and to give support to norm-based face processing when people make similarity judgements.
- This paper looks primarily at outlying data (deviation from the norm less than the 5^{th} percentile and more than the 95^{th} percentile) as suggested by Unnikrishnan. We also see how the amount of discriminatory information in other ranges differs (Fig. 6). Further work is required to see whether better recognition could be achieved by using the percentile values which provide the best performance individually and combining the data i.e. are there certain *super-percentiles* which contain more discriminatory information than others?
- Investigate why the discriminatory information dips towards the 25th/75th percentile as shown in Fig. 6 before rising again.

5. Conclusion

This paper has provided evidence that outlying data contains disproportionately more discriminatory information which is useful for face recognition. Discarding 90% of the data typically results in only a 24% decrease in recognition performance on 3D surface normal data. This lends direct support to Unnikrishnan's [2] hypothesis, but it is unlikely that this particular implementation reflects any particular process of the HVS as images of the outliers are not easily recognisable by humans and no own-race or own-sex effects were observed (although alternative explanations are explored). Additionally we show that 3D surface normal data leads to better recognition than 2D photographs. Future work will look into the suborganisation of face space to see whether there are discrete subspaces for under-represented groups.

6. References

[1] B. S. Manjunath, R. Chellappa, C. von der Malsburg, A feature based approach to face recognition, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1992, pp. 373–378.
[2] M. K. Unnikrishnan, How is the individuality of a face recognized?, Journal of theoretical biology.
[3] T. Kanade, Computer recognition of human faces, Birkhauser, Basel, Switzerland and Stuttgart, Germany, 1973.
[4] M. D. Kelly, Visual identification of people by computer. Techical Report AI-130, Stanford AI Project, Stanford, CA, 1970.
[5] W. W. Bledsoe, Man-machine facial recognition, Panoramic Research Inc., Palo Alto, CA.
[6] M. Turk, A. Pentland, Eigenfaces for recognition, Journal of Cognitive Neuroscience 3 (1) (1991) 71–86.
[7] P. Belhumeur, J. Hespanha, D. Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (7) (1997) 711–720.
[8] A. M. Martínez, A. C. Kak, PCA versus LDA, IEEE Transactions on Pattern Analysis and Machine Intelligence (2001) 228–233.
[9] M. Lades, J. C. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R. P. Wurtz, W. Konen, Distortion invariant object recognition in the dynamic link architecture, IEEE Transactions on Computers 42 (3) (1993) 300–311.
[10] L. Wiskott, J. M. Fellous, N. Krüger, C. von der Malsburg, Face recognition by elastic bunch graph matching, IEEE Transactions on Pattern Analysis and Machine Intelligence (1997) 775–779.
[11] D. Gabor, Theory of communication, IEE j, Comm. Eng 93 (1946) 429–457.
[12] D. H. Hubel, T. N. Wiesel, Sequence regularity and geometry of orientation columns in the monkey striate cortex., The Journal of Comparative Neurology 158 (3) (1974) 267–293.

[13] J. G. Daugman, Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters, Journal of the Optical Society of America A 2 (7) (1985) 1160–1169.
[14] P. J. Phillips, H. Moon, S. A. Rizvi, P. J. Rauss, The FERET evaluation methodology for Face-Recognition algorithms, IEEE Transactions on Pattern Analysis and Machine Intelligence (2000) 1090–1104.
[15] K. Messer, J. Kittler, M. Sadeghi, M. Hamouz, A. Kostin, F. Cardinaux, S. Marcel, S. Bengio, C. Sanderson, N. Poh, Face authentication test on the BANCA database, in: Proc. of International Conference on Pattern Recognition, 2004, pp. 523–532.
[16] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition., Vol. 1, IEEE Computer Society, 2001.
[17] T. Ahonen, A. Hadid, M. Pietikäinen, Face description with local binary patterns: Application to face recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence (2006) 2037–2041.
[18] C. Shan, S. Gong, P. W. McOwan, Facial expression recognition based on local binary patterns: A comprehensive study, Image and Vision Computing 27 (6) (2009) 803–816.
[19] P. Sinha, B. Balas, Y. Ostrovsky, R. Russell, Face recognition by humans: Nineteen results all computer vision researchers should know about, Proceedings of the IEEE 94 (11) (2006) 1948–1962.
[20] R. Mauro, M. Kubovy, Caricature and face recognition, Memory & Cognition 20 (4) (1992) 433–440.
[21] G. Rhodes, S. Brennan, S. Carey, Identification and ratings of caricatures: Implications for mental representations of faces, Cognitive Psychology 19 (4) (1987) 473–497.
[22] C. A. Meissner, J. C. Brigham, Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review, Psychology, Public Policy, and Law 7 (1) (2001) 3–35.
[23] K. Chang, K. Bowyer, P. Flynn, Face recognition using 2D and 3D facial data, in: ACM Workshop on Multimodal User Authentication, 2003, pp. 25–32.
[24] M. Hüsken, M. Brauckmann, S. Gehlen, C. V. der Malsburg, Strategies and benefits of fusion of 2D and 3D face recognition, in: IEEE workshop on face recognition grand challenge experiments, 2005, p. 174.
[25] M. F. Hansen, G. A. Atkinson, L. N. Smith, M. L. Smith, 3D face reconstructions from photometric stereo using near infrared and visible light, Computer Vision and Image Understanding 114 (8) (2010) 942–951.
[26] R. J. Woodham, Photometric method for determining surface orientation from multiple images, Opt. Eng. 19 (1) (1980) 139–144.
[27] B. Gökberk, M. O. İrfanoğlu, L. Akarun, 3D shape-based face representation and feature extraction for face recognition, Image and Vision Computing 24 (8) (2006) 857–869.
[28] D. A. Forsyth, J. Ponce, Computer Vision: A modern approach, Prentice Hall Professional Technical Reference, 2002.
[29] R. T. Frankot, R. Chellappa, A method for enforcing integrability in shape from shading algorithms, IEEE Transactions on pattern analysis and machine intelligence 10 (4) (1988) 439–451.

Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 00 (2010) 1–9
Procedia Computer Science

ICEBT 2010

A efficient and practical 3D face scanner using near infrared and visible photometric stereo

Gary A. Atkinson, Mark F. Hansen, Melvyn L. Smith, Lyndon N. Smith

The University of the West of England, Bristol BS16 1QY

Abstract

This paper is concerned with the acquisition of model data for automatic 3D face recognition applications. As 3D methods become progressively more popular in face recognition research, the need for fast and accurate data capture has become crucial. This paper is motivated by this need and offers three primary contributions. Firstly, the paper demonstrates that four-source photometric stereo offers a potential means for data capture that is computationally and financially viable and easily deployable in commercial settings. We have shown that both visible light and less intrusive near infrared light is suitable for facial illumination. The second contribution is a detailed set of experimental results that compare the accuracy of the device to ground truth, which was captured using a commercial projected pattern range finder. Importantly, we show that not only is near infrared light a valid alternative to the more commonly exploited visible light, but that it actually gives more accurate reconstructions. Finally, we assess the validity of the Lambertian assumption on skin reflectance data and show that better results may be obtained by incorporating more advanced reflectance functions, such as the Oren-Nayar model.

Keywords: 3D reconstruction, face recognition, photometric stereo

1. Introduction

The variability of face images with lighting and pose result in one of the most significant pitfalls of current 2D face recognition methods. This is the prime motivation for the move of recent years to incorporate 3D information [1]. The 3D shape of a face is completely invariant to the illumination distribution and pose changes can more readily be corrected for if the 3D structure of the face is known.

One problem with 3D face recognition however, is the difficulty of acquiring the necessary 3D face models in order to perform the actual recognition. This problem is the motivation for our work. The paper aims to use a relatively inexpensive and easily deployable system to capture 3D face geometries and assess the accuracy of reconstructions for face recognition. The device uses four-source photometric stereo (PS) synchronised at high-speed to a camera and triggered by an ultrasound proximity sensor [2]. We apply our experiments using both visible-light flash-gun sources and the more covert near-infrared (NIR) LED cluster sources. The results in this paper make the system attractive for use in many commercial and industrial settings such as at entrances to high security areas, airport check-in and border control.

Email address: gary.atkinson@uwe.ac.uk (Gary A. Atkinson)

1.1. Related Work

The use of 3D information for face recognition has been attracting increasing attention in recent years [1, 3, 4]. Popular methods include the 3D morphable model of Blanz and Vetter [5] and the geodesic representations of Bronstein et al. [6] and Mpiperis et al. [7]. Research that directly compares 2D and 3D recognition frequently reports improved success rates for 3D recognition and that the best results occur when 2D and 3D information is fused [1]. As demand for practical face recognition systems is likely to increase, it is important that the most accurate methods are used and that the acquisition devices are both practical and affordable. There are a number of existing ways to capture and reconstruct 3D face information and the benefits and limitations of the most common approaches will now be discussed with the aim of putting our device into context.

Structured light scanning is perhaps the best known approach to generating 3D models of faces. This was used for generating the morphable head model in [5] and also for all the 3D faces used in the renowned FRGC2.0 dataset [8]. The Minolta Vivid 910 device [9] used for FRGC2.0 has a quoted accuracy of ±0.10mm. However, these devices take about 2.5s to capture the data, during which time the subject could move, thus distorting the reconstruction and are sensitive to high levels of ambient illumination. A much faster alternative is to project patterns across the entire face, as in the 3dMD system [10], which is used in this paper to acquire ground truth models. This device uses a number cameras to take images of an object from different positions and uses the projected patterns to solve the correspondence problem between the images. The benefits of this system are its high accuracy (reported as <0.2mm) and the speed of image acquisition (1.5ms). However, the processing time is much longer, the system is expensive and it requires a time consuming calibration procedure.

A much cheaper method to acquire the 3D shape of a face is to use shape-from-shading (SFS) where the naturally occuring intensity patterns of a face are used to extract the 3D geometry from a single image [11]. However the problem is ill-posed, meaning that there is no guarantee of a unique solution for a given image [12]. Smith and Hancock [13] use a statistical model in the SFS paradigm to recover geometry. An alternative is to photograph the object multiple times under different illumination. This technique is known as Photometric Stereo (PS) and was first devised by Woodham [14] who showed that for any Lambertian surface, three differently illuminated images are sufficient to remove the ambiguity associated with a single image. Several researchers have attempted to address issues of shadows, specularities and non-Lambertian reflectance with PS [15, 16, 17, 18, 19, 20].

Of the vast amount of research into automatic face recognition during the last two decades [21], relatively little work has involved PS. Kee et al. investigate the use of 3-source PS under dark room conditions [22]. They were able to determine the optimal light source arrangement and demonstrate a working recognition system. Zhou, Chellappa and Jacobs apply rank, integrability and symmetry constraints to adapt PS to face-specific applications [23]. Zhou et al. extended a PS approach to unknown light sources [24]. Georghiades et al. show how reconstructions from PS can be used to form a generative model to synthesise images under novel pose and illumination [25].

Comparing point clouds, the shape index, depth maps, profiles and surface normals in terms of face recognition performance, Gökberk et al. [4] concluded that surface normals provide the best features for face recognition. It is surprising therefore, that so few applications to date utilise PS, which inherently generates surface normals.

1.2. Contributions

The first contribution of this paper is a qualitative demonstration of the capabilities of our PS-based face scanner. The device, shown in Fig. 1, is suitable for practical recognition environments and consists of four illumination sources placed evenly around a high-speed camera. Subjects walk through the archway towards the camera located on the back panel and exit through the side. Compared to existing technologies, our device is cheap to build and involves exceptionally short image capture and processing times. The device is also able to operate at high resolution, is robust to ambient illumination and requires only minimal calibration. All images are captured in approximately 20ms, resulting in only very small misalignment between frames. This allows subjects to be imaged as they casually walk through the archway. We have tested our device using both visible and NIR illumination sources.

The second contribution is a thorough analysis of the accuracy of the device using both visible and NIR illumination. We found that the latter yields more accurate reconstructions when compared with ground truth. To the best of our knowledge, no published research has looked at using NIR light sources in PS for the purpose of face recognition. All ground truth experiments are based on models generated by a commercial 3dMD scanner [10] and are presented using RMS height errors and ℓ_2-norm errors.

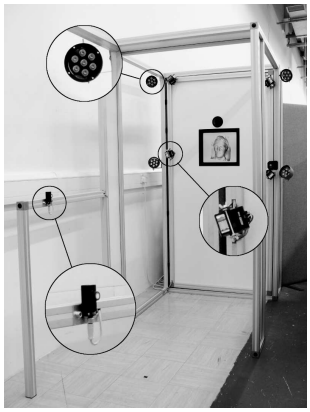


Figure 1: The geometry capture device. Enlarged areas from top to bottom: a NIR light source, a visible light source and the ultrasound proximity trigger. The camera can be seen on the back panel.

The final contribution is a quantitative analysis of the validity of the Lambertian assumption on skin reflectance. The extent of the discrepancies between the measured skin reflectance and Lambert's Law are demonstrated graphically and shown to be relatively minor for non-grazing angles. We also show that skin is more Lambertian under NIR illumination. Lastly, the reflectance analysis demonstrates the possibilities of improving the reconstructions by incorporating the Oren-Nayar reflectance model into the method.

2. Data capture

Figure 1 shows a photograph of the device that we have constructed for data acquisition. The person walks towards camera on the back panel from the left. An ultrasound proximity switch detects the presence of the individual and triggers the acquisition procedure. Four images are captured using the high speed camera on the back panel with the face illuminated by four light sources in sequence. Our rig allows either Jessops 100M visible-light flashguns to be used (colour temperature 5600K) or stripped down NIR X-vision VIS080IR lensed 7-LED clusters ($\approx$ 850nm).

It is generally expected that the face will be moving at the time of acquisition. For this reason, it is necessary to use a high speed camera to rapidly acquire the images as the light sources change before significant motion is possible. We therefore use a Basler A504kc 1280×1024 pixel camera operating at 200 fps. It was found experimentally that this was the frame rate necessary to avoid significant face movement between images. The light sources are synchronised to the camera frames using FPGA interfacing [2]. All interfacing and synchronisation is programmed in LabVIEW (although the image processing and shape estimation are performed in MATLAB). The sensor used to initiate the entire process is a highly directional Baumer ultrasound proximity switch.

Four greyscale images are captured by the camera with each corresponding to one of the four visible or NIR light sources. The regions containing the actual face are extracted from the background using the method described by Lienhart and Maydt [26]. We then estimate of the field of surface normals using a standard photometric stereo technique and assuming known light source directions [27, §5.4]. Finally, we integrate these surface normals using the well-established Frankot-Chellappa method [28] to recover the height map estimate. Figure 2 shows an example of four raw images and the resultant height estimate. For each test presented in this paper, we either use the four visible light sources or the four NIR sources.

One disadvantage of the visible light set-up is that the firing of flashguns is obvious to the subject and possibly intrusive to any surrounding people. NIR light by contrast, is more covert for a face recognition environment and subjects are less inclined to "pose" for the camera, meaning that more neutral expressions are likely. It is also worth noting the advantage that many camera sensors are inherently more sensitive to NIR light. One disadvantage of NIR illumination is the relative difficulty in obtaining the necessary brightness for the required short exposure times. While the flashguns were easily bright enough with an exposure time of 1ms, an exposure of 5ms was needed for the NIR

Figure 2: Example of four raw images and the resultant surface reconstruction.

