UNIVERSITY OF SOUTHAMPTON

On Probabilistic Methods for Object Description and Classification

By

Alexander Ian Bazin

A thesis submitted for the degree of Doctor of Engineering

School of Electronics and Computer Science,

University of Southampton,

United Kingdom.

November, 2006

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS
SCHOOL OF ELECTRONICS AND COMPUTER SCIENCE

Doctor of Engineering

ON PROBABILISTIC METHODS FOR OBJECT DESCRIPTION AND CLASSIFICATION

by Alexander Ian Bazin

This thesis extends the utility of probabilistic methods in two diverse domains: multimodal biometrics and machine inspection. The attraction for this approach is that it is easily understood by those using such a system; however the advantages extend beyond the ease of human utility. Probabilistic measures are ideal for combination since they are guaranteed to be within a fixed range and are generally well scaled.

We describe the background to probabilistic techniques and critique common implementations used by practitioners. We then set out our novel probabilistic framework for classification and verification, discussing the various optimisations and placing this framework within a data fusion context.

Our work on biometrics describes the complex system we have developed for collection of multimodal biometrics, including collection strategies, system components and the modalities employed. We further examine the performance of multimodal biometrics; particularly examining performance prediction, modality correlation and the use of imbalanced classifiers. We show the benefits from score fused multimodal biometrics, even in the imbalanced case and how the decidability index may be used for optimal weighting and performance prediction.

In examining machine inspection we describe in detail the development of a complex system for the automated examination of ophthalmic contact lenses. We demonstrate the performance of this system and describe the benefits that complex image processing techniques and probabilistic methods can bring to this field.

We conclude by drawing these two areas together, critically evaluating the work and describing further work that we feel is necessary in the field.

Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank my supervisors Mark Nixon and Brian Kett for their guidance and support during my Engineering Doctorate. I would also like to thank the staff at Neusciences, especially Trevor Cole, for their involvement in the industrial inspection work and for agreeing to sponsor my Engineering Doctorate. I am grateful to John Carter, Lee Middleton, Galina Veres, David Wagg, Silvia Wong and Alex Buss for their work on the DTC8.11 project that has supported so much of this thesis. Finally I would like to thank my family and friends for supporting me through these last years, particularly for tolerating me whilst writing up.

Definitions and Abbreviations Used

Biometrics Identification of a person by an observed biological or behavioural characteristic.

Posterior $P(C|x)$ Probability of data coming from a particular class in light of all data.

Likelihood $P(x|C)$ Probability of observing the data given that the data belongs to the specified class.

Evidence $P(x)$ Probability of observing data irrespective of the data's class.

Prior $P(C)$ Probability of observing a particular class irrespective of the data obtained.

Intra-class variance Variance between recorded data belonging to the same class.

Inter-class variance Variance between recorded data belonging to differing classes.

Data fusion Combination of multiple sources of information in order to make a more accurate or more robust decision.

Feature fusion Combination of data sources at the feature level, i.e. before classification.

Score fusion Combination of data sources at the score level, i.e. after classification with classifiers providing continuous outputs.

Decision fusion Combination of data sources at the decision level, i.e. after each piece of information has had a classification assigned.

Balanced classifiers Two or more classifiers where the performance of the worst classifier is no more than half that of the best classifier.

Imbalanced classifiers Two or more classifiers that do not fulfil the definition of balanced classifiers.

Score transformation Process of altering scores from disparate classifiers such that all conform to the same range and distribution.

False match rate Percentage of those impostors who are falsely matched to a client at a given threshold, also known as a false acceptance.

False non-match rate Percentage of genuine clients who are falsely identified as impostors at a given threshold, also known as a false rejection.

Equal error rate Percentage value at which the false match rate and false non-match rate become equal whilst varying the threshold.

Client Genuine enrolled user of a biometric system attempting to gain authorised access.

Impostor Malicious user of a biometric system attempting to gain access despite not having permission to do so.

Gait Unique, repeatable, observable pattern produced by a subject as they walk.

Modality Single biometric method used for identification.

Agent Part of a system that performs information preparation and exchange on behalf of a client or server.

Voxel Volume pixel, the smallest distinguishable box-shaped part of a three-dimensional space.

SQL Structured query language.

XML Extendable mark-up language.

Industrial inspection Visual based task of determining faults in manufactured goods in a production environment.

Commercial of the shelf process control device Device for interfacing industrial inspection system with the production line.

Symbols Used

Chapter 2 (First Use)

$P(C\|x)$, $P(C\|d)$	Posterior probability given x or d
$P(x\|C)$, $P(d\|C)$	Class likelihood based on x or d
$P(x)$, $P(d)$	Evidence of x or d
$P(C)$	Prior probability of class C
N	Number of classes
C_i, C_j	Class i or j
C	The client class
I	The impostor class
t	Threshold for a verification decision
R	Number of classifiers under combination
w_i	Weight of classifier i
E_i	Error rate of classifier i
Λ	Covariance of feature vectors
M	Length of feature vector (also number of training examples in eigenface technique leading to feature of length M)
σ^2	Variance of feature
Ω	Feature vector from eigenface technique
μ	Mean of feature
d	Distance between new measured feature and reference vector
K	Number of example images per class
L	Number of likelihoods produced in training set
HL	Vector of example likelihoods from training set
HF	Mapping vector to a flat histogram of likelihoods
HG	Mapping vector to a Gaussian histogram of likelihoods
$S(C)$	Degree of support to a proposition C
s_i	Evidence in support of $S(C)$

Chapter 3 (First Use)

Γ	Face image

Ψ	Average (mean) face
Φ	Difference between face image and average (mean) face
A	Matrix of training examples
u_l	Eigenface
U	Matrix of all eigenfaces
ω_i	Component of feature vector
K_j	LDA client mean
S	LDA between class scatter
Y_i	LDA class covariance
Σ_j	LDA within class scatter
a_j	Client specific Fisher face
$\hat{p}$	Measurand error
n	Total number of samples
N_s	Number of subjects
n_g	Number of samples per subject
Z_{ij}	Binary error value
p	Expected or desired error value
α	Confidence level
d'	Decidability index

Chapter 5 (First Use)

IR	Dispersion
R_{max}	Maximum chord across feature
R_{min}	Minimum chord across feature
$M1$-$M4$	Hu rotation invariant moments (1-4)
η_{pq}	Scale invariant moment order pq
μ_{pq}	Location invariant moment order pq
P_{xy}	Binary pixel value at location xy
$\bar{x}, \bar{y}$	Centre of mass of feature in direction x or y

Chapter 1

Context and Contributions

1.1 Overview of Research

Probabilistic methods are a group of classification techniques marked by the fact that their output is a probabilistic measure of similarity between the object under test and some hypothesised class or classes. The output of this kind of measure compared with a hard decision or unconstrained score has numerous advantages explained here. Initially the attraction for such a measure is that it is easily understood by those using such a system; however the advantages extend beyond the ease of human utility. For example probabilistic measures are ideal for combination with other probabilistic outputs since they are guaranteed to be within a fixed range and are generally well scaled, in certain formulations we may also transparently bias the classification in favour of certain outcomes. We are interested in examining the utility of probabilistic methods in two diverse computer vision application domains: multimodal biometrics and machine inspection.

Biometrics in the automated recognition of a subject by biological or behavioural characteristics; whilst the field is over fifty years old the majority of this work has focused on the use of single biological traits, usually termed modalities, captured in a controlled environment at short range. Recently interest has turned both to recognition at a distance and the concurrent use of multiple modalities, so called multimodal biometrics. Recognition at a distance is of obvious benefit in the era of CCTV surveillance and for covert identification; it also has significant benefits in

terms of throughput and of user acceptance where contact devices have proven unpopular for hygiene and other health related fears. However identification at a distance often suffers from occlusion and hence strengthens the case for multimodal biometrics. Investigators of multimodal biometrics have claimed significant performance improvements over individual modalities in addition to their greater flexibility. Probabilistic methods are well suited to the combination techniques used in multimodal biometrics; and in this thesis we examine the utility of these techniques, the range of their use, technical issues with their implementation and the prediction of when such techniques should be employed. In order to perform such analysis we have constructed an automated system for the collection of multimodal biometric data, this is a highly complex system incorporating significant technical challenges and considerable research effort. We describe the system, the processing algorithms used, and the modalities employed; we also describe the collection methodology and the statistical basis for these decisions.

In comparison with biometrics, machine inspection is a relatively mature field; however we have examined the sub-field of ophthalmic contact lens inspection at the request of a commercial organisation. This is field that has had little to no intervention from complex image processing techniques nor from probabilistic methods. Ophthalmic contact lens inspection involves the examination of images of contact lenses on a production line for detection and classification of any faults on the lenses. Since contact lenses are classified as medical devices the standards for fault tolerance are very tightly controlled by medical regulators and all decision must be carefully recorded for subsequent auditing; in addition since inspection takes place in a production environment processing time is strongly constrained. In this thesis we develop an automated contact lens inspection system that is suitable for use in a production environment, we demonstrate the performance of this system and describe the benefits that complex image processing techniques and probabilistic methods can bring to this field.

1.2 Contributions

This thesis documents several key contributions made to the fields of probabilistic methods, biometrics and machine inspection.

In the field of probabilistic methods these contributions may be summarised as:

1. The use of global variance estimation for homogeneous sets of classes;
2. The modelling of class likelihoods by a logistic function;
3. The formulation of the verification problem as a two class problem modelled by intra and inter-class logistic functions;
4. The demonstration of a real improvement in both equal error rate and score distribution by the use of our probabilistic framework;
5. Examination of claims of optimal weighting for probabilistic fusion.

In the field of biometrics these contributions may be summarised as:

1. The development of an automated system for the collection and processing of multimodal biometric data;
2. The examination of the use of footfall data as a viable modality for biometric verification;
3. The demonstration of performance improvements using weighted fusion on highly imbalanced modalities;
4. The examination of the effect of modality correlation on biometric fusion, and the conclusion that reduction in correlation is a good indicator of improvement in performance;
5. The demonstration that the decidability index after fusion may be accurately be predicted, and further more that calculating the maximal decidability provides an optimal weighing scheme in multimodal biometrics.

In the field of machine inspection these contributions may be summarised as:

1. The development of a system for automatically inspecting medical devices within a time-constrained environment.
2. The application of complex image processing techniques to ophthalmic lens inspection;
3. The demonstration of the reliability of our probabilistic classification framework for classifying faults in medical devices.

1.3 Document Structure

The overall structure separates the industrial inspection application from the biometrics application, starting with basic probabilistic tenets common to both application domains.

Chapter 2 provides a background to popular probabilistic techniques and methods of combining probabilistic confidence measures for different sources. It also details the theoretical background behind the probabilistic framework we have developed, which forms an underpinning for the work in the remainder of this thesis. In the final section of this chapter we perform a comparison between our probabilistic formulation and Dempster-Shafer theory.

Chapter 3 describes a broad range of novel contributory methodologies, technologies and systems for biometric recognition that have been used in the Data Information Fusion, Defence Technology Centre 8.11 (DTC 8.11) contract. Our aim in the DTC 8.11 contract, that forms the basis of much of the work in this Chapter 3, was to construct a system for collecting multimodal biometric data from subjects and automatically verify their identity. Chapter 3 explains the modalities that are targeted by the system, reviews the background to these modalities and common extraction techniques before describing in detail the extraction methods we use. It then describes the collection and verification system we have developed, focusing on the following areas: the hardware used to construct the system, the pre-processing stages carried out on the captured data, and the storage solutions for the large volume of data collected. The collection strategy for our system is explained along with the testing methodologies we use in the system, as well as detailed descriptions of specific tests.

Chapter 4 examines the simple case of whether score fusion based on our probabilistic framework is an effective method for improvement of performance. It then continues, to examine whether multimodal biometrics are an effective tool when the performance of the modalities are imbalanced. In using the weighted fusion schemes described in Chapter 2, we state that *"Equation 16 describes the optimal weights, w_i, ... where E_i is the error in addition to the Bayes error from classifier i."* in Chapter 4 we seek to explore whether this is the optimal weight when

approximated by the Equal Error Rate; we expand this to examine the role correlation may have on performance and optimal weighting. Finally Chapter 4 considers the how we may predetermine any performance improvement we may see and provide a quantitative assessment of when score fusion is of benefit.

Chapter 5 provides an overview of our developed ophthalmic lens inspection system including its interaction with the manufacturing equipment and human operators. This is more industrial in nature than the other work described in this thesis and fulfils much of the commercial focus elements of the Engineering Doctorate scheme. This high level overview describes both the inspection system and allied control and monitoring software. We then describe in detail the methods used for processing the lens image, extracting relevant feature metrics, classifying fault types and comparing these classified features with the customer's inspection standards. Finally the testing regime that has been implemented is discussed both with reference to the accuracy of the algorithms and the performance of the system as a whole.

Chapter 6 summarises the findings and contributions made in this thesis. It then discusses the further work that is desirable to complete outstanding tasks in this thesis or to provide further development of our key work.

1.4 Publications

The following publications by have ensued from this research programme:

[1] Bazin, A.I., Cole, T., Kett, B., Nixon, M.S. (2006) An Automated System for Contact Lens Inspection. In *Proceedings of 2nd International Symposium on Visual Computing (ISVC)*, In Press, Lake Tahoe, NV, USA.

[2] Middleton, L., Wagg, D. K., Bazin, A. I., Carter, J. N. and Nixon, M. S. (2006) A smart environment for biometric capture. In *Proceedings of IEEE Conference on Automation Science and Engineering*, In Press, Shanghai, China.

[3] Middleton, L., Wagg, D. G., Bazin, A. I., Carter, J. N., and Nixon, M. S. (2006) Developing a non-intrusive biometric environment. In *Proceedings of*

IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), In Press, Beijing, China.

[4] Middleton, L., Buss, A. A., Bazin, A. I. and Nixon, M. S. (2005) A floor sensor system for gait recognition. In *Proceedings of Fourth IEEE Workshop on Automatic Identification Advanced Technologies*, pp. 171-176, Buffalo, New York, USA.

[5] Bazin, A. I., Middleton, L. and Nixon, M. S. (2005) Probabilistic Fusion of Gait Features for Biometric Verification. In *Proceedings of Eighth International Conference of Information Fusion*, pp. 124-131, Philadelphia, PA, USA.

[6] Bazin, A. I. and Nixon, M. S. (2005) Probabilistic combination of static and dynamic gait features for verification. In *Proceedings of Biometric Technology for Human Identification II, SPIE Defense and Security Symposium 5779*, pp. 23-30, Orlando (Kissimmee), Florida USA.

[7] Bazin, A. I. and Nixon, M. S. (2005) Gait Verification Using Probabilistic Methods. In *Proceedings of 7th IEEE Workshop on Applications of Computer Vision*, pp. 60-65, Breckenridge, CO.

[8] Bazin, A. I. and Nixon, M. S. (2004) Facial Verification Using Probabilistic Methods. In *Proceedings of British Machine Vision Association Workshop on Biometrics*, London.

The work in [8] describes our initial attempts at constructing a novel probabilistic framework and this work is described in Appendix B. In [7] we explain our novel probabilistic framework based on logistic functions, the first such use of logistic functions in this manner, we demonstrate the performance of this approach on the gait verification task. Our work in [5, 6] extend our novel probabilistic method into multimodal biometrics, including the consideration of decidability and correlation as measures of efficacy; this is the first application directly combining the output of probabilistic classifiers for biometric verification.

We describe in [4] a new biometric modality, footfall, and the construction of a suitable sensor system. The work in [2, 3] explains our work in constructing the first system for multimodal biometric capture at a distance. Finally [1] gives an overview of our innovative industrial inspection system utilising modern computer vision techniques combined with our novel probabilistic framework.

The following publications were also produced by the author during the course of their Engineering Doctorate, but do not contribute to the content of this thesis:

Damper, R. I., Marchand, Y., Marsters, J. D. S. and Bazin, A. I. (2005) Aligning text and phonemes for speech technology applications using an EM-like algorithm. *International Journal of Speech Technology*, 8(2) pp. 149-162.

Damper, R. I., Marchand, Y., Marsters, J. D. S. and Bazin, A. I. (2004) Aligning letters and phonemes for speech synthesis. In *Proceedings of 5th International Speech Communication Association (ISCA) Workshop on Speech Synthesis*, pp. 209-214, Pittsburgh, PA.

1.5 Declaration

This thesis describes the research undertaken by the author while working within a collaborative research environment at the Information: Signals, Images, Systems research group under the support of the Engineering and Physical Sciences Research Council. This report documents the original work of the author except in: Chapter 4 which was conducted in conjunction with Drs. Lee Middleton, Galina Veres, David Wagg and Mr. Alex Buss under the Data Information Fusion, Defence Technology Centre 8.11 project, and Chapter 6 which was developed in conjunction with Mr Trevor Cole at Neusciences under contract for CooperVision.

Chapter 2

Probabilistic Methods

2.1 Introduction

Probabilistic methods are a group of classification techniques that, when comparing a piece of query data to a set of possible classifications, produce a probabilistic measurement of similarity between the query data and each class. This can be contrasted with the distance based (dissimilarity) metrics or hard classification (decision) based produced by many other popular classification schemes. There are a number of obvious advantages with the use of probabilistic methods which return class assignment with accompanying measures of class certainty, especially in areas where one may wish to combine disparate sources of data or where one would like further understanding as the confidence in a classification decision.

This chapter provides a background to popular probabilistic techniques and methods of combining probabilistic confidence measures for different sources. It also details the theoretical background behind the probabilistic framework we have developed, which forms an underpinning for the work in the remainder of this thesis. In the final section of this chapter we perform a comparison between our probabilistic formulation and Dempster-Shafer theory.

2.2 Bayesian Classification

Bayesian classifiers are perhaps the best known method of obtaining a probabilistic output from a classifier. The naïve Bayesian classifier and its variants [9, 10] have gained interest for use in face recognition [11-15]. The Bayesian classifier calculates the posterior probability, *P(C|x)*, of a class, *C*, given data, *x*. This is based on the likelihood of the data given the class, *P(x|C)*, the evidence of the data, *P(x)*, and the prior probability of the class, *P(C)*, as shown in equation 1.

$$P(C \mid x) = \frac{P(x \mid C)P(C)}{P(x)} \tag{1}$$

The prior probability, *P(C)*, is usually assigned such that all classes are equiprobable; and the evidence is taken as the weighted sum of the likelihoods over all classes:

$$P(C) = \frac{1}{N} \tag{2}$$

$$P(x) = \sum_{j}^{N} P(C_j)P(x \mid C_j) \tag{3}$$

where *i* is the number of classes. By combining equations 1, 2 and 3, this leads to:

$$P(C_i \mid x) = \frac{P(x \mid C_i)}{\sum_{j}^{N} P(x \mid C_j)} \tag{4}$$

Here, *x* will be a feature vector describing an unclassified object and C_i will be one possible class identity. In the classification task we assign the data from an unknown object to the identity C_i if the posterior probability for that class is maximum:

$$\begin{aligned} &\text{assign } x \to C_i \text{ if} \\ &P(C_i \mid x) = \max_k P(C_k \mid x) \end{aligned} \tag{5}$$

It also aids our future work to consider the more restricted verification problem, which is more typically used in biometrics. In this problem one already has a claimed class identity for an object descriptor. Here the object descriptor is an extraction of measurements of the subject and the class claim provides a claim for the identity of a single subject. This makes the process somewhat simplified with two possible classes, either: Client, *C*, where the individual subject is who they claim to be or Impostor, *I*, where they are not. The two classes can be assumed to be equally likely (in the absence of other evidence) and are mutually exclusive, *P(C|x) = 1 - P(I|x)*. Hence simplifying equation 4 we obtain:

$$P(C \mid x) = \frac{P(x \mid C)}{P(x \mid C) + P(x \mid I)} \tag{6}$$

We would then accept the individual identity claim if $P(C|x) > t$ where t is a threshold that may be chosen based on the desired security of the system.

The likelihood, *P(x|C)*, will be estimated from the distribution of x for each subject; in many examples [12, 13] this is assumed to be Gaussian distributed. The multivariate Gaussian for likelihood estimation is given in equation 7; where μ_i and Λ_i are the values for the mean and covariance of the feature vectors of class C_i.

$$P(x \mid C) = \frac{e^{-\frac{1}{2}(x-\mu_i)^T \Lambda^{-1}(x-\mu_i)}}{(2\pi)^{M/2}\left|\Lambda_i\right|^{1/2}} \tag{7}$$

Moghadden et al. [13] argue for two global distributions; one to describe the distribution of variance between measurements of the same subject (*intra-class variation*) and another to describe variation between subjects (*inter-class variation*). Liu and Wechsler's work [12] describes only intra-class variation but hypothesises that this variance is consistent across subjects, leading to the ability to estimate variance globally. In section 2.4 we evaluate the use of global covariance estimation

and the suitability of Gaussian likelihood estimation. We also investigated the statement by Liu and Wechsler that the posterior probability is not a significantly better metric than the likelihood.

2.3 Data Fusion

Data fusion may be defined as the combination of two or more feature vectors, classification schemes or identification decisions with the aim of providing a more robust estimate of class identity. Fusion may occur on one of three levels [16]: feature, score or decision. This section will give a brief overview of all methods but shall focus primarily on score fusion in a probabilistic environment.

For fusion at the feature level, combinations of feature vectors are usually produced by simple vector concatenation before being passed to a classifier of choice (usually a simple Euclidean distance classifier), this type of fusion is exemplified by Kyong's work on combining face and ears [17]. Whilst this method is simple and when used with sufficient training data should be optimal, given the usual paucity of data it is unlikely to be as effective as late fusion. This is especially true if simple classifiers are used which do not take into account the varying performance of the modalities being fused. More complex methods may involve the use of feature set selection or transformation after combination to yield the most discriminant feature vector [18]. In practice good results from feature level fusion are difficult to achieve due to incompatible feature types or unknown relationships between feature spaces. Additionally this technique does not scale well due to increasing demands for more complex classifiers and increased storage to deal with rapidly expanding dimensionality.

Decision fusion methods are attractive since they need little or no training and so cope well with the lack of data often available. Simple decision fusion rules are merely logical functions such as AND or OR, slightly more complex rules may also include weighted voting algorithms. Rank based rules [19] such as rank summation, Borda count or minimum rank can also considered decision fusion schemes. These methods are simple but do not take into account the scores underlying the initial decisions; voting rules are also prone to ambiguity especially when there are few

inputs or classes. The output of these rules is a hard classification and hence unappealing in a probabilistic framework or if further analysis may be needed.

Score fusion schemes are broadly of two types: those that regard the output of the initial classifiers as feature vectors that may be used as inputs to further classifiers [20, 21]; and those that treat the outputs of the initial classifiers as scores that may be combined using mathematical rules [19, 22-28].

Many classifiers have been used in score fusion (e.g. support vector machines, multilayer perceptrons, Bayesian classifiers, Fisher's linear discriminant analysis based classifiers and C4.5 decision trees); with Ben-Yacoub et al. [20] claiming that Bayesian classifiers or support vector machines give the greatest improvement. The major drawback of using a classifier based fusion scheme is the requirement for training; in many tasks data is at a premium, especially where many examples of one class are needed. This lack of training data makes a classifier based fusion scheme unsuitable for the applications envisaged in this thesis.

Kittler et al. [22] propose a common theoretical framework for the combination of scores based on the Bayesian decision rule (equation 5), this assumes the combination of posterior probabilities (or approximations thereof). The rules described by Kittler are shown in equations 8-12; where $P(C|x_i)$ is the posterior probability from a signal classifier and $P(C|x_1,...,x_R)$ is the posterior probability from the R fused classifiers.

