



Integrated Shape and Pose Modelling

M. Benjamin Dias and Bernard F. Buxton

Department of Computer Science

University College London

Gower Street, London, WC1E 6BT, UK

Ben.Dias@cs.ucl.ac.uk

<http://www.cs.ucl.ac.uk/staff/B.Dias/>

Abstract

Flexible Shape Models (FSMs), have been widely used for modelling shape variations of deformable objects [4]. A major limitation of this approach is that it assumes a near fronto-parallel view. Recently, representing images as a Linear Combination of Views (LCV) has become a popular approach for modelling 3D pose variations in a 2D image context (e.g. [7]). This technique, however, only works for images of rigid objects. Apart from explicit 3D model approaches, most previous models that can cope with both shape and pose variations have either used a relationship between pose and shape parameters (e.g. [3]) or modelled the pose variations as variations in shape (e.g. FSMs [4]). However, variations of an object's pose and shape are *extrinsic* and *intrinsic* degrees of freedom, respectively. The two should, therefore, not be confounded since they are independent and, in general, not correlated. We have, therefore, developed an *integrated* approach that uses a coupled-view FSM [3] to represent the shape of the entire face as seen from two different view-points, and uses the LCV technique to deal with pose variations. A preliminary comparison is made to the conventional FSM [4].

1 Introduction

Over the past few years statistical shape modelling, via point distribution models and Principal Component Analysis (PCA), has been an active area of research in computer vision. Cootes et al [4] have shown how Flexible Shape Models (FSMs) can be built to model the shape of different objects, including mechanical parts such as resistors and in particular, human body parts such as hands and faces. Knowledge of the allowed variability within a class of shapes, which may be learnt from a set of training examples, is used to build the FSM and to apply constraints to it. The main limitation of this model is that it assumes a near fronto-parallel view and breaks down when the imaged object is allowed to rotate through a large angular range about axes parallel to the image plane.

Cootes et al [3] attempted to overcome this by building coupled-view models and by learning the relationship between the pose angle and the model parameters. They also learnt the correlation between the parameters of the two coupled models, enabling the system to switch between them. By use of such FSMs built on three views (frontal, half-profile and profile), Cootes et al were able to model shape and pose (and appearance)



variations of the face. However, variations of the viewpoint of an object's pose are *extrinsic* degrees of freedom, whilst variations of its physical shape are *intrinsic* degrees of freedom. Therefore, the two should not be confounded since they are independent and though they may sometimes seem to be, they are, in general, not correlated.

Recently there has also been a growing interest in two-dimensional *view-centered* representation of three-dimensional objects. In this approach, three-dimensional objects are represented by a linear combination of two-dimensional images or line drawings. Ullman and Basri [16] were the first to show how novel views could be created by a Linear Combination of Views (LCV). Several others have taken this concept further, from the artificial images of Ullman and Basri to real images [13, 7, 10]. Hansard and Buxton [7] have also shown how the LCV coefficients can be related to the camera viewing angle. Although this technique produces very good, realistic-looking results for rigid objects, it breaks down when confronted with objects that can change shape.

Koufakis and Buxton [10] used a hybrid approach in modelling the human face, utilizing a separate PCA technique to model the eyes and mouth. This enabled them to model some of the intrinsic (shape) and extrinsic (pose) parameters of the face separately by appropriately *blending* the two techniques. Such an approach, however, should only be applied to objects whose shape changes are localized. Movements of the jaw, for example, mean that this is not really the case for the human head. We have, therefore, developed an *integrated* approach that uses a coupled-view FSM [3] to model the shape of the entire face, as seen from two different view-points, and uses a reformulated version of LCV to deal with pose variations. The resulting Integrated Shape and Pose Model (ISPM) is thus an image or view-based model of a non-rigid object in which 3D effects are implicit. Building an explicit 3D model would, of course, be an alternative, but not doing so means that we can avoid poorly conditioned 3D reconstruction steps and therefore generate virtual images with less noise [1]. It is also notable that psychophysical and neuropsychical findings support a two-dimensional, view-centered representation of three-dimensional objects in human vision (for a survey see [12]).

The ISPM described in this paper integrates flexible shape modelling and LCV. Thus, we commence with a brief recap of these techniques.