Figure 3: Example raw images and reconstructions using visible (top) and NIR light sources for four subjects. For these experiments only, the subjects were asked to rest their chin on a support in order to ensure that all subjects are compared to each other in fair conditions.

LEDs (i.e. the maximum possible exposure for the given frame rate). Although this was adequate for our experiments, we had to use LED lenses that provided a narrow divergence angle, meaning that the face had to be more precisely positioned to obtain full illumination. For the visible light sources, the images were bright enough even for large diversion angles, removing the need for accurate positioning of apparatus and allowing subjects to pass through the archway without having to consider their exact location with respect to the camera.

To account for ambient illumination, a control image is taken after the final light source and subtracted from the other four images. The resultant images are then normalised in terms of intensity before reconstruction takes place. This is done by linearly scaling the greylevels of each image so that the mean intensity was equal for each image.

3. Reconstruction Analysis

Figure 3 shows a set of reconstructions using visible and NIR light. The general 3D structures of the faces have clearly been well estimated qualitatively. In order to quantitatively assess the accuracy of the reconstructions, we scanned eight different faces using a commercial 3dMD projected pattern range finder [10]. The 3dMD models were rescaled so that the distance between tear ducts was the same as in the corresponding PS reconstruction. All reconstructions were then cropped to 160×200px regions centred on the nose tip that encompass the eyebrows and mouth. Part of the forehead is omitted by this choice of cropping region as it is frequently occluded by hair and is therefore deemed unreliable for face recognition. An example of the face regions used for comparison can be seen in Fig. 4, which also shows a ground truth reconstruction acquired using a 3dMD scanner. The face regions from visible and NIR light sources are then aligned to ground truth using the Iterative Closest Point (ICP) algorithm [29].

Individual RMS and ℓ_2-norm error results between the reconstructions and ground truth are displayed in Fig. 5. The eight subjects consist of 6 males and 2 females and a mixture of Caucasian and Asian ethnicities. The variations in residual errors and ℓ_2-norm distances between visible and NIR reconstructions are significant according to paired t-tests ($p = 0.05$). This demonstrates that PS using NIR as a light source is a perfectly valid approach and leads to more accurate reconstructions.

In order to obtain an indication of the regions where the greatest differences occur between ground truth and PS reconstructions, the residuals and ℓ_2-norm errors at each pixel are plotted in Fig. 6. Typically, the largest variations occur in regions with the highest curvatures, such as eye sockets, nose tips and the sides of the nose.

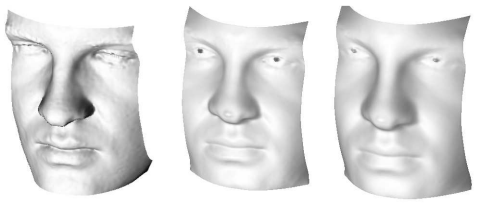


Figure 4: 3D Reconstructions for one subject from a 3dMD scanner (left) which is used as ground truth, PS using visible light sources (middle), and PS using NIR sources (right).

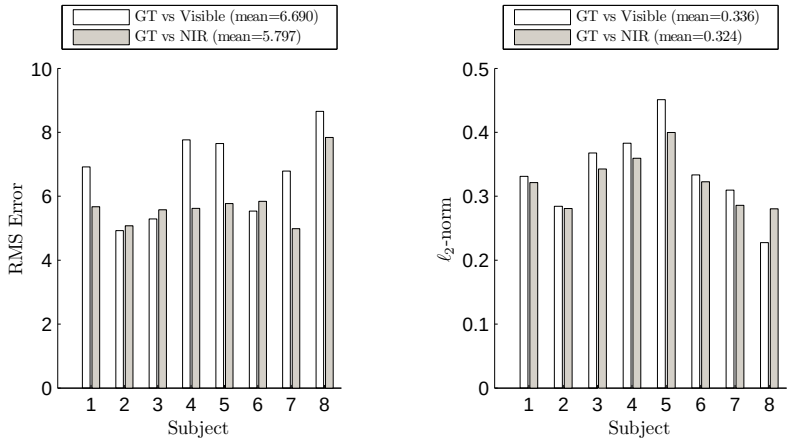


Figure 5: RMS (left) and ℓ_2-norm errors between Ground Truth (GT) and visible PS and NIR PS for each subject. The order of subjects is arbitrary.

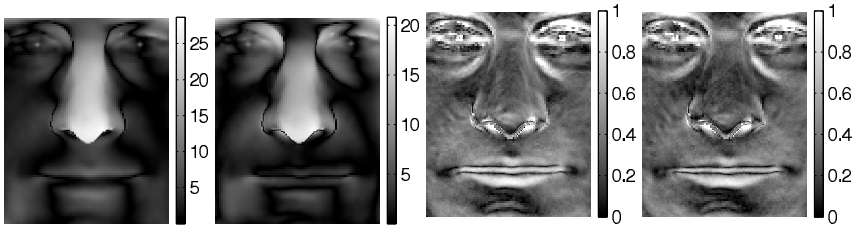


Figure 6: Representative examples of the residuals and the ℓ_2-norm errors at each pixel. Left to right: residuals for visible and NIR respectively, ℓ_2-norm errors for visible and NIR respectively. Lighter areas represent larger errors.

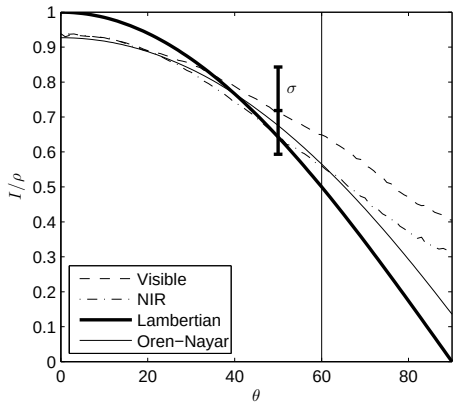


Figure 7: Mean I/ρ values averaged over 8 subjects against θ. To the right of the vertical line at $\theta = 60°$, data were too sparse to be of significance. For reference, *one* standard deviation is shown to give an indication of spread.

In attempting to produce the most accurate reconstructions possible via PS, it was found that the estimated surface normals could be enhanced by using normals acquired by re-differentiating the reconstructed height map estimate. It is unclear as to why this should be the case but preliminary analysis indicates that the reason may be due to the imposition of integrability constraints and the fitting of limited basis functions in the Fourier domain [30], as required by our adopted integration method. These factors may cause errant normals to be "smoothed out" leading to a more accurate reconstruction. However, if this method of improving reconstructions is used, a second integration step would be needed thus removing one of the benefits of PS for face recognition: that the surface normals (and hence distinctive differential surface features) are recovered directly. More research is required into this area in order to confirm that the improvements result from the imposed integrability constraints.

4. Reflectance Analysis

To determine whether Lambert's law is obeyed more strictly under NIR light than visible light, we have plotted graphs of I/ρ against the angle between the light source and the normal vector, θ. For a purely Lambertian surface, the relationship between the two should follow a cosine law. The results can be seen in Fig. 7. To generate the graph, values of I, ρ and θ were estimated for each pixel of each image for each of eight faces. The angle θ is calculated for each point of the face from the 3dMD scan data and the known light source vectors. The average values of I/ρ are used for each 1° increment in θ. The line at $\theta = 60°$ indicates a reasonable cut-off point after which data points become too sparse to be significant. The RMS difference between the measured curves and the cosine curve in the range of $0 \leq \theta \leq 60$ is 0.04 (s.d. 0.11) for NIR light and 0.06 (s.d. 0.12) for visible. For completeness, the RMS difference across the whole curve is 0.11 (s.d. =0.13) for NIR light and 0.17 (s.d. =0.12) for visible. The figure demonstrates that skin under NIR light is marginally more Lambertian than under visible light.

Although the data suffers from significant noise levels, the NIR condition has a lower RMS error and is therefore closer to the Lambertian curve than for visible light. This difference is significant given the large numbers of pixels and subjects used in the trials. This represents an average pixel intensity error of 10 greylevels for NIR and 15 for visible light across the image, assuming a maximum of 256 grey level intensities. We believe that this result is related to the fact that NIR light penetrates more deeply into the skin than visible light [31], which facilitates a more uniform scattering than surface reflection. Note however, that neither the Lambertian model nor the Oren-Nayar model (see below) take account of internal scattering or Fresnel effects. The results above demonstrate that the more Lambertian behaviour associated with NIR light also leads to more accurate reconstructions. A more detailed analysis for two individual subjects is shown in Fig. 8 and Table 1. There are small differences in the I/ρ curves caused by different light sources but this appears to have little negative impact on the reconstructions and is likely to be due to environmental effects.

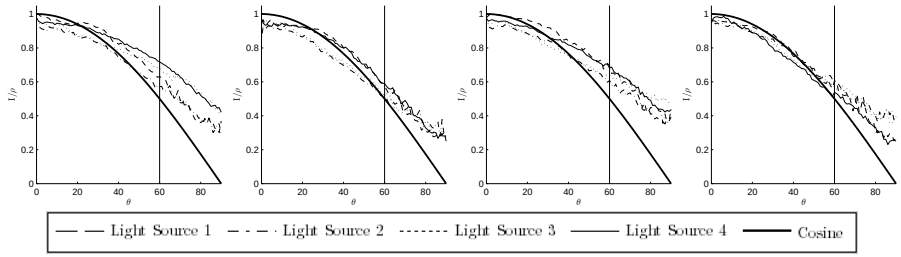


Figure 8: I/ρ values from individual light sources plotted against θ for the first two reconstructions shown in Fig. 3. Left to right: Subject 1 under visible, Subject 1 under NIR, Subject 2 under visible, Subject 2 under NIR. The light sources are labelled clockwise from the bottom-left in Fig. 1.

	Visible		NIR	
	RMS, $\theta \leq 60°$	RMS, overall	RMS, $\theta \leq 60°$	RMS, overall
All Faces	0.06, ($\sigma = 0.11$)	0.16 ($\sigma = 0.12$)	0.04 ($\sigma = 0.12$)	0.11 ($\sigma = 0.13$)
Subject 1	0.07, ($\sigma = 0.09$)	0.16 ($\sigma = 0.18$)	0.05 ($\sigma = 0.12$)	0.10 ($\sigma = 0.22$)
Subject 2	0.07, ($\sigma = 0.10$)	0.17 ($\sigma = 0.18$)	0.04 ($\sigma = 0.13$)	0.12 ($\sigma = 0.21$)

Table 1: The RMS collective error across all eight reconstructions and for the first two reconstructions shown in Fig. 3 separately. The standard deviations are shown in brackets.

Comparison to the Oren-Nayar model

We have also compared our reflection measurements to the Oren-Nayar reflectance model [32], as shown in Fig. 7. The Oren-Nayar model represents the reflecting surface as an array of V-shaped groves of random orientation, commonly called "microfacets". The distribution of microfacet orientations is characterised by a roughness parameter and each facet is assumed to act as perfect Lambertian reflector. This model is able to account for the common feature of limb-brightening and is itself based on the earlier Torrance-Sparrow model [33] where each microfacet is assumed to be mirror-like.

We have chosen to use the Oren-Nayar model as skin is not a smooth surface and it has been shown previously to be successful on a range of materials of varying degrees of roughness [32]. We do not believe that the microscopic structure of skin closely matches the Oren-Nayar model, but are merely demonstrating how alternate methods for reflection may improve our framework in future work. Investigating the various degrees of freedom of the BRDFs is also reserved for future work. Furthermore, there are additional models for skin reflectance which take account of a huge range of physical phenomena [34, 35], but these are out of the scope of this paper.

The Oren-Nayar curve in Fig. 7 represents an example intensity profile for reference with a roughness parameter of 0.2. Clearly, this model fits the measured reflectance data significantly more accurately than the Lambertian curve, suggesting that the model could be incorporated into the method in the future. This will however, add significant complexity and computation time to the algorithm. This is because a minimisation method must be implemented in order to recover all the model parameters and to accommodate the increased number of angular degrees of freedom in the model.

5. Discussion

The results presented in this paper demonstrate that PS is an effective method for producing 3D facial reconstructions for automatic recognition. Using the device with a standard PS algorithm, LabVIEW interfacing, MatLab processing and a typical modern PC, the time between device trigger and the reconstructed height map was approximately four seconds. The construction of the hardware also lends itself well to relatively unobtrusive data capture with a minimum amount of effort from the subject.

Our system offers several benefits over commonly used existing laser triangulation and projected pattern 3D shape capture devices:

1. It is significantly cheaper to construct.

2. Acquisition time is shorter than laser triangulation systems.
3. Data processing time is shorter than projected pattern systems.
4. The method is robust to typical ambient illumination conditions.
5. It is very robust against accidental collisions (because it is tolerant to errors in the light source vectors).
6. Very fine details of the face can be reconstructed.
7. Calibration is very quick and simple and only needs to be performed after the initial light source positioning.
8. Although our system cannot reconstruct hair with high levels of accuracy, it can at least provide some details of its overall shape (see Fig. 3, for example). In contrast, laser triangulation and projected pattern systems usually fail completely with hair.

At present, the 3D reconstructions are not yet as accurate as those from projected pattern range finders. They do however provide extremely fine detail of a face such as wrinkles and pores. The reconstructions under NIR were shown to be more accurate than those under visible light and diminish the need for flashing lights, making the system less intrusive compared to visible light. Zivanov et al. [36] offer an alternative argument to ours, stating that shorter wavelength light gives better results. Their justification is that shorter wavelengths undergo less internal scattering and thus provide a crisper, more defined reconstruction. It would appear therefore that a compromise must be reached in deciding between fine detail (using Zivanov's short wavelength suggestion) and overall geometry and covertness (using our NIR method).

One current limitation of the hardware described in this paper is that it does not cope with large deviations of peoples' height. Extremely tall or short people, or wheelchair bound persons would probably trigger the device correctly, but the location of the face could be outside of the field of view of the camera. Two possible solutions for this are (1) to use two cameras and trigger sensors at different heights or (2) to increase the field of view of the camera. Another improvement which could be made involves detecting the coordinates of the face and adjusting the light source vectors accordingly to improve the accuracy of the PS reconstruction. In the current system, the light source unit vectors are calculated from a point at the centre of the camera's field of view and this is used for all reconstructions regardless of where the face is actually located. For this reason, the light source unit vectors are less accurate if the person walking through the device does not locate their face near the centre of the camera's field of view. The exact error caused by this inaccuracy is unknown, but amending the light source angles on a per person basis will improve the surface normal estimates.