Sum rule: $$P(C \mid x_1,...,x_R) = \sum_i^R P(C \mid x_i) \quad (8)$$

Product rule: $$P(C \mid x_1,...,x_R) = \prod_i^R P(C \mid x_i) \quad (9)$$

Minimum rule: $$P(C \mid x_1,...,x_R) = \min_i^R P(C \mid x_i) \quad (10)$$

Maximum rule: $$P(C \mid x_1,...,x_R) = \max_i^R P(C \mid x_i) \quad (11)$$

Median rule: $$P(C \mid x_1,...,x_R) = \underset{i}{med}^R P(C \mid x_i) \quad (12)$$

The minimum and maximum rules are approximations to the product and sum rules respectively; the median rule considers that the sum rule can be considered as computing the mean posterior probability and hence approximates this behaviour using the median as a robust estimate of the mean.

The assumption of posterior probabilities poses problems for non-Bayesian classifiers, especially when attempting to combine classifiers with disparate distributions of ranges and scores. Where fusion of non-Bayesian classifiers has been attempted then score transformation techniques have been proposed [19, 21, 29, 30] to allow these classifiers to fit into the framework proposed by Kittler. It was the intention of our research to provide a probabilistic framework for data fusion where score transformation is not necessary; for this reason these methods will not be discussed further, other than to note that this transformation may introduce further errors into the classification process.

These rules work well in the case where classifiers are balanced (the error rates from each classifier are approximately equal); however when classifiers are imbalanced, use of the sum or product rules can lead to performance that is worse than the best individual classifier [25]. One solution is the use of Behaviour-Knowledge Space as proposed by Huang and Suen [31] which formulates a look-up table to translate classifier outputs to a class label with attached confidence. This again reduces the fusion to a rank rather than score based system, and requires a very large training set to populate the knowledge space. A more straightforward solution is the use of weighted sum and product rules (Linear and Logistic Opinion Pools) proposed by Benediktsson and Swain [32], these rules are shown in equations 13 and 14.

Weighted sum: $$P(C \mid x_1,...,x_R) = \sum_i^R w_i P(C \mid x_i) \qquad (13)$$

Weighted product: $$P(C \mid x_1,...,x_R) = \prod_i^R P(C \mid x_i)^{w_i} \qquad (14)$$

where: $$\sum_{i}^{R} w_i = 1 \qquad (15)$$

The setting of the weights, w_i, still represents a training requirement but a very much smaller requirement than other trained methods discussed above. Equation 16 describes the optimal weights, w_i, as stated in [25] where E_i is the error in addition to the Bayes error from classifier i.

$$w_k = \left(\sum_{i}^{R} \frac{1}{E_i} \right)^{-1} \frac{1}{E_k} \qquad (16)$$

Other methods [33, 34] base the decision to fuse on the perceived expertise of a given classifier for a given situation. If the classifier is considered an expert then the decision of that classifier is used, otherwise fusion rules are used that reflect the confusion over the classification. "Arrogant" classifiers (those that tend towards a certain decision regardless of collaborating evidence) as described in [34] may be discounted since they do not yield outputs suitable for combination.

Also of interest are papers describing those situations where classifiers are suitable for fusion, some interesting rules of thumb appear [19]:

"1. Combining data from multiple inaccurate sensors (having an individual probability of correct inference of less than 0.5) does not provide a significant overall advantage.

2. Combining data from multiple highly accurate sensors (having an individual probability of correct inference of more than 0.95) does not provide a significant increase in inference accuracy.

3. When the number of sensors becomes large (e.g., greater than 8 to 10), adding additional identical sensors does not provide a significant improvement in inference accuracy. Note, however, that adding a new sensor type may have a very

significant impact in inference capability, because of an added dimensionality of observational data.

4. The greatest marginal improvement in sensor fusion occurs for a moderate number of sensors (i.e., one to seven), each having a reasonable probability of correct identification."

In addition Daugman [28] states that in the case of imbalanced classifiers the error rate of the weaker classifier "*must be smaller than twice the cross-over [equal error] rate of the stronger test*". However Roli et al. [25] have demonstrated that the use of the weighted sum rule does give an improvement in performance when fusing imbalanced classifiers in face recognition. We will seek to explore the accuracy of these assertions in Chapter 4 since we have the means to test these in a probabilistic multimodal setting.

2.4 Global Variance Estimation

As described in section 2.2, obtaining accurate likelihoods is dependent on good estimates of the mean and variance of class data. We are often very constrained on the availability of data, particularly multiple examples of the same class. Given a set of M dimensional feature vectors Ω from class C_i, it is trivial to calculate the mean vector, μ_i, even from small numbers of examples. However when we attempt to calculate the covariance matrix of the feature vectors Λ_i we find that unless the number of examples used to calculate Λ_i is greater than M, the covariance matrix is likely to be singular and hence unsuitable for use in calculations. Where there can be a reasonable assumption that classes have similar distributions such as in the case of biometrics; we considered the proposal by Liu and Wechsler [12] that the covariance could be assumed to be uniform across all classes and further that the covariance could be considered as a diagonal matrix of the M variances, σ^2,of the elements in Ω.

$$\Lambda = diag\{\sigma_1^2, \sigma_2^2, ..., \sigma_M^2\} \tag{17}$$

For a training set T consisting of K examples from N classes we estimate the variance by subtracting the class mean, μ_i, from each training vector belonging to class C_i.

$$d_k = \left|\Omega_k - \mu_i\right|, \quad k = 1,...,K, \quad \Omega \in C_i \tag{18}$$

This forms a set of differences $D = \{d_1, d_2, .., d_K\}$ and hence:

$$\sigma^2 = \frac{1}{K}\sum_{k=1}^{K}(d_i)^2, \quad k = 1,...,K \tag{19}$$

This then yields a non-singular outcome. Tests on the validity of these assumptions may be found in the chapter describing biometric data and systems (Chapter 3). This method would clearly be inappropriate in circumstances where both the location (mean) and structure (variance) of the data differ greatly between each class.

2.5 Gaussian Based Likelihood Models

In this section we examine the use of Gaussian based likelihood models for classification in high dimensional feature space. The multivariate Gaussian likelihood is given by equation 20:

$$P(x \mid C) = \frac{e^{-\frac{1}{2}(x-\mu_i)^T \Lambda^{-1}(x-\mu_i)}}{(2\pi)^{M/2}\left|\Lambda_i\right|^{1/2}} \tag{20}$$

We further found that with no loss of performance we may approximate the covariance matrix with the diagonal variance matrix as described in 2.4.

In order to test the use of this Gaussian probabilistic frame work, we used the Principal Component Analysis technique to extract facial feature vectors from a data set of the Notre-Dame HID database [35]. Having pre-processed the images by

centring and cropping to uniform size, transforming to 8-bit greyscale and reducing them to a 78-dimensional feature space we then constructed training, gallery and probe sets. We trained the global covariance estimate using 595 images from 119 subjects. We approximated the class means, μ_i, using a gallery of single images of 200 subjects not used in training the covariance matrix.

From equation 20 we estimated the likelihoods for 200 probe images of the subjects in the gallery using both the local and global covariance estimates. We then used equation 4 to calculate the posterior probabilities of each probe image over all classes and formed an identification decision using equation 5. For comparison we also used a Euclidean distance classifier for subject identification using the same probe and gallery sets.

The rank one recognition rate of the Bayesian classifier using a global estimation of the covariance matrix was 78%; this compares favourably with the Euclidian distance classifier with a rank one recognition rate of 61%. However, looking at the probabilistic outputs we realised that there appeared to be a significant difficulty with Bayesian classifiers that had not been reported in the biometrics literature. It appears that the likelihoods derived from the PCA data are badly scaled, spanning fifty or more orders of magnitude, due to the Gaussian functions becoming very narrow in high dimensions. Likewise the posterior probabilities from Bayes rule tend to cluster near 0 or 1. These properties could make it difficult to obtain a reasonable threshold for verification and may reduce the effectiveness of the proposed data fusion algorithms; they also remove much of the intuitive nature of probabilistic recognition. This problem appears to be not yet covered in the literature; we suspect since there has as yet been no consideration of subsequent use of the classification data, research performance has been satisfied by recognition performance alone.

2.5.1 Histogram Scaling

In order to rectify the problem of poorly scaled outputs we propose a new method for scaling the likelihoods such that they are well distributed between 0 and 1 [8]. This method is based on using histogram equalisation to map the likelihoods obtained through the multivariate Gaussian to one of three histograms: a uniform

histogram, a Gaussian histogram with a centre at 0.5, and a twin Gaussian histogram with one centre at 0.25 for impostors and another at 0.75 for clients.

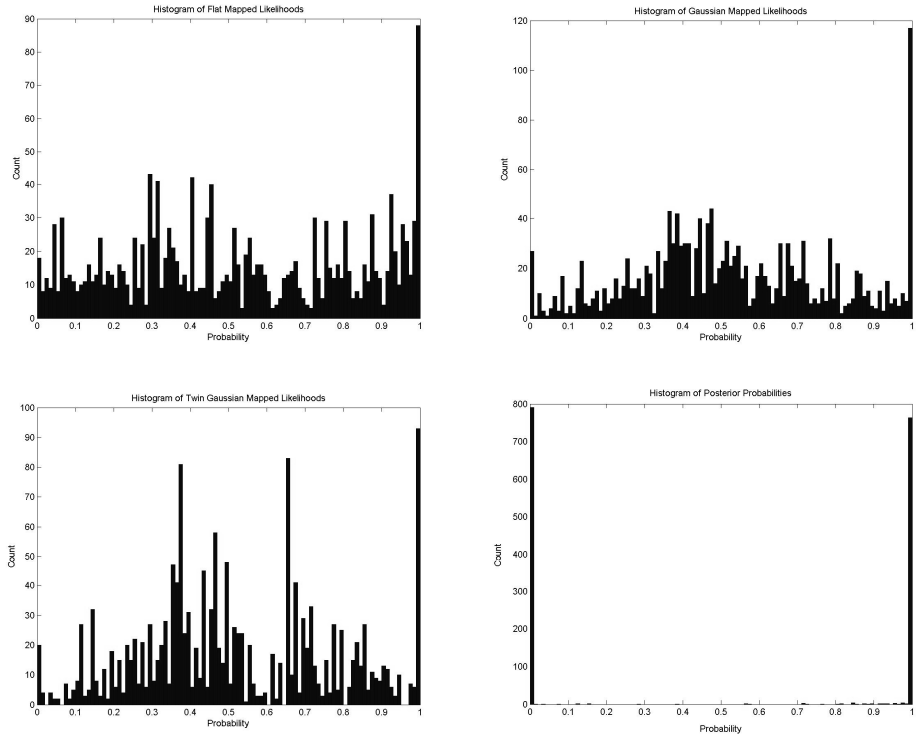

Figure 2-1 Results of Histogram Scaling of Likelihoods

Using likelihoods from clients and impostors we form an input vector of length *L, HL*, ranked in ascending order. We create two ordered output vectors drawn from flat, *HF*, and Gaussian distributions, *HG*. For the Gaussian distribution the mean, *v*, is chosen as 0.5 with a variance, *s*, of 0.25; the length of these two vectors, *J*, is set so that there may be a unique mapping between the input vector and both of the output vectors, these are constructed according to equations 21 and 22.

$$HF(j) = \frac{1}{J} j, \;\; j = 1,...,J \tag{21}$$

$$HG(j) = \frac{1}{2}\left[1 + erf\left(\frac{j - v}{s\sqrt{2}}\right)\right], \;\; j = 1,...,J \tag{22}$$

Each point in vectors *HF* and *HG* is multiplied by the number of points, *L*, in the example vector *HL* and rounded down to the nearest integer to give an index for *HL*. The value in the mapping vector, *M{G|F}*, at this point is the value of *HL* corresponding to the cumulative variance at the same point in *H{G|F}*. This provides us with a mapping from the input histogram to the output histogram, as per equation 23.

$$M\{G|F\}(j) = HL(floor(H\{G|T\}(j) \times L)), \quad j = 1,...,J \tag{23}$$

Using this method we also produce a twin Gaussian mapping where two *HG* vectors of length *J/2* are formed; the first with a centre at 0.25 and the second with a centre at 0.75. These vectors are then formed into mappings using equation 23 with the vector *HL* being drawn entirely from impostors in the first instance, and entirely from clients for the second mapping. By concatenating these two mappings we form a mapping *HT* consisting of two Gaussians trained on client and impostor data.

Again using the Notre-Dame database we performed verification between a gallery image of a subject and four probe images of the same subject, for each subject we also presented four impostor images to the system; this test was carried out across 200 subjects in all. For each verification the posterior probability, together with the flat, Gaussian and twin Gaussian mapped likelihoods were obtained giving four measures of identity for each subject. The EERs for each of these is given in Table 2-1 with the resultant class distributions shown in Figure 2-1.

Method	Equal Error Rate %
Posterior Probability	17.2
Uniform (Flat) Histogram	14.8
Gaussian Histogram	14.5
Twin Gaussian Histogram	14.9

Table 2-1 Equal Error Rates for the Histogram Scaling Experiments

From our experiment we found that the histogram techniques showed improvement in verification performance over the posterior probability method. The performances of the histogram mapped techniques are significantly better than the

posterior probability method at the 1% significance level using a McNemar's test however the performance difference between the three histogram techniques is not significant.

After careful consideration of our findings we decided that a direct method for calculating the likelihoods without any score normalisation was a more intellectually robust approach, therefore this work was abandoned in favour of the logistic based likelihood which coincidentally gave improved results.

2.6 Logistic Function Based Likelihood Models

After initial tests using Gaussian likelihood models for face recognition, we found that the Gaussian approach did not produce suitable probabilistic outputs for use in fusion. Since we had set out to construct a schema that directly produced well scaled probabilistic outputs suitable for use in fusion, it was clearly appropriate to seek an improved method for likelihood estimation that required no post processing.

When dealing with two class problems, such as biometric verification, there is a clear advantage in following the methods of Moghaddam et al. [13] in calculating the evidence based on intra- and inter-class likelihoods, since this significantly simplifies the calculations.

In keeping with our probabilistic framework we seek to find a function that tends to unity where the difference between the class mean and feature vector is zero and tends to zero as the distance between the class mean and feature vector becomes larger than the class variance.

A suitable model for these distributions is a logistic function [36] such that:

$$P(d \mid C_i) = \frac{1}{1 + e^{\frac{d - \mu_{C_i}}{b_{C_i}}}} \tag{24}$$

$$b_{C_i} = \sqrt{\frac{3\sigma_{C_i}^{\ 2}}{\pi^2}} \tag{25}$$

Once likelihoods have been calculated for all classes we may then use equation 4 to find the posterior probability for any given class.

For a two class problem such as biometric verification we may make further refinements to our method. In this case, if we have a distance, *d*, between a new feature and the feature vector of a claimed identity we wish to calculate the client and impostor likelihoods, *P(d|C)*, *P(d|I)*, given measurements of the intra and inter-class means and variances, μ_C, μ_I, σ_C^2, σ_I^2. Here we wish the client likelihood to tend to one as the difference between the new feature and reference vector tend to zero and tend to zero and the difference becomes larger than the intra-class mean. Conversely we would wish the impostor likelihood to tend to one when the difference is larger than the inter-class mean, and tend to zero as the difference nears zero. If the distributions of *d* for clients and impostors are slightly overlapping then the desired behaviour of the function and the underlying distributions are shown in Figure 2-2. The overlapping area is that where client and impostor feature sets are of similar distances from the template; it is within this region that errors occur. The amended functions are shown as equations 26-28.

$$P(d \mid C) = \frac{1}{1 + e^{\frac{d-\mu_C}{b_C}}} \tag{26}$$

$$P(d \mid I) = \frac{1}{1 + e^{\frac{-(d-\mu_I)}{b_I}}} \tag{27}$$

$$b_l = \sqrt{\frac{3\sigma_l^{\ 2}}{\pi^2}} \text{ where } l = C \mid I \tag{28}$$

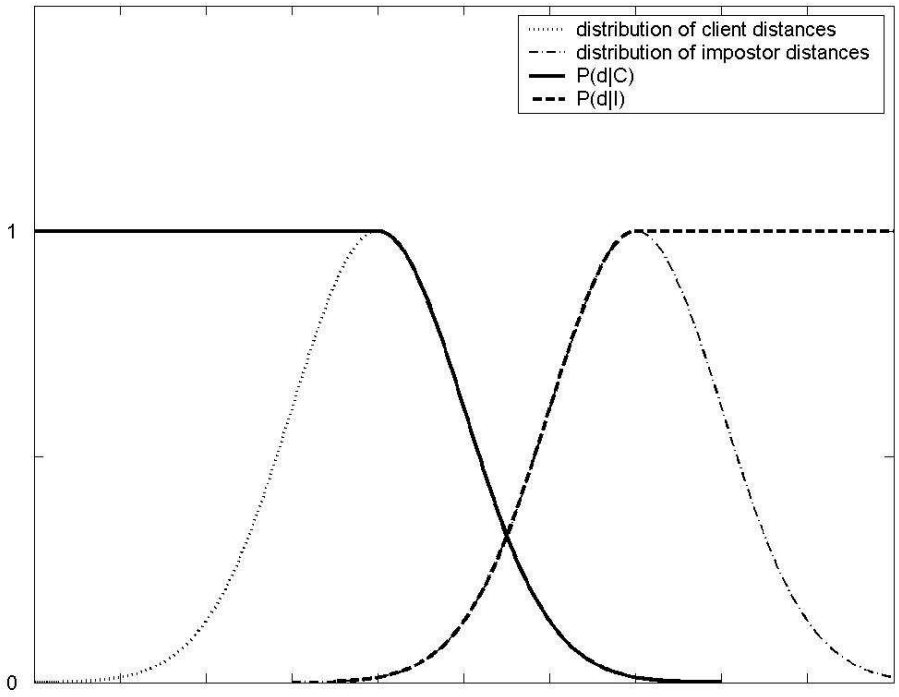

Figure 2-2 Inter and Intra Class Distributions and Likelihoods

These two functions conform to our requirements set out above that they take into account knowledge of the variations of *d*, that they are well distributed and guaranteed to produce outputs between zero and one.

Having found the intra and interclass likelihoods using equations 30 to 32 we may calculate the posterior probability *P(C|d)* from equation 6.

Importance sampling [37, 38] could also be used to describe the distribution of data, however we found that likelihood measures were sufficient to accurately describe the data so felt this unnecessary.

2.7 Dempster-Shafer Theory

Dempster-Shafer theory [39] provides an alternative probabilistic, which is claimed to include Bayes' rule as a restrictive special case. Their formulation considers a frame of discernment which is a finite set of all possible outcomes (or in

our case classes). The belief in a given possibility, *Bel(C)*, is given by combining the orthogonal sums of all pieces of evidence, m_x, in support of this possibility. Belief is described as *"the degree of support a body of evidence provides for a proposition"*, and hence is akin to our view of the posterior probability, *P(C|x)*. Given two pieces of evidence, s_1, s_2, for a possibility *C*; the degree of support *S(C)* is given by:

$$S(C) = s_1 \oplus s_2 = 1 - (1 - s_1) - (1 - s_2) \tag{33}$$

It is also noted that *n* pieces of evidence may be pooled using pairwise orthogonal sums:

$$(((s_1 \oplus s_2) \oplus s_3)...) \oplus s_n \tag{34}$$

Should instead we have two pieces of evidence pointing to conflicting beliefs such as s_1 pointing to outcome *A* and pointing to s_2 outcome *B*, then we erode the belief in both outcomes:

$$S(A) = \frac{s_1(1 - s_2)}{1 - s_1 s_2} \tag{35}$$

$$S(B) = \frac{s_2(1 - s_1)}{1 - s_1 s_2} \tag{36}$$

By way of comparison the evidence in our formulation of Bayes rule would be combined as follows:

$$S(A) = \frac{s_1}{s_1 + s_2} \tag{37}$$

$$S(B) = \frac{s_2}{s_1 + s_2} \tag{38}$$

The authors claim that their method is preferable to Bayes rule in the case of conflicting evidence for mutually exclusive classes since it retains the "*representation of ignorance*" implicit in our estimates of the support for each belief. Whilst this holds

some sway, when compared to our formulation of Bayes rule it also confuses matters in marginal cases and provides the counter-intuitive situation where our belief in a complete set of mutually exclusive possibilities does not sum to unity.

2.8 Conclusions

In this chapter we have set out the role of probabilistic techniques in classification. We have discussed the formulation of Bayes' rule, which we intend to use for our probabilistic framework, in order to yield posterior probabilities that we may make decisions on. We then describe various methods of data fusion, focusing particularly on score fusion. Having concluded that score fusion has the greatest potential for our applications, we expand on the use of mathematical rules for score combination; these rules contain the ability to weight these inputs based on classifier efficacy. The theoretical optimum for classifier weighting is briefly discussed.

Having set out background techniques we then considered two specific improvements to our probabilistic framework which dealt specifically with problems we had identified. Firstly we looked at global covariance estimation for homogeneous sets of classes in order to overcome a paucity of data. Then we considered the most appropriate likelihood model for our framework, settling on the logistic function as especially suitable for the two class problem and those applications with high dimensional feature vectors. Finally we considered an alternative probabilistic framework for combining evidence, Dempster-Shafer theory, and highlighted key differences with our framework.

In future chapters we illustrate the use of these techniques in disparate application domains and evaluate some of the claims and assumptions that we have made in this chapter.

In summary our contributions to knowledge from this chapter are:

1. The use of global variance estimation for homogeneous sets of classes;
2. The modelling of class likelihoods by the logistic function;
3. Formulating the verification problem as a two class problem modelled by intra and inter-class logistic functions.

Chapter 3

Biometric Data and Systems

3.1 Introduction

This chapter describes a broad range of contributory methodologies, technologies and systems for biometric recognition that could be used in the Data Information Fusion, Defence Technology Centre 8.11 (DTC 8.11) contract. Our aim in the DTC 8.11 contract, that has formed the basis of much of the work in this chapter, was to pioneer a secure access portal to improve building security by constructing a system for collecting multimodal biometric data from subjects and automatically verifying their identity.

We have built a 'tunnel' environment to aquire multiple biometric modalities at a distance (at least two metres; such as face and gait, rather than fingerprint or iris). The tunnel is a self contained system with automatic calibration, subject enrolment, feature capture and extraction, storage and identification. Automation was considered important in order to develop a system that could be deployed in a live environment and also to reduce the very large human burden in collecting a very large biometric database. The focus on biometrics that may be used at a distance is threefold: firstly this reduces social factors such as contact with unfamiliar devices that others have used; secondly these systems may be used covertly and possibly incorporated into existing surveillance systems; and thirdly subject throughput should be improved since no interaction with the system is necessary.

The increased possibility of occlusion or other failures to acquire accurate data when capturing at a distance is a primary concern leading to the preference of multiple modalities. By capturing a number of biometric modes we can be more confident that we have useable biometric samples to identify a subject and often will be able to combine these to improve identification performance in addition to reducing failure to acquire rates. In a mass transportation system or other large public installation, a multi-modal system also provides us an opportunity to include users who would usually be unable to use a conventional unimodal system due to disability or cultural sensitivity; since we can select only the appropriate modalities for their use whilst maintaining reliability for other users.