2 Flexible Shape Modelling

Given a set of images of an object in different positions, orientations and/or scales, we may use the iterative GPA algorithm [6] to align them in a common frame of reference.

Thus, suppose we begin with a set of images, \mathbf{X}_i , ($i = 1, \dots, m$) with landmarks $(x_i(j), y_i(j))$, $j = 1, \dots, n$, written as

$$\mathbf{X}_i = \begin{bmatrix} x_i(1) & x_i(2) & \cdots & x_i(n) \\ y_i(1) & y_i(2) & \cdots & y_i(n) \end{bmatrix} \quad (1)$$

If the images are all centered, the GPA algorithm generally finds the transformations

$$\mathbf{S}_i = \begin{bmatrix} r_i \cos \theta_i & -r_i \sin \theta_i \\ r_i \sin \theta_i & r_i \cos \theta_i \end{bmatrix} \quad (2)$$

that minimize the total squared Procrustes error:



$$\Delta^2 = \frac{1}{2} \sum_{i \neq j} (\mathbf{S}_i \mathbf{X}_i - \mathbf{S}_j \mathbf{X}_j)^2 \quad (3)$$

This problem is usually solved iteratively via a simple four-step algorithm [4].

Given a set of aligned images we can build an FSM to model the shape variations of the imaged object by PCA [4] so that the landmark points may be represented as:

$$\mathbf{x}_i = \bar{\mathbf{x}} + \mathbf{P} \mathbf{b}_i \quad (4)$$

where \mathbf{P} is the matrix of principal components and \mathbf{b}_i is the weight vector for image i . Here, we represent each image as a column-vector as follows:

$$\mathbf{x}_i = (x_i(1), \dots, x_i(n), y_i(1), \dots, y_i(n))^T \quad (5)$$

The elements of a weight vector \mathbf{b} can be used to manipulate the shape of the object and can be constrained via the Mahalanobis distance measure [4]. These shape parameters can be retrieved from a given aligned, centred image \mathbf{x} via:

$$\mathbf{b} = \mathbf{P}^T (\mathbf{x} - \bar{\mathbf{x}}) \quad (6)$$

2.1 The Coupled-View FSM

Suppose we have images of example shapes, taken from two different view-points. We can build two FSMs for these sets of images and also learn the relationship between the shape parameters of the two models [3]. We proceed in a similar manner to Cootes et al [3], except that we utilize the new Extended Procrustes Alignment (EPA) algorithm (see Section 4.2) to achieve pose alignment *prior to building the two individual FSMs*. Suppose the shape in two corresponding images \mathbf{x}_{i1} & \mathbf{x}_{i2} can be represented by \mathbf{b}_{i1} and \mathbf{b}_{i2} via the two FSMs. Then we may form the combined parameter vectors

$$\mathbf{j}_i = \begin{pmatrix} \mathbf{b}_{i1} \\ \mathbf{b}_{i2} \end{pmatrix} \quad (7)$$

and model the relationship between the two FSMs, using PCA:

$$\mathbf{j}_i = \bar{\mathbf{j}} + \mathbf{P}_b \mathbf{c}_i \quad (8)$$

3 Linear Combination of Views (LCV)

In the LCV system, each image is represented, under affine viewing conditions, as a linear combination of two or more basis views [16, 13, 10, 7]. In this paper we deal exclusively with two basis views, though the LCV system easily extends to more if necessary, both in principle [9] and in practice [10]. In LCV, the j^{th} landmark, $(x(j), y(j))$, in an image is related to the j^{th} landmark in the two basis views $(x'(j), y'(j))$ and $(x''(j), y''(j))$ by two linear relationships which, if the images are centred, each with origin chosen at their centroids, may be written as:

$$x(j) = a_1 x'(j) + a_2 y'(j) + a_3 x''(j) + a_4 y''(j) \quad (9)$$



$$y(j) = b_1x'(j) + b_2y'(j) + b_3x''(j) + b_4y''(j) \quad (10)$$

Provided there are at least four corresponding landmark points in all three images, we can use Equations (9) & (10) to estimate the LCV coefficients, a_k & b_k ($k = 1, \dots, 4$) using standard least squares techniques [5, 9]. Novel view reconstruction can also be achieved via Equations (9) & (10), given a set of valid LCV coefficients.