6. References

[1] K. W. Bowyer, K. Chang, P. Flynn, A survey of approaches and challenges in 3D and multi-modal 3D+ 2D face recognition, Comput. Vis. Image Understand. 101 (1) (2006) 1–15.
[2] G. A. Atkinson, A. R. Farooq, M. L. Smith, L. N. Smith, Facial reconstruction and alignment using photometric stereo and surface fitting, in: Proc. IbPRIA, 2009, pp. 88–95.
[3] W. Zhao, R. Chellappa, P. J. Phillips, A. Rosenfeld, Face recognition: A literature survey, ACM Comput. Surv. (2003) 399–458.
[4] B. Gökberk, M. O. İrfanoğlu, L. Akarun, 3D shape-based face representation and feature extraction for face recognition, Image and Vision Comput. 24 (8) (2006) 857–869.
[5] V. Blanz, T. Vetter, Face recognition based on fitting a 3D morphable model, IEEE Trans. Patt. Anal. Mach. Intell. (2003) 1063–1074.
[6] A. M. Bronstein, M. M. Bronstein, R. Kimmel, Three-Dimensional face recognition, Intl. J. Comp. Vis. (2005) 5–30.
[7] I. Mpiperis, S. Malassiotis, M. G. Strintzis, 3-D face recognition with the geodesic polar representation, IEEE Trans. Information Forensics and Security (2007) 537–547.
[8] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, Overview of the face recognition grand challenge, in: Proc. CVPR, Vol. 1, 2005.
[9] www.konicaminolta.com/sensingusa/products/3d/non-contact/vivid910 [9 June 2010].
[10] www.3dmd.com/3dmdface.html [9 June 2010].
[11] B. K. P. Horn, Shape from shading: A method for obtaining the shape of a smooth opaque object from one view, Ph.D. thesis, MIT (1970).
[12] P. N. Belhumeur, D. J. Kriegman, A. L. Yuille, The bas-relief ambiguity, Intl. J. Comp. Vis. 35 (1999) 33–44.
[13] W. Smith, E. Hancock, Facial shape-from-shading and recognition using principal geodesic analysis and robust statistics, Intl. J. Comp. Vis. 76 (2008) 71–91.
[14] R. J. Woodham, Photometric method for determining surface orientation from multiple images, Opt. Eng. 19 (1) (1980) 139–144.
[15] E. N. Coleman, R. Jain, Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry, Comput. Vision and Image Proc. (1982) 309–328.
[16] F. Solomon, K. Ikeuchi, Extracting the shape and roughness of specular lobe objects using four light photometric stereo, IEEE Trans. Patt. Anal. Mach. Intell. 18 (1996) 449–454.

[17] S. Barsky, M. Petrou, The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows, IEEE Trans. Patt. Anal. Mach. Intell. 25 (2003) 1239–1252.
[18] A. S. Georghiades, Recovering 3-D shape and reflectance from a small number of photographs, Leuven, Belgium, 2003, pp. 230–240.
[19] J. Sun, M. Smith, L. Smith, S. Midha, J. Bamber, Object surface recovery using a multi-light photometric stereo technique for non-Lambertian surfaces subject to shadows and specularities, Image Vision Comput. 25 (2007) 1050–1057.
[20] C. Hernández, G. Vogiatzis, R. Cipolla, Shadows in three-source photometric stereo, in: Proc. ECCV, 2008, pp. 290–303.
[21] W. Zhao, R. Chellappa, Face processing: advanced modeling and methods, Academic Press, 2006.
[22] S. C. Kee, K. M. Lee, S. U. Lee, Illumination invariant face recognition using photometric stereo, IEICE T. Inf. Syst. E Ser. D 83 (7) (2000) 1466–1474.
[23] S. K. Zhou, R. Chellappa, D. W. Jacobs, Characterization of human faces under illumination variations using rank, integrability, and symmetry constraints, Proc. ECCV (2004) 588–601.
[24] S. K. Zhou, G. Aggarwal, R. Chellappa, D. W. Jacobs, Appearance characterization of linear lambertian objects, generalized photometric stereo, and illumination-invariant face recognition, IEEE Trans. Patt. Anal. Mach. Intell. 29 (2) (2007) 230–245.
[25] A. S. Georghiades, P. N. Belhumeur, D. J. Kriegman, From few to many: Illumination cone models for face recognition under variable lighting and pose, IEEE Trans. Patt. Anal. Mach. Intell. (2001) 643–660.
[26] R. Lienhart, J. Maydt, An extended set of Haar-like features for rapid object detection, in: IEEE ICIP, 2002, pp. 900–903.
[27] D. A. Forsyth, J. Ponce, Computer Vision, A Modern Approach, Prentice Hall, Upper Saddle River, NJ, 2003.
[28] R. T. Frankot, R. Chellappa, A method for enforcing integrability in shape from shading algorithms, IEEE Trans. Patt. Anal. Mach. Intell. 10 (1988) 439–451.
[29] P. J. Besl, H. D. McKay, A method for registration of 3-D shapes, IEEE Trans. Patt. Anal. Mach. Intell. 14 (2) (1992) 239–256.
[30] R. T. Frankot, R. Chellappa, A method for enforcing integrability in shape from shading algorithms, IEEE Trans. Patt. Anal. Mach. Intell. 10 (4) (1988) 439–451.
[31] C. Fredembach, N. Barbuscia, S. Süsstrunk, Combining visible and near-infrared images for realistic skin smoothing, in: In Proc. IS&T/SID Color Imaging Conference, 2009.
[32] M. Oren, S. K. Nayar, Generalization of the Lambertian model and implications for machine vision, Intl. J. Comp. Vis. 14 (1995) 227–251.
[33] K. Torrance, M. Sparrow, Theory for off-specular reflection from roughened surfaces, J. Opt. Soc. Am. 57 (1967) 1105–1114.
[34] C. Donner, H. W. Jensen, Light diffusion in multi-layered translucent materials, ACM Trans. Graphics 24 (2005) 1032–1039.
[35] L. Li, C. S. Ng, Rendering human skin using a multi-layer reflection model, Int'l J. Mathematics and Computers in Simulation 3 (2009) 44–53.
[36] J. Zivanov, P. Paysan, T. Vetter, Facial normal map capture using four lights - an effective and inexpensive method of capturing the fine scale detail of human faces using four point lights, in: GRAPP, 2009, pp. 13–20.

Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 2 (2010) 20–25
Procedia Computer Science
www.elsevier.com/locate/procedia

ICEBT 2010

Baseline face recognition using photometric stereo data

Stefanos Zafeiriou[a], Mark F. Hansen[b], Gary A. Atkinson[b], Maria Petrou[a], Melvyn L. Smith[b]

[a]*Communications and Signal Processing Group, Imperial College London. SW7 2AZ.*
[b]*Machine Vision Laboratory, University of the West of England, Bristol BS16 1QY*

Abstract

This research is motivated by the need for face recognition in uncontrolled environments. In other words, we are interested in face recognition arrangements whereby the users do not need to interact with the recognition technology. The contribution of this paper is to perform a range of recognition experiments on face image data as people casually enter a building, without any instructions about expression. Specifically, we capture four images per session in rapid succession (all within 20ms). The four images are synchronised to different light sources to enable photometric stereo processing to estimate albedo images, surface normals and depth maps. Additional capture sessions then take place over periods of many weeks. Our recognition experiments are on each of the three modalities as well as a fusion technique for the albedo and depth. Using a variety of photometric stereo methods, surface integration methods (to recover depth) and recognition algorithms such as principal component analysis and nonnegative matrix factorisation, we acquire a maximum recognition rate of 86% for 96 subjects.

Keywords: 3D face recognition, photometric stereo

1. Introduction

In recent years, face recognition [1] has undergone two major developments. First, a few basic systems have become commercialised. These are typically 2D systems that require close co-operation with the users and assume fixed pose, expression and lighting. Second, and partly in an attempt to relax these assumptions, researchers have began incorporating 3D information into the recognition process [2]. This paper aims to assist in both of these developments by presenting 3D face recognition experiments in uncontrolled conditions.

The paper applies a range of existing face recognition algorithms to data captured using a photometric stereo (PS) [3] device placed at the entrance of a busy workplace. The device that we adopted [4] captures four images of each subject per session and reconstructs the 3D geometry of their face using several different algorithms. The device works by detecting the presence of an individual with an ultrasound switch and using this to trigger a high speed (200fps) camera to acquire the raw images. The camera is hardware connected to four different flashgun light sources, which are positioned around an archway, through which the subjects pass as they enter their offices. The physical location of the device meant that we were able to test the algorithms on more challenging situations than the current commercial systems allow. This is because the data for the experiments were captured as subjects "casually" pass through the archway rather than having to make a specific effort to co-operate with the system. This arrangement accurately

Email address: gary.atkinson@uwe.ac.uk (Gary A. Atkinson)

1877-0509
doi:10.1016/j.procs.2010.11.004

simulates one of the ultimate goals for access-control face recognition, where there is no interaction required between the subjects and the technology.

A comparative study of 3D face recognition/verification methods was recently published in [5], where the authors implemented a large inventory of 3D recognition methods and tested them on various representations of facial geometry (i.e. depth images, curves and normal fields). Moreover, the authors applied various fusion strategies to the results of 3D face recognition methods. A recent study on the fusion of information of intensity and depth images was presented in [6]. In that paper the authors demonstrated that: 1) the recognition results of intensity and depth images are approximately the same, 2) fusion of intensity and depth information produce considerably better results than using only one modality (i.e, only intensity images or only depth maps).

The experiments in this paper are based on each of the albedo images (2D), the recovered 3D depth and normal images [2],[5]. In summary, the contributions of this paper are 1) it demonstrates how different methods in the pipeline of PS affect the recognition rate, and 2) it verifies that a similar conclusion to [6] can be drawn for the modalities derived from PS methods. We applied three different PS methods in order to compute the normal field and the albedo image and five different integration methods that compute the height map from the normal field.

2. Photometric Stereo and Surface Reconstruction

All of our experiments are performed on PS data acquired using the capture device described in [4]. This section of the paper summarises the standard PS method [7, §5.4], which we implemented using both three and four sources. We used an implementation of the PS method of Woodham [3] and have mainly concentrated on a four-source version of the technique, although we have also compared our results with those of methods using three sources.

Standard PS involves imaging an object using three or more light source directions and assumes that the surface reflectance follows Lambert's law. The following matrix equation is then constructed from the captured greyscale images using Lambert's law for each pixel $\mathbf{x} = [x, y]$:

$$\begin{bmatrix} I_1(\mathbf{x}) \\ I_2(\mathbf{x}) \\ \vdots \\ I_N(\mathbf{x}) \end{bmatrix} = \rho(\mathbf{x}) \begin{bmatrix} \mathbf{L}_1^T \\ \mathbf{L}_2^T \\ \vdots \\ \mathbf{L}_N^T \end{bmatrix} \mathbf{n}(\mathbf{x}) \tag{1}$$

where $I_i(\mathbf{x})$ is the ith measured pixel brightness, $\mathbf{L}_i$ is the ith light source vector, N is the number of light sources, $\rho(\mathbf{x})$ is the reflectance albedo and $\mathbf{n}(\mathbf{x})$ is the unit surface normal. The light source positions are calculated once by measuring the highlights on a specular sphere. The albedos and surface normal components can then be calculated by solving (1) using these known light source vectors and the measured intensity values. In addition to this basic PS method, we also applied the PS method proposed in [8] that aims to address the negative impacts of shadows and specularities.

We next describe the problem of reconstructing a surface $(\mathbf{x}, f(\mathbf{x}))$ from the surface normals estimated from PS. The normals can be expressed in terms of surface gradients using

$$\tilde{\mathbf{n}}(\mathbf{x}) = \frac{1}{\sqrt{1 + \left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}} \left(-\frac{\partial f}{\partial x}, -\frac{\partial f}{\partial y}, 1\right)^T \tag{2}$$

Let the computed value of the unit normal at some point $\mathbf{x}$ be $\mathbf{n}(\mathbf{x}) = [a(\mathbf{x}), b(\mathbf{x}), c(\mathbf{x})]$, as calculated by (1). We can then say that

$$\frac{\partial f}{\partial x} = \frac{a(\mathbf{x})}{c(\mathbf{x})} \quad \frac{\partial f}{\partial y} = \frac{b(\mathbf{x})}{c(\mathbf{x})}. \tag{3}$$

Here, we also perform another check on the data set. Let the images $\mathbf{P}(\mathbf{x}) = [\frac{a(\mathbf{x})}{c(\mathbf{x})}]$ and $\mathbf{Q}(\mathbf{x}) = [\frac{b(\mathbf{x})}{c(\mathbf{x})}]$. If the surface is integrable, we can say that

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x} \tag{4}$$

and, therefore

$$A(\mathbf{x}) \equiv \frac{\partial(\mathbf{P}(\mathbf{x}))}{\partial y} - \frac{\partial(\mathbf{Q}(\mathbf{x}))}{\partial x} \tag{5}$$

must be close to zero at each point $\mathbf{x}$.

Assuming that the partial derivatives satisfy condition (5), we can reconstruct the surface up to some constant error in depth. The partial derivatives give the change in surface height with a small step in either the x or the y direction. This means that we can get the surface by summing these changes in height along some path. In particular, we have

$$f(\mathbf{x}) = \oint_C \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) \cdot \mathrm{d}\mathbf{l} + c \tag{6}$$

where C is a curve starting at some fixed point and ending at $\mathbf{x}$, $\mathrm{d}\mathbf{l}$ is the infinitesimal element along the curve and c is a constant of integration, which represents the unknown height of the surface at the starting point. Choosing the correct integration path and enforcing condition (4) is not a trivial task. In our experiments we applied the surface reconstructions proposed in [9, 10, 11, 12, 13].

3. Face recognition using Albedo and Depth Images

Using the methods outlined in the previous section, we have at our disposal a range of modalities in which to apply recognition algorithms: raw images, albedo images, surface normals and depth maps. In this section we outline a range of methods used for feature extraction from albedo and depth images for face recognition. We have chosen to apply a family of methods that aim at extracting features using linear projections (also referred to as subspace methods). This family includes, for example, Principal Component Analysis (PCA) and Nonnegative Matrix Factorization (NMF). In our experiments, NMF [14] produced the best recognition result. In subspace methods such as NMF, the facial images are lexicographically scanned in order to form feature vectors.

Let m be the number of samples in the image database $\mathcal{U} = \{\mathbf{u}_1, \mathbf{u}_2, .., \mathbf{u}_m\}$ where $\mathbf{u}_i \in \Re^n$ is a database image. A linear transformation of the original n-dimensional space onto a subspace with m dimensions ($m << n$) is a matrix $\mathbf{W}^T \in \Re^{m \times n}$. The new feature vectors $\mathbf{y}_k \in \Re^m$ are given by:

$$\mathbf{y}_k = \mathbf{W}^T(\mathbf{u}_k - \bar{\mathbf{u}}), \quad k \in \{1, 2, \ldots, m\} \tag{7}$$

where $\bar{\mathbf{u}} \in \Re^n$ is the mean image of all samples. Classification is performed using a simple distance measure and a nearest neighbour classifier using the normalized correlation.

In this paper we also adopt a specific method for recognition that uses the orientation of the normals, as this is particularly well suited to our available data. The method is simple to implement and is based on a novel representation of faces, the so-called NormalFaces. Using $\mathbf{P}$ and $\mathbf{Q}$ calculated using PS from image $\mathbf{u}$ we proceed to compute

$$\mathbf{\Phi}(\mathbf{x}) = \operatorname{atan}\frac{\mathbf{Q}(\mathbf{x})}{\mathbf{P}(\mathbf{x})} \tag{8}$$

which is an image that contains the normal orientations. The orientations are measured in the interval $[-\frac{\pi}{2}, \frac{\pi}{2}]$. For two normal images $\mathbf{\Phi}_1(\mathbf{x})$ and $\mathbf{\Phi}_2(\mathbf{x})$ we then use the following dissimilarity measure:

$$d(\mathbf{\Phi}_1(\mathbf{x}), \mathbf{\Phi}_2(\mathbf{x})) = 1 - \frac{1}{mn\pi} \sum_{i=1}^{mn} |\mathbf{\Phi}_1(\mathbf{x}_i) - \mathbf{\Phi}_2(\mathbf{x}_i)|. \tag{9}$$

This measure is then used to extract features using pseudo-Euclidean embedding as described in [15] and classification is performed using the normalized correlation in the new space.