An equally important aim for this system was to collect a large scale multimodal database for human identification, at a distance and in a controlled environment. Collection of a large database of this kind is vital due to the lack of a single source of biometric data of this kind. This leads to poor quality evaluation data since many studies restrict themselves to small number of samples and subject created from amalgam databases where different modalities actually come from different subjects captured under differing conditions. This leads to myriad problems in effectively evaluating the data, since one cannot assume that orthogonality of data nor covariate factors are not artefacts of conflicting experimental protocols between the combined databases. Uneven protocols between merged databases also rules out many study types such as those of temporal or environmental effects.

This system is the first multimodal biometric system in the world to be based on collection of modalities at a distance, there is also contemporaneous work for distance modalities based on ‘Iris on the Move’ being performed at Sarnoff Corp [40].

We begin this chapter by describing the modalities that are targeted by the system, we review the background to these modalities and common extraction techniques before describing in detail the extraction methods we use. We will then describe the collection and verification system we have developed, focusing on the following areas: the hardware used to construct the system, the pre-processing stages carried out on the captured data, and the storage solutions for the large volume of data collected. The collection strategy for our system is explained. We finish the chapter

by explaining the testing methodologies we use in the system as well as detailed descriptions of specific tests.

3.2 Modalities

This section describes the modalities that we chose to use in the tunnel. As explained in the introduction, all of these are capable of automatic capture and extraction at a distance. For each modality we give a background to the history of the modality before discussing the technique or techniques that we have selected.

3.2.1 Face

Face recognition from still images is now some 30 years old and a number of comprehensive review papers have been written describing various techniques to extract feature vectors [41, 42]. Broadly, techniques may be split into feature based and holistic techniques, with holistic techniques in the ascendance in recent years. The baseline holistic technique for face recognition is the eigenface method proposed by Turk and Pentland [43]. This technique, based on Principal Component Analysis, transforms an image (of length N^2, in vector form), Γ, to a new lower dimensional vector, Ω. Given a set of M training images $\Gamma_1, \Gamma_2, \Gamma_3, \ldots, \Gamma_M$, the mean face, Ψ, is given by equation 39, and the difference between each training example and the mean face is $\Phi_i = \Gamma_i - \Psi$.

$$\Psi = \frac{1}{M}\sum_{n=1}^{M}\Gamma_n \tag{39}$$

Since the calculation of eigenvectors from a N^2 by N^2 matrix would be computationally impossible for typical image sizes, we use a reduced covariance matrix, A^TA, where, $A = [\Phi_1\ \Phi_2\ \ldots\ \Phi_M]$, which is of a more manageable size of M by M. We then find the M eigenvectors, v_i, of A^TA. These vectors are linearly combined by equation 40 to form the M eigenfaces, u_l, which can also be denoted as a matrix $U = [u_1, u_2, \ldots u_M]$.

$$u_l = \sum_{k=1}^{M} v_{lk} \Phi_k, \; l = 1,..., M \tag{40}$$

If we sort the M eigenvectors by their eigenvalues in descending order we may choose to only use the largest few vectors or those that account for a specified percentage of the variation. This provides a trade off between noise immunity, vector size and accurate description of the variation. When using the eigenface technique for recognition, it is useful to ignore the first eigenvector (i.e. that with the largest eigenvalue) since it typically represents variation in illumination [44].

A new image, Γ, may then be transformed to the new lower dimensional vector, Ω, where $\Omega = [\omega_1 \; \omega_2 ... \; \omega_M]$ and ω_k is calculated by equation 41. In this case M is either the original number of training images or a lower number based on the choices set out in the proceeding paragraph.

$$\omega_k = u_k^T (\Gamma - \Psi) \tag{41}$$

The eigenface technique is by no means the most effective; in comparative tests it performs 5-10% worse than the best algorithm [45-47]. However it is well understood and useful for forming a baseline test of face recognition. The eigenface technique was used in Chapter 2 to test the use of Gaussian models for distribution estimation. Other still-image techniques of interest are those with a probabilistic or Bayesian element [11-13, 48-50] and some of these methods have informed our probabilistic techniques described in Chapter 2.

Pre-processing techniques are also an important factor in face recognition systems and tools are available to enable this [51]. Face recognition from video is a more recent area, again of particular interest for our work are those using probabilistic or Bayesian techniques [52, 53].

Having looked carefully at the eigenface technique with the publicly available pre-processing tools we concluded that a more robust method was required for fusion

with our other techniques. For this we looked to the Software Development Kit from OmniPerception Ltd. which grew out of work at the University of Surrey, UK.

The OmniPerception code is based on client specific Fisher faces [54], which builds upon the work of Belhumeur [55] on Linear Discriminant Analysis for face recognition. After pre-processing with proprietary algorithms and PCA dimensionality reduction as described above, the client specific approach used by OmniPerception proceeds as follows:

C_j is the claimed identity of client j and I is the impostor class. Whilst we are dealing with the verification problem with a single claimed identity, the impostor class is built from the other enrolled users in the system during training. If each client has M_j example images projected into PCA space ($\Omega_{j1}, \Omega_{j2}, \ldots \Omega_{jMj}$) in the training set of size M, then the mean for each client K_j is given by equation 42.

$$K_j = \frac{1}{M_j} \sum_{n=1}^{M_j} \Omega_{j_n} \tag{42}$$

The impostor mean K_I based on client j can then be calculated using equation 43, where Ψ_j is the mean vector for that client. The impostor mean will stay close to the origin regardless of the client.

$$K_I = -\frac{M_i}{M - M_i} \Psi_j \tag{43}$$

The between class scatter, S_j, for the client-impostor case is given by equation 44.

$$S_j = -\frac{M_i}{M - M_i} K_j K_j^T \tag{44}$$

The covariance of the impostor class, Y_I, given by equation 46 is related to the covariance of the client class, Y_j, calculated using equation 45.

$$Y_j = \frac{1}{M_j}\sum_{n=1}^{M_j}\left(\Omega_n - K_j\right)\left(\Omega_n - K_j\right)^T \ \Omega_n \in C_j \tag{45}$$

$$Y_I = \frac{1}{M - M_j}\left[MY - M_j Y_j - \frac{MM_j}{M - M_j}K_j K_j^T\right] \tag{46}$$

The within class scatter matrix, Σ_j, is given by equation 47.

$$\Sigma_j = Y - S_j \tag{47}$$

The only non-zero eigenvector, v, can now be found directly from equation 48.

$$v = \Sigma_j K_j \tag{48}$$

Therefore the overall client specific linear discriminant transformation from pre-processed image for client j is given by equation 49, and hence is the client specific Fisher face, a_j, for this identity is the product of the eigenvector found in equation 48 and the matrix of eigenfaces from equation 40.

$$a_j = Uv \tag{49}$$

For a new pre-processed image, z, the similarity decision score, d_j, for client j may be calculated by projecting it onto the Fisher face for that client and subtracting the weighted class mean also projected onto the client Fisher face, this is given by equation 50. In the case where the total number of training examples is very much larger than the number of client specific example then the second term will tend to zero and the similarity score is simply the absolute score of the image projected onto the Fisher face.

$$d_j = \left| a_i z - \frac{M_j}{M - M_j}\Psi_j \right| \tag{50}$$

This score is then transformed, again in a proprietary way, to give a well distributed output between zero and one; the transformed score may be used directly for fusion.

3.2.2 Gait

Gait recognition is defined as the identification of a person through the pattern produced by walking. This field has produced significant interest over recent years, and through this work it has been shown that a subject's gait pattern is sufficiently unique for identification [56]. Gait has particular advantages over other biometrics: it can be used at a distance, uses no additional skills on the part of the subject, and may be performed without subjects being trained to interact with the system. All of these advantages make it particularly valuable in surveillance or security systems.

A recent review of gait recognition techniques has been produced by Nixon et al [57]. Recognition methods can be broadly divided into two groups, silhouette based techniques and model based techniques. Silhouette based techniques [58, 59] tend to offer speed and simplicity, but are only indirectly linked to gait and are difficult to normalise for noise or variations caused by covariate factors such as clothing. Model based techniques [60-62] use the shape and dynamics of gait to guide the extraction of a feature vector. Static and dynamic measurements can be extracted directly whilst the constraints of the model ensure that only plausible human shape and motion is permitted. The constraints of the model also dramatically reduce the effects of variance due to clothing or noise.

Veres et al [63] describes two silhouette based methods based on analysis of a sequence comprising one complete gait cycle. After correction for radial distortion, background subtraction is performed and a complete binary silhouette for each frame is extracted by connected component analysis and morphological operators. The silhouettes are then downsampled and normalised for height and location to give a common centre of mass. To extract a full signature the silhouettes are combined over the whole gait cycle. The average silhouette, $A_{x,y}$, is obtained by calculating the point average of the whole sequence as per equation 51:

$$A_{x,y} = \frac{1}{t}\sum_{i=0}^{t} P(i)_{x,y} \tag{51}$$

Where $P(i)_{x,y}$ is the binary pixel value at point x,y of the i-th silhouette in a gait sequence t frames in length. Usually a silhouette of 64x64 pixels is used giving a feature vector of length 4,096.

The differential silhouette, $D_{x,y}$, is formed by a differencing operation on all silhouettes in the sequence to capture motion and again yields a 4,096 dimensional vector. This is seen in equation 52.

$$D_{x,y} = P(t)_{x,y} - P(t-1)_{x,y} - \ldots - P(0)_{x,y} \tag{52}$$

Wagg and Nixon [64, 65] propose a method of model based estimation. The gait signature derives from bulk motion and shape characteristics of the subject, articulated motion estimation using an adaptive model and motion estimation using deformable contours. After extraction of the edge images via the Canny edge detector; a motion compensated temporal accumulation algorithm [66] is used to extract the bulk motion of the subject in the horizontal plane. This is then filtered using template matching, leaving only motion due to the subject. Shape estimation is then performed using a more accurate model of the subject's shape.

Articulated motion is estimated as sinusoidal models of hip, knee, ankle and pelvic rotation. These provide a starting point for model adaptation of the subject's limb movements. An adaptive process for joint location is then applied to the sequence to form a more accurate and robust model of limb movement. This adaptive process is based on an iterative gradient descent model repeated until no changes occur over the entire sequence. Example images for each of these processing stages are shown in Figure 3-1.

The processes described in [64] yield 45 parameters based on joint rotation models for the hip, knee and ankle (e.g. rotational range and period) and 18

parameters describing the subject's speed, gait frequency and body proportions (e.g. torso to leg ratio, stride period, heel to toe strike time). A further 10 parameters are extracted from the processes described in [65]. All of these parameters are normalised to make them size invariant. More recent experiments have found that for controlled environments adding height as another parameter yields an additional improvement over the height normalised feature vector, we will explore the options of treating the height parameter as an additional measure to be fused in section 3.5.2.

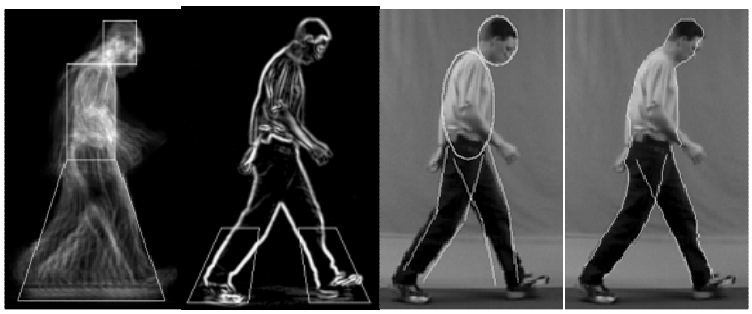

Figure 3-1 Stages of dynamic gait extraction

In our collection environment described in section 3.3 the pre-processing methods used to obtain fronto-parallel silhouettes vary from that described above, which are used for our experiments (as seen in Figure 3-2) described in section 3.5.2 and chapter 4. These differences are due to the capture of three dimensional data and the methods are fully described in section 3.3.2.

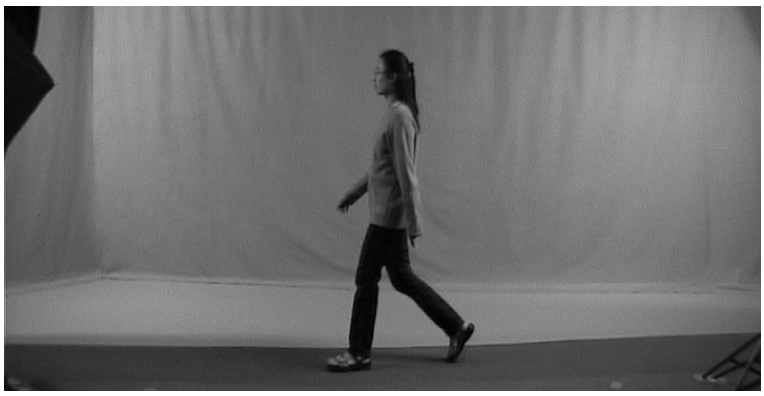

Figure 3-2 Example fronto-parallel image from a gait sequence

3.2.3 Ear

Biometrics based on ears are little explored in the literature, though three different techniques of interest appear in a review paper by Pun and Moon [67]. The first, proposed by Burge and Burger [68] describes the use of edge data from still images of ears. Edge relaxation is used to form curve segments, which are combined using a Voronoi neighbourhood graph model; finally an error correcting graph matching algorithm performs classification. The second method described in the literature is an "eigen-ear" approach [17] almost identical to the eigenface method described in section 3.2.1.

The third approach, proposed by Hurley et al. [69], uses a force field functional technique where based on the intensity of surrounding pixels and ellipse of test pixels are drawn along through a potential energy surface over the ear image until they reach potential wells. Though Hurley reports excellent recognition results using the force field functional technique, this is achieved by exhaustive template matching rather than a feature vector based match process. This significantly complicates the incorporation of this method into our probabilistic framework and so has not been progressed further at this stage. Hurley also proposes using the location of the potential wells as a feature vector, however this produces only a four dimensional feature vector and does not yield sufficiently high recognition rates to prove useable.

Because of the difficulties with the force field functional approach, we considered the use of a PCA approach for feature vector extraction. The performance of PCA dependes strongly on accurate cropping [69]; thus following on our desire to have an entirely automated processing chain we sought to devise an accurate and timely method for cropping ear images prior to PCA.

Our solution first used the Sobel edge detector to find all edges in an image containing an ear. Once an edge detected image has been obtained, we then use a Hough transform for ellipses [70] to find the ear. Since an ear in the image will usually be of a definable maximum and minimum size and within a small degree of rotation from vertical it is possible to severely constrain the search space for the Hough transform. This makes operating on high resolution images containing an ear a

tractable problem. Once the size and centre of the ellipse has been found the original image is cropped at that location. We then downsample the cropped image to a common size and perform PCA as described above. Accounting for 75% of the variation and removing the largest eigenvector to compensate for lighting variations we derive an 119 dimensional feature vector.

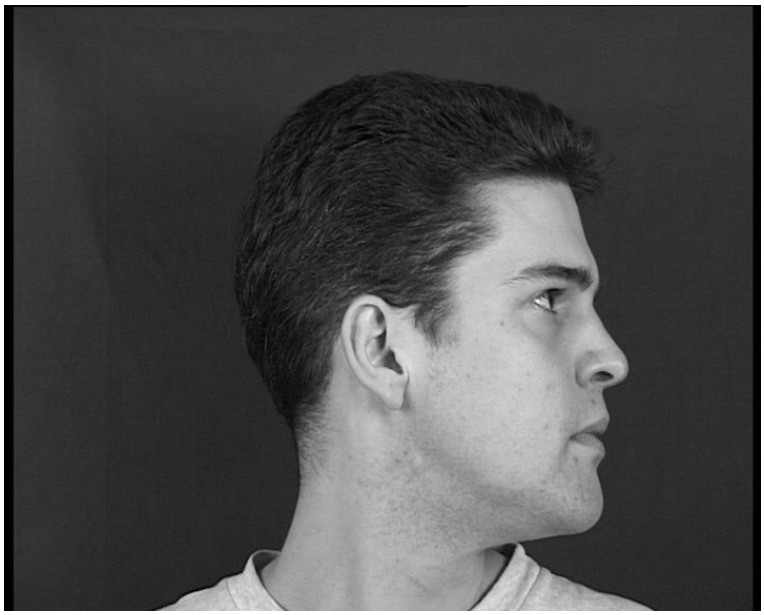

Figure 3-3 Example image for ear recognition

3.2.4 Footfall

There is strong desire to complement the use of video based gait recognition with that based on gait cadence or footfall by means of a sensor floor. Indeed the historical justification often used for gait recognition is a reference Shakespeare makes to gait cadence "*Great Juno, comes; I know her by her gait*"[1], which refers to a character offstage. A small number of sensor floor systems have been developed although few are specifically for identification. A key use of these systems is the study of pathological gait by physiologists, such as in the diagnosis of age related disease [71]; some commercial companies such as Tekscan (http://www.tekscan.com) supply systems for this end. Unfortunately these systems are prohibitively expensive for a large surface area and use proprietary interfaces that make adapting their use to

[1] Ceres in The Tempest Act 4 Scene 1 by William Shakespeare

recognition problematic. Also with a view to investigating pathological gait Reilly and Soames [72] describe a delay line based approach where a delay line lies orthogonal to a conductor carrying a pulsed current. Subjects wearing permeable shoes step on points where the delay line and conductor cross induce a large current in the delay line. This solution is unsuitable for recognition for two reasons; firstly it is unreasonable for subjects to wear special equipment to be identified and secondly the authors expressed problems in producing the delay line.

Recognitions systems (or those that may be adapted as such) have been partially developed in a research environment. Cattin uses footfall and video is his system [26]; this technique uses a sensor floor to measure the ground reaction force across an array of twelve pizeo force sensors, one at each corner of three 60cm x 60cm wooden plates. The feature vector is comprised of the windowed power spectral density of the reaction force in the range 0-20Hz. Orr [73] proposes a system based on load cells to measure ground reactive force of a single footstep for identification, ten features are extracted from the load profile and used for recognition. The ORL active floor [74] is also a load cell system, though this time a single large plate with load cells in the corner; using hidden Markov models they were also able to demonstrate recognition capability. Whilst each of these methods report reasonable recognition performance, we would also like to be able to use the sensor floor to locate the subject within the tunnel to aid with video processing.

Non-recognition systems that can inform this aim include the MIT 'Magic Carpet' developed by Paradiso [75]. In this work, grids of piezoelectric cable monitored approximately 60 times per second have been used with 10cm accuracy. Whilst suitable for their use of tracking, this lacks sufficient information to be useful for recognition. The same group then developed into the z-tile design [76] which uses twenty force sensing hexagonal tiles with an accuracy of 40^2mm. This group also examined the use of optical range finders [77] to give 40^2mm accuracy, but this solution lacks the ability to sense the subject's force profile and hence reduces the scope for biometric identification.

In designing our system [4] we wished to have a resolution of $30mm^2$, through knowledge of the mechanics of gait we calculated that a minimum sample frequency

of 7Hz was required and hence selected a frequency significantly higher than this (22Hz). Using a resistive grid, N frames are captured to form a series of binary images, I_n, where n is the frame number. $I_n(x,y)$ has the value 0 when the switch at location (x,y) is open and 1 when closed. From this image sequence we produced three footfall metrics. First we obtain the aggregate the aggregate image, A, through equation 53. This cannot be used for recognition because it is not position invariant. We then sum across A to give a footfall profile, f, as given by equation 54, where Y is the number of sensors across the width of the track; again this is not position invariant but is related to the force applied by the subject as they walk.

$$A(x,y) = \sum_{i=0}^{N-1} i_n(x,y) \tag{53}$$

$$f(x) = \sum_{y=0}^{Y-1} A(x,y) \tag{54}$$

From the footfall profile we may then find the heelstrikes, h_i, and use these to calculate the stride length, L_S, via equation 55, where s is the sensor resolution.

$$L_S = (h_{i+1} - h_i)s \tag{55}$$

Similarly we may find the frame at which each heelstrike first occurred and calculate the step time, T_s. Our final metric is the average ratio of time spent on the heel versus that spent on the toe, R_{HT}, over the M strikes recorded in the sequence. The number of frames spent on each is given by $f(h_i)$ and $f(t_i)$, hence the ratio is given by equation 56.

$$R_{HT} = \frac{1}{M} \sum_{i=0}^{M} \frac{f(h_i)}{f(t_i)} \tag{56}$$

3.3 Biometric Data Collection System

In order to construct the tunnel to capture the modalities described above, a number of key elements needed to be completed [3]. Broadly these were: the physical hardware to construct the tunnel and sensors to capture the data, the software to aggregate the data and perform pre-processing such that the modality extraction techniques performed above would work, and the storage system in order that data would be automatically labelled and could be easily retrieved for matching or further study. Each of these areas is described in detail in this section and includes key design decisions and innovations.

3.3.1 Physical Structure and Hardware

The structure of the tunnel is constructed of a lightweight aluminium framework which allows easy mounting of cameras and backdrops. The floor space within the tunnel is 3m x 3m (plus a lead in and lead out area of 1m at either end), with a 70cm wide track running down the centre. The walls are constructed with panels of 2m high fibre board built on a lightweight aluminium frame. In order to aid the pre-processing stages described in section 3.3.2 a pseudorandom non-repeating pattern covers the side walls and floor; this consists of adjoining 40cm squares of three colours (blue, grey and green, used as standard chromakey colours). The track is coloured green to reduce reflection. The tunnel is lit with six standard fluorescent lighting units positioned to give even lighting over the entire track. The current tunnel and a synthetic rendering of the tunnel can be seen in Figure 3-4.

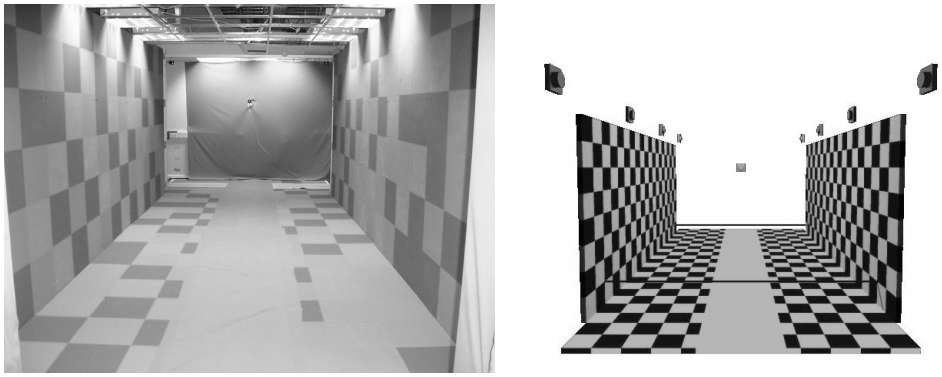

Figure 3-4 Actual and Synthetic Views of the Tunnel

The tunnel contains nine firewire video cameras running at 30fps. Eight of these have a resolution of 640x480 and are positioned equidistantly around the tunnel to capture the subject's gait, the other camera is a high resolution (1024x768) positioned at the end of the tunnel to record the face. All of these cameras are fully synchronised across multiple firewire-busses using proprietary synchronisation units (http://www.ptgrey.com/). The configuration of the eight gait cameras can be seen in Figure 3-5.