The LCV technique has recently been implemented via a couple of *over-complete* equations [10, 7], such as Equations (9) & (10), since it allows for a symmetric treatment of the two (or more) basis images. We can reformulate the LCV technique via the Centered Affine Trifocal Tensor (CATT) [2], which is introduced next.

3.1 The Centered Affine Trifocal Tensor (CATT)

The geometric relationship between any three images of a rigid object, taken from different viewpoints is defined by the trifocal tensor [8]. It encapsulates the required 3D geometry (structure) of the imaged object and provides a means of implicitly modelling three-dimensional objects in a (*two-dimensional*) image context. Under the assumption of affine or weak-perspective projection, the relationships between the images are linear and are defined by the affine trifocal tensor [11, 15]. If the images are centred, this tensor has twelve non-zero entries [2] that may be written in matrix notation as follows:

$$\mathbf{T}_1^{**} = \begin{bmatrix} t_1 & t_2 & 0 \\ t_3 & t_4 & 0 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{T}_2^{**} = \begin{bmatrix} t_5 & t_6 & 0 \\ t_7 & t_8 & 0 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{T}_3^{**} = \begin{bmatrix} 0 & 0 & -t_9 \\ 0 & 0 & -t_{10} \\ -t_{11} & -t_{12} & 0 \end{bmatrix} \quad (11)$$

We adopt this notation of the CATT [2] to represent the (*linear*) relationship between three images with corresponding landmark points $(x(j), y(j))$, $(x'(j), y'(j))$ and $(x''(j), y''(j))$ as:

$$t_1x(j) + t_5y(j) + t_9x''(j) + t_{11}x'(j) = 0 \quad (12)$$

$$t_2x(j) + t_6y(j) + t_9y''(j) + t_{12}x'(j) = 0 \quad (13)$$

$$t_3x(j) + t_7y(j) + t_{10}x''(j) + t_{11}y'(j) = 0 \quad (14)$$

$$t_4x(j) + t_8y(j) + t_{10}y''(j) + t_{12}y'(j) = 0 \quad (15)$$

The elements of the CATT satisfy the following constraints [15]:

$$\begin{vmatrix} t_1 & t_2 & -t_9 \\ t_3 & t_4 & -t_{10} \\ -t_{11} & -t_{12} & 0 \end{vmatrix} = 0 \quad \text{and} \quad \begin{vmatrix} t_5 & t_6 & -t_9 \\ t_7 & t_8 & -t_{10} \\ -t_{11} & -t_{12} & 0 \end{vmatrix} = 0 \quad (16)$$

The homogeneous Equations (12) - (15) only have a non-trivial solution if we apply a further constraint on the over-all scale, such as the following, which also makes the solution unique:

$$\sum_{i=1}^{12} t_i^2 = 1 \quad (17)$$



3.2 Reformulating The LCV Technique

We have reformulated the LCV technique by using Equations (12) - (15), along with the constraints given in Equations (16) & (17), in place of Equations (9) & (10) to define the linear relationship between the three images. In this version of the LCV technique, the elements of the CATT are the LCV coefficients. Given a new image and the two basis views, we can estimate the corresponding CATT, given at least three corresponding landmark points in each of the three images. For this purpose we introduce the CATT matrix, \mathbf{T}_i , corresponding to an image \mathbf{X}_i and the two basis images \mathbf{X}' & \mathbf{X}'' as:

$$\mathbf{T}_i = \begin{bmatrix} t_1(i) & t_5(i) & t_9(i) & 0 & t_{11}(i) & 0 \\ t_2(i) & t_6(i) & 0 & t_9(i) & t_{12}(i) & 0 \\ t_3(i) & t_7(i) & t_{10}(i) & 0 & 0 & t_{11}(i) \\ t_4(i) & t_8(i) & 0 & t_{10}(i) & 0 & t_{12}(i) \end{bmatrix} \quad (18)$$

Thus, to compute the corresponding CATT, we form $\mathbf{Y}_i = (\mathbf{X}_i, \mathbf{X}'', \mathbf{X}')^T$ (the joint image [14]), and compute the CATT matrix \mathbf{T}_i that minimizes

$$e_i^2 = \|\mathbf{T}_i \mathbf{Y}_i\|^2 \quad (19)$$

subject to the constraints given in Equations (16) & (17).