4. Recognition Experiments

Our experiments are performed on subjects' face images that have been captured with more than a week's interval for 126 people. For the majority of images (90 people in total) the interval was greater than one month. The database contains 96 persons with three or more images per person. For the experiments presented here, we tested using three scenarios: (a) using a single training image, (b) using two training images, and (c) using a multi-modal approach. For the second scenario, the testing image for all 96 subjects was the same one used for the first scenario. This realization was implemented in order to test whether or not recognition using two samples of the same modality is better than fusing information across different modalities. The first of these scenarios is a particularly challenging experimental procedure, especially since most of the training and testing images display a different facial expression.

4.1. Face recognition from Albedo Images

Four source, three source and ray trace-based PS methods were employed for albedo computation. These methods are abbreviated as 4L-PS, 3L-PS and RAY-PS, respectively. The recognition rates using one albedo image for training and one for testing for all the tested PS methods are shown in Figure 1(a). As it can be seen, the recognition rate is affected by the PS method applied and noticeably better recognition performance is achieved by PS methods that use all four illuminants. The best recognition rate was 78%.

For the case of the two sample experiment we used a decision fusion strategy similar to [16]. That is, we combined the match scores for each person across the two samples of 2D albedo images and ranked the subjects based on the combined scores. Scores from each modality are linearly normalized to the range of $[0, 100]$ before combining. We explored various confidence-weighted versions of the sum, product and minimum rules. Among the fusion rules that we tested, the sum rule provided the best performance overall. The recognition rates for the two sample experiments are shown in Figure 1(b). As it can be seen, the use of more than two samples increases the recognition performance. Moreover, the methods which use all four illuminants achieved better recognition rates than those using only three. The best recognition rate was 85%.

4.2. Face recognition from Depth and NormalFace Images

We applied five different methods for surface reconstruction from the normal field. For the reconstruction methods we use the following abbreviations: 1) 'at' for the method in [12], which enforces condition (4) with a linear system of equations over the image, 2) dctFC for the DCT Frankot-Chellappa method [11], 3) FC for the original Frank-Chellappa method [9], 4) 'ls' for the least square solution of the Poison equation [17] and 5) 'me' for the reconstruction based on M-estimator [13]. The recognition rates for the one sample experiment and for all reconstruction and PS methods are plotted in Figure 1(c). The best recognition result obtained was 74%. As it can be seen, PS and reconstruction methods greatly affect the recognition performance. More precisely, four source PS methods always achieve better recognition results. Moreover, the depth maps that were produced by dctFC constantly outperformed the performance of the depth maps produced by all other reconstruction methods.

Experiments using two samples for training and one sample for testing were conducted in a similar manner. The results of these experiments are shown in Figure 1(d). The best recognition result was 86%. Finally, the results of the experiments using NormalFace for all tested PS methods are shown in Figures 1(e) and 1(f) for one sample and two sample recognition, respectively.

4.3. Fusion 2D and 3D

Multimodal decision fusion is performed by combining the match scores for each person across the modalities of 2D albedo and depth image and ranking the subjects based on the combined scores in a similar manner as in the two sample experiments. The sum rule provided the best performance. We performed fusion only on depth images derived from the DCT-FC method. Fusion of intensity and geometry information was conducted only on the subset of persons that have more than 2 samples available in order to be directly comparable with the single modality two sample experiments. The recognition results from multimodal fusion using various PS methods are summarized in Figure 2. The best recognition result was 85%. A summary of the best recognition results for the various modalities and multimodal fusion is given in Table 1.

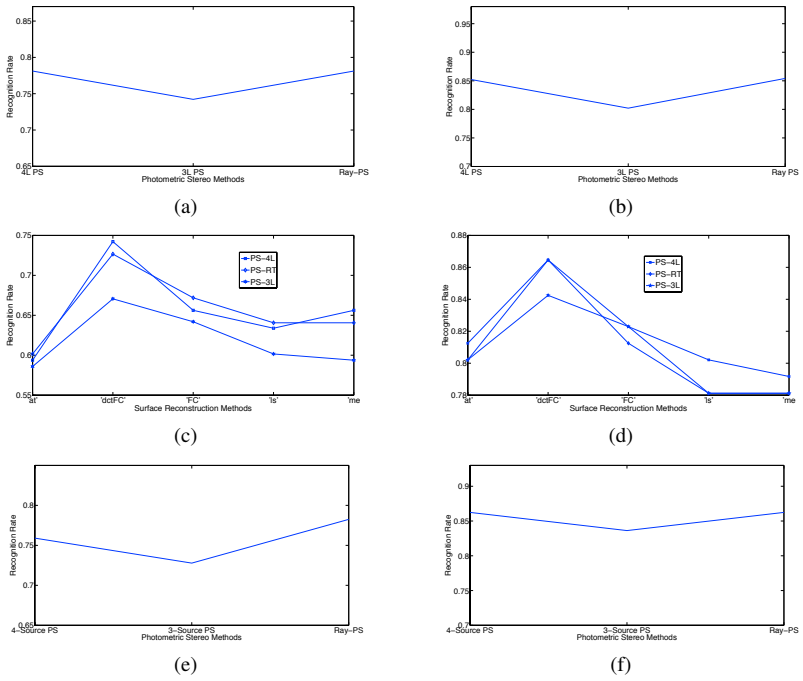


Figure 1: Experiments using, (a) one albedo image for training and one for testing; (b) two albedo images for training and one for testing; (c) one depth image for training and one for testing; (d) two depth images for training and one for testing; (e) one NormalFace for training and one for testing; (f) two NormalFaces for training and one for testing

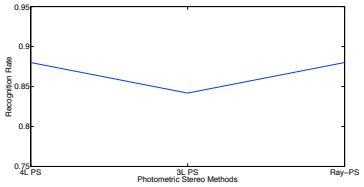


Figure 2: Multimodal fusion results for recognition.

5. Conclusions

In this paper, we presented face recognition results based on PS data captured in a real life commercial setting. We presented experiments which demonstrate how different methods in the pipeline of PS affect the recognition performance and concluded that four-source PS methods produce facial samples (albedo, normals) that achieve constantly high performance regardless of the reconstruction method applied. However, we also showed that the reconstruction methods greatly affect the recognition rates.

This paper also verified most of the findings of [16]. Specifically, in most cases, the best recognition results

Table 1: A summary of the best percentage of recognition for all the conducted experiments across different modalities.

One Sample			Two Samples			Modality Fusion
Albedo	Depth	Normal	Albedo	Depth	Normal	Albedo + Depth
78	74	78	85	86	86	85

of recovered albedo, normals and the reconstructed depth maps achieve approximately the same results, while in some cases the recovered albedo produces better results. The fusion of albedo and reconstructed surfaces produce significantly better results than using only the albedo or the depth images. Fusion of two albedo images in the same way that we fused the results of albedo and depth map gave approximately the same recognition results. The best recognition rate that we obtained was approximately 86%.

6. References

[1] W. Zhao, R. Chellappa (eds), Face Processing: Advanced Modeling and Methods, Elsevier, 2006.
[2] K. W. Bowyer, K. Chang, P. Flynn, A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition, Comp. Vis. Im. Understanding 101 (2006) 1–15.
[3] R. J. Woodham, Photometric method for determining surface orientation from multiple images, Optical Engineering 19 (1980) 139–144.
[4] M. F. Hansen, G. A. Atkinson, L. N. Smith, M. L. Smith, 3D face reconstructions from photometric stereo using near infrared and visible light, Comp. Vis. Im. Understanding to appear.
[5] B. Gokberk, H. Dutagaci, A. Ulas, L. Akarun, B. Sankur, Representation plurality and fusion for 3-D face recognition, IEEE Transactions on Systems Man and Cybernetics-Part B-Cybernetics 38 (1) (2008) 155–173.
[6] K. Chang, K. Bowyer, P. Flynn, An evaluation of multimodal 2D+ 3D face biometrics, IEEE Transactions on Pattern Analysis and Machine Intelligence (2005) 619–624.
[7] D. A. Forsyth, J. Ponce, Computer Vision, A Modern Approach, Prentice Hall, Upper Saddle River, NJ, 2003.
[8] V. Argyriou, M. Petrou, Recursive photometric stereo when multiple shadows and highlights are present, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1–6.
[9] R. T. Frankot, R. Chellappa, A method for enforcing integrability in shape from shading algorithms, IEEE Trans. Patt. Anal. Mach. Intell. 10 (1988) 439–451.
[10] T. Simchony, R. Chellappa, M. Shao, Direct analytical methods for solving Poisson equations in computer vision problems, IEEE Transactions on Pattern Analysis and Machine Intelligence 12 (5) (1990) 435–446.
[11] A. S. Georghiades, P. N. Belhumeur, D. J. Kriegman, From few to many: illumination cone models for face recognition under variable lighting and pose, IEEE Trans. Patt. Anal. Mach. Intell. 23 (2001) 643–660.
[12] A. Agrawal, R. Chellappa, R. Raskar, An algebraic approach to surface reconstruction from gradient fields, in: Proc. ICCV, 2005, pp. 174–181.
[13] A. Agrawal, R. Raskar, R. Chellappa, What is the range of surface reconstructions from a gradient field?, Lecture Notes in Computer Science (European Conference on Computer Vision (ECCV 2006)) 3951 (2006) 578.
[14] D. Lee, H. Seung, Algorithms for non-negative matrix factorization, in: NIPS, 2000, pp. 556–562.
[15] E. Pekalska, P. Paclik, R. Duin, A generalized kernel approach to dissimilarity-based classification, The Journal of Machine Learning Research 2 (2002) 175–211.
[16] K. Chang, K. Bowyer, P. Flynn, An evaluation of multimodal 2D+ 3D face biometrics, IEEE Transactions on Pattern Analysis and Machine Intelligence (2005) 619–624.
[17] T. Simchony, R. Chellappa, M. Shao, Direct analytical methods for solving Poisson equations in computer vision problems, IEEE Trans. Patt. Anal. Mach. Intell. 12 (1990) 435–446.

Appendix C

Photoface - 61 subjects, 1000 Sessions

These are the sessions used throughout the thesis when 1000 sessions are referred to. An electronic copy is available from the author.

1001\2008-03-01_07-58-49	1001\2008-03-02_11-25-38	1001\2008-03-20_15-10-39
1001\2008-04-05_11-22-53	1001\2008-04-19_08-34-45	1001\2008-05-14_11-01-27
1001\2009-06-26_14-48-20	1001\2009-07-09_13-25-40	1001\2009-07-10_15-03-01
1001\2009-07-28_15-19-04	1002\2008-02-25_17-34-06	1002\2008-03-28_14-11-04
1002\2008-04-01_18-52-11	1002\2008-04-01_18-54-30	1002\2009-07-08_08-18-43
1002\2009-07-30_14-57-27	1002\2009-08-17_16-00-09	1002\2009-08-26_10-33-57
1002\2009-08-27_08-20-34	1002\2009-08-27_16-41-09	1003\2008-02-18_17-40-16
1003\2008-02-18_17-40-29	1003\2008-02-22_16-42-59	1003\2008-02-22_16-43-26
1003\2008-02-25_17-24-07	1003\2008-03-04_17-36-56	1003\2008-03-07_11-25-56
1003\2008-03-12_17-50-55	1003\2008-03-31_17-11-17	1003\2008-04-02_17-27-25
1003\2008-04-22_16-39-32	1003\2008-05-20_17-51-27	1003\2008-05-28_17-29-57
1003\2008-06-10_17-22-02	1003\2008-06-19_17-05-00	1003\2009-07-08_17-16-59
1003\2009-07-24_16-05-35	1003\2009-10-07_15-44-24	1003\2009-10-23_15-36-47
1004\2008-02-21_11-07-38	1004\2008-02-21_11-08-02	1004\2008-02-21_11-39-28
1004\2008-02-21_13-38-47	1004\2008-02-21_14-07-06	1004\2008-02-21_17-12-02
1004\2008-02-21_18-11-37	1004\2008-02-25_17-33-16	1004\2008-03-05_12-23-16
1004\2008-03-05_12-45-05	1007\2008-03-28_15-28-02	1007\2008-03-28_15-28-34
1007\2008-04-21_11-28-50	1007\2008-05-28_10-11-10	1007\2009-02-26_10-47-19
1007\2009-06-25_13-56-34	1007\2009-07-14_14-35-14	1007\2009-07-17_10-13-25
1007\2009-07-20_14-17-31	1007\2009-07-22_15-06-41	1007\2009-07-28_13-55-55

1007\2009-07-31_11-24-16
1007\2009-08-13_11-26-19
1007\2009-08-13_11-28-37
1007\2009-08-13_11-29-47
1007\2009-08-13_11-56-03
1007\2009-08-17_09-35-28
1007\2009-08-17_13-40-48
1007\2009-08-18_16-19-44
1007\2009-08-19_14-14-32
1007\2009-08-20_07-43-54
1007\2009-08-20_17-11-11
1007\2009-08-21_17-36-03
1007\2009-08-24_15-44-30
1007\2009-08-26_14-57-02
1007\2009-08-27_14-51-34
1007\2009-09-01_10-20-10
1007\2009-09-04_09-33-18
1007\2009-09-08_08-43-25
1007\2009-09-08_08-44-52
1007\2009-09-08_13-12-31
1007\2009-09-14_08-49-24
1007\2009-09-14_16-58-11
1007\2009-09-18_16-28-36
1007\2009-09-28_13-24-48
1007\2009-10-02_16-07-57
1007\2009-10-06_17-08-32
1007\2009-10-08_07-02-38
1007\2009-10-15_17-44-04
1007\2009-10-22_15-30-01
1008\2008-02-18_13-42-16
1008\2008-02-22_13-53-10
1008\2008-04-14_08-30-11
1008\2008-04-15_11-17-30
1008\2008-04-16_16-46-12
1008\2008-04-18_13-43-00
1008\2008-05-29_08-15-48
1008\2008-06-18_16-48-33
1008\2009-06-26_15-45-00
1008\2009-07-02_09-52-00
1008\2009-07-09_17-13-23
1008\2009-07-13_12-32-38
1008\2009-07-20_13-58-12
1008\2009-07-23_08-09-50
1008\2009-07-24_11-17-46
1008\2009-08-18_12-35-35
1008\2009-08-24_08-24-10
1008\2009-09-01_12-39-24
1008\2009-09-01_12-40-13
1008\2009-09-03_16-24-10
1008\2009-09-14_10-58-44
1008\2009-09-17_16-43-43
1008\2009-09-23_08-07-50
1008\2009-09-25_08-27-19
1008\2009-10-09_08-27-55
1008\2009-10-12_08-28-42
1009\2008-02-18_08-50-39
1009\2008-02-18_11-10-59
1009\2008-02-19_08-54-56
1009\2008-02-21_10-13-44
1009\2008-02-29_08-39-56
1009\2008-02-29_13-49-12
1009\2008-03-12_12-02-37
1009\2008-03-13_14-44-23
1009\2008-03-17_14-15-57
1009\2008-04-22_09-06-04
1009\2009-06-26_14-20-48
1009\2009-06-30_14-55-41
1009\2009-07-07_14-29-51
1009\2009-07-08_14-19-18
1009\2009-07-09_09-26-58
1009\2009-07-13_12-39-01
1009\2009-08-19_14-05-50
1009\2009-09-11_10-59-32
1009\2009-09-23_08-34-34
1009\2009-09-30_08-30-21
1009\2009-10-23_09-07-09
1010\2009-07-08_20-24-55
1010\2009-07-23_18-22-42
1010\2009-07-23_18-56-20
1010\2009-08-04_19-05-32
1010\2009-08-20_19-39-40
1010\2009-09-11_17-30-34
1010\2009-09-29_18-53-13
1010\2009-10-24_14-55-27
1012\2008-02-18_06-46-13
1012\2008-02-19_17-24-46
1012\2008-02-19_17-25-08
1012\2008-02-22_07-39-59
1012\2008-02-22_07-40-08
1012\2008-02-22_07-40-19
1012\2008-02-22_17-02-36
1012\2008-02-22_17-02-47
1012\2008-02-22_17-02-58
1012\2008-02-22_17-03-10
1012\2008-02-29_06-27-29
1012\2008-02-29_17-41-40
1012\2008-03-03_07-51-58
1012\2008-03-04_10-30-22
1012\2008-03-06_16-13-55
1012\2008-03-08_11-10-42
1012\2008-03-08_15-43-08
1012\2008-03-08_15-43-24
1012\2008-03-08_15-43-40
1012\2008-03-09_08-18-42
1012\2008-03-09_14-26-14
1012\2008-03-27_10-23-04
1012\2008-04-04_20-13-14
1012\2008-04-04_20-13-26
1012\2008-04-04_20-13-37
1012\2008-04-10_11-34-07
1012\2008-04-23_16-45-59
1012\2008-04-23_16-46-14
1012\2008-04-23_16-46-28
1012\2008-05-10_09-38-18