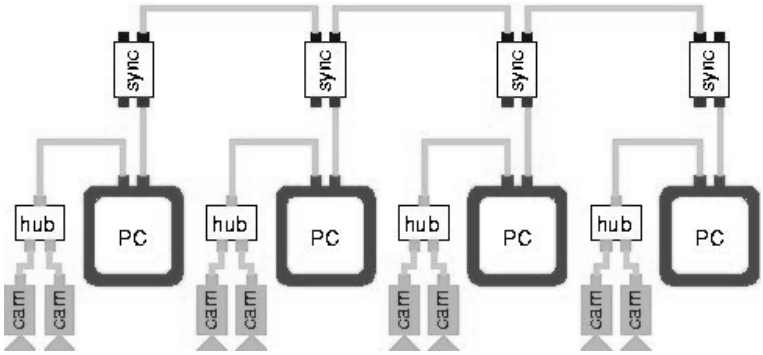

Figure 3-5 Gait Cameras Configuration

As mentioned in section 3.2.4 our sensor floor was constructed as a resistive grid. Basing our 30^2mm resolution requirement over the complete track we needed 96 by 16 sensors over the sensor floor. This is achieved by separating a grid of wires by a deformable material such as foam; when a force is applied at any point on the grid the wires at that point will come into contact, closing the switch. To read this sensor a voltage is scanned down the rows, with a voltage applied to a particular row it is read off each of the columns in turn. A microcontroller is used to control the scan and to transfer the results to a PC for processing. In order to avoid ghosting, where more than two paths could have been followed to give the same output, we used two techniques: firstly we used four electrically isolated grids with lower resolution so that multiple switches on the same grid would not be simultaneously depressed, and secondly we used a second offset layer to double resolution without risking ghosting. Fuller details may be found in [4].

Recording, processing, matching and storage of all data are performed on a cluster of eight modern PCs. Four of these are used for capture and initial processing

of gait images, one is used for the capture and processing of the facial data, one is interfaced with the sensor floor and triggering system, one is used as a system controller and server, with the final machine having been upgraded to 1.5TB of storage for archiving of data.

In order for the tunnel to be fully automated, a simple entry and exit detection system has been produced based on break beam sensors. This laser based system is controlled and monitored through the parallel port of a PC. In the recording phase of this project a barcode scanner is used to allow subjects to identify themselves to the system for automatic data labelling.

3.3.2 Software, Agents and Processing

A system diagram is given by Figure 3-6; firstly calibration is performed to ready the system for data acquisition, and then during data capture various processing tasks are performed before storage occurs as described in 3.3.3. Processing tasks are divided into two categories, local and global. Local tasks are performed entirely on a single computer using only local information, global tasks by contrast use distributed processing and disparate data locations. In order to effectively utilise available processing time and to effectively manage global tasks, a distributed architecture is used, mediated by an agent framework [78].

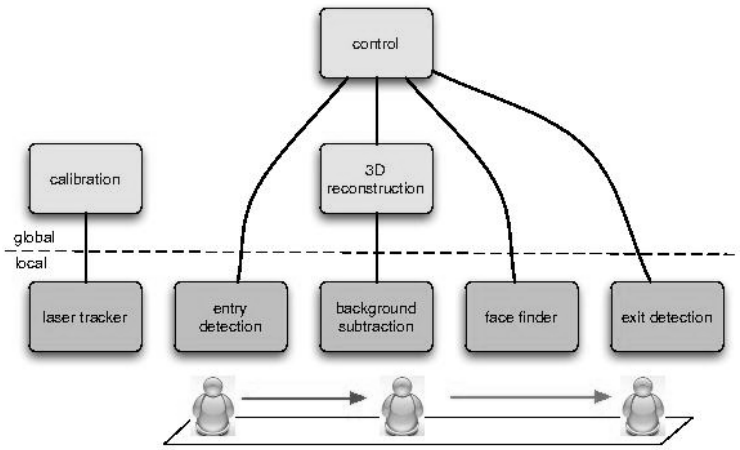

Figure 3-6 System Diagram for the Processing Stages

This agent framework has a number of features particularly relevant to imaging systems such as the tunnel environment: it is a lightweight framework, allowing the majority of CPU time to be utilised for image processing; the framework also permits locking to prevent processing from stateful devices being interrupted mid-session. It also has multi-language support (presently C++, Java and Python) allowing code reuse from previous research work. Agents are capable of automatic discovery of middleware components on a TCP/IP network and can query other agents to utilise services provided elsewhere. Agents communicate via the router which acts as a broker between agents requiring and providing services. In our framework an agent can contain one or more remote agents, allowing it to act simultaneously as both a client and server. Communication facilitated via the router is in the form of XML messaging which is used to control actions of agents and the router as well as set and read agents' input and output ports. For the transfer of large amounts of data, which would swamp the router, direct connection between agents can be initiated via streamers; streamers are again mediated by the router, though with the data passing directly between sockets on the connected agents. Agents are implemented to perform each processing task described in this section as well as other administrative, control and display tasks.

Calibration must be performed prior to any session of data collection. For 3D calibration it is necessary to find a model, K, pose, R, and position, t, for the camera. This allows the projection of 3D world space coordinates, X, into 2D image coordinates, x, (and vice versa) where:

$$x = K[R \mid t]X \tag{57}$$

Additionally correction must be made for radial distortion due to curvature of the camera lens. The distorted coordinates, x_d, are given by equation 58 and are based on the lens' optical centre, x_c, the distortion parameters, κ_i, and the distance from the optical centre, r.

$$x_d = x_c + \left(1 + \kappa_1 r + \kappa_2 r^2 + \ldots\right)x \tag{58}$$

Once all of the parameters for 3D calibration and radial distortion correction are calculated for each camera then the tunnel can be considered calibrated. Our calibration technique involves four steps: radial distortion parameterisation, intrinsic parameterisation, extrinsic parameterisation and global optimisation.

From an image of the tunnel, Figure 3-7 a), we use a Hough transform [79] on the Canny edge detected image, Figure 3-7 b), to find all long curves. By straightening these we may calculate the radial distortion terms. In most cases it is sufficient to use a single term, κ_1, for correction. Using the radially corrected image and the same set of lines, we use the vanishing points to calculate the intrinsic parameters in a manner similar to Cipolla [80]. Using knowledge of the geometric properties of the environment we may then calculate the extrinsic parameters. Finally using the spatially unique pattern on the tunnel, optimisation over all cameras may be performed by minimising projection errors between the known world points and image coordinates. Parameter extraction for single cameras are carried out by local agents with only global optimisation carried out remotely, therefore parallelising the task and reducing the load on the network. The final active volume projected onto a single image is shown in Figure Figure 3-7 c).

Once a recording has been made of a subject walking through the tunnel, background subtraction must be performed on the sequence from each camera (including the face camera). Because of we have computed the active volume for each camera during calibration, we may reduce the search space for background subtraction. To perform the subtraction we use a modified two step process [81]. The background is estimated in RGB space using the median image since this is more robust to moving objects and illumination variance. Once the background is estimated the majority of background pixels are removed by image differencing, then the remaining background pixels are removed by a process of shadow suppression. Shadow detection is performed by detecting a decrease in saturation in HSV space. All of these processes are performed locally on a frame by frame basis, and the resulting background subtracted images are sent across the network for further processing.

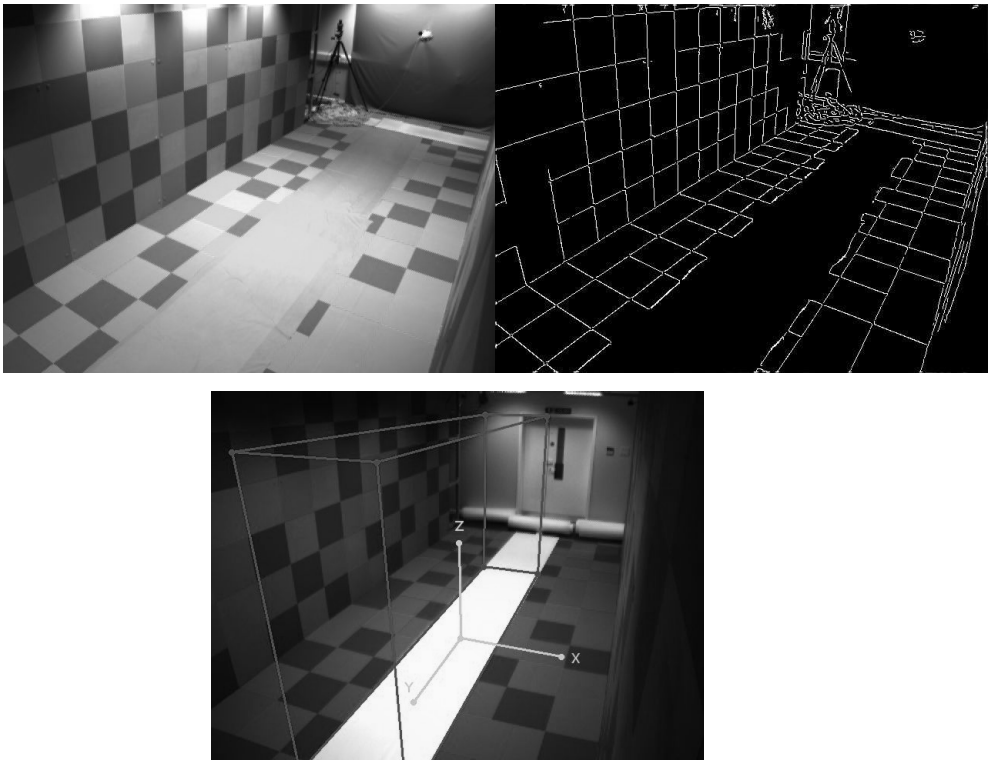

Figure 3-7 a) Image From Tunnel b) Edge Detected Image c) Active Volume Projected onto Image

Once the background subtracted images have been centrally received a 3D reconstruction of the subject must be created for each frame. We use a method termed voxel-based shape from silhouette [82] which is a well established technique for projecting multiple 2D images into 3D space. Simply this method involves the inverse of equation 57, each voxel is projected into each camera's image space. If this projection is within the 2D silhouette for all cameras then the voxel is accepted as a true point in the 3D silhouette. Obviously this is a computationally intensive task; we reduce the computational load by pre-computing the image coordinates and performing a two pass scan, one at low resolution and then another high resolution scan within the low resolution bounds.

Face detection is performed subsequently to the background subtraction stage on the face camera. The head is found by searching for a step change in the silhouette width at the shoulders, the region above this can be assumed to be the head. Further checks are then carried out to ensure that the head fits the expected anatomical model

and is well centred and of sufficient size for a good template extraction to be performed.

3.3.3 Storage

Because of the desire to use the biometric tunnel as an ongoing collection environment for biometric data there is a need to be able to re-evaluate previously collected data in the face of new techniques for pre-processing and feature extraction. It is therefore necessary to store all collected data in a lossless manner, together with relevant biographical data of the subject.

The data storage requirement per subject record is approximately 70MB for the eight gait cameras, another 7MB for the face camera, with an additional 15MB for extracted vectors and other information. This brings the total storage per run to around 100MB. Given a server capacity of 1.5TB we can store some 15,000 records.

To avoid the labelling and search problems encountered with previous databases [83] we use an SQL database to store biographical information about the subject and to point to relevant recording files and extracted templates. For the first recording of each subject, biographical information is entered via a web interface and the subject is assigned an anonymous identification number which is printed as a bar code onto a record card. For each subsequent recording the subject needs only to scan this record card to ensure that their data is correctly labelled. Run identifiers are added to file names and the database by the agents as processing stages are completed. This format may be easily exported to a flat file or XML format for distribution if necessary.

3.4 Collection strategy

Having previously constructed a large gait database [83], and with other large biometric databases having been produced [35, 84] we became interested in how much data we should collect to make the tunnel data collection statistically significant.

Veres has produced work giving a novel approach to the calculation of necessary database size which we précis here. In a database of n total samples, comprising of N_s subjects and n_g samples per subject, the error rate per sample for a given individual i will be given by equation 59. Z_{ij} is a binary value representing a recognition error for the j-th sample of the i-th subject.

$$\hat{p}_i = \frac{1}{n_g}\sum_{j}^{n_g} Z_{ij} \tag{59}$$

Therefore the total recognition error across the whole database is given by equation 60.

$$\hat{p} = \frac{1}{N_s}\sum_{i}^{N_s} \hat{p}_i \tag{60}$$

Assuming that the data is identical and independently distributed the recognition errors, Z_{ij}, can be described as Bernoulli trials. The total number of errors, s, in n trials is distributed according to the binomial distribution:

$$P_{n,p}(s) = \binom{n}{p} p^s (1-p)^{n-s} \tag{61}$$

The expected value of the error rate is, $p = s/n$ with the empirical value of the error rate on the data set given by $\hat{p}$. With a certain confidence *(1-α)* we wish the expected value of the error rate, p, not to exceed a given value.

$$p < \hat{p} + \varepsilon(n,\alpha) \tag{62}$$

Where $\varepsilon(n, \alpha)=\beta p$, hence it is fixed to a given fraction β of p. Therefore the null hypothesis is given by:

$$H_0 : p - \hat{p} < \beta p \tag{63}$$

Thus with confidence *(1-α)* we can state:

$$\text{Prob}\left(p \geq \hat{p} + \varepsilon(n,\alpha)\right) = \sum_{np-s \geq \varepsilon n} P_{n,p}(s) \leq \alpha \tag{64}$$

We are interested in finding the values of n, N_s and n_g that fulfil equation 64.

Using the Chernoff bound [85] we can state that the lower bound becomes:

$$\varepsilon(n,\alpha) = \sqrt{-2\ln\alpha}\sqrt{\frac{p}{n}} \tag{65}$$

Since we stated above that we wish to have $\varepsilon(n, \alpha)=\beta p$ we can then use equation 65 to find a value of n based on, α, and a fixed fraction, β, of, p.

$$n = \frac{-2\ln\alpha}{\beta^2 p} \tag{66}$$

Having obtained a new database achieving an error rate of $\hat{p}$ using the best matcher, we wish to test the statistical significance of this result. Under small values of p and assuming a normal distribution, the hypothesis given in equation 63 becomes that of equation 67 with $z_\alpha = \sqrt{-2\ln\alpha}$.

$$p - \hat{p} = z_\alpha \sqrt{\frac{p}{n}} \tag{67}$$

Solving for $p - \hat{p}$ we pass H_0 with confidence *(1-α)* if we meet the following test:

$$\frac{z_\alpha^2}{2n}\left(1 + \sqrt{1 + \frac{4n\hat{p}}{z_\alpha^2}}\right) < \beta p \tag{68}$$

We now calculate the number of subjects, N_s, required. We call σ^2 the inter-subject error variance, which is estimated by equation 69:

$$\sigma^2 \approx \frac{\sum_{i=1}^{N_S} (\hat{p}_i - \hat{p})^2}{N_S} \tag{69}$$

Therefore we can obtain with a confidence *(1-α)* equation 70.

$$p - \hat{p} = z_\alpha \frac{\sigma}{\sqrt{N_S}} \tag{70}$$

Thus using equation 63 we can obtain the number of subjects via equation 71, noting that σ is largely independent of n_g when $n_g \gg 1/p$.

$$N_s = \left(\frac{z_\alpha \sigma}{\beta p} \right)^2 \tag{71}$$

The number of samples per subject can be expressed as a function of γ, the ratio between inter-subject error variance, σ^2, and the intra-subject error variance, ω^2, which can be estimated from previous data. Since γ cannot be less than unity by definition $\gamma \approx \max(1, \hat{\gamma})$ and since $n' = \gamma n$ and $n' = n_g N_s$ then:

$$n_g = \frac{\gamma n}{N_s} \tag{72}$$

For collecting a new dataset we wish for the expected error rate p to be no more than 1.25 times the error rate of the best matcher with a confidence of 95%; we may use the large gait database collected at the University of Southampton [83] and successful gait extraction techniques [63, 65] to estimate other necessary values. The values needed to calculate the required size of the dataset are shown in Table 3-1. Using equations 66, 70 and 72 we may then find the required dataset size for various typical expected error rates, these are given in Table 3-2.

α	0.05
β	0.2
z_α	1.65
$\hat{p}$	0.0153
σ^2	0.0019
γ	3

Table 3-1 Values Used to Calculate Dataset Size

p	0.01	0.02	0.03
n	14975	7490	5000
N_s	1294	324	144
n_g	35	70	105

Table 3-2 Values for Number of Samples, Number of Subjects and Number of Samples Per Subject for Various Expected Error Rates

When collecting large numbers of samples per subject we must also take care to ensure that subjects do not tire during recording sessions, nor allow too much time (more than a couple of days) to elapse between recording sessions to ensure that our estimates of variance hold [86]. This presents a significant logistical challenge when constructing a large database, but previous experience indicates this is possible provided a suitable pool of volunteers is available. As mentioned in section 3.3.3 we have sufficient storage to accommodate all of the above scenarios. The particular size of the database shall be decided at a later stage and will be dependent on the performance of the modalities as well as the practicalities of collecting such datasets; in any case we will attempt to exceed the numbers given above.

For situations where the size of the population being modelled, N, is of a similar size to the sample population, N_s; we use a corrected estimate for the number of subjects, N_f. This is given by equation 73 and assumes uniform sampling of the population.

$$N_f = \frac{N_s}{1 + \frac{N_s}{N}} \tag{73}$$

3.5 Testing

Having designed our system and chosen the algorithms we wish to use; we must have the capability to test whether our system performs as we would wish, and be able to compare these results with other techniques. In this section we discuss three topics: firstly we explain a number of statistical techniques that we shall use to assess the performance of our techniques and compare them with one another; secondly we look at the performance of each modality individually; and finally we look at the performance of the collection system. Discussion of multimodal performance is left until Chapter 4.

In all testing we should be guided by the standard protocols developed for the evaluation of biometric algorithms [44, 45, 51]; of particular interest is the recently published ISO/IEC 19795-1 [87] which contains detailed recommendations for data collection, evaluation types and protocols. Some of our work in defining international standards is discussed in Appendix A.

3.5.1 Statistical tests and measures

The key metric we use in assessing performance of biometric systems is the Equal Error Rate. This is the value of the False Match Rate, when the acceptance threshold is adjusted to make the FMR equal to the False Non-Match Rate. This value may also be read directly from a Receiver Operator Characteristic curve.

It is also of interest to consider how the various classification and fusion process improve the separability of the clients and impostors; this can be measured by Daugman's decidability index [28] and is given by:

$$d' = \frac{|\mu_1 - \mu_2|}{\sqrt{\frac{1}{2}\left(\sigma_1^2 + \sigma_2^2\right)}} \tag{74}$$

Where μ_1 and μ_2 are the mean values of the client and impostor posterior probabilities respectively, and σ^2_1 and σ^2_2 are the variances for the client and impostor posterior probabilities. The decidability concerns the area of overlap between the two distributions; if this area is large, decidability is low. It is obvious that there will be a strong relationship between decidability and error rate. By way of comparison the decidability index for an experiment with a classification rate of 99.2% on 252 examples was 3.43 [69]. This is similar to the Fisher ratio but since we are simply dealing with a two class verification problem we have chosen not to consider more complex additions to this theory.

We also need to consider our ability to evaluate our methods in relationship to other techniques. Beveridge et al. [88] provide a comprehensive review of those statistics most suited for evaluating biometric systems, especially binomial theory, McNemar's tests and information about bootstrapping and sampling. Used throughout this work, McNemar's test is a sign test, based on those probes where the two classifiers fail to agree. The output of this sign test is a p-value describing the likelihood of the performance of the two classifiers being identical; when this value is below a chosen threshold we may say the difference between the two classifiers is statistically significant.

3.5.2 Modality testing

In order to asses the performance of our biometric algorithms in advance of collecting a large new dataset, we evaluated them on previously collected data which approximated the data we would collect. In this section we discuss the performance of each modality individually. The performance of all of the modalities can be seen in Table 3-3.

In assessing the face recognition algorithm we used a subset of the XM2VTS database of frontal face images [89] without any occlusion of the face. Using the inbuilt OmniPerception model for face data, together with appropriate eye spacing information we allowed the SDK to perform automated face location, feature extraction and comparison. We used four images for each of 197 subjects (listed in

Appendix B), from these images we cross compared all images of the same subject to form a client set of 1182 comparisons (six comparisons per subject excluding self-comparisons of images). We then compared each subject with six randomly selected images not from the same subject; this formed our impostor set of 1182 comparisons. We were granted access to reconstructed feature vectors that we could use with our probabilistic framework, however because of the way these are constructed these vectors perform much more poorly than the direct method described in section 3.2.1. Looking at the distribution of match scores from the direct method shown in Figure 3-8 we see that these still meet our requirements for scale and regularity (that the are well distributed across the full range of zero to one) and so may still be used for fusion in Chapter 4.

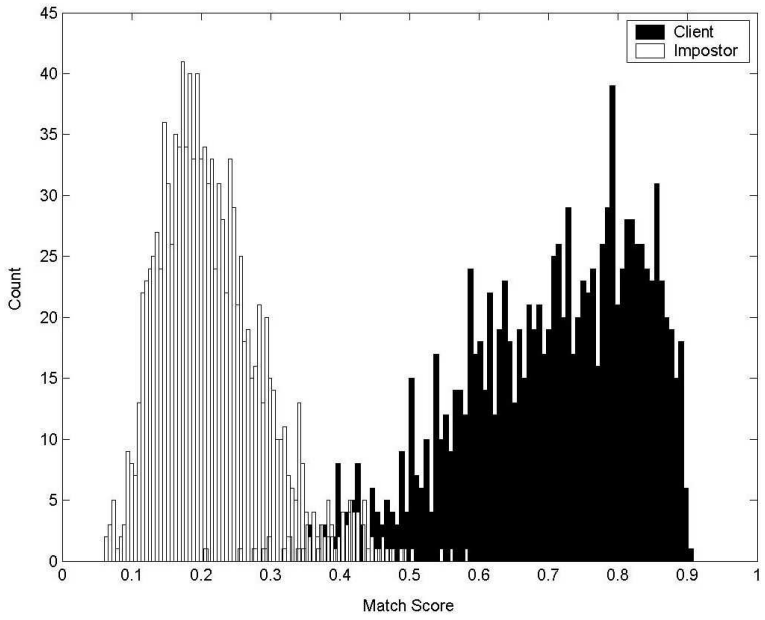

Figure 3-8 Distribution of Client and Impostor Scores for the OmniPerception Face Recognition Algorithm and Matcher

To assess the performance of our gait algorithms we used the Southampton HiD database [83] consisting of 1,079 sequences from 115 subjects walking to the left we were able to construct training, gallery, client and impostor sets; these sets were converted to the dynamic and both static feature vectors as described in section 3.2.2. The training set consisted of 145 sequences of 15 subjects that could be used to estimate the intra and inter-class mean and variance; the gallery consisted of single sequences from 100 subjects; the client set consisted of 834 sequences each matched

to a subject in the gallery set; the impostor set consisted of 834 sequences where the sequences were not matched to a subject in the gallery.

For verification we use our probabilistic framework described in Chapter 2, we also performed verification using a simple Euclidean distance classifier in order to verify the performance improvements expected by using our method. For comparison the EER for the dynamic method using a Euclidean distance classifier is 5.7%; using the McNemar's test we can see that the improvement due to our framework is statistically significant at the 95% confidence level. More importantly the distribution of match scores for the Euclidean distance classifier span five orders of magnitude and extremely poorly distributed (making setting a verification threshold extremely difficult). By contrast the distribution of match scores based on the probabilistic framework are shown in Figure 3-9 and we can see that these clearly fulfil our requirements set out in Chapter 2, that they span the full range and are well distributed.

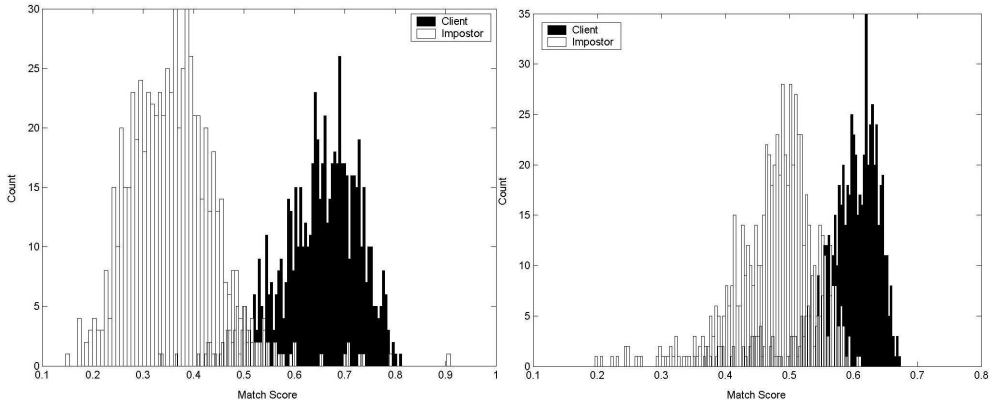

Figure 3-9 Client and Impostor Distributions for Dynamic and Static Gait Distributions Using the Probabilistic Classifer

To evaluate the automated extraction and verification from our ear recognition algorithm we again used the XM2VTS database, this time with the left most head rotation image. Using four images each of 114 subjects (listed in Appendix B) we compile a client set of 684 comparisons and by comparing client images to random images selected from clients in the dataset we produce an impostor set of 684 comparisons. The remainder of the clients are used for training. Again the probabilistic framework described in Chapter 2 is employed; the distributions of client

and impostor scores are shown in Figure 3-10. There is clearly a concern over both the performance of the algorithm and the resultant distribution of client scores, the effect of this will be considered in Chapter 4 to influence whether further work is expended on this modality. For the moment it is sufficient to note that after manual inspection of the extracted ear images, the cropping seems to be the primary difficulty in gaining acceptable performance levels. The PCA technique is (as noted in section 3.2.1) is particularly sensitive to proper centring, masking and rotation; and it therefore seems sensible to consider either a better extraction technique or less sensitive algorithm.

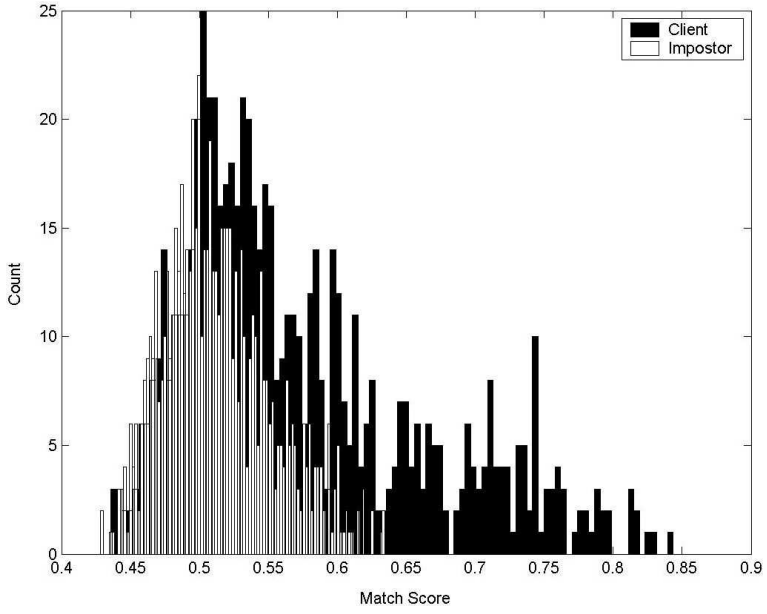

Figure 3-10 Distribution of Client and Impostor Scores for the Ear Modality

Given the novel nature of footfall sensor there did not exist a suitably large database for initial evaluation. For this reason we recorded a small initial database of fifteen subjects with eight records each. We use five of these subjects for training and the remaining ten as test data. As with the other modalities we compare all records of a subject with their other records to produced 280 client comparisons, we then compare each client record with randomly selected non-client records to produce an impostor set of 280 comparisons. Again the probabilistic framework described in Chapter 2 is employed and the distribution of scores shown in Figure 3-11. The results of the footfall sensor are promising given such a small training population and limited feature vector.

Modality	EER (%)	Decidability
Face	2.9	4.47
Gait (Dynamic)	5.2	3.40
Gait (Static 1)	14.2	1.86
Gait (Static 2)	21.6	1.61
Ear	35.4	0.87
Footfall	22.3	1.49

Table 3-3 Equal Error Rates and Decidability Indices for Modalities Under Test

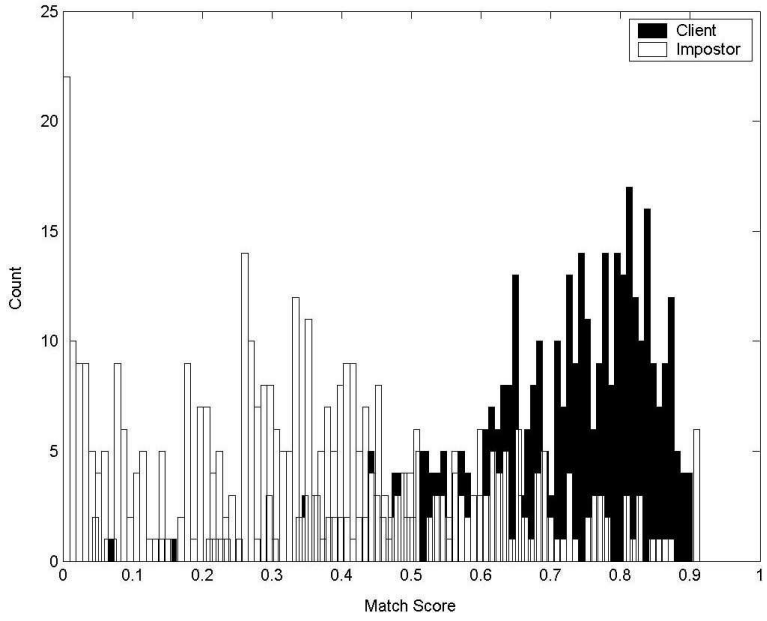

Figure 3-11 Distribution of Client and Impostor Scores for the Footfall Data

As we can see there is a great deal of difference in the performance of the various modalities, the effect of this will be fully explored in Chapter 4.

3.5.3 System testing

Whilst we have yet to collect sufficient subject data from the tunnel for meaningful recognition performance evaluation, we have performed a number of systems tests to evaluate the throughput and quality of data we can achieve. Using approximately one hundred trials we have obtained the processing times required for a single frame of data; these are shown in Table 3-4.

Component	Time (ms)
Capture	33
Background Subtraction	270
Transmission	12
Reconstruction	250
Face Finding	385
Save Image	60
Save Voxel Data	1300
Total	**2310**

Table 3-4 Timing of System Components

Given that each sequence is approximately 90 frames, this gives a processing time of about 3.46 minutes per run. This gives a throughput of 15 subjects per hour which is sufficient to record our data, alternately raw data may be saved directly to disk and processed offline whilst the tunnel is not being used (i.e. overnight). There is significant scope for more efficiency to be built into the algorithms in order to speed this process.

In addition to throughput calculations, we have manually inspected each sequence at all stages in order to spot defects. We have also performed small scale feature extraction for all modalities in order to verify that the features are of the expected format and are useable for subject verification.

3.6 Conclusions

In this chapter we have described in detail the algorithms and processes we will use for biometric data collection. We have discussed each modality that we intend to use in our system, given an overview of that modality and then given a detailed description of the algorithm or algorithms that we have chosen to implement.

We then explained the hardware and software decisions we have made in constructing our biometric collection system. Particularly we discussed the hardware

used, the role of the agent framework in producing a flexible processing system and the particular methods for pre-processing in the 3D data environment.

Finally we gave results for biometric algorithms used for subject verification on publicly available large databases in order to asses their relative performance and suitability for data collection and the multimodal biometric assessments that will be performed in Chapter 4. We have concerns over the robustness of the ear finding algorithm which appear to be causing significant degradation of expected performance for ear recognition. In a similar manner the performance of the footfall sensor is not as impressive as had been hoped; this is likely due to the simplicity of the features extracted thus far, and the scarcity of training data. We recommend that these modalities are recorded and stored when the tunnel is used for collection, but further work is invested into producing robust algorithms that are capable of comparable performance to other biometric modalities.

In the next chapter we discuss how these modalities may be used in multimodal biometrics to provide greater performance than the individual modalities. Particularly we look at the effect of weighting and classifier correlation, and consider how we may predict the benefit of fusion given the performance of the individual modalities.

In summary the key contributions to knowledge from this chapter are:

1. The demonstration of a real improvement in both equal error rate and score distribution by the use of our probabilistic framework;
2. The development of an automated system for the collection and processing of multimodal biometric data;
3. The examination of the use of footfall data as a viable modality for biometric verification.

Chapter 4

Multimodal Biometrics

4.1 Introduction

There are applications for biometric systems that require greater performance than can be achieved by a single modality system. This may be in terms of error rate, system accessibility, throughput, circumvention protection and others [90]. These various system requirements present a complex trade-off between single and multiple modality systems. Whilst we touch on some of these benefits and trade-offs in this thesis we focus primarily on the improvement in error rate that may be achieved using multimodal biometric systems.

As discussed in Chapter 2, we feel that score fusion is the most effective method of biometric fusion both in terms of performance benefit and user understanding. This chapter focuses on evaluating score fusion as a tool for making biometric systems more suitable for deployment in secure environments where a single biometric offers insufficient performance. It is then necessary to determine under what circumstances score fusion may be used to deliver a performance benefit and how best to bias the fusion process to achieve the optimal performance.

In this chapter we first examine the simple case of whether score fusion based on our probabilistic framework is an effective method for improvement of performance. We then build on this to examine whether the use of multimodal biometrics is an effective tool when the performance of the modalities are

imbalanced. In using the weighted fusion schemes described in Chapter 2, we stated that *"Equation 16 describes the optimal weights, w_i, as stated in [25] where E_i is the error in addition to the Bayes error from classifier i."* in this chapter we will seek to explore whether this is indeed the optimal weight when approximated by the Equal Error Rate; we expand this to examine the role correlation may have on performance and optimal weighting. Finally in this chapter we shall consider the how we may predetermine any performance improvement we may see and provide a quantitative assessment of when score fusion is of benefit.

Whilst wherever possible fusion is carried out on identical datasets (i.e. face and ear from XM2VTS [89] or multiple gait modalities from the Southampton Large Database [83]), due to the paucity of data as yet collected from our biometric tunnel described in Chapter 3 this has not always been possible. Where multiple datasets are needed in order to be able to examine combinations of multiple modalities, we have ensured that the same numbers of records per subject are used and these are 'matched' to create synthetic subjects based on single similar subjects in each database. This avoids most complications due to mixed subjects, though does not allow us to evaluate as entirely as we might like the effect of correlation or lack thereof between modalities. The problems arising from creating synthetic subjects are one of our prime motivations for starting to create a multimodal database as described in Chapter 3. We discuss the possible effects of synthetic subjects in our conclusion during Chapter 6.

For the reader's convenience when comparing performance of our multimodal experiments with the base performance, the performance of individual modalities is repeated here as Table 3-3.

Modality	EER (%)	Decidability
Face	2.9	4.47
Gait (Dynamic)	5.2	3.40
Gait (Static 1)	14.2	1.86
Gait (Static 2)	21.6	1.61
Ear	35.4	0.87
Footfall	22.3	1.49

Table 4-1 Equal Error Rates and Decidability Indices for Modalities Under Test (Repeated)

4.2 Fusion of face and gait

Our first task in evaluating the fusion of biometrics within a probabilistic framework was to evaluate the use of score fusion in combining two similarly performing modalities. In this case we chose to combine face and dynamic gait, using 684 impostor and 684 client comparisons derived from 114 synthetic subjects created in an amalgam of features from a subset of the XM2VTS and the Southampton Large databases.

We evaluated the performance of the fused modalities using the weighted and unweighted product and sum rules; we discard the other fusion rules discussed in Chapter 2 since they do not have the ability to be easily weighted. The weights were calculated by the using equation 16 and the Equal Error Rates measured in Chapter 3. The EER and decidability index for each modality is shown in Table 4-2, with the Receiver Operator Characteristic curve shown in Figure 4-1.

Fusion Method	**EER (%)**	**Decidability**
Static Product	0.8%	5.66
Static Sum	0.8%	5.60
Weighted Product	0.9%	5.41
Weighted Sum	0.7%	5.43

Table 4-2 Equal Error Rates and Decidability Indices for Fusion of Face and Dynamic Gait Scores

The improvement in EER seen over the best performing modality (Face, 2.9%) is statistically significant at the 5% level using McNemar's test. There is no significant difference between EER for the various combination schemes. Whilst there is some variability in the decidability indices, where a higher value indicates greater noise immunity, such small differences are unlikely to prove significant.

Having seen that we can see a significant improvement in performance by combining highly accurate sensors with static and weighted fusion rules in a probabilistic framework, we need to consider whether there is benefit to be seen from using weighted fusion with less accurate sensors.

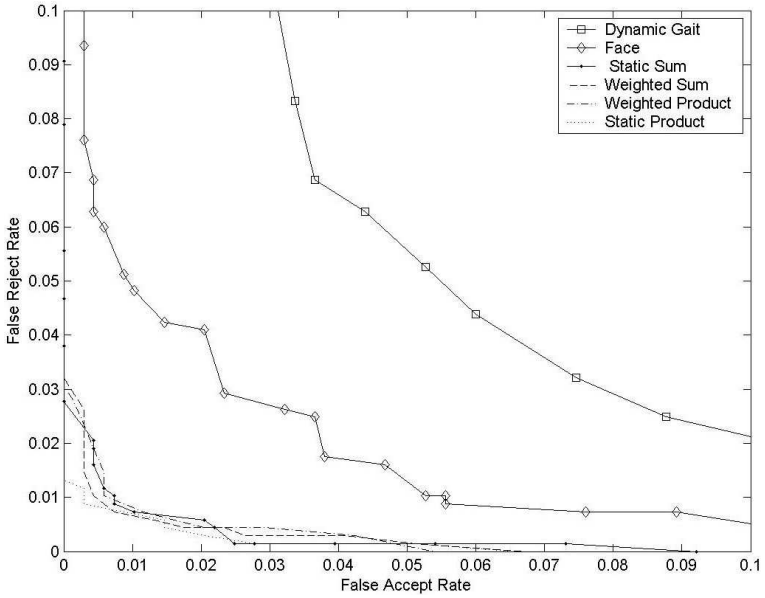

Figure 4-1 Receiver Operator Characteristic Curve for Face and Dynamic Gait Fusion

4.3 Combination of imbalanced classifiers

Having assessed the case of balanced highly accurate classifiers, we now seek to evaluate the assertion made by Daugman [28] that in the case of imbalanced classifiers the error rate of the weaker classifier "*must be smaller than twice the cross-over [equal error] rate of the stronger test*". For this examination we use the three gait modalities described in Chapter 3; we use these modalities since they are imbalanced modalities under Daugman's definition.

In order to understand the effect of weighted and unweighted fusion on imbalanced modalities we combine the best performing modality (Dynamic Gait, 5.2%) with each static modality in turn and then with both static modalities together. We determine the weights using equation 16 and the EERs determined in Chapter 3. As in section 4.2 we use 684 client and 684 impostor comparisons which are a subset of sequences from the Southampton Large database.

As can be seen from Table 4-3 the performance of fusion of highly imbalanced modalities is somewhat confused, with some fusion methods reducing the performance in certain situations. What is clear is that as the imbalance grows, or becomes more complex, the greatest benefit can then be achieved through weighting

the fusion schemes. The improvement seen in using weighted fusion is statistically significant at the 5% level for both the Dynamic & Static 2 combination and the Dynamic & Static 1 & 2 combination. Whilst these improvements seem small, we must remember that we are dealing with already effective classifiers and a reasonably large number of tests; hence this improvement is unlikely to have come from feature space noise.

	Static Product		**Static Sum**		**Weighted Product**		**Weighted Sum**	
Combination	**EER**	**d'**	**EER**	**d'**	**EER**	**d'**	**EER**	**d'**
Dynamic & Static 1	4.2	3.65	4.0	3.67	4.1	3.64	3.7	3.67
Dynamic & Static 2	6.0	3.17	6.2	3.15	4.7	3.54	4.7	3.54
Dynamic & Static 1 & 2	4.2	3.44	4.3	3.44	3.6	3.74	3.4	3.76

Table 4-3 Equal Error Rates and Decidability Indices for Combination of All Gait Modalities

This shows that the claim Daugman makes in reference to imbalanced classifiers does not hold when examining imbalanced modalities in a weighted probabilistic fusion scenario. It is likely that this claim is invalid for our experiments due to the increased flexibility of the score fusion methods we are using. We can also see proof of our claim in Chapter 3 that decidability is obviously closely linked with the error rate. From our measurements in Table 4-3 we may calculate the correlation between the decidability and EER as -0.94 which is strongly inversely correlated. This is intuitive since greater class separation gives greater noise immunity to the system.

From this experiment we can draw the clear conclusion that when fusing imbalanced classifiers it is beneficial to use our weighted probabilistic framework. We now need to determine what the optimal settings for these weights are, and whether they can be determined easily in advance.

4.4 Optimal weighting

Having shown that there is a clear benefit to using weighted combinations of modalities during score fusion of imbalanced classifiers, we now consider what the

optimal weighting scheme is for classifiers combined using the weighted sum rule. We have chosen to illustrate the weighting calculations with this rule because it performed marginally better in the previous experiment, however our tests have shown that very similar outcomes are observed using the Weighted Product rule, with very marginally higher observed EERs.

In order to asses the effect of weighting on performance we cross combined each modality in our test set creating ten separate fusion experiments. As with the experiments above we created 114 synthetic subjects that became 684 client and 684 impostor comparisons. Match scores were generated using the methods described in Chapter 3 and our probabilistic framework discussed in Chapter 2.

Using the weighted sum rule, we heuristically found the lowest EER, associated decidability index and the weighting of the strongest modality that would achieve that performance. We also found the EER and decidability at the weighing calculated using equation 16 and the EERs found in Chapter 3. These results are shown in Table 4-4. For comparison the weighting if the decidability indices were used is also shown; this is calculated using equation 75. In both cases this data is found from a training set and the tested on a separate set of subject features.

$$w_i = \frac{d_i'}{\sum_{j=0}^{N} d_j'} \tag{75}$$

Considering the results in Table 4-4 it is clear that whilst the weights calculated using equation 16 are often of similar performance to the optimal weight, there is some difference between the two. There are insufficient results to determine if these are statistically significant, however we can make some observations. Firstly we should note that (unsurprisingly) the performance of the static sum rule is better than using the calculated weight in the four cases where the optimal weight is close to 0.5. We also note that in the cases where the calculated weight performs better than the static rule the advantage gained is twice that of the converse situation.

It would appear that in the majority of cases the optimal weight biases the fusion in favour of the least accurate modality, this result can be partially achieved through the use of the decidability index for weighting as shown in equation 74. This would increase the advantage over the static rule by a further 10% whilst still retaining the principled method of pre-calculating weights, we return to the consideration of the decidability index in section 4.6.

Combination	EER (Opt)	d' (Opt)	Weight (Opt)	EER (Calc)	d' (Calc)	Weight (Calc)	Weight (d')
Face & Gait (Dynamic)	0.7	5.46	0.64	0.7	5.44	0.64	0.56
Face & Gait (Static 1)	1.5	4.80	0.50	2.2	4.64	0.83	0.70
Face & Gait (Static 2)	1.7	4.72	0.63	2.3	4.63	0.88	0.73
Face & Ear	2.4	4.31	0.53	2.7	4.50	0.92	0.83
Gait (Dynamic) & Gait (Static 1)	3.6	3.71	0.57	3.9	3.67	0.72	0.64
Gait (Dynamic) & Gait (Static 2)	4.3	3.52	0.73	4.7	3.54	0.80	0.67
Gait (Dynamic) & Ear	4.3	3.48	0.67	4.8	3.48	0.86	0.79
Gait (Static 1) & Gait (Static 2)	10.6	2.3	0.70	11.3	2.29	0.60	0.53
Gait (Static 1) & Ear	11.8	2.03	0.75	12.3	2.03	0.71	0.68
Gait (Static 2) & Ear	18.8	1.8	0.53	18.9	1.8	0.62	0.64

Table 4-4 Equal Error Rates and Decidability Indices for Optimal and Calculated Weights

4.5 Classifier Correlation

We also wish to consider how the correlation of additional modalities affects the performance of the fused modalities. We calculate the correlation using the methodology described in [91]. The correlation ρ_{nc} is given by:

$$\rho_{nc} = \frac{nN_c^f}{N - N_c^t - N_c^f + nN_c^f} \tag{76}$$

Where n is the number of classifiers under test, N is the total number of sequences (1,368), N_C^f is the number of sequences where all classifiers have an incorrect output at threshold C, and N_C^t is the number of sequences where all classifiers have a correct output at a threshold C. The paper proposes adding additional modalities in descending order of accuracy, calculating the correlation each time; only if the correlation is reduced is it acceptable to add this modality.

We chose to sequentially fuse each modality and at each point calculate optimal performance and the correlation. We also highlight the optimal and calculated weights. The results of the fusion of the 684 client and 684 impostor match scores for each modality are shown in Table 4-5.

We can see from Table 4-5 that the proposal in [91] of only adding modalities that reduce the correlation is borne out. Other smaller initial experiments performed fusing the less accurate modalities also anecdotally support these results. What is notable again is that the optimal weights for the fusion strongly bias the fusion towards the weaker classifier. Unfortunately there is not sufficient data to methodically examine the interrelation between the correlation of modalities and the optimal weighting.

Combination	EER %	d'	Correlation	Optimal Weight	Calculated Weight
Face & Gait (Dyn)	0.7	5.46	0.047	0.63	0.64
& Gait (Stat 1)	0.4	5.48	0.011	0.55	0.88
& Gait (Stat 2)	0.4	5.49	0.024	0.99	0.92
& Ear	0.2	5.30	0.003	0.73	0.95

Table 4-5 Equal Error Rates, Decidability Indices and Weightings for the Correlation Experiment

4.6 Prediction of performance

It is useful in a fusion environment to be able to a priori predict the performance of the fused system. We feel that this is most achievable by predicting the decidability index. The reason for targeting the decidability index is twofold; firstly we are convinced by experiments above that it is a good analogue for the Equal Error Rate,

secondly as we shall see below, it is trivial to predict the decidability index given good measures or estimates of the class distributions. Whilst the EER should also succumb to similar analysis, efforts thus far have been disappointing; this has generally been due to the tails of the distributions being not quite Gaussian.

Given the decidability index, equation 74, is predicated on the client and impostor match scores we can substitute for the means and variances as follows:

$$\mu_l = \sum_{i=0}^{n} w_i \mu_{li} \tag{77}$$

$$\sigma_l^2 = \sqrt{\sum_{i=0}^{n} w_i^2 \sigma_{li}^2} \tag{78}$$

where $l = C|I$.