1012\2008-05-11_08-19-02
1012\2008-05-11_15-40-22
1012\2008-05-12_15-35-46
1012\2008-05-12_19-19-17
1012\2008-05-13_15-19-52
1012\2008-05-14_15-18-40
1012\2008-05-14_16-44-37
1012\2008-05-14_17-07-25
1012\2008-05-14_20-48-04
1012\2008-05-15_16-47-47
1012\2008-05-16_16-52-30
1012\2008-05-16_19-36-47
1012\2008-05-19_07-02-58
1012\2008-05-22_12-07-22
1012\2008-05-22_13-59-58
1012\2008-05-29_21-16-41
1012\2008-06-02_16-00-56
1012\2008-06-03_07-00-06
1012\2008-06-06_11-49-50
1012\2008-06-12_07-49-50
1012\2008-06-12_13-26-44
1012\2008-06-18_12-34-12
1012\2008-06-20_07-19-55
1012\2008-06-25_17-46-20
1012\2008-06-27_17-43-58
1012\2008-06-29_12-03-09
1012\2009-06-25_16-36-34
1012\2009-06-26_18-20-06
1012\2009-06-30_07-31-26
1012\2009-07-08_21-52-08
1012\2009-07-14_07-04-13
1012\2009-07-22_14-43-01
1012\2009-07-23_18-48-50
1012\2009-07-23_18-59-12
1012\2009-08-06_16-01-34
1012\2009-08-13_07-02-34
1012\2009-09-09_13-16-13
1012\2009-09-10_19-07-09
1012\2009-10-01_20-18-46
1012\2009-10-14_20-57-47
1013\2008-02-25_14-50-46
1013\2008-02-29_09-11-20
1013\2008-02-29_10-46-05
1013\2008-03-03_09-29-39
1013\2008-03-03_15-30-23
1013\2008-03-10_08-46-05
1013\2008-03-11_07-57-36
1013\2008-03-13_09-04-05
1013\2008-03-18_09-59-14
1013\2008-03-27_15-22-28
1013\2008-03-28_10-43-50
1013\2008-05-12_15-03-32
1013\2008-05-14_09-51-58
1013\2008-05-19_08-28-00
1013\2008-05-19_11-26-44
1013\2008-05-20_15-29-42
1013\2008-05-28_08-27-27
1013\2008-05-28_11-41-51
1013\2008-05-28_12-07-58
1013\2008-05-29_08-07-48
1013\2008-05-29_09-06-51
1013\2008-05-29_10-07-55
1013\2008-06-02_11-06-49
1013\2008-06-02_14-56-34
1013\2008-06-03_08-33-09
1013\2008-06-04_07-51-58
1013\2008-06-04_16-37-07
1013\2008-06-05_08-02-09
1013\2008-06-06_13-22-02
1013\2008-06-09_09-55-04
1013\2008-06-09_14-53-39
1013\2008-06-10_16-19-40
1013\2008-06-11_09-12-11
1013\2008-06-17_09-09-14
1013\2008-06-18_10-45-04
1013\2008-06-19_08-15-24
1013\2008-06-20_08-25-49
1013\2008-06-23_08-17-21
1013\2008-06-26_10-13-23
1013\2009-07-01_09-23-18
1013\2009-07-07_12-27-27
1013\2009-07-09_10-22-07
1013\2009-07-17_12-27-06
1013\2009-07-21_14-54-55
1013\2009-07-22_14-57-41
1013\2009-07-28_15-31-10
1013\2009-08-03_10-54-34
1013\2009-09-03_10-18-53
1013\2009-09-28_10-25-38
1013\2009-10-15_07-02-32
1013\2009-10-22_10-35-12
1014\2008-02-19_09-05-27
1014\2008-02-19_16-45-02
1014\2008-02-20_17-14-18
1014\2008-02-21_16-56-08
1014\2008-02-22_17-33-58
1014\2008-02-25_12-36-10
1014\2008-02-28_16-08-11
1014\2008-02-29_16-48-59
1014\2008-03-03_11-20-42
1014\2008-03-03_16-38-12
1014\2008-03-03_16-39-10
1014\2008-03-03_17-20-56
1014\2008-03-04_17-15-05
1014\2008-03-05_12-50-38
1014\2008-03-11_16-32-37
1014\2008-04-11_14-37-01
1014\2008-04-11_16-17-40
1014\2008-04-15_17-13-27
1014\2008-04-16_17-53-09
1014\2008-05-09_16-17-10
1014\2008-05-09_16-48-08
1014\2008-05-28_15-37-14
1014\2008-06-16_11-57-06

1014\2008-06-16_11-57-15
1014\2008-06-17_17-14-24
1014\2009-06-26_15-11-07
1014\2009-06-30_08-34-24
1014\2009-07-02_16-36-59
1014\2009-07-09_16-41-58
1022\2008-02-29_08-33-54
1022\2008-02-29_12-27-23
1022\2008-02-29_12-27-38
1022\2008-04-22_10-46-57
1022\2008-04-23_10-20-28
1022\2008-04-23_12-08-34
1022\2008-04-23_15-05-30
1022\2008-04-24_10-12-45
1022\2008-04-25_09-41-49
1022\2008-05-22_10-23-44
1022\2008-05-23_10-10-35
1022\2008-06-20_12-16-43
1022\2009-06-29_11-53-08
1022\2009-06-29_12-13-03
1024\2008-02-21_08-11-44
1024\2008-03-04_17-02-18
1024\2008-03-05_17-29-19
1024\2008-03-19_18-30-44
1024\2008-04-21_16-01-42
1024\2008-04-28_18-50-17
1024\2009-07-30_08-17-52
1024\2009-08-20_16-16-29
1024\2009-09-11_14-54-11
1024\2009-10-06_10-09-47
1025\2008-02-15_17-23-58
1025\2008-02-15_18-37-13
1025\2008-02-15_20-40-13
1025\2008-02-17_14-37-02
1025\2008-02-25_17-08-28
1025\2008-02-28_17-09-26
1025\2008-02-29_18-03-54
1025\2008-03-06_19-40-01
1025\2008-03-07_20-11-51
1025\2008-03-07_20-50-03
1025\2008-03-11_21-01-00
1025\2008-03-16_09-13-06
1025\2008-04-08_20-56-04
1025\2008-04-23_07-57-53
1025\2008-05-09_20-59-58
1025\2008-05-14_09-26-43
1025\2008-05-19_20-51-01
1025\2008-05-25_14-44-52
1025\2008-05-29_12-49-01
1025\2008-06-14_09-33-00
1025\2009-06-29_18-42-23
1025\2009-07-09_07-17-01
1025\2009-07-18_09-25-49
1025\2009-08-14_16-12-59
1027\2008-02-21_11-58-26
1027\2008-02-21_13-44-09
1027\2008-03-14_10-27-02
1027\2008-03-19_11-14-54
1027\2009-06-29_17-06-25
1027\2009-06-29_17-18-54
1027\2009-07-22_08-26-01
1027\2009-08-04_11-02-23
1027\2009-08-05_15-26-13
1028\2008-02-19_12-00-35
1028\2008-02-22_15-44-02
1028\2008-06-02_08-01-02
1028\2008-06-03_08-11-18
1028\2008-06-04_07-57-15
1028\2009-06-29_16-51-57
1028\2009-06-30_14-37-46
1028\2009-07-02_16-53-46
1028\2009-07-03_14-27-19
1028\2009-07-08_16-28-24
1028\2009-07-18_14-36-39
1032\2008-02-29_12-54-29
1032\2008-05-16_12-48-49
1032\2009-07-02_18-10-33
1032\2008-03-04_11-09-36
1032\2008-05-23_15-53-35
1032\2009-09-30_16-04-03
1032\2008-06-04_16-09-25
1032\2009-09-30_17-33-13
1032\2008-04-29_16-15-00
1032\2008-06-05_13-19-07
1032\2009-09-30_17-33-45
1033\2008-03-19_11-15-47
1033\2008-03-20_15-11-36
1033\2008-03-20_15-12-44
1033\2008-04-09_17-48-48
1033\2008-06-19_11-32-22
1033\2009-07-14_18-15-16
1033\2009-07-17_12-19-16
1037\2008-02-22_12-36-03
1037\2008-02-26_10-46-47
1037\2008-03-18_10-50-58
1037\2008-04-07_15-21-11
1037\2008-04-22_12-39-48
1037\2009-06-25_13-46-31
1037\2009-07-01_15-14-16
1037\2009-07-01_15-14-46
1038\2008-02-23_11-34-55
1038\2008-02-23_13-39-59
1038\2008-02-24_09-22-37
1038\2008-02-24_09-22-50
1038\2008-02-29_08-54-47
1038\2009-07-10_12-19-31
1038\2009-07-21_12-21-23
1038\2009-09-17_13-11-57
1040\2008-04-01_18-20-43
1040\2008-04-01_18-21-57
1040\2008-04-01_18-22-42
1040\2008-04-01_18-24-45
1040\2008-04-02_09-31-07
1040\2009-08-28_11-26-32

1042\2008-02-20_12-59-58
1042\2008-02-20_15-27-22
1042\2008-02-20_15-27-31
1042\2008-02-20_15-27-38
1042\2008-02-20_17-12-12
1042\2008-02-20_17-12-30
1042\2008-02-21_17-10-04
1042\2008-02-22_11-03-39
1042\2008-02-22_12-06-16
1042\2008-02-22_12-06-36
1042\2008-02-22_12-06-57
1042\2008-02-26_12-52-40
1042\2008-02-28_17-18-18
1042\2008-02-28_17-18-29
1042\2008-02-29_08-35-37
1042\2008-03-03_13-15-44
1042\2008-03-05_12-33-24
1042\2008-03-06_12-28-59
1042\2008-03-06_12-29-37
1042\2008-03-07_12-00-06
1042\2008-03-11_17-24-05
1042\2008-03-25_12-12-19
1042\2008-04-14_17-39-16
1042\2008-04-15_17-14-45
1042\2008-04-16_09-13-43
1042\2008-05-19_10-47-01
1042\2008-05-19_10-47-27
1042\2008-05-19_10-47-32
1042\2008-05-27_12-18-00
1042\2008-05-27_12-18-05
1042\2008-05-28_18-40-44
1042\2008-05-29_09-52-06
1042\2008-06-03_08-58-08
1042\2008-06-03_08-58-26
1042\2008-06-04_09-20-27
1042\2008-06-12_17-21-01
1042\2008-06-13_12-36-22
1042\2008-06-26_08-27-26
1042\2008-06-27_16-08-22
1042\2008-06-30_12-03-34
1042\2009-06-26_16-09-32
1042\2009-07-14_16-12-36
1043\2008-02-15_15-52-05
1043\2008-02-18_07-55-00
1043\2008-02-18_07-55-19
1043\2008-02-18_09-47-05
1043\2008-02-18_10-55-10
1043\2008-02-19_14-25-45
1043\2008-02-19_16-03-27
1043\2008-02-20_09-18-54
1043\2008-02-20_14-52-56
1043\2008-02-22_07-53-50
1043\2008-02-22_11-59-00
1043\2008-02-25_15-00-22
1043\2008-02-26_09-03-51
1043\2008-03-03_10-44-45
1043\2008-03-04_13-49-23
1043\2008-03-11_15-00-19
1043\2008-03-17_12-51-11
1043\2008-03-20_14-46-13
1043\2008-03-26_13-30-23
1043\2008-05-12_11-11-43
1043\2008-05-12_11-12-00
1043\2008-05-15_14-14-39
1043\2008-05-19_09-38-34
1043\2008-05-22_16-24-15
1043\2008-05-30_13-30-52
1043\2008-06-05_12-52-00
1043\2008-06-06_11-16-07
1043\2008-06-10_15-23-59
1043\2009-06-26_11-56-47
1043\2009-06-29_15-42-35
1043\2009-06-30_14-09-38
1043\2009-06-30_15-38-32
1043\2009-07-02_16-11-57
1043\2009-07-07_13-56-07
1043\2009-07-09_09-52-51
1043\2009-07-31_12-29-36
1044\2009-07-07_15-43-30
1044\2009-07-09_13-44-47
1044\2009-07-13_14-03-41
1044\2009-07-28_13-56-39
1044\2009-07-30_14-40-53
1044\2009-07-31_12-40-28
1044\2009-08-14_12-59-46
1044\2009-08-18_17-01-40
1044\2009-09-02_10-02-52
1047\2008-02-18_17-35-15
1047\2008-02-21_18-07-01
1047\2008-02-28_18-12-43
1047\2008-03-25_16-35-53
1047\2008-03-26_17-34-16
1047\2008-04-14_17-52-56
1047\2008-04-16_17-20-57
1047\2008-04-21_18-04-43
1047\2008-04-21_18-04-43
1047\2009-06-29_12-29-47
1047\2009-07-02_11-19-21
1047\2009-07-03_16-11-35
1047\2009-07-13_15-33-30
1047\2009-07-24_14-50-27
1047\2009-08-12_15-27-04
1050\2007-12-03_19-18-23
1050\2007-12-03_19-18-51
1050\2007-12-04_08-55-53
1050\2007-12-04_08-58-36
1050\2007-12-04_08-58-57
1050\2007-12-04_08-59-22
1050\2007-12-04_09-00-00
1050\2007-12-04_09-09-23
1050\2007-12-04_09-21-11
1050\2007-12-04_09-25-45
1050\2007-12-04_09-35-03
1050\2007-12-04_09-36-40