Hence we can find the decidability index based upon the weighted fusion of the two modalities. This also provides a possible solution to our discussion on optimal weighting in section 4.4, since it is very simple to find the maximum decidability index for given client and impostor distributions across the range of possible weights. We can then choose the weight that provides the maximum decidability index as the optimal weight for fusion.

To test the accuracy of the prediction of the decidability index and the suitability of this for providing optimal fusion weights, we repeated the experiment described in section 4.4; however this time we predicted the maximum decidability index and used this to weight our fusion. The results of this experiment together with a reminder of the value of the optimal EER based on heuristic methods are shown in Table 4-6.

As can be seen there is no significant difference between the optimal and decidability weighted performance metrics. This method would be especially useful, and more efficient, if we could predict the client and impostor means and variances

for each modality directly from the probabilistic framework rather than having to produce an example set of match scores. This step is an area of ongoing research.

Combination	EER %	d' (Measured)	d' (Predicted)	Weight	EER % (Optimal)
Face & Gait (Dynamic)	0.8	5.60	5.61	0.5	0.7
Face & Gait (Static 1)	1.5	4.82	4.83	0.57	1.5
Face & Gait (Static 2)	1.8	4.76	4.74	0.71	1.7
Face & Ear	2.7	4.54	4.55	0.76	2.4
Gait (Dynamic) & Gait (Static 1)	3.7	3.71	3.87	0.56	3.6
Gait (Dynamic) & Gait (Static 2)	4.3	3.51	3.76	0.71	4.3
Gait (Dynamic) & Ear	4.4	3.51	3.51	0.76	4.3
Gait (Static 1) & Gait (Static 2)	10.8	2.31	2.45	0.66	10.6
Gait (Static 1) & Ear	12.3	2.03	2.05	0.71	11.8
Gait (Static 2) & Ear	18.9	1.80	1.83	0.56	18.8

Table 4-6 Equal Error Rates and Decidability Indices for Weights Determined by Predicted Decidability

4.7 Conclusions

This chapter has examined the role of score fusion in multimodal biometrics. We have looked at two cases to prove the value of fusion: the case of balanced modalities to illustrate that highly performing classifiers benefit from fusion; and the case of highly imbalanced classifiers to show that weighting is necessary to achieve continued improvements. Having shown the value of weighted fusion we then continued to consider the optimal weighting using the sum rule. We showed that the proposal illustrated by equation 16 is not optimal but is on balance preferable to static fusion.

We then discussed the effect of correlation on fusion of modalities. We found that as expected, the reduction in correlation was a good indicator of performance

improvement. We note that there should be some attempt at examining the interrelation between correlation and optimal weighting with a larger collection of subjects and modalities.

Finally in this chapter we have discussed the prediction of performance for fused modalities. We tackle the prediction of decidability index since this is a more tractable problem. We show that the prediction from pre-calculated class means and variances is simple and accurate. Further we demonstrate how we may pre-compute optimal weights for fusion by maximising the predicted decidability.

In summary the key contributions to knowledge from this chapter are:

1. Demonstration of performance improvements using weighted fusion on highly imbalanced classifiers;
2. Indication that the optimal weights are not given by equation 16 as stated in [25];
3. Examination of the effect of modality correlation on biometric fusion, and the conclusion that reduction in correlation is a good indicator of improvement in performance;
4. Demonstration that the decidability index after fusion may be accurately be predicted, and further more that calculating the maximal decidability provides an optimal weighing scheme.

Chapter 5

Ophthalmic Lens Inspection

5.1 Introduction

As part of the Engineering Doctorate's focus on industrially relevant research outcomes, we sought to expand the use of our probabilistic methods to other areas of interest to commercial organisations. In the work in this chapter we apply modern computing vision techniques to a computer vision application domain that has not as yet benefited from such improvements; we more importantly examine the use of our probabilistic method in an N class classification problem, showing applicability to another novel domain.

Industrial inspection is a vital part of the manufacturing process, especially in safety critical products such as medical devices. We have implemented a novel system for the inspection of ophthalmic contact lenses in a time constrained production line environment. Ophthalmic contact lenses are formed by injecting a monomer into a single use hard plastic mould which has been formed to give the required lens curvature. Once the monomer had been cured in an oven, a manufacturing machine breaks open the moulds and separates the lens from the mould base. It is then transferred to an individual window for inspection before packaging, as shown in Figure 5-1.

Due to the mechanical nature of the removal from the mould, together with occasional defects in the moulding process, lenses are prone to a number of

manufacturing defects. These include: bubbles within the monomer, splits or chips in the lens due to poor forming or damage in removal from the mould, attached monomer or rough edge due to poor removal from the mould, and contamination with particles of dust or debris. Since ophthalmic contact lenses are medical devices, the size and number of these defects must be strictly monitored and controlled. These inspection standards are laid down by government regulators and vary depending on the type and envisaged longevity of the lens.

We seek to produce a system that will perform automated inspection of ophthalmic contact lenses in a manufacturing environment. It is required to perform this inspection task at the accuracy level of a trained human operator whilst maintaining production line speeds. There have been a number of partial contact lens inspection or characterisation systems described in the literature [92-94], as well as fault detection systems for other lens types [95]. However none of the systems described in the literature report the accurate fault detection and performance required for this system. Additionally we could find no published ophthalmic inspection systems using probabilistic classification techniques or complex image processing methods.

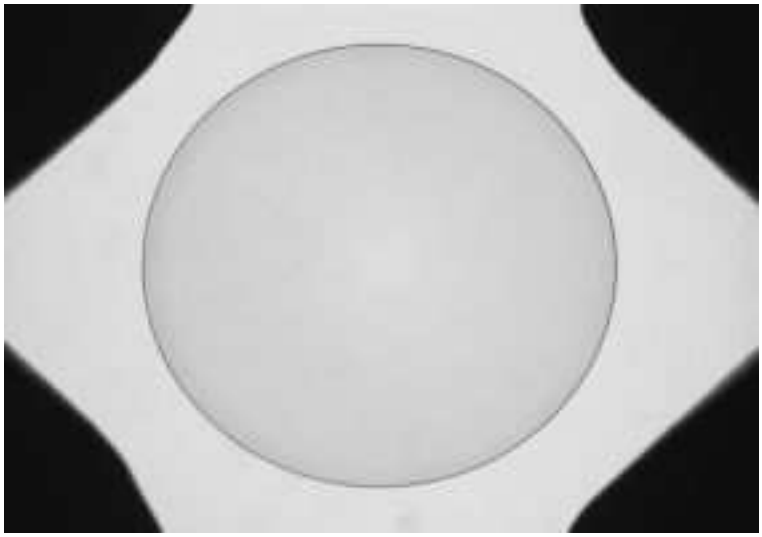

Figure 5-1 An Example Lens Image From the Inspection System

This chapter firstly provides an overview of the developed system including its interaction with the manufacturing equipment and human operators. This high level

overview describes both the inspection system and allied control and monitoring software. We then describe in detail the methods used for processing the lens image, extracting relevant feature metrics, classifying fault types and comparing these classified features with the customer's inspection standards. Finally the testing regime that has been implemented is discussed both with reference to the accuracy of the algorithms and the performance of the system as a whole.

5.2 System overview

The system is divided into two distinct processes designed to be run on separate machines. This allows monitoring and reporting to be separated from inspection; enabling remote working and multiple inspections to be running in parallel. The image processor contains modules to perform the full range of inspection activity and a separate process is instigated for each camera. On a single manufacturing line it is anticipated that there will be multiple cameras (and hence image processors) inspecting lenses in parallel. Provided there is sufficient processing power it is not necessary that this translates into a one image processor per CPU requirement; this decision would be taken after fully considering both the desired performance of the software and the hardware specification of the servers available. These multiple image processors are designed to be under the control of a single workstation process, run on a separate machine. The workstation process is responsible for set-up, display and reporting for the system. This workstation connects to the image processors remotely via TCP/IP and hence those deploying or monitoring the system do not need to be co-located with the manufacturing line.

This chapter focuses primarily on the function of the image processor software; however we believe it is useful for the reader to understand the operation of the full system and its interaction with the wider manufacturing environment. A system diagram for our working prototype system is shown in Figure 5-2.

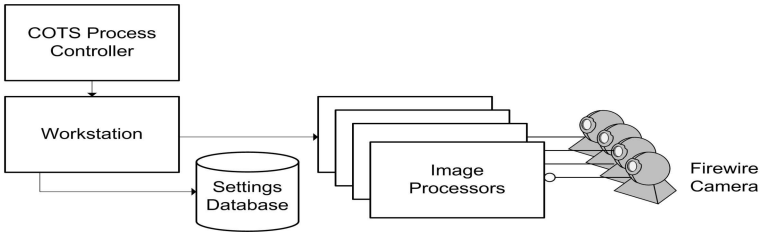

Figure 5-2 System Diagram for Inspection System

Before a new 'batch' of lenses is to be inspected, the user must initialise the system. This involves firstly choosing which process modules are to be used, adjusting the settings for each module, loading stored initialisation files and creating the inspection standards for the lenses to be compared against. In the first instance these setups will be created by a supervisor and on subsequent runs the operator will simply select the appropriate setup for the type of lens on the manufacturing line.

Once the system has been set up it may begin inspecting lenses. On the manufacturing line, once the lenses have been removed from their moulds but prior to being placed into packaging, they pass below high resolution grey-scale cameras where an image of the lens is captured for inspection. The timing of this process is synchronised with the production process and is controlled by a Commercial Off The Shelf (COTS) process control device. This device tells the servers when a lens is under the camera and ready to be inspected, triggering the image processor to acquire the image and begin inspection. Whilst the image processor is inspecting the acquired image the process controller monitors the elapsed inspection time to avoid schedule overrun, should an overrun occur a signal is sent to abort the inspection of that image and reject the lens (in these cases the image would be queued for an offline inspection to diagnose the system fault that may have occurred). In the typical case where inspection is successfully completed within the stipulated time, the process controller is informed of the pass/fail decision and the lens is either transferred to packaging or rejected as appropriate. The pass/fail decision as well as relevant statistics (feature counts and sizes etc) are passed via XML to the workstation for collation and reporting. A flow diagram of the system is shown in Figure 5-3.

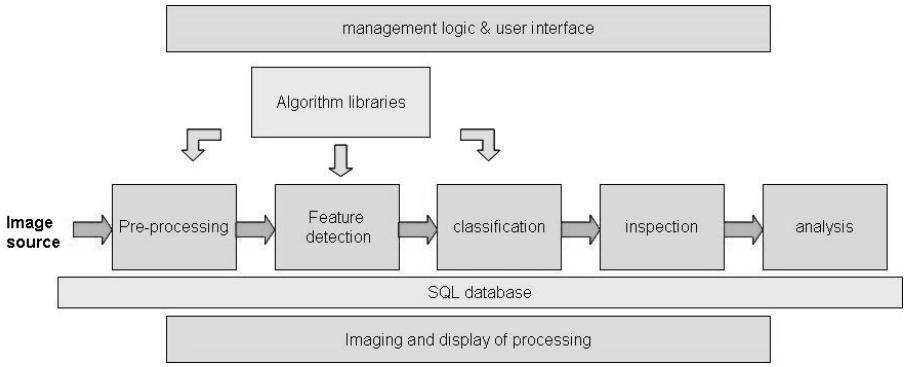

Figure 5-3 Flow Diagram of Inspection System

After a run is completed, the operator can use the workstation to review the fault profile for that run and may reprocess any images that timed out in order to diagnose the system fault that caused this. During the run the workstation can be used to monitor the current and historic yield and identify recurring faults that may be indicative of a systemic manufacturing problem.

5.3 Modules

In order to maximise future flexibility the image processor is divided into separate modules. Each module typically implements one task or algorithm with a well defined set of inputs and outputs. This design methodology allows new techniques or additional functionality to be quickly added to the system. Each of the modules developed for the current system are described in this section.

5.3.1 Image and lens pre-processing

This module comprises of a number of algorithms which must be performed immediately after image acquisition to make the lens image ready for feature detection and further processing. Before processing of the lens occurs a check is made on the image at a number of points where clear background is expected to be visible. The mean intensity and standard deviation for each patch is calculated and compared to standard values. If these patches diverge from expected values then this highlights either an obstructed view (i.e. debris on the window) or a failing illumination source or camera. Should more than one patch fail this check then the processing line is halted and the operator is warned of this problem.

The initial processing step is calculating the centre of the lens. This is achieved by detecting the edge transition at spaced points around the lens. Once a number of points have been found then the centre may be converged upon in an iterative process using simple trigonometry. If no centre can be reliably found the software concludes that the lens is either not present or is suffering from some gross defect; in either case the steps described in below and in sections 5.3.2 through 5.3.6 are not performed and instead the algorithm described in section 5.3.7 is invoked.

Having detected an accurate centre for the lens it is now necessary to fit appropriate ellipse parameters to describe the edge. Initially we considered using an active contour approach [96], however this proved overly complex for the regular shape of the lens. Direct fitting [97] and Constrained Hough Transforms [98, 99] were also judged computationally inefficient. The method found to be both sufficiently accurate and efficient was the Randomised Hough Transform which has been variously described [100, 101]. Since the normal size and shape of the lens will be known for any given batch of lenses and given that the centre has already been accurately calculated, it is possible to strongly constrain the RHT to converge on accurate parameters very rapidly.

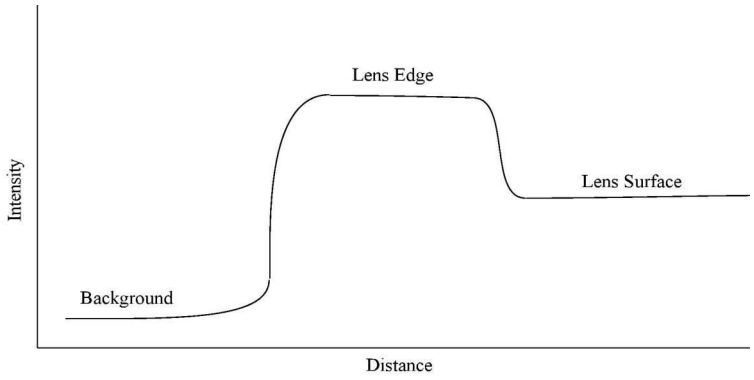

Figure 5-4 Intensity Profile for a Lens Edge

Once the centre and ellipse parameters have been accurately estimated, the real outer and inner edges of the lens are extracted. This is achieved by finding the transition from the darker edge to the lighter inner lens (the inner edge) and from the

darker edge to the much lighter background (the outer edge); a likely edge intensity profile can be seen in Figure 5-4.

5.3.2 Surface feature detection

The system defines the surface area as a circular region covering the centre 90% of the lens. It is in this region that surface features are searched for, the special case where a feature extends between the surface and edge region is dealt with as part of section 5.3.3.

To find surface features of interest, a modified Canny operator [102] with a 5x5 window is run over the entire surface region. In order to prevent small gaps creating multiple features out of a single poorly defined feature, the hysteresis thresholding stage is allowed to consider pixels in a 5x5 neighbourhood rather than simply adjoining pixels. The Canny operator produces a binary image of feature points that may be of interest.

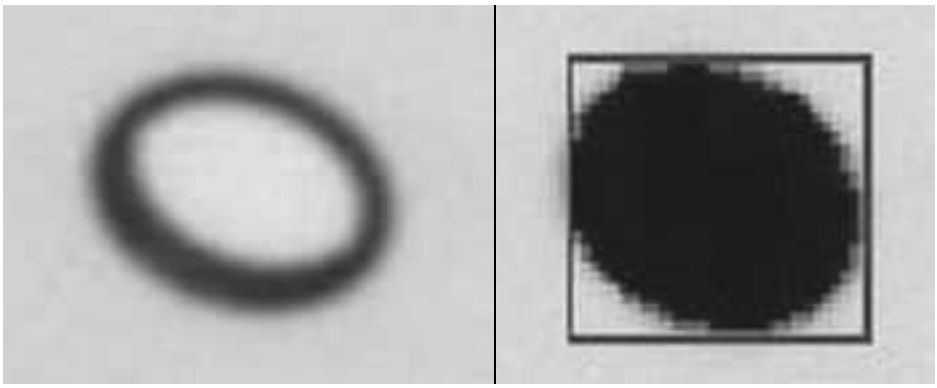

Figure 5-5 Bubble in Monomer Before and After Extraction

Once the Canny operator has been used; spatially separate features are extracted for feature description. Starting with the uppermost pixel in the surface region we scan left to right working progressively downwards until we find a pixel that has been marked by the Canny operator as a feature pixel. This then becomes a seed point for a new feature. Any feature points within a 5x5 neighbourhood of this pixel are also added as seed points for the feature and their neighbourhood is examined. This neighbourhood search continues iteratively until no neighbouring pixels remain. The

scan for feature pixels then continues until a new feature pixel is found and the extraction of neighbours is repeated to yield another feature. This is repeated until all surface features have been extracted into separate array lists containing the pixel locations. An example of a found feature (bubble in the monomer) before and after processing, as described in this section and in section 5.3.4, can be seen in Figure 5-5.

5.3.3 Edge feature extraction

The edge region of the lens is defined as an annulus covering the outer 10% of the lens and for our purposes we also consider a small region outside of the outer edge to search for debris attached to the lens edge.

Using the extracted ellipse parameters we 'unwrap' the annulus to form a rectangular image. This is achieved by mapping the ellipse points onto the midline of a rectangular image and work (radially) outwards and inwards from this line to translate the remainder of the edge region. Pixels with insufficient resolution to be uniquely translated are interpolated from neighbouring pixels. Having formed the unwrapped image we then perform checks along the outer and inner edge to find small edge faults. These tests look for trends in the spacing between the outer and inner edges, absolute deviation of the edge from the fitted ellipse and variations in intensity within the region bordered by the outer and inner edges. For illustration Figure 5-6(a) shows an edge fault in the original image and Figure 5-6(b) shows the same edge fault after extraction in the unwrapped edge image.

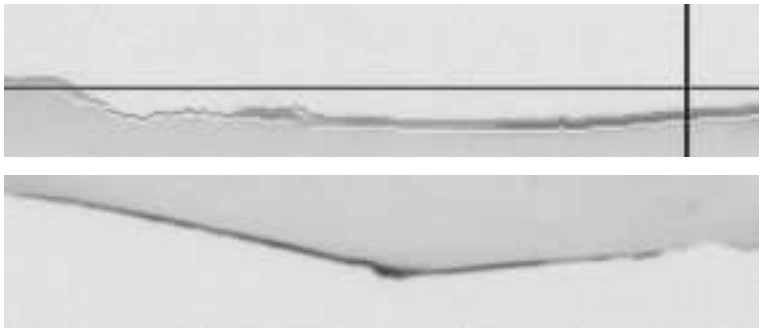

Figure 5-6 Edge Fault (a) Before Extraction (b) After Extraction

Rule based heuristic checks for faults are performed; subsequently features in the edge band are extracted in the same manner as described in section 5.3.2 with one important exception. If a feature extends into the surface the edge feature extractor searches the interface region for connecting features and merges these into one, this is done in an iterative manner to ensure that large features are fully connected rather than appearing as several smaller features.

5.3.4 Feature description

Once we have a set of features, all stored as ordered array lists of pixels we process each feature to extract mathematical descriptors for classification.

We first extract the perimeter of the feature (i.e. identify those pixels that fully enclose the feature). We achieve this by starting with the upper left pixel of the feature and progressing in a clockwise direction to find the next neighbouring pixel. By structuring our neighbour search in a clockwise direction we can guarantee that we always find the outermost neighbouring pixel.

Having extracted the perimeter of the feature we then fill it for use in further mathematical descriptors. The fill is performed by working clockwise and filling between the perimeter in either an upwards or downward direction as appropriate. Checks are made to ensure the perimeter is not a single line at this point to ensure that the fill does not escape the feature perimeter.

Given two collections of pixels, one representing the perimeter and another representing the filled feature we can then extract mathematical measures of the shape for classification. We firstly calculate gross shape measures: perimeter length (P), area (A), maximum chord (R_{max}), minimum chord (R_{min}), dispersion (IR) and compactness (C) [103]. Where:

$$IR = \frac{R_{\max}}{R_{\min}} \tag{79}$$

$$C = \frac{4\pi A}{P^2} \tag{80}$$

Perimeter and area are simply the size of the relevant array lists, with the maximum and minimum chords quickly determined by simple geometric operations. Compactness (80) is a measure of the perimeter relative to the area and dispersion (79) is the ratio of the largest circle enclosed by the feature to the smallest circle enclosing the feature.

More complex measures are produced by calculating the first four rotation invariant moments (*M1-M4*) [104] given by equations 81-84, with η_{pq} is given by equation 85. These moments are invariant to position, size and rotation.

$$M1 = \eta_{20} + \eta_{02} \tag{81}$$

$$M2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \tag{82}$$

$$M3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \tag{83}$$

$$M4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \tag{84}$$

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}} \text{ where } \gamma = \frac{p+q}{2} + 1 \tag{85}$$

$$\mu_{pq} = \sum_{x=1}^{N}\sum_{y=1}^{M}(x - \bar{x})^p (y - \bar{y})^q P_{xy} \tag{86}$$

In equation 86, P_{xy} is a binary value denoting the presence of a feature pixel at position x,y with $\bar{x}$ and $\bar{y}$ being the centre of mass of the feature on the respective axes.

We also extract information about the grey-scale intensity of the feature; mean intensity and standard deviation. Additionally features in the edge region have

Boolean information appended to describe their position in the region and whether they extend outside of the lens or into the surface region.

5.3.5 Feature classification

Once we have extracted mathematical information to describe our feature we then must classify which fault type the feature most closely resembles. To simplify this we split the features into three types based on their location within the lens: surface feature, edge feature, and surface features in edge band. We do this in order to remove implausible classification possibilities from the set of outcomes and because the surface and edge features have different feature vectors due to additional Boolean tests on the edge. The list of fault types is shown in Table 5-1.

Edge Faults	**Surface Faults**	**Non-Fault**
Particle	Surface Split	Particle in Saline
Hole	Particle	
Flash	Hole	
Rough Edge	Blemish	
Edge Chip	Scratch	
Edge Particle	Distorted	
Edge Split		
Blemish		
Scratch		
Distorted		
Thick or Thin		

Table 5-1 Fault Types for Edge and Surface Classification

The two groups of surface features are classified using the probabilistic framework described in Chapter 2. This is modified to perform a classification task rather than a verification task as described in the previous chapters. Given a feature vector of a suspected fault, d, we model the likelihood for each class, $P(d \mid C_i)$, using the logistic function given by equation 87.

$$P(d \mid C_i) = \frac{1}{1 + e^{\frac{d - \mu_{C_i}}{b_{C_i}}}} \tag{87}$$

Once all class likelihoods have been calculated we then combine these likelihoods using equation 88 to yield posterior probabilities for each class. It would be possible to include prior probabilities in equation 88 to take into account the frequency of observed faults, though we have not done so in our prototype system.

$$P(C_i \mid x) = \frac{P(x \mid C_i)}{\sum_{j}^{N} P(x \mid C_j)} \tag{88}$$

The fault classification is then determined by equation 89.

$$\begin{aligned} &\text{assign } x \rightarrow C_i \text{ if} \\ &P(C_i \mid x) = \max_k P(C_k \mid x) \end{aligned} \tag{89}$$

This classification system provides a probabilistic method for determining fault types and indicating to the system the confidence in the decision. The probabilistic outputs can be used to identify uncertain classifications which may require manual intervention or increased training.

The edge classifier is implemented as a C4.5 decision tree [105] trained to identify those faults that may be found in the edge region and other non-fault artefacts that may also be detected. A different implementation to the surface classifier was used due to the Boolean values in the edge feature vectors, making a Bayesian classifier unsuitable. The classifier is implemented as a java bean from Neuscience's NeuJDesk range, and is trained offline using hand labelled faults that have been extracted in the manner described in sections 5.3.1 through 5.3.4.

5.3.6 Inspection standards comparison

As discussed in section 5.1 there exist strict criteria for the size and number of defects that may be present in any ophthalmic contact lens and as with most other medical regulations the outcome of these comparisons must be deterministic, strictly adhered to and carefully documented.

Having determined the fault type of each feature, as described in section 5.3.5, and the size of the feature, from calculations described in section 5.3.4, we may then compare each feature against the predefined inspection standard for the lens type under examination. Every feature is recorded according to whether it causes an outright failure, whether it could contribute to a cumulative failure, or whether it is of a type or size to not be significant for our decision.

Once every feature has been compared against the standard, the whole standard is checked to see if any failures have been recorded; either cumulative or outright. If there are one or more failures then the COTS process controller is instructed to reject the lens and the major failure mode is recorded; otherwise then the COTS process controller is instructed to pass the lens for packaging and an entry of 'no failure' is entered into the system logs.

We have also made it possible that inspections against multiple standards are possible for regulatory or commercial reasons, however only the primary standard is used to instruct the COTS process controller. It is conceived that this information could be used to instruct a more complex COTS process controller to allow multiple packaging decisions to be made from the multiple standard decisions.

5.3.7 Gross fault detection

Should a valid lens centre or ellipse not be detected as described in section 5.3.1, rather than processing the lens in a way which is likely to fail in a catastrophic manner, we instead perform a high level examination of the image in order to determine one of three gross failure modes: no lens present, lens fragment, or shattered lens. There is also the possibility that large debris could have obscured the window though this would likely cause the illumination check to fail. An example of a lens suffering from a gross failure can be seen in Figure 5-7.

To perform this check we accumulate pixels over the entire image into three 'bins'. These are: pixels of about background intensity, pixels of about lens surface intensity, and pixels of about lens edge intensity. By comparing these with the number expected of a complete lens we can judge how much of a lens is present. Furthermore

by comparing the ratio of edge intensity to surface intensity pixels we can determine the extent to the deformation of the lens.

In any case the COTS processor is instructed to reject the lens and the type of lens deformation is recorded as the major failure mode in the system log.

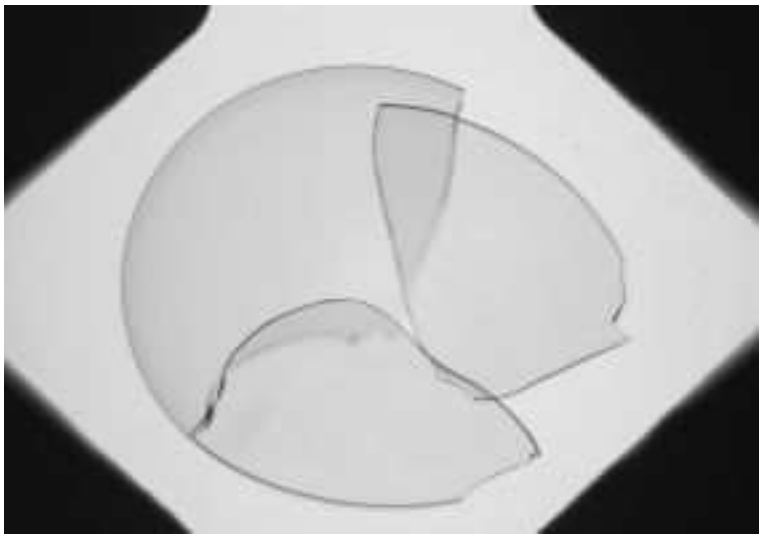

Figure 5-7 Shattered Lens

5.4 Testing

In evaluating the system against the requirements of the project we have considered a number of tests both at the module and system level.

5.4.1 Module tests

We have tested each module sequentially and compared the outputs with expert opinion and the performance of other systems. In the pre-processing stage we compared the extracted centre coordinates and ellipse parameters with hand marked lenses to ensure pixel level accuracy in the extraction. For feature extraction steps we have consulted widely with experts in the field to ensure that the system detects all features and artefacts that are detected by a human expert.

The classifiers have been trained and tested on separate hand-labelled features and perform at a very high level of accuracy. We have also ensured that the feature

extraction and inspection standards processes perform as intended by careful comparison with reference implantations.

5.4.2 System tests

Having ensured that all system components are performing as expected we have performed tests on the whole system to ensure that timing and yields are as expected. In initial tests on a few thousand images we can achieve correct reject/accept decisions on 100% of lenses including classifying the correct largest failure mode. Current trials indicate that processing times of approximately one second are achievable on standard Pentium D 3.00 GHz, 2 GB RAM, Windows 2003 Enterprise Server and there is scope for further compiler optimisation. Process timings for a lens with five faults is shown in Table 5-1. The use of comparable exhaustive established techniques for feature detection would fail to meet these time constraints.

Process	**Timing (seconds)**
Pre-processing	0.141
Surface	0.047
Edge	0.328
Feature Description	0.514
Feature Classification	0.016
Standard Comparison	0.009
Total	**1.055**

Table 5-2 Timings for Lens Inspection Processing Steps

5.5 Conclusions

In this chapter we have described a novel method for the industrial inspection of ophthalmic contact lenses in a time constrained production line environment. In describing this system we have discussed the requirement for a fast an accurate inspection system for fault detection in regulated medical devices. We have given an overview of the system including interfaces to other systems and with operators. We also have described in detail the modules that comprise the inspection system and the tests that these modules have undergone. Finally we briefly describe the full system tests we have performed to establish that our system meets the specifications laid down.

Further this work has applicability to a wider field than inspection of ophthalmic contact lenses; there are many products that need rapid and accurate fault detection with similar fault profiles to those seen in this work. This is especially relevant to those situations where immediate feedback of such results can be used to adjust process parameters. Additionally the processes developed here may find uses in non industrial inspection applications, such as pathological screening applications and object recognition systems.

In summary our contributions to knowledge from this chapter are:

1. Development of a system for automatically inspecting medical devices within a time-constrained environment.
2. Application of complex image processing techniques to ophthalmic lens inspection;
3. Demonstration of the reliability of our probabilistic classification framework for classifying faults in medical devices.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

This thesis has shown the application of probabilistic methods to two distinct areas of computer vision. In this section we discuss the findings from each chapter and relate these to the more general premise of the thesis. We also highlight shortcomings and new avenues in our work which will inform our description of future work in section 6.3.

6.1.1 Probabilistic Methods

Chapter 2 sets out the role of probabilistic techniques in classification and laid the baseline techniques for us to build upon during the remainer of this thesis. To achieve this we discussed the formulation of Bayes' rule, which forms the bedrock for our probabilistic framework; yielding posterior probabilities that we may make decisions on. We then described various methods of data fusion, focusing particularly on score fusion methods since this combination level is most appropriate for a probabilistic approach to object description and classification. Having concluded that score fusion has the greatest potential for our applications, we expanded on the use of mathematical rules for score combination; these rules contain the ability to weight inputs based on classifier efficacy. The theoretical optimum for classifier weighting is discussed before testing on this assertion is performed in Chapter 4.

Having set out background techniques we then considered two specific improvements to our probabilistic framework which dealt specifically with problems in probabilistic classification which we had identified. Firstly we looked at global covariance estimation for homogeneous sets of classes in order to overcome a paucity of data. Then we considered the most appropriate likelihood model for our framework, settling on the logistic function as especially suitable for the two class problem and also for those applications with high dimensional feature vectors. Finally we considered an alternative probabilistic framework for combining evidence, Dempster-Shafer theory, and highlighted key differences with our framework.

Our key contributions to knowledge from Chapter 2 are summarised below:

We described the use of global variance estimation for homogeneous sets of classes, allowing accurate estimation of class means and variances from small datasets, especially those datasets with few examples per subject though with many subjects.

The modelling of class likelihoods by the logistic function is a significant improvement over Gaussian based likelihood models in the two class problem and when using high dimensional feature vectors. The better distributed outputs provide a more scaled response allowing the resultant scores to span the whole output range of zero to one. Appendix B discusses the problems with Gaussian likelihoods in more detail.

By formulating the verification problem as a two class problem modelled by intra and inter-class logistic functions we were able to greatly diminish the amount of training data required and reduce the size of the classification models. Additional processing benefits are achieved by removing the necessity to perform comparison with all known subjects and removing the need to retrain the classifier when a new subject is added to the population.

6.1.2 Biometric Data and Systems

In Chapter 3 we described in detail the algorithms and processes used for biometric data collection. We discussed each modality used in our system, giving an overview of that modality and a detailed description of the algorithm or algorithms that we have chosen to implement.

We then explained the hardware and software decisions made in constructing our biometric collection system. Particularly we discussed the hardware used, the role of the agent framework in producing a flexible processing system and the particular methods for pre-processing in the 3D data environment.

Finally we gave results for biometric algorithms used for subject verification on publicly available large databases in order to asses their relative performance and their suitability for both data collection and the multimodal biometric assessments performed in Chapter 4. We expressed concerns over the robustness of the ear finding algorithm which appear to be causing significant degradation of the expected performance for ear recognition. In a similar manner the performance of the footfall sensor is not as impressive as had been hoped; this is likely due to the simplicity of the features extracted thus far, and the scarcity of training data. We recommended that these modalities are recorded and stored when the tunnel is used for collection, but further work is invested into producing robust algorithms that are capable of comparable performance to the other biometric modalities.

Our key contributions to knowledge from Chapter 3 are summarised below:

We demonstrated a real improvement in both equal error rate and score distribution by the use of our probabilistic framework. The improvement in performance through more efficient classification techniques is of obvious benefit; however in our opinion the demonstration of a robust probabilistic classifier is of more significance in the context of using biometrics in a multimodal environment.

We describe the development of an automated system for the collection and processing of multimodal biometric data. This is important for the progress of

biometric research which has suffered from insufficiently large datasets to properly evaluate multimodal biometrics on fully contemporaneous subjects. This lack of suitable datasets has been especially apparent in 'at a distance' modalities and for studies involving covariates such as time, clothing, race or gender.

Our introduction of the use of footfall data as a viable modality for biometric verification is a significant expansion of the gait modality, allowing the use of data that should be less correlated than parallel processing of video to multiple gait features. The method shows a good deal of promise and is an area that warrants much greater investigation.

6.1.3 Multimodal Biometrics

Chapter 4 examined the role of probabilistic score fusion in multimodal biometrics. We considered two cases to prove the value of fusion: the case of balanced modalities to illustrate that highly performing classifiers benefit from fusion; and the case of highly imbalanced classifiers to show that weighting is necessary to achieve continued improvements in this situation. Having shown the value of fusion we then continued to consider the optimal weighting for score fusion. We showed that the proposal illustrated by equation 16 is not optimal but is on balance preferable to static fusion.

We then discussed the effect of correlation on fusion of modalities. We found that as expected, the reduction in correlation was a good indicator of performance improvement. We noted that there should be some attempt at examining the interrelation between correlation and optimal weighting with a larger collection of modalities.

Finally we discussed the prediction of performance for fused modalities. We tackle the prediction of decidability index since this is a more tractable problem. Here we show that the prediction from pre-calculated class means and variances is simple and accurate. Further we demonstrate how we may pre-compute optimal weights for fusion by maximising the predicted decidability.

Our key contributions to knowledge from Chapter 4 are summarised below:

The demonstration of performance improvements using weighted fusion on highly imbalanced classifiers demonstrates the robustness of the score fusion techniques used, and show that assertions made by others over the limitations of biometrics fusion are inaccurate.

We have given and indication that the optimal weights are not given by equation 16 as stated in [25] and that the optimal weighting tends to be more skewed to the weaker classifier than would be expected.

Our examination of the effect of modality correlation on biometric fusion has led to the conclusion that reduction in correlation is a good indicator of improvement in performance for biometric fusion. This is significant in deciding which modalities will be most appropriate to fuse. Further consideration needs to be given to correlation once sufficiently large contemporaneous datasets have been produced.

We have demonstrated that the decidability index after fusion may be accurately be predicted from pre-fusion data, and further more that calculating the maximal decidability index achievable from a given combination of modalities provides an optimal weighing scheme for fusion.

6.1.4 Ophthalmic Lens Inspection

In Chapter 5 we described a novel method for the industrial inspection of ophthalmic contact lenses in a time constrained production line environment. In describing this system we discussed the requirement for a fast and accurate inspection system for fault detection in regulated medical devices. We gave an overview of the system including interfaces to other systems and with system operators. We also have described in detail the modules that comprise the inspection system and the tests that these modules have undergone. Finally we briefly describe the full system tests we have performed to establish that our system meets the specifications laid down by the customer.

Our key contributions to knowledge from Chapter 5 are summarised below:

We have developed a system for automatically inspecting medical devices within a time-constrained environment. This is a important contribution to the field and the reduction in wastage due to increased accuracy will provide a significant cost saving to manufacturers.

The application of complex image processing techniques to ophthalmic lens inspection is a new use of these techniques. We have found that these techniques have improved the performance over previous systems both in terms of accuracy and speed.

In demonstrating of the reliability of our probabilistic classification framework for determining fault types in medical devices, we have show the applicability of probabilistic methods to new fields and in particular demonstrated the strength of our probabilistic framework across diverse application areas.

6.1.5 General Findings

By demonstrating the utility of probabilistic methods, and particularly of our probabilistic framework, across disparate application areas we strengthen the case for more widespread adoption of probabilistic classifiers. They are most suited to areas where fusion or operator feedback may utilise the probabilistic output, however examination of this output will also provide feedback that may guide optimal thresholds or indicate insufficient training.

We have also developed in Chapter 3 a system architecture that may be well suited to other fields such as medical or behavioural analysis. Additionally data from this system will be available to guide the improvement of biometric processing and related techniques.

The conclusions drawn in Chapter 4 equally have wider ramifications in that these results should be applicable to the output of any set of probabilistic data that one

may wish to fuse. Whilst biometrics is one of the fastest growing area for the use of fusion; one can easily envisage applicability in the military, medical, information management and financial spheres.

The work in Chapter 5 has applicability to a wider field than inspection of ophthalmic contact lenses; there are many products that need rapid accurate fault detection with similar fault profiles to those seen in this work. This is especially relevant to those situations where immediate feedback of such results can be used to adjust process parameters. Additionally the processes developed here may find uses in non-industrial inspection applications, such as pathological screening applications and object recognition systems.

6.2 Critical Appraisal

This section discusses briefly reviews the full scope of work undertaken in the preparation of this thesis in order to critically appraise the effort and draw lessons from those activities.

In the area of probabilistic methods although we feel we have strongly contributed to the field, we spent too much time examining the use of Gaussian methods and attempting to correct their shortcomings rather than seeking other avenues which ultimately proved the successful course of action. Dempster-Shafer theory also proved a distraction which although useful to provide a contrast to our approach did not yield particularly useful results.

In constructing the biometric collection system, we did not allow sufficient time to construct both the physical tunnel and the software. Primarily this was due to an underestimation of the complexity of this task. We also undertook a great deal of prototyping work and investigated simpler solutions for many tasks, such as naïve two dimensional image stitching, which proved unsuccessful. Future projects in this area would be well advised to exclude the development of complex engineering systems from the scope of the doctoral work and plan such systems more carefully.

Because of the length of time taken in this doctorate to complete the construction of the tunnel, other research activities had to be curtailed. Details of further work that would have been performed are discussed in section 6.3. The work on biometrics also could have made earlier use of others techniques and software giving more rapid access to data that could be used for data fusion.

Our work on multimodal biometrics also suffered from the lateness of the tunnel and lack of available multimodal data. This required synthetic subjects to be constructed from available unimodal datasets, and will require further work to validate on true multimodal data in order to safeguard against correlation effects. It would also have been advantageous to have a larger more diverse subject population to examine multimodal biometrics across a more representative group.

The ophthalmic lens inspection, whilst strongly in the spirit of the Engineering Doctorate and valuable work, does make this thesis somewhat fragmented and prevented much further work from occurring in multimodal biometrics. This also delayed the overall progress due to the requirement of getting up to speed in a second (albeit related) field. Our final concern, which was noted elsewhere, is that the commercial nature of this work prevents as much disclosure of the system and results as one would like.

Overall this thesis is an accurate summary of the work undertaken during this Engineering Doctorate, and whilst there are a number of areas where one may wish to revaluate the decisions made in order to maximise progress, it is nevertheless a valuable contribution to the field.

6.3 Future work

This section briefly discusses issues that have not been addressed in the main body of the thesis or remain to be completed. In this section we also discuss possible extensions to this work and expectations for the direction these diverse topics will take.

6.3.1 Probabilistic Methods

Whilst our work on probabilistic methods is reasonably complete, there remains some work where it would be prudent to re-examine our assumptions or explore theories in greater depth. We would find it useful to consider again Dempster-Shafer theory and particularly consider any extension that would make this more compatible with our casting of the fusion problem. We should also consider more complex or hierarchical fusion rules that could provide benefit in terms of performance or processing speed.

The largest theoretical topic still to be considered is that of class distribution and its effect on performance prediction. If this topic could be more rigorously examined it is likely that we would find two benefits: firstly we should be able to more intimately examine the relationship between EER and decidability and hence predict both from our current understanding; secondly we may be able to predict performance and optimal weighting directly from training vectors, in advance of modality evaluation tests.

6.3.2 Biometric Data and Systems

The key task still to be performed on the system is the collection of data across a sufficiently large populous and over a significant period of time. This collection is likely to take at least six months, though thanks to the automated nature of the system should not be too laborious. Such collection should begin as soon as a suitable cohort of subjects has been recruited; this is unlikely to be possible before October 2006 since undergraduate students will be required to get sufficient subject numbers.

There remains work to be carried out on full automation of the system, particularly allowing automation of the verification and fusion process as well as automatic collection. The requirement for more efficient processing and data transfer techniques are also necessary in order to ease collection and allow for real time verification to be performed by the system.

Finally for the system, improvements are still necessary in the extraction of the various modalities; this is particularly important in the newer modalities of footfall and ear recognition. Whilst these tasks are both substantial research topics in their own right, some progress is being made and this will be greatly aided by the large volume of high quality data to be collected by our system. In ear recognition Arbab-Zavar [106] is undertaking promising work on the XM2VTS database, examining feature set selection using the SIFT algorithm [107]. This is significant both in terms of good performance and because, unlike Force Field Functionals [69], this technique produces feature vectors that may be used in our probabilistic framework. There exists a great deal of work to be done on the footfall data, both as a modality in its own right and as an adjuvant to improve localisation for video based gait recognition.

There exists opportunity for much of the work in sensor rich environments such as the tunnel to cross-over into other fields e.g. medical, smart rooms and behavioural sensing. A wealth of work seems apparent in bringing Human Computer Interaction research into the field of biometric systems such as the one developed in this thesis.

6.3.3 Multimodal Biometrics

Our work on multimodal biometrics is reasonably complete, however it would be worth re-performing the tests we have undertaken (especially those involving correlation) with the larger contemporaneous dataset collected by the biometric tunnel. This would have two benefits: firstly it would ensure that none of the results are affected by unexpected interactions between templates of the synthetic subjects; secondly it would serve to reinforce the statistical significance of our results and clarify those results on the edge of our significance tests.

Building on our ideas in section 6.3.1 we should seek to exploit the better understanding of class distribution and performance prediction and evaluate this on larger databases. We may also wish to expand on this evaluation to explore more complex interactions within and between modalities; especially those based on covariates such as time, clothing, race and sex. More complex impostor profiles may also be built to distinguish between active and passive impostor attacks.

Finally we must consider improving the flexibility of our fusion scheme, since this is one of the key drivers of multimodal biometrics. This would involve the investigation of techniques such as: personalised fusion profiles describing individualised weights and modality selection; the use of soft fusion to incorporate other detail that may be extracted during profile such as sex, height, weight or other characteristics; and the introduction of trust ontologies to provide more understandable decisions with influence from other information such as access time, behaviour, or previous verification attempts.

6.3.4 Ophthalmic Lens Inspection

The most pressing remaining task for ophthalmic lens inspection is the deployment of the system into the production environment and the integration with the customer's batch control and reporting systems. In addition we would like to continue work on the understanding of fault types and their location and occurrence profiles; this would allow 'on the fly' adjustment of class weighting to produce more accurate classification of faults.

As mentioned in section 6.1.5 there is significant applicability of our system and algorithms to other medical and non-medical inspection tasks. It is desirable that some effort is expended in producing a generic object inspection system for sale to other manufacturing customers which could be rapidly adapted to their needs. Indeed the flexibility of our system and the ability to rapidly switch algorithms within the system would provide a useful framework for computer vision research since researchers could quickly develop and test new algorithms without the need to design and build an entirely new processing chain.

Appendix A

Biometric Standardisation

A.1 Introduction

As the field of biometrics has matured and become commercially viable there has been an increasing need for interoperability between the systems and subsystems of various vendors, as well as defined testing schemes, language definitions and usage scenarios. In order to facilitate these aims the International Organisation for Standardisation formed a subcommittee to examine the possible scope for standardisation in biometrics. This committee met for the first time in 2003 as ISO/IEC JTC1/SC37; shortly before this first meeting, the British Standards Institute formed the IST/44 to coordinate United Kingdom input into the standardisation process. SC37 now consists of twenty-four member countries, with a further six observer countries and six international liaison organisations.

The working groups of the committees focus on six distinct areas of standardisation interest:

1. Harmonised Biometric Vocabulary
2. Biometric Technical Interfaces
3. Biometric Data Interchange Formats
4. Biometric Functional Architecture and Related Profiles
5. Biometric Testing and Reporting
6. Cross-Jurisdictional and Societal Aspects

Currently there are nine published standards, with twenty-eight projects at various stages of completion. The most well known uses of these standards are the Biometric Interchange Formats [108] being used in the new International Civil Aviation Organisation e-Passports which are currently under adoption by countries, and the International Labour Organisation's Identity Document for Seafarers, for which a biometric profile is being developed [109].

We have become involved as a UK expert within IST/44 and to SC37 focusing on the standardisation effort of multimodal biometrics. This has taken place primarily within working groups 1 and 2, discussing standardised definitions for multimodal biometrics [110] and producing a technical report on multimodal biometrics [111].

A.2 Vocabulary Harmonisation

Whilst in this thesis we have used the phrase multimodal biometric to generically refer to all combinations of biometrics, as is the current academic tradition, we have become aware of the shortfall this contains in properly describing biometric systems utilising fusion techniques. The need to settle on a fully descriptive set of terms is of paramount importance for the progress of the technical report described in section A.3 and further standardisation work mapped out in A.4.

Key amongst this work has been the decision to move to the descriptor *Multibiometric* to describe a biometric system containing any form of data fusion. This is then further divided into five categories:

1. *Multimodal* – The combination of two or more independent biometric characteristics (e.g. face and gait) irrespective of sensor type or processing method;
2. *Multisensoral* – The combination of biometric data from two or more sensors, all examining the same modality;
3. *Multialgorithmic* – The combination of biometric data extracted using different algorithms, but having been obtained from a single sensor and hence single biometric characteristic;

4. *Multiinstance* – The combination of multiple instances of the same biometric characteristic obtained by identical sensor types and processed by identical algorithims (e.g. the combination of right and left iris images);
5. *Multipresentational* – The combination of repeated presentation of the same instance of a biometric characteristic obtained by a single sensor type (e.g. multiple face images).