1050\2007-12-04_09-38-47
1050\2007-12-04_10-08-09
1050\2007-12-04_10-22-02
1050\2007-12-04_10-52-25
1050\2007-12-04_10-54-15
1050\2007-12-04_11-12-42
1050\2007-12-04_11-42-52
1050\2007-12-04_11-44-14
1050\2007-12-04_11-58-35
1050\2007-12-04_12-08-01
1050\2007-12-04_13-06-56
1050\2007-12-04_13-07-40
1050\2007-12-04_13-09-03
1050\2007-12-04_13-22-12
1050\2007-12-04_13-22-56
1050\2007-12-04_13-24-15
1050\2007-12-04_13-25-51
1050\2007-12-04_13-27-06
1050\2007-12-04_13-33-22
1050\2007-12-04_13-33-41
1050\2007-12-04_13-35-33
1050\2007-12-04_13-35-48
1050\2007-12-04_14-49-42
1050\2007-12-04_14-50-50
1050\2007-12-04_18-06-57
1050\2007-12-05_09-10-34
1050\2007-12-05_09-17-29
1050\2007-12-05_09-17-48
1050\2007-12-05_09-22-04
1050\2007-12-05_09-23-57
1050\2007-12-05_09-24-05
1050\2007-12-05_09-28-11
1050\2007-12-05_09-53-18
1050\2007-12-05_10-07-29
1050\2007-12-05_10-07-47
1050\2007-12-05_10-21-00
1050\2007-12-05_10-48-21
1050\2007-12-05_11-18-47
1050\2007-12-05_13-05-46
1050\2007-12-05_13-06-14
1050\2007-12-05_13-56-47
1050\2007-12-05_14-18-18
1051\2008-04-03_13-29-30
1051\2008-04-16_18-10-55
1051\2008-04-18_15-02-56
1051\2008-05-15_13-05-18
1051\2009-08-18_08-22-13
1051\2009-10-16_16-21-30
1051\2009-10-21_18-21-53
1066\2008-02-18_13-55-25
1066\2008-02-20_07-37-12
1066\2008-02-21_11-05-40
1066\2008-03-05_10-26-09
1066\2008-03-06_07-35-35
1066\2008-03-13_07-17-53
1067\2008-02-18_08-51-34
1067\2008-02-18_08-51-46
1067\2008-02-18_17-08-36
1067\2008-02-20_09-54-32
1067\2008-02-20_17-24-09
1067\2008-02-21_17-44-49
1067\2008-02-22_16-52-11
1067\2008-02-28_17-18-04
1067\2009-06-25_17-39-19
1067\2009-06-25_13-02-21
1074\2008-02-19_17-38-09
1074\2008-03-04_19-07-21
1074\2008-03-04_19-09-36
1074\2008-03-11_18-30-45
1074\2008-03-13_13-37-12
1074\2008-03-13_13-37-24
1074\2008-03-13_13-37-35
1074\2008-03-13_13-37-47
1074\2008-03-13_18-43-23
1074\2008-03-14_17-22-58
1074\2008-04-08_18-50-41
1074\2008-06-03_20-07-28
1074\2009-06-30_19-26-57
1074\2009-08-20_17-22-24
1075\2008-02-15_15-41-14
1075\2008-02-20_13-06-34
1075\2008-02-21_17-57-51
1075\2008-02-28_17-36-25
1075\2008-02-28_17-37-10
1075\2008-03-10_12-54-50
1075\2008-03-14_11-36-40
1075\2008-03-17_14-36-37
1075\2008-03-17_14-36-54
1075\2008-03-26_13-06-08
1075\2008-04-01_18-45-27
1075\2008-04-03_17-50-00
1075\2008-04-03_19-19-47
1075\2008-04-04_13-02-38
1075\2008-04-07_19-02-24
1075\2008-04-08_12-40-01
1075\2008-04-08_19-00-21
1075\2008-04-09_18-59-45
1075\2008-04-10_10-47-11
1075\2008-04-10_13-16-20
1075\2008-04-15_10-06-53
1075\2008-04-16_18-09-44
1075\2008-04-21_18-56-19
1075\2008-04-23_18-58-53
1075\2008-04-25_13-42-55
1075\2008-04-28_19-31-12
1075\2008-05-09_13-11-59
1075\2008-05-12_18-45-29
1075\2008-05-14_18-53-40
1075\2008-05-15_11-29-59
1075\2008-05-15_15-36-06
1075\2008-05-16_16-15-29
1075\2008-05-16_16-15-44
1075\2008-05-29_16-29-53
1075\2008-06-03_18-52-51

1075\2008-06-05_17-28-10
1075\2009-08-28_15-11-09
1080\2008-03-06_10-38-23
1080\2008-03-19_09-04-14
1080\2008-03-28_08-47-26
1080\2008-04-21_07-34-05
1080\2008-05-09_13-35-49
1080\2009-07-22_12-51-22
1080\2009-10-15_08-57-37
1085\2008-02-20_16-19-11
1085\2008-02-22_11-06-12
1085\2008-02-22_11-06-30
1085\2008-02-22_12-18-32
1085\2008-02-22_14-44-32
1085\2008-02-22_16-49-47
1085\2008-02-25_11-33-49
1085\2008-02-25_14-30-34
1085\2008-02-29_10-15-34
1085\2008-03-03_12-13-24
1085\2008-03-03_16-48-48
1085\2008-03-04_11-40-57
1085\2008-03-04_16-44-26
1085\2008-03-04_19-03-34
1085\2008-03-11_13-09-37
1085\2008-03-12_15-39-24
1085\2008-03-13_13-52-21
1085\2008-03-14_10-09-27
1085\2008-03-14_10-09-46
1085\2008-03-18_14-37-18
1085\2008-03-18_16-52-14
1085\2008-03-19_11-28-34
1085\2008-04-07_11-38-32
1085\2008-04-08_15-08-58
1091\2008-02-19_10-09-41
1091\2008-02-19_10-11-09
1091\2008-02-19_10-11-20
1091\2008-02-19_10-11-38
1091\2008-04-21_17-47-00
1091\2009-06-29_11-12-38
1119\2008-02-29_10-16-20
1119\2008-03-05_14-06-19
1119\2008-03-11_15-09-00
1119\2008-03-11_16-17-55
1119\2008-04-08_17-15-47
1119\2008-05-29_16-49-20
1119\2009-06-29_12-28-13
1132\2008-03-31_16-27-46
1132\2008-03-31_16-28-28
1132\2008-03-31_16-29-45
1132\2008-03-31_16-30-01
1132\2008-04-03_18-43-37
1132\2008-04-03_18-45-19
1132\2008-04-03_19-59-34
1132\2008-04-09_20-07-22
1132\2008-04-09_20-15-32
1132\2008-04-21_20-01-48
1132\2008-05-12_20-23-30
1132\2008-05-13_18-55-10
1132\2008-05-14_20-00-35
1132\2008-05-15_19-54-15
1132\2008-05-23_19-37-02
1132\2008-05-28_20-05-20
1151\2008-02-18_10-59-51
1151\2008-02-18_17-33-08
1151\2008-02-19_10-42-34
1151\2008-02-19_14-10-03
1151\2008-02-19_14-10-15
1151\2008-02-20_09-07-06
1151\2008-02-22_13-15-04
1151\2008-02-22_14-42-03
1151\2008-04-04_13-40-12
1151\2008-05-16_16-27-17
1173\2007-12-04_11-55-10
1173\2007-12-04_13-08-20
1173\2007-12-04_13-57-47
1173\2007-12-04_13-58-14
1173\2007-12-04_17-53-11
1173\2007-12-04_17-57-37
1173\2007-12-05_13-06-49
1173\2007-12-05_20-34-11
1173\2007-12-06_17-16-06
1173\2007-12-06_20-22-00
1173\2007-12-06_20-22-07
1173\2007-12-07_13-42-55
1173\2007-12-11_15-34-13
1173\2007-12-11_15-34-29
1173\2007-12-12_15-42-24
1174\2007-12-04_17-50-47
1174\2007-12-04_17-50-54
1174\2007-12-04_17-52-56
1174\2007-12-05_11-02-06
1174\2007-12-05_14-37-51
1174\2007-12-05_14-38-17
1174\2007-12-05_14-38-48
1174\2007-12-05_18-30-09
1174\2007-12-07_11-55-54
1174\2007-12-11_19-50-38
1174\2007-12-12_17-23-00
1183\2009-03-31_10-42-53
1183\2009-03-31_10-43-30
1183\2009-03-31_10-46-23
1183\2009-03-31_10-46-42
1183\2009-03-31_10-47-01
1183\2009-03-31_10-49-23
1186\2007-12-04_17-45-44
1186\2007-12-05_11-15-41
1186\2007-12-06_12-19-10
1186\2007-12-06_16-34-53
1186\2007-12-06_16-37-56
1186\2007-12-06_16-42-11
1186\2007-12-11_11-30-50
1186\2007-12-12_16-43-30
1192\2007-12-06_11-38-22
1192\2007-12-06_11-55-00

1192\2007-12-06_17-00-44
1192\2007-12-07_13-54-56
1192\2007-12-11_12-32-50
1192\2007-12-12_15-24-54
1201\2007-12-04_09-22-52
1201\2007-12-04_11-10-55
1201\2007-12-04_11-12-10
1201\2007-12-04_15-24-45
1201\2007-12-05_11-36-01
1201\2007-12-05_14-09-43
1201\2007-12-07_16-26-45
1201\2007-12-07_16-26-58
1201\2007-12-07_16-27-15
1201\2007-12-11_14-48-03
1201\2007-12-11_14-53-48
1201\2007-12-11_14-56-57
1201\2007-12-11_15-00-56
1201\2007-12-11_15-01-24
1201\2007-12-11_15-01-41
1201\2007-12-13_11-49-45
1212\2008-11-19_15-03-26
1212\2009-06-25_12-35-27
1212\2009-06-25_12-35-51
1212\2009-07-07_12-05-52
1212\2009-07-07_12-17-00
1212\2009-07-07_12-17-27
1212\2009-09-04_15-24-00
1213\2007-12-03_19-17-54
1213\2007-12-04_08-56-04
1213\2007-12-04_09-45-01
1213\2007-12-04_09-45-16
1213\2007-12-04_09-45-29
1213\2007-12-04_12-32-56
1213\2007-12-04_16-43-24
1213\2007-12-04_16-50-20
1213\2007-12-04_16-50-33
1213\2007-12-05_09-31-38
1213\2007-12-05_10-02-33
1213\2007-12-05_15-59-30
1213\2007-12-07_13-24-01
1213\2007-12-07_14-58-56
1217\2007-12-04_09-20-33
1217\2007-12-04_09-23-03
1217\2007-12-04_09-30-09
1217\2007-12-04_09-30-22
1217\2007-12-04_09-30-47
1217\2007-12-04_10-37-52
1217\2007-12-04_13-46-32
1217\2007-12-04_13-59-43
1217\2007-12-04_16-18-19
1217\2007-12-04_17-32-29
1217\2007-12-04_18-06-05
1217\2007-12-05_17-11-20
1217\2007-12-05_18-10-27
1217\2007-12-06_11-28-14
1217\2007-12-06_12-44-32
1217\2007-12-06_13-03-01
1217\2007-12-06_14-30-44
1217\2007-12-06_16-07-59
1217\2007-12-06_16-11-22
1217\2007-12-07_09-20-58
1217\2007-12-07_15-50-08
1217\2007-12-10_16-53-49
1217\2007-12-10_18-06-35
1217\2007-12-11_09-50-16
1217\2007-12-11_12-41-26
1217\2007-12-12_13-49-27
1217\2007-12-12_15-02-49
1217\2007-12-13_12-57-03
1220\2007-12-04_10-19-31
1220\2007-12-04_14-10-45
1220\2007-12-04_18-04-01
1220\2007-12-05_09-48-20
1220\2007-12-05_09-48-32
1220\2007-12-05_09-48-43
1220\2007-12-05_10-49-57
1220\2007-12-05_10-50-35
1220\2007-12-06_16-10-04
1220\2007-12-06_17-21-55
1220\2007-12-07_10-53-34
1220\2007-12-07_11-23-52
1220\2007-12-07_12-53-13
1220\2007-12-11_12-28-01
1220\2007-12-11_12-28-09
1220\2007-12-11_14-12-55
1220\2007-12-11_17-15-46
1220\2007-12-12_10-07-23
1220\2007-12-12_16-32-26
1220\2007-12-13_12-48-15
2001\2008-02-18_11-00-29
2001\2008-02-19_11-31-48
2001\2008-02-19_12-53-11
2001\2008-02-19_14-02-34
2001\2008-02-20_15-55-13
2001\2008-02-25_11-57-05
2001\2008-03-04_13-02-53
2001\2008-03-05_16-41-36
2001\2008-03-05_17-23-17
2001\2008-03-12_16-58-43
2001\2008-03-13_13-28-51
2001\2008-03-14_10-57-39
2001\2008-03-18_09-57-38
2001\2008-03-18_10-49-30
2002\2008-02-19_15-20-58
2002\2008-02-19_15-58-06
2002\2008-02-20_09-36-31
2002\2008-02-20_11-20-24
2002\2008-02-20_16-39-01
2002\2008-02-21_09-37-09
2002\2008-02-21_10-44-32
2002\2008-02-21_14-15-22
2002\2008-02-26_08-40-14
2002\2008-03-17_10-01-11
2002\2008-03-17_13-41-17

2004\2008-02-19_17-41-10
2004\2008-02-21_18-10-25
2004\2008-02-25_17-39-07
2004\2008-03-14_18-06-00
2004\2008-03-14_18-06-19
2004\2008-03-17_08-41-19
2004\2008-03-18_08-46-57
2004\2008-03-19_14-42-13
2004\2008-03-19_14-42-29
2004\2008-03-19_18-11-31
2004\2008-03-20_08-18-03
2004\2008-03-20_17-15-17
2004\2008-03-26_16-41-07
2004\2008-03-26_18-38-11
2004\2008-04-02_16-59-06
2004\2008-04-04_10-43-23
2004\2008-04-11_18-04-43
2004\2008-04-11_18-04-56
2004\2008-04-24_17-52-21
2005\2008-02-19_12-34-11
2005\2008-02-19_12-34-23
2005\2008-02-19_12-35-12
2005\2008-02-19_12-35-34
2005\2008-02-19_12-35-49
2005\2008-02-19_12-36-06
2005\2008-02-19_12-36-34
2005\2008-02-19_12-36-47
2005\2008-02-19_12-36-59
2005\2008-02-19_12-37-18
2006\2008-02-18_12-32-23
2006\2008-02-19_16-52-17
2006\2008-02-22_07-36-44
2006\2008-02-26_10-27-30
2006\2008-03-03_12-09-02
2006\2008-03-04_16-50-47
2006\2008-03-12_16-45-08
2006\2008-03-18_07-35-44
2006\2008-05-22_13-51-28
2006\2008-05-30_07-39-31
2006\2008-06-03_15-44-48
2006\2008-06-10_14-40-00
2006\2008-06-10_16-13-25
2011\2008-03-28_13-18-47
2011\2008-04-08_18-33-30
2011\2008-04-08_18-33-44
2011\2008-04-08_18-33-58
2011\2008-04-09_16-51-33
2011\2008-04-09_18-10-30
2011\2008-04-11_11-35-31
2011\2008-04-14_17-56-23
2011\2008-04-15_17-39-52
2011\2008-04-16_18-03-14
2011\2008-04-17_15-39-07
2011\2008-04-17_17-43-39
2011\2008-04-23_17-01-45
2011\2008-04-30_17-39-10
2011\2008-05-09_14-43-25
2011\2008-05-13_13-21-01
2011\2008-05-27_10-43-21
2011\2008-05-27_17-34-55
2011\2008-06-02_17-17-16
2011\2008-06-09_12-29-49
2011\2008-06-10_13-18-26
2012\2008-02-21_15-23-56
2012\2008-03-18_11-15-12
2012\2008-04-01_17-15-51
2012\2008-04-17_11-31-22
2012\2008-04-17_17-16-15
2012\2008-04-22_17-13-04
2012\2008-05-19_11-09-44
2012\2008-06-11_09-54-42
2012\2008-06-27_09-59-27
2014\2008-03-27_07-36-57
2014\2008-04-30_07-34-16
2014\2008-04-30_07-34-35
2014\2008-05-01_07-18-04
2014\2008-05-16_07-21-24
2014\2008-05-16_07-21-35
2014\2008-05-16_07-21-46
2015\2008-02-20_11-38-23
2015\2008-02-22_10-29-21
2015\2008-02-25_09-20-11
2015\2008-02-28_17-07-06
2015\2008-03-07_11-33-36
2015\2008-03-11_09-56-17
2015\2008-03-31_14-36-32
2023\2008-03-05_17-09-51
2023\2008-03-12_17-13-06
2023\2008-03-13_17-06-56
2023\2008-03-14_17-00-24
2023\2008-03-17_17-06-41
2023\2008-03-19_17-03-41
2023\2008-03-20_16-12-58
2023\2008-03-28_16-34-05
2023\2008-03-31_17-08-18
2023\2008-04-07_17-08-11
2023\2008-04-08_17-09-39
2023\2008-04-11_16-36-01
2023\2008-04-18_16-31-13
2023\2008-04-23_17-06-29
2023\2008-05-13_16-08-39
2026\2008-02-18_17-52-39
2026\2008-02-28_15-07-02
2026\2008-02-28_15-38-17
2026\2008-02-28_15-39-08
2026\2008-05-18_13-11-39
2026\2008-05-18_13-11-45
2026\2008-05-18_13-11-54
2026\2008-05-18_15-50-08
2028\2008-02-21_17-29-09
2028\2008-02-25_17-00-47
2028\2008-02-28_17-17-33
2028\2008-03-03_13-14-12
2028\2008-03-06_17-24-53

2028\2008-03-07_16-49-23	2028\2008-03-18_13-27-07	2028\2008-04-11_16-42-29
2028\2008-04-23_16-45-15	2028\2008-05-20_17-37-31	2064\2007-12-04_09-37-23
2064\2007-12-04_12-06-58	2064\2007-12-04_15-20-26	2064\2007-12-05_13-42-35
2064\2007-12-05_13-43-26	2064\2007-12-05_17-31-10	2064\2007-12-06_12-32-13
2064\2007-12-07_11-12-45	2064\2007-12-10_17-42-25	2064\2008-05-16_11-06-33
2064\2008-05-16_11-07-20	2064\2008-05-16_11-10-13	2064\2008-05-16_11-13-41
2064\2008-05-16_11-17-22	2064\2008-05-16_11-21-54	2064\2008-05-16_11-34-25
2188\2007-12-06_00-15-02	2188\2007-12-06_00-15-35	2188\2007-12-06_03-55-25
2188\2007-12-06_03-56-14	2188\2007-12-06_03-56-43	2188\2007-12-06_03-57-46

Appendix D

Photoface - 40 subjects, 400 Sessions

These are the sessions used throughout the thesis when 400 sessions of Photoface data are referred to. An electronic copy is available from the author.

1001\2008-03-01_07-58-49	1001\2008-03-02_11-25-38	1001\2008-03-20_15-10-39
1001\2008-04-05_11-22-53	1001\2008-04-19_08-34-45	1001\2008-05-14_11-01-27
1001\2009-06-26_14-48-20	1001\2009-07-09_13-25-40	1001\2009-07-10_15-03-01
1001\2009-07-28_15-19-04	1002\2008-02-25_17-34-06	1002\2008-03-28_14-11-04
1002\2008-04-01_18-52-11	1002\2008-04-01_18-54-30	1002\2009-07-08_08-18-43
1002\2009-07-30_14-57-27	1002\2009-08-17_16-00-09	1002\2009-08-26_10-33-57
1002\2009-08-27_08-20-34	1002\2009-08-27_16-41-09	1003\2008-02-18_17-40-16
1003\2008-02-18_17-40-29	1003\2008-02-22_16-42-59	1003\2008-02-22_16-43-26
1003\2008-02-25_17-24-07	1003\2008-03-04_17-36-56	1003\2008-03-07_11-25-56
1003\2008-03-12_17-50-55	1003\2008-03-31_17-11-17	1003\2008-04-02_17-27-25
1004\2008-02-21_11-07-38	1004\2008-02-21_11-08-02	1004\2008-02-21_11-39-28
1004\2008-02-21_13-38-47	1004\2008-02-21_14-07-06	1004\2008-02-21_17-12-02
1004\2008-02-21_18-11-37	1004\2008-02-25_17-33-16	1004\2008-03-05_12-23-16
1004\2008-03-05_12-45-05	1007\2008-03-28_15-28-02	1007\2008-03-28_15-28-34
1007\2008-04-21_11-28-50	1007\2008-05-28_10-11-10	1007\2009-02-26_10-47-19
1007\2009-06-25_13-56-34	1007\2009-07-14_14-35-14	1007\2009-07-17_10-13-25
1007\2009-07-20_14-17-31	1007\2009-07-22_15-06-41	1008\2008-02-18_13-42-16
1008\2008-02-22_13-53-10	1008\2008-04-14_08-30-11	1008\2008-04-15_11-17-30
1008\2008-04-16_16-46-12	1008\2008-04-18_13-43-00	1008\2008-05-29_08-15-48
1008\2008-06-18_16-48-33	1008\2009-06-26_15-45-00	1008\2009-07-02_09-52-00

1009\2008-02-18_08-50-39
1009\2008-02-18_11-10-59
1009\2008-02-19_08-54-56
1009\2008-02-21_10-13-44
1009\2008-02-29_08-39-56
1009\2008-02-29_13-49-12
1009\2008-03-12_12-02-37
1009\2008-03-13_14-44-23
1009\2008-03-17_14-15-57
1009\2008-04-22_09-06-04
1012\2008-02-18_06-46-13
1012\2008-02-19_17-24-46
1012\2008-02-19_17-25-08
1012\2008-02-22_07-39-59
1012\2008-02-22_07-40-08
1012\2008-02-22_07-40-19
1012\2008-02-22_17-02-36
1012\2008-02-22_17-02-47
1012\2008-02-22_17-02-58
1012\2008-02-22_17-03-10
1013\2008-02-25_14-50-46
1013\2008-02-29_09-11-20
1013\2008-02-29_10-46-05
1013\2008-03-03_09-29-39
1013\2008-03-03_15-30-23
1013\2008-03-10_08-46-05
1013\2008-03-11_07-57-36
1013\2008-03-13_09-04-05
1013\2008-03-18_09-59-14
1013\2008-03-27_15-22-28
1014\2008-02-19_09-05-27
1014\2008-02-19_16-45-02
1014\2008-02-20_17-14-18
1014\2008-02-21_16-56-08
1014\2008-02-22_17-33-58
1014\2008-02-25_12-36-10
1014\2008-02-28_16-08-11
1014\2008-02-29_16-48-59
1014\2008-03-03_11-20-42
1014\2008-03-03_16-38-12
1022\2008-02-29_08-33-54
1022\2008-02-29_12-27-23
1022\2008-02-29_12-27-38
1022\2008-04-22_10-46-57
1022\2008-04-23_10-20-28
1022\2008-04-23_12-08-34
1022\2008-04-23_15-05-30
1022\2008-04-24_10-12-45
1022\2008-04-25_09-41-49
1022\2008-05-22_10-23-44
1024\2008-02-21_08-11-44
1024\2008-03-04_17-02-18
1024\2008-03-05_17-29-19
1024\2008-03-19_18-30-44
1024\2008-04-21_16-01-42
1024\2008-04-28_18-50-17
1024\2009-07-30_08-17-52
1024\2009-08-20_16-16-29
1024\2009-09-11_14-54-11
1024\2009-10-06_10-09-47
1025\2008-02-15_17-23-58
1025\2008-02-15_18-37-13
1025\2008-02-15_20-40-13
1025\2008-02-17_14-37-02
1025\2008-02-25_17-08-28
1025\2008-02-28_17-09-26
1025\2008-02-29_18-03-54
1025\2008-03-06_19-40-01
1025\2008-03-07_20-11-51
1025\2008-03-07_20-50-03
1028\2008-02-19_12-00-35
1028\2008-02-22_15-44-02
1028\2008-06-02_08-01-02
1028\2008-06-03_08-11-18
1028\2008-06-04_07-57-15
1028\2009-06-29_16-51-57
1028\2009-06-30_14-37-46
1028\2009-07-02_16-53-46
1028\2009-07-03_14-27-19
1028\2009-07-08_16-28-24
1032\2008-02-29_12-54-29
1032\2008-05-16_12-48-49
1032\2009-07-02_18-10-33
1032\2008-03-04_11-09-36
1032\2008-05-23_15-53-35
1032\2009-09-30_16-04-03
1032\2008-06-04_16-09-25
1032\2009-09-30_17-33-13
1032\2008-04-29_16-15-00
1032\2008-06-05_13-19-07
1042\2008-02-20_12-59-58
1042\2008-02-20_15-27-22
1042\2008-02-20_15-27-31
1042\2008-02-20_15-27-38
1042\2008-02-20_17-12-12
1042\2008-02-20_17-12-30
1042\2008-02-21_17-10-04
1042\2008-02-22_11-03-39
1042\2008-02-22_12-06-16
1042\2008-02-22_12-06-36
1043\2008-02-15_15-52-05
1043\2008-02-18_07-55-00
1043\2008-02-18_07-55-19
1043\2008-02-18_09-47-05
1043\2008-02-18_10-55-10
1043\2008-02-19_14-25-45
1043\2008-02-19_16-03-27
1043\2008-02-20_09-18-54
1043\2008-02-20_14-52-56
1043\2008-02-22_07-53-50
1047\2008-02-18_17-35-15
1047\2008-02-21_18-07-01
1047\2008-02-28_18-12-43
1047\2008-03-25_16-35-53

1047\2008-03-26_17-34-16
1047\2008-04-14_17-52-56
1047\2008-04-16_17-20-57
1047\2008-04-21_18-04-43
1047\2008-04-21_18-04-43
1047\2009-06-29_12-29-47
1050\2007-12-03_19-18-23
1050\2007-12-03_19-18-51
1050\2007-12-04_08-55-53
1050\2007-12-04_08-58-36
1050\2007-12-04_08-58-57
1050\2007-12-04_08-59-22
1050\2007-12-04_09-00-00
1050\2007-12-04_09-09-23
1050\2007-12-04_09-21-11
1050\2007-12-04_09-25-45
1067\2008-02-18_08-51-34
1067\2008-02-18_08-51-46
1067\2008-02-18_17-08-36
1067\2008-02-20_09-54-32
1067\2008-02-20_17-24-09
1067\2008-02-21_17-44-49
1067\2008-02-22_16-52-11
1067\2008-02-28_17-18-04
1067\2009-06-25_13-02-21
1067\2009-06-25_17-39-19
1074\2008-02-19_17-38-09
1074\2008-03-04_19-07-21
1074\2008-03-04_19-09-36
1074\2008-03-11_18-30-45
1074\2008-03-13_13-37-12
1074\2008-03-13_13-37-24
1074\2008-03-13_13-37-35
1074\2008-03-13_13-37-47
1074\2008-03-13_18-43-23
1074\2008-03-14_17-22-58
1075\2008-02-15_15-41-14
1075\2008-02-20_13-06-34
1075\2008-02-21_17-57-51
1075\2008-02-28_17-36-25
1075\2008-02-28_17-37-10
1075\2008-03-10_12-54-50
1075\2008-03-14_11-36-40
1075\2008-03-17_14-36-37
1075\2008-03-17_14-36-54
1075\2008-03-26_13-06-08
1085\2008-02-20_16-19-11
1085\2008-02-22_11-06-12
1085\2008-02-22_11-06-30
1085\2008-02-22_12-18-32
1085\2008-02-22_14-44-32
1085\2008-02-22_16-49-47
1085\2008-02-25_11-33-49
1085\2008-02-25_14-30-34
1085\2008-02-29_10-15-34
1085\2008-03-03_12-13-24
1132\2008-03-31_16-27-46
1132\2008-03-31_16-28-28
1132\2008-03-31_16-29-45
1132\2008-03-31_16-30-01
1132\2008-04-03_18-43-37
1132\2008-04-03_18-45-19
1132\2008-04-03_19-59-34
1132\2008-04-09_20-07-22
1132\2008-04-09_20-15-32
1132\2008-04-21_20-01-48
1151\2008-02-18_10-59-51
1151\2008-02-18_17-33-08
1151\2008-02-19_10-42-34
1151\2008-02-19_14-10-03
1151\2008-02-19_14-10-15
1151\2008-02-20_09-07-06
1151\2008-02-22_13-15-04
1151\2008-02-22_14-42-03
1151\2008-04-04_13-40-12
1151\2008-05-16_16-27-17
1173\2007-12-04_11-55-10
1173\2007-12-04_13-08-20
1173\2007-12-04_13-57-47
1173\2007-12-04_13-58-14
1173\2007-12-04_17-53-11
1173\2007-12-04_17-57-37
1173\2007-12-05_13-06-49
1173\2007-12-05_20-34-11
1173\2007-12-06_17-16-06
1173\2007-12-06_20-22-00
1174\2007-12-04_17-50-47
1174\2007-12-04_17-50-54
1174\2007-12-04_17-52-56
1174\2007-12-05_11-02-06
1174\2007-12-05_14-37-51
1174\2007-12-05_14-38-17
1174\2007-12-05_14-38-48
1174\2007-12-05_18-30-09
1174\2007-12-07_11-55-54
1174\2007-12-11_19-50-38
1201\2007-12-04_09-22-52
1201\2007-12-04_11-10-55
1201\2007-12-04_11-12-10
1201\2007-12-04_15-24-45
1201\2007-12-05_11-36-01
1201\2007-12-05_14-09-43
1201\2007-12-07_16-26-45
1201\2007-12-07_16-26-58
1201\2007-12-07_16-27-15
1201\2007-12-11_14-48-03
1213\2007-12-03_19-17-54
1213\2007-12-04_08-56-04
1213\2007-12-04_09-45-01
1213\2007-12-04_09-45-16
1213\2007-12-04_09-45-29
1213\2007-12-04_12-32-56
1213\2007-12-04_16-43-24
1213\2007-12-04_16-50-20