Similar work has been performed in harmonising the vocabulary describing the levels of biometric fusion. The development of these concepts, although in the early stage of adoption by the community, allows much greater precision when describing and evaluating multibiometric systems. This process ensures that ambiguity is removed from the standardisation process, which is of great importance for successful deployment of international standards.

A.3 Technical Report on Multimodal and Other Multibiometric Fusion

We have had responsibility for preparing the United Kingdom position on this technical report [111]. The technical report contains descriptions of current practice on multimodal and other multibiometric fusion systems focusing on possible standardisation activities for these systems.

The report discusses various possible architectures and levels for the fusion of multiple biometrics. It particularly focuses on decision and score level fusion, since these are the more popular techniques and are believed to be the most effective. Various score normalisation and fusion techniques are described in detail to aid the readers understanding of the field. The report also contains information on terminology as discussed in A.2 and an extensive bibliography of related literature to introduce the reader to the topic. Finally the report attempts to identify possible areas for standardization, these include: further work in the area of record formats for multibiometric systems; development of suitable frameworks and Application Programming Interfaces; application profiles to describe appropriate uses of multibiometrics; and testing methodologies for performance evaluation and standards compliance.

This technical report has recently been submitted for publication ballot, and is expected to be published by ISO/IEC JTC1 early in 2007.

A.4 Progress towards standards

As discussed in section A.3 many areas of standardisation are possible for multibiometric systems. At present two projects are in their formative stages with input from the US and UK national bodies; these relate to interchange formats and APIs.

There exists a number of interchange formats for single biometric modalities, with four published and seven more in progress. These formats allow systems from various vendors to properly interpret and process the biometric data contained therein, leading to interoperable systems such as the e-Passport application. It is considered that this type of format would also be important for allowing multibiometric systems to exchange data with subsystems or other systems. Current ideas focus on the packaging of statistical information on score distribution, performance information, fusion level and method etc to ensure sufficient data is accrued to enable the proper functioning of multibiometric systems containing components from different vendors.

SC37 have developed a successful API for use in biometrics, known as BioAPI [112]. Currently this is only suitable for use in single biometric application, though does support components from multiple vendors. It is thought likely that amendments to BioAPI could be made to enable multibiometric operation. These amendments will focus on introducing new primitive functions and Biometric Information Records to enable the processing, fusion, verification and decision making on multiple biometric records, collected in so called Auxiliary BIRs. Such amendments would allow developers to produce interoperable multibiometric systems or system components without revealing proprietary techniques or worrying about unexpected interaction between components.

Both of these activities are likely to be introduced to SC37 during the latter half of 2007.

Appendix B

Lists of Subjects

B.1 Face Recognition

000	030	059	084	115	153	188	221	256	287
001	031	060	085	116	155	190	222	257	288
002	032	061	086	122	158	192	223	258	291
003	033	062	088	124	159	193	225	259	293
004	034	063	089	126	160	194	226	259	295
008	035	064	091	128	162	198	227	261	297
009	036	066	092	129	163	201	229	262	298
010	037	067	093	130	166	203	231	263	299
012	038	069	096	131	171	204	233	265	304
014	039	070	097	132	174	205	237	266	310
015	045	071	101	137	175	206	238	267	311
016	046	072	102	138	177	208	239	269	314
018	047	073	105	142	178	209	242	271	320
019	048	074	106	143	179	211	243	274	324
021	049	076	107	145	180	212	245	275	360
022	053	078	111	147	181	213	246	279	361
024	054	080	111	149	182	214	248	281	371
026	055	081	112	150	184	216	249	282	
028	056	082	114	151	185	217	251	285	
029	057	083	114	152	187	218	254	286	

B.2 Ear Recognition

000	022	045	067	105	142	179	216	251	297
001	024	046	071	106	149	181	217	254	304
002	026	048	074	111	151	182	218	256	324
003	028	053	078	111	152	188	221	257	360
004	029	054	080	114	155	190	222	259	361
008	030	055	082	114	159	193	225	261	371
012	031	056	083	115	160	194	226	263	
014	032	057	084	116	162	201	231	266	
015	033	059	085	126	171	205	233	274	
018	034	060	091	129	174	206	243	281	
019	036	063	092	132	177	209	245	287	
021	038	064	093	138	178	213	248	288	

References

[1] A. I. Bazin, et al. "An Automated System for Contact Lens Inspection", *Proc. 2nd Int'l Symposium on Visual Computing (ISVC)*, In Press, (2006)

[2] L. Middleton, et al. "A smart environment for biometric capture", *Proc. IEEE Conf. Automation Science and Engineering*, In Press, (2006)

[3] L. Middleton, et al. "Developing a non-intrusive biometric environment", *Proc. IEEE/RSJ Int'l Conf. Intelligent Robots and Systems (IROS)*, In Press, (2006)

[4] L. Middleton, et al. "A Floor Sensor System for Gait Recognition", *Proc. 4th IEEE Workshop Automatic Identification Advanced Technologies (AutoID 05)*, 171-176, (2005)

[5] A. I. Bazin, L. Middleton, M. S. Nixon. "Probabilistic fusion of gait features for biometric verification", *Proc 8th Int'l Conf. Information Fusion*, **2**, 124-131, (2005)

[6] A. I. Bazin, M. S. Nixon. "Probabilistic combination of static and dynamic gait features for verification", *Proc. SPIE Biometric Technology for Human Identification II*, **5779**, 23-30, (2005)

[7] A. I. Bazin, M. S. Nixon. "Gait Verification Using Probabilistic Methods", *Proc. 7th IEEE Workshop on Applications of Computer Vision (WACV/MOTION '05)*, 60-65, (2005)

[8] A. I. Bazin, M. S. Nixon. "Facial Verification Using Probabilistic Methods", *Proc. British Machine Vision Association Workshop on Biometrics*, (2004)

[9] P. Langley, W. Iba, K. Thompson. "Analysis of Bayesian classifiers", *National Conf. Artificial Intelligence*, 223-228, (1992)

[10] G. H. John, P. Langley. "Estimating continuous distributions in Bayesian classifiers", *Proc. 11th Conf. Uncertainty in Artificial Intelligence*, 338-345, (1995)

[11] L. Chengjun. "A Bayesian discriminating features method for face detection", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **25**, 725-740, (2003)

[12] C. Liu, H. Wechsler. "Probabilistic reasoning models for face recognition", *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, 827-832, (1998)

[13] B. Moghaddam, T. Jebara, A. Pentland. "Bayesian face recognition", *Pattern Recognition*, **33**, 1771-1782, (2000)

[14] P. Verlinde, et al. "Applying Bayes based classifiers for decision fusion in a multi-modal identity verification system", *Int'l Symposium In Memoriam Pierre Devijver*, (1999)

[15] S. Eickeler, M. Jabs, G. Rigoll. "Comparison of confidence measures for face recognition", *Proc. 4th Int'l Conf. Automatic Face and Gesture Recognition*, 257-262, (2000)

[16] A. Ross, A. K. Jain. "Multimodal Biometrics: An Overview", *Proc. 12th European Signal Processing Conference (EUSIPCO'04)*, 1221-1224, (2004)

[17] C. Kyong, et al. "Comparison and combination of ear and face images in appearance-based biometrics", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **25**, 1160-1165, (2003)

[18] A. Ross, R. Govindarajan. "Feature level fusion using hand and face biometrics", *Proc. of SPIE Biometric Technology for Human Identification II*, **5779**, 196-204, (2005)

[19] B. Achermann, H. Bunke, *Combination of Classifiers at the Decision Level for Face Recognition*, Insitut fur Informatik und angewandte Mathematik, Universitat Bern, (1996)

[20] S. Ben-Yacoub, Y. Abdeljaoued, E. Mayoraz. "Fusion of face and speech data for person identity verification", *IEEE Trans. Neural Networks*, **10**, 1065-1074, (1999)

[21] R. Brunelli, D. Falavigna. "Person identification using multiple cues", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **17**, 955-966, (1995)

[22] J. Kittler, et al. "On combining classifiers", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **20**, 226-239, (1998)

[23] A. Josang. "The consensus operator for combining beliefs", *Artificial Intelligence*, **141**, 157-170, (2002)

[24] N. Poh, S. Bengio, J. Korczak. "A multi-sample multi-source model for biometric authentication", *Proc. IEEE Signal Processing Society Workshop Neural Networks for Signal Processing XII*, 375-384, (2002)

[25] F. Roli, G. Fumera, J. Kittler. "Fixed and trained combiners for fusion of imbalanced pattern classifiers", *Proc. 5th Int'l Conf. Information Fusion*, **vol.1**, 278-284, (2002)

[26] P. C. Cattin, D. Zlatnik, R. Borer. "Sensor fusion for a biometric system using gait", *Proc. Int'l Conf. Multisensor Fusion and Integration for Intelligent Systems*, 233-238, (2001)

[27] N. Cuntoor, A. Kale, R. Chellappa. "Combining multiple evidences for gait recognition", *IEEE Int'l Conf. Accoustics, Speech, and Signal Processing (ICASSP 2003)*, **3**, 33-36, (2003)

[28] J. Daugman, *Biometric decision landscapes*, University of Cambridge Computer Laboratory, (1999)

[29] G. Shakhnarovich, T. Darrell. "On probabilistic combination of face and gait cues for identification", *Proc. 5th IEEE Int'l Conf. Automatic Face and Gesture Recognition*, 169-174, (2002)

[30] L. Wang, et al. "Fusion of static and dynamic body biometrics for gait recognition", *IEEE Trans. Circuits and Systems for Video Technology*, **14**, 149-158, (2004)

[31] Y. S. Huang, C. Y. Suen. "A method of combining multiple experts for the recognition of unconstrained handwritten numerals", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **17**, 90-94, (1995)

[32] J. A. Benediktsson, P. H. Swain. "Consensus theoretic classification methods", *IEEE Trans. Systems, Man and Cybernetics*, **22**, 688-704, (1992)

[33] L. I. Kuncheva. "Switching between selection and fusion in combining classifiers: an experiment", *IEEE Trans. Systems, Man and Cybernetics B*, **32**, 146-156, (2002)

[34] A. L. Magnus, M. E. Oxley. "Fusing and filtering arrogant classifiers", *Proc. 5th Int'l Conf. Information Fusion*, **vol.1**, 388-395, (2002)

[35] P. J. Flynn, K. W. Bowyer, P. J. Phillips. "Assessment of Time Dependency in Face Recognition: An Initial Study", *Proc. Audio- and Video-Based Biometric Person Authentication (AVBPA 03)*, 44-51, (2003)

[36] D. von Seggern. "CRC Standard Curves and Surfaces", 250, (1992)

[37] J. Cheng, M. J. Druzdzel. "AIS-BN: an adaptive importance sampling algorithm for evidential reasoning in large Bayesian networks", *Journal of Artificial Intelligence Research*, **13**, 155-188, (2000)

[38] P. H. Borcherds. "Importance sampling: an illustrative introduction", *European Journal of Physics*, **21**, 405-411, (2000)

[39] G. Shafer. "A Mathematical Theory of Evidence", *Princeton University Press*, (1976)

[40] J. R. Matey. "Iris on the Move ", *Proc. IEEE*, (2006)

[41] W. Zhao, et al. "Face recognition: A literature survey", *ACM Computing Surveys*, **35**, 399-459, (2003)

[42] R. Chellappa, C. L. Wilson, S. Sirohey. "Human and machine recognition of faces: a survey", *Proceedings of the IEEE*, **83**, 705-740, (1995)

[43] M. Turk, A. Pentland. "Eigenfaces for Recognition", *Journal of Cognitive Neuroscience*, **3**, 71-86, (1991)

[44] P. J. Phillips, et al. "The FERET evaluation methodology for face-recognition algorithms", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **22**, 1090-1104, (2000)

[45] P. Grother, R. J. Micheals, P. J. Phillips. "Face recognition vendor test 2002 performance metrics", 937-945, (2003)

[46] M. Sadeghi, et al. "A comparative study of automatic face verification algorithms on the BANCA database", *Proc. Audio- and Video-Based Biometric Person Authentication (AVBPA 03)*, **2688**, 35-43, (2003)

[47] S. A. Rizvi, P. J. Phillips, H. Moon. "Verification protocol and statistical performance analysis for face recognition algorithms", *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, 833-838, (1998)

[48] A. Colmenarez, B. Frey, T. S. Huang. "A probabilistic framework for embedded face and facial expression recognition", *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition (CVPR 99)*, **Vol. 1**, 592-597, (1999)

[49] B. Moghaddam, A. Pentland. "Probabilistic visual learning for object representation", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **19**, 696-710, (1997)

[50] B. Moghaddam, W. Wahid, A. Pentland. "Beyond eigenfaces: probabilistic matching for face recognition", *Proc. 3rd IEEE Int'l Conf. Automatic Face and Gesture Recognition*, 30-35, (1998)

[51] D. S. Bolme, et al. "The CSU face identification evaluation system: its purpose, features, and structure", *Proc. 3rd Int'l Conf. Computer Vision Systems (ICVS 03)*, 304-313, (2003)

[52] G. Shakhnarovich, J. W. Fisher, T. Darrell. "Face recognition from long-term observations", *Proc. 7th European Conf. Computer Vision Computer Vision (ECCV 2002)*, **3**, 851-865, (2002)

[53] R. Chellappa, S. Zhou, B. Li. "Bayesian methods for face recognition from video", *Proc. Int'l Conf. Acoustics, Speech and Signal Processing (CASSP 02)*, **vol.4**, 4068-4071, (2002)

[54] J. Kittler, Y. P. Li, J. Matas. "Face verification using client specific fisher faces", *Statistics of Directions, Shapes and Images*, 63-66, (2000)

[55] P. N. Belhumeur, J. P. Hespanha, D. J. Kriegman. "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **19**, 711-720, (1997)

[56] M. S. Nixon, et al. "Automatic recognition by gait", *Sensor Review*, **23**, 323-331, (2003)

[57] M. S. Nixon, T. Tan, R. Chellappa. "Human Identification Based on Gait", *Springer*, (2005)

[58] J. B. Hayfron-Acquah, M. S. Nixon, J. N. Carter. "Automatic gait recognition by symmetry analysis", *Pattern Recognition Letters*, **24**, 2175-2183, (2003)

[59] R. T. Collins, R. Gross, J. Shi. "Silhouette-based human identification from body shape and gait", *Proc. 5th IEEE Int'l Conf. Automatic Face Gesture Recognition*, 366-371, (2002)

[60] H. Ning, et al. "Articulated model based people tracking using motion models", *Proc. 4th IEEE Int'l Conf. Multimodal Interfaces*, 383-388, (2002)

[61] D. Meyer, J. Posl, H. Niemann. "Gait classification with HMMs for trajectories of body parts extracted by mixture densities", *Proc. British Machine Vision Conf. (BMVC 98)*, **vol.2**, 459-468, (1998)

[62] D. Cunado, M. S. Nixon, J. N. Carter. "Automatic extraction and description of human gait models for recognition purposes", *Computer Vision and Image Understanding*, **90**, 1-41, (2003)

[63] G. V. Veres, et al. "What image information is important in silhouette-based gait recognition?" *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition (CVPR 2004)*, **2**, 776-782, (2004)

[64] D. K. Wagg, M. S. Nixon. "On automated model-based extraction and analysis of gait", *Proc. 6th IEEE Int'l Conf. Automatic Face and Gesture Recognition.*, 11-16, (2004)

[65] D. K. Wagg, M. S. Nixon. "Automated markerless extraction of walking people using deformable contour models", *Computer Animation and Virtual Worlds*, **15**, 399-406, (2004)

[66] D. K. Wagg, M. S. Nixon. "Model-Based Gait Enrolment in Real-World Imagery", *Proc. Multimodal User Authentication*, 189-195, (2003)

[67] K. H. Pun, Y. S. Moon. "Recent advances in ear biometrics", *Proc. 6th IEEE Int'l Conf. Automatic Face and Gesture Recognition*, 164 - 170, (2004)

[68] M. Burge, W. Burger. "Ear biometrics in computer vision", *Proc. 15th Int'l Conf. Pattern Recognition*, **vol.2**, 822-826, (2000)

[69] D. J. Hurley, M. S. Nixon, J. N. Carter. "Force field feature extraction for ear biometrics", *Computer Vision and Image Understanding*, **Accepted**, (2004)

[70] A. S. Aguado, M. E. Montiel, M. S. Nixon. "On using Directional Information for Parameter Space Decomposition in Ellipse Detection", *Pattern Recognition*, **29**, 369-381, (1996)

[71] J. Perry. "Gait analysis: Normal and pathalogical function", *SLACK Inc.*, (1992)

[72] R. E. Reilly, R. W. Soames. "An imaging walkway for gait analysis", *IEE Colloquium on `Medical Imaging: Transduction and Parallel Processing' (Digest No.050)*, 5-1, (1992)

[73] R. J. Orr, G. D. Abowd. "The smart floor: A mechanism for natural user system identification and tracking", *Proc. Conf. Human Factors in Computing Systems*, (2000)

[74] M. D. Addlesee, et al. "The ORL active floor [sensor system]", *IEEE Personal Communications*, **4**, 35-41, (1997)

[75] J. A. Paradiso, et al. "Sensor systems for interactive surfaces", *IBM Systems Journal*, **39**, 892-914, (2000)

[76] L. McElligott, et al. "ForSe FIElds - force sensors for interactive environments", *Proc. 4th Int'l Conf. Ubiquitous Computing (UbiComp 02)*, 168-175, (2002)

[77] M. Fernstrom, N. Griffith. "LiteFoot-auditory display of footwork", *Proc. 5th Int'l Conf. Auditory Display (ICAD 98)*, 16, (1998)

[78] L. Middleton, et al. "Lightweight agent framework for camera array applications", *Proc. 9th Int'l Conf. Knowledge-Based Intelligent Information and Engineering Systems (KES 2005), Part IV*, **LNCS 3684**, 150-156, (2005)

[79] M. S. Nixon, A. S. Aguado. "Feature Extraction & Image Processing", *Newnes*, (2002)

[80] R. Cipolla, T. Drummond, D. Robertson. "Camera calibration from vanishing points in images of architectural scenes", *Proc. 10th British Machine Vision Conference*, **vol.2**, 382-391, (1999)

[81] G. K. M. Cheung, et al. "Real time system for robust 3D voxel reconstruction of human motions", *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, **2**, 714-720, (2000)

[82] N. Ahuja, J. Veenstra. "Generating octrees from object silhouettes in orthographic views", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **11**, 137-149, (1989)

[83] J. D. Shutler, et al. "On a Large Sequence-based Human Gait Database", *Proc. 4th Int'l Conf. Recent Advances in Soft Computing*, 66-71, (2002)

[84] R. Kumar, et al. "Representation of scenes from collections of images", *Proceedings of the 1995 IEEE Workshop on Representation of Visual Scenes, Jun 24 1995*, 10-17, (1995)

[85] H. Chernoff. "A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on the Sums of Observations", *Annals of Mathematical Statistics*, **23**, 493-509, (1952)

[86] G. V. Veres, M. S. Nixon, J. N. Carter. "Modelling the Time-Variant Covariates for Gait Recognition", *Proceedings of 5th International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA)*, **3546**, 597-606, (2005)

[87] ISO/IEC19795-1, *Information Technology - Biometric Performance Testing and Reporting - Part 1: Principles and Framwork*, (2006)

[88] J. R. Beveridge, et al. "Parametric and nonparametric methods for the statistical evaluation of human id algorithms", *Proc. 3rd Workshop on the Empirical Evaluation of Computer Vision Systems*, (2001)

[89] K. Messer, et al. "XM2VTS: The Extended M2VTS Database", *Proc. 2nd Conf. on Audio- and Video-based Person Authentication (AVBPA '99)*, 72--77, (1999)

[90] H. Korves, et al. "Multibiometric Fusion: From Reseach to Operations", 39-48, (2005)

[91] K. Goebel, Y. Weizhong, W. Cheetham. "A method to calculate classifier correlation for decision fusion", *Information, Decision and Control*, 135-140, (2002)

[92] C. J. Elliott. "Automatic optical measurement of contact lenses", *Proc. SPIE Automatic Optical Inspection*, **654**, 125-129, (1986)

[93] J. M. Pladellorens, et al. "Analysis and characterization of surface defects in ophthalmic lenses", *Proc. SPIE Surface Scattering and Diffraction for Advanced Metrology II*, **4780**, 99-106, (2002)

[94] P. C. D. Hobbs. "Ideas for fast and cheap object capture", *Proc. SPIE Three-Dimensional Imaging, Optical Metrology, and Inspection IV*, **3520**, 133-137, (1998)

[95] J. H. Cho, M. W. Cho, M. K. Kim. "Computer-aided design, manufacturing and inspection system integration for optical lens production", *International Journal of Production Research*, **40**, 4271-4283, (2002)

[96] S. R. Gunn, M. S. Nixon. "A robust snake implementation; a dual active contour", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **19**, 63-68, (1997)

[97] A. Fitzgibbon, M. Pilu, R. B. Fisher. "Direct least square fitting of ellipses", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **21**, 476-480, (1999)

[98] C. E. Olson. "Constrained Hough transforms for curve detection", *Computer Vision and Image Understanding*, **73**, 329-345, (1999)

[99] Y. Xie, Q. Ji. "A new efficient ellipse detection method", *Proc. 16th Int'l Conf. Pattern Recognition*, **2**, 957-960 BN - 950 7695 1695 X, (2002)

[100] Z. Cheng, Y. Liu. "Efficient technique for ellipse detection using restricted randomized Hough transform", *Proc.. Int'l Conf. Information Technology: Coding and Computing (ITCC 04)*, **2**, 714-718 BN - 710 7695 2108 7698, (2004)

[101] R. A. McLaughlin. "Randomized Hough transform: improved ellipse detection with comparison", *Pattern Recognition Letters*, **19**, 299-305, (1998)

[102] J. Canny. "A computational approach to edge detection", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **8**, 679-698, (1986)

[103] D. J. Hurley, M. S. Nixon, J. N. Carter. "Force field energy functionals for image feature extraction", *Image and Vision Computing*, **20**, 311-317, (2002)

[104] M. K. Hu. "Visual Pattern Recognition by Moment Invariants", *IRE Trans. Information Theory*, **8**, 179-187, (1962)

[105] J. R. Quinlan. "Simplifying decision trees", *Int'l Journal of Man-Machine Studies*, **27**, 221-234, (1987)

[106] B. Arbab-Zavar, *A Model-Based Analysis of Ear Biometrics*, School of Electronics and Computer Science, University of Southampton, (2006)

[107] D. G. Lowe. "Distinctive image features from scale-invariant keypoints", *Int'l Journal of Computer Vision*, **60**, 91-110, (2004)

[108] ISO/IEC-19794, *Information technology - Biometric data interchange formats*, (2006)

[109] ISO/IEC-24713-3, *Information technology - Biometric Profiles for Interoperability and Data Interchange - Part 3: Biometric Based Verification and Identification of Seafarers*, (2006)

[110] *ISO/IEC JTC1/SC37 Standing Document 2, version 5 – Harmonized Biometric Vocabulary*, (2006)

[111] ISO/IEC-24722, *Information technology - Technical Report on Multimodal and Other Multibiometric Fusion*, (2006)

[112] ISO/IEC-19784-1, *Information technology - Biometric application programming interface - Part 1: BioAPI specification*, (2006)