1213\2007-12-04_16-50-33
1213\2007-12-05_09-31-38
1217\2007-12-04_09-20-33
1217\2007-12-04_09-23-03
1217\2007-12-04_09-30-09
1217\2007-12-04_09-30-22
1217\2007-12-04_09-30-47
1217\2007-12-04_10-37-52
1217\2007-12-04_13-46-32
1217\2007-12-04_13-59-43
1217\2007-12-04_16-18-19
1217\2007-12-04_17-32-29
1220\2007-12-04_10-19-31
1220\2007-12-04_14-10-45
1220\2007-12-04_18-04-01
1220\2007-12-05_09-48-20
1220\2007-12-05_09-48-32
1220\2007-12-05_09-48-43
1220\2007-12-05_10-49-57
1220\2007-12-05_10-50-35
1220\2007-12-06_16-10-04
1220\2007-12-06_17-21-55
2001\2008-02-18_11-00-29
2001\2008-02-19_11-31-48
2001\2008-02-19_12-53-11
2001\2008-02-19_14-02-34
2001\2008-02-20_15-55-13
2001\2008-02-25_11-57-05
2001\2008-03-04_13-02-53
2001\2008-03-05_16-41-36
2001\2008-03-05_17-23-17
2001\2008-03-12_16-58-43
2002\2008-02-19_15-20-58
2002\2008-02-19_15-58-06
2002\2008-02-20_09-36-31
2002\2008-02-20_11-20-24
2002\2008-02-20_16-39-01
2002\2008-02-21_09-37-09
2002\2008-02-21_10-44-32
2002\2008-02-21_14-15-22
2002\2008-02-26_08-40-14
2002\2008-03-17_10-01-11
2004\2008-02-19_17-41-10
2004\2008-02-21_18-10-25
2004\2008-02-25_17-39-07
2004\2008-03-14_18-06-00
2004\2008-03-14_18-06-19
2004\2008-03-17_08-41-19
2004\2008-03-18_08-46-57
2004\2008-03-19_14-42-13
2004\2008-03-19_14-42-29
2004\2008-03-19_18-11-31
2005\2008-02-19_12-34-11
2005\2008-02-19_12-34-23
2005\2008-02-19_12-35-12
2005\2008-02-19_12-35-34
2005\2008-02-19_12-35-49
2005\2008-02-19_12-36-06
2005\2008-02-19_12-36-34
2005\2008-02-19_12-36-47
2005\2008-02-19_12-36-59
2005\2008-02-19_12-37-18
2006\2008-02-18_12-32-23
2006\2008-02-19_16-52-17
2006\2008-02-22_07-36-44
2006\2008-02-26_10-27-30
2006\2008-03-03_12-09-02
2006\2008-03-04_16-50-47
2006\2008-03-12_16-45-08
2006\2008-03-18_07-35-44
2006\2008-05-22_13-51-28
2006\2008-05-30_07-39-31
2011\2008-03-28_13-18-47
2011\2008-04-08_18-33-30
2011\2008-04-08_18-33-44
2011\2008-04-08_18-33-58
2011\2008-04-09_16-51-33
2011\2008-04-09_18-10-30
2011\2008-04-11_11-35-31
2011\2008-04-14_17-56-23
2011\2008-04-15_17-39-52
2011\2008-04-16_18-03-14
2023\2008-03-05_17-09-51
2023\2008-03-12_17-13-06
2023\2008-03-13_17-06-56
2023\2008-03-14_17-00-24
2023\2008-03-17_17-06-41
2023\2008-03-19_17-03-41
2023\2008-03-20_16-12-58
2023\2008-03-28_16-34-05
2023\2008-03-31_17-08-18
2023\2008-04-07_17-08-11
2028\2008-02-21_17-29-09
2028\2008-02-25_17-00-47
2028\2008-02-28_17-17-33
2028\2008-03-03_13-14-12
2028\2008-03-06_17-24-53
2028\2008-03-07_16-49-23
2028\2008-03-18_13-27-07
2028\2008-04-11_16-42-29
2028\2008-04-23_16-45-15
2028\2008-05-20_17-37-31
2064\2007-12-04_09-37-23
2064\2007-12-04_12-06-58
2064\2007-12-04_15-20-26
2064\2007-12-05_13-42-35
2064\2007-12-05_13-43-26
2064\2007-12-05_17-31-10
2064\2007-12-06_12-32-13
2064\2007-12-07_11-12-45
2064\2007-12-10_17-42-25
2064\2008-05-16_11-06-33

Appendix E

FRGCV2.0 - 40 subjects, 400 Sessions

These are the sessions used throughout the thesis when 400 sessions of FRGCv2.0 data are referred to. An electronic copy is available from the author.

Spring2003range\02463d452	Spring2003range\02463d454	Spring2003range\02463d456
Spring2003range\02463d458	Spring2003range\02463d460	Spring2003range\02463d462
Spring2003range\02463d464	Spring2003range\02463d466	Fall2003range\02463d546
Fall2003range\02463d548	Spring2003range\04202d344	Spring2003range\04202d346
Spring2003range\04202d348	Spring2003range\04202d350	Spring2003range\04202d352
Spring2003range\04202d354	Spring2003range\04202d356	Spring2003range\04202d358
Fall2003range\04202d438	Fall2003range\04202d440	Spring2003range\04203d340
Spring2003range\04203d342	Spring2003range\04203d344	Spring2003range\04203d346
Spring2003range\04203d348	Spring2003range\04203d350	Spring2003range\04203d352
Spring2003range\04203d354	Fall2003range\04203d436	Fall2003range\04203d438
Spring2003range\04217d331	Spring2003range\04217d333	Spring2003range\04217d335
Spring2003range\04217d337	Fall2003range\04217d399	Fall2003range\04217d401
Fall2003range\04217d403	Fall2003range\04217d405	Spring2004range\04217d455
Spring2004range\04217d457	Spring2003range\04221d343	Spring2003range\04221d345
Spring2003range\04221d347	Spring2003range\04221d349	Fall2003range\04221d429
Fall2003range\04221d431	Fall2003range\04221d433	Fall2003range\04221d435
Fall2003range\04221d437	Spring2004range\04221d541	Spring2003range\04222d345
Spring2003range\04222d347	Spring2003range\04222d349	Spring2003range\04222d351
Spring2003range\04222d353	Spring2003range\04222d355	Spring2003range\04222d357
Spring2003range\04222d359	Fall2003range\04222d391	Fall2003range\04222d393
Spring2003range\04225d207	Spring2003range\04225d209	Spring2003range\04225d211
Fall2003range\04225d291	Fall2003range\04225d293	Fall2003range\04225d295
Fall2003range\04225d297	Fall2003range\04225d299	Spring2004range\04225d396

Spring2004range\04225d398
Spring2003range\04233d308
Spring2003range\04233d310
Spring2003range\04233d312
Spring2003range\04233d314
Spring2003range\04233d316
Spring2003range\04233d318
Spring2003range\04233d320
Fall2003range\04233d390
Fall2003range\04233d392
Fall2003range\04233d394
Spring2003range\04239d302
Spring2003range\04239d304
Spring2003range\04239d306
Spring2003range\04239d308
Spring2003range\04239d310
Fall2003range\04239d378
Fall2003range\04239d380
Fall2003range\04239d382
Spring2004range\04239d480
Spring2004range\04239d482
Spring2003range\04265d211
Spring2003range\04265d213
Fall2003range\04265d261
Fall2003range\04265d263
Fall2003range\04265d265
Fall2003range\04265d267
Spring2004range\04265d337
Spring2004range\04265d339
Spring2004range\04265d345
Spring2004range\04265d347
Spring2003range\04286d184
Spring2003range\04286d186
Spring2003range\04286d188
Spring2003range\04286d190
Spring2003range\04286d192
Spring2003range\04286d194
Fall2003range\04286d263
Fall2003range\04286d265
Fall2003range\04286d267
Spring2004range\04286d367
Spring2003range\04288d180
Spring2003range\04288d182
Spring2003range\04288d184
Spring2003range\04288d186
Spring2003range\04288d188
Spring2003range\04288d190
Fall2003range\04288d252
Fall2003range\04288d254
Fall2003range\04288d256
Fall2003range\04288d258
Spring2003range\04297d208
Spring2003range\04297d210
Spring2003range\04297d212
Spring2003range\04297d214
Spring2003range\04297d216
Spring2003range\04297d218
Fall2003range\04297d261
Fall2003range\04297d263
Fall2003range\04297d265
Spring2004range\04297d305
Spring2003range\04301d156
Spring2003range\04301d158
Spring2003range\04301d160
Fall2003range\04301d240
Fall2003range\04301d242
Fall2003range\04301d244
Fall2003range\04301d246
Fall2003range\04301d248
Spring2004range\04301d349
Spring2004range\04301d357
Spring2003range\04309d83
Spring2003range\04309d85
Spring2003range\04309d87
Fall2003range\04309d161
Fall2003range\04309d163
Fall2003range\04309d165
Fall2003range\04309d167
Spring2004range\04309d245
Spring2004range\04309d247
Spring2004range\04309d251
Spring2003range\04311d174
Spring2003range\04311d176
Spring2003range\04311d178
Spring2003range\04311d180
Spring2003range\04311d182
Fall2003range\04311d226
Fall2003range\04311d228
Fall2003range\04311d230
Fall2003range\04311d232
Spring2004range\04311d280
Spring2003range\04319d120
Spring2003range\04319d122
Spring2003range\04319d124
Fall2003range\04319d186
Fall2003range\04319d188
Fall2003range\04319d190
Fall2003range\04319d192
Spring2004range\04319d264
Spring2004range\04319d266
Spring2004range\04319d268
Spring2003range\04320d198
Spring2003range\04320d200
Spring2003range\04320d202
Spring2003range\04320d204
Spring2003range\04320d206
Spring2003range\04320d208
Fall2003range\04320d270
Fall2003range\04320d272
Fall2003range\04320d274
Spring2004range\04320d340
Spring2003range\04324d203
Spring2003range\04324d205
Spring2003range\04324d207
Spring2003range\04324d209
Spring2003range\04324d211
Spring2003range\04324d213
Fall2003range\04324d276
Fall2003range\04324d278
Fall2003range\04324d280
Fall2003range\04324d282
Spring2003range\04327d216
Spring2003range\04327d218
Spring2003range\04327d220
Spring2003range\04327d222
Fall2003range\04327d290
Fall2003range\04327d292
Fall2003range\04327d294
Fall2003range\04327d296

Spring2004range\04327d392
Spring2004range\04327d394
Spring2003range\04334d214
Spring2003range\04334d216
Spring2003range\04334d218
Spring2003range\04334d220
Fall2003range\04334d300
Fall2003range\04334d302
Fall2003range\04334d304
Fall2003range\04334d306
Fall2003range\04334d308
Spring2004range\04334d410
Spring2003range\04336d207
Spring2003range\04336d209
Spring2003range\04336d211
Fall2003range\04336d291
Fall2003range\04336d293
Fall2003range\04336d295
Fall2003range\04336d297
Fall2003range\04336d299
Spring2004range\04336d393
Spring2004range\04336d395
Spring2003range\04343d230
Spring2003range\04343d232
Spring2003range\04343d234
Spring2003range\04343d236
Spring2003range\04343d238
Fall2003range\04343d319
Fall2003range\04343d321
Fall2003range\04343d323
Fall2003range\04343d325
Fall2003range\04343d327
Spring2003range\04344d201
Spring2003range\04344d203
Spring2003range\04344d205
Spring2003range\04344d207
Fall2003range\04344d245
Fall2003range\04344d247
Spring2004range\04344d335
Spring2004range\04344d337
Spring2004range\04344d339
Spring2004range\04344d349
Spring2003range\04347d207
Spring2003range\04347d209
Spring2003range\04347d211
Spring2003range\04347d213
Spring2003range\04347d215
Fall2003range\04347d289
Fall2003range\04347d291
Fall2003range\04347d293
Fall2003range\04347d295
Fall2003range\04347d297
Spring2003range\04349d224
Spring2003range\04349d226
Spring2003range\04349d228
Spring2003range\04349d230
Spring2003range\04349d232
Fall2003range\04349d312
Fall2003range\04349d314
Fall2003range\04349d316
Fall2003range\04349d318
Fall2003range\04349d320
Spring2003range\04350d191
Spring2003range\04350d193
Spring2003range\04350d195
Spring2003range\04350d197
Spring2003range\04350d199
Spring2003range\04350d201
Fall2003range\04350d258
Fall2003range\04350d260
Fall2003range\04350d262
Fall2003range\04350d264
Spring2003range\04370d155
Spring2003range\04370d157
Spring2003range\04370d159
Spring2003range\04370d161
Fall2003range\04370d223
Fall2003range\04370d225
Fall2003range\04370d227
Fall2003range\04370d229
Spring2004range\04370d295
Spring2004range\04370d297
Spring2003range\04372d194
Spring2003range\04372d196
Spring2003range\04372d198
Spring2003range\04372d200
Spring2003range\04372d202
Fall2003range\04372d269
Fall2003range\04372d271
Fall2003range\04372d273
Fall2003range\04372d275
Spring2004range\04372d331
Spring2003range\04379d192
Spring2003range\04379d194
Spring2003range\04379d196
Spring2003range\04379d198
Spring2003range\04379d200
Fall2003range\04379d280
Fall2003range\04379d282
Fall2003range\04379d284
Fall2003range\04379d286
Fall2003range\04379d288
Spring2003range\04385d237
Spring2003range\04385d239
Spring2003range\04385d241
Spring2003range\04385d243
Spring2003range\04385d245
Spring2003range\04385d247
Spring2003range\04385d249
Fall2003range\04385d323
Fall2003range\04385d325
Fall2003range\04385d327
Spring2003range\04387d241
Spring2003range\04387d243
Spring2003range\04387d245
Spring2003range\04387d247
Spring2003range\04387d249
Spring2003range\04387d251
Spring2003range\04387d253
Spring2003range\04387d255
Fall2003range\04387d322
Fall2003range\04387d324
Spring2003range\04388d189
Spring2003range\04388d191
Spring2003range\04388d193
Spring2003range\04388d195
Spring2003range\04388d197
Spring2003range\04388d199
Spring2003range\04388d201

Spring2003range\04388d203
Fall2003range\04388d283
Fall2003range\04388d285
Spring2003range\04394d227
Fall2003range\04394d295
Fall2003range\04394d297
Fall2003range\04394d299
Fall2003range\04394d301
Spring2004range\04394d395
Spring2004range\04394d397
Spring2004range\04394d399
Spring2004range\04394d401
Spring2004range\04394d411
Spring2003range\04397d246
Spring2003range\04397d248
Spring2003range\04397d250
Fall2003range\04397d332
Fall2003range\04397d334
Fall2003range\04397d336
Fall2003range\04397d338
Fall2003range\04397d340
Spring2004range\04397d444
Spring2004range\04397d446
Spring2003range\04400d216
Spring2003range\04400d218
Spring2003range\04400d220
Spring2003range\04400d222
Spring2003range\04400d224
Spring2003range\04400d226
Fall2003range\04400d294
Fall2003range\04400d296
Fall2003range\04400d298
Fall2003range\04400d300
Spring2003range\04408d190
Spring2003range\04408d192
Spring2003range\04408d194
Spring2003range\04408d196
Spring2003range\04408d198
Fall2003range\04408d266
Fall2003range\04408d268
Fall2003range\04408d270
Fall2003range\04408d272
Spring2004range\04408d360
Spring2003range\04418d203
Spring2003range\04418d205
Spring2003range\04418d207
Spring2003range\04418d209
Spring2003range\04418d211
Fall2003range\04418d285
Fall2003range\04418d287
Fall2003range\04418d289
Fall2003range\04418d291
Fall2003range\04418d293
Spring2003range\04419d174
Spring2003range\04419d176
Spring2003range\04419d178
Spring2003range\04419d180
Spring2003range\04419d182
Fall2003range\04419d250
Fall2003range\04419d252
Fall2003range\04419d254
Fall2003range\04419d256
Spring2004range\04419d318
Spring2003range\04427d174
Spring2003range\04427d176
Spring2003range\04427d178
Spring2003range\04427d180
Spring2003range\04427d182
Spring2003range\04427d184
Fall2003range\04427d264
Fall2003range\04427d266
Fall2003range\04427d268
Fall2003range\04427d270