A serious limitation to the LCV approach is that the LCV coefficients cannot be estimated without reference to the target coordinates [7] and hence only existing images can be reconstructed. However, we can learn the relationship between the imaged object's 3D pose and the LCV coefficients from some labelled training data. For example, Hansard and Buxton [7] derived a relationship between the LCV coefficients and the camera viewing angle, by fitting polynomial functions to them. If desired, we can do the same here & fit such polynomials to the elements of the CATT in a similar manner.

3.3 Selecting The Best Pair of Basis Views From a Set of Images

Previously [10, 7], the basis views used in the LCV model have been intuitively selected, from the available training images. We propose, instead, the use of the following error criterion, to select the *best* pair of basis views from a given set of images of a rigid object.

Given m images (\mathbf{X}_i), there are ${}^m C_2$ possible choices of two basis views. Thus, we choose two images, \mathbf{X}' & \mathbf{X}'' , at a time, and compute the resulting total squared LCV error:

$$\epsilon^2 = \sum_{i=1}^m e_i^2 \quad (20)$$

where e_i^2 is given in Equation (19). We then select the pair with the least error, ϵ^2 .

The above error criterion may not always be sufficient for choosing the '*best*' pair of basis views, since it could result in the chosen basis views being very nearly the same, even when the set of training images sample large pose variations of the imaged object. To prevent this, we compute the total difference between the two chosen basis views

$$d^2 = \|\mathbf{X}' - \mathbf{X}''\|^2 \quad (21)$$

and add its inverse to ϵ^2 , which gives a combined error

$$\epsilon^2 = \epsilon^2 + \frac{1}{831d^2} \quad (22)$$



that prevents the system from choosing nearly identical basis images, unless the training set requires it to do so. As it stands Equation (22) is dimensionally inhomogeneous, so we normalize the images prior to computing the errors. Provided the images are always centered, the normalization used here is a scaling [8], so that the RMS distance of the landmark points from the origin is equal to $\sqrt{2}$. This makes Equation (22) insensitive to different choices of coordinates in the image.

4 Extending The GPA Algorithm

The GPA algorithm is a tool for removing extrinsic position, orientation and scale variations from a set of training images. If all the images are taken from a fixed view, this allows us to model the remaining intrinsic shape variations of the imaged object. However, pose variations due to rotation of a 3D object about axes parallel to the image plane, are not removed by the GPA algorithm. Thus, they remain *'mixed up'* with the shape variations and, for example, the first mode of an FSM built for faces, using images aligned via the GPA algorithm, frequently models head rotation mixed with the opening and closing of the mouth (e.g. see [4]). This is an affect we wish to avoid, since pose variations should not, in principle, be modelled as shape variations.

The similarity group of transformations (Equation (2)) which only deals with orientation and scale, is frequently used to align centred images, in implementations of the GPA algorithm. With such a transformation, a full pose alignment cannot be expected. In order to rectify this, we utilize the CATT, to perform a pose alignment prior to applying GPA.

4.1 Pose Alignment in 2D via The CATT

Pose alignment in a 2D context can be achieved by using the CATT to reconstruct images of a given shape, \mathbf{X}_i , as seen from the two basis views, via the corresponding fundamental matrices. Here, we start with the mean shape as seen from the basis views, $\bar{\mathbf{X}}'$ & $\bar{\mathbf{X}}''$, and update their shape, until they match the shape of \mathbf{X}_i .

We commence by computing the best approximation of the CATT corresponding to \mathbf{X}_i , $\bar{\mathbf{X}}'$ & $\bar{\mathbf{X}}''$ as described in Section 3.2. The resulting error, e_i^2 (in Equation (19)), will then be due, largely, to the difference in shape between the three images and can be reduced further by changing the shape of $\bar{\mathbf{X}}'$ & $\bar{\mathbf{X}}''$. This shape update has to be computed separately for each basis view, since any attempt to map a shape from one view to two becomes degenerate. Therefore, we proceed by extracting the two fundamental matrices, \mathbf{F}_{i1} & \mathbf{F}_{i2} , that link \mathbf{X}_i to each of the basis views from the trifocal tensor (via *Algorithm 14.1, on page 366* of [8]). We then, generate the new images \mathbf{X}_i' & \mathbf{X}_i'' , which are images of the shape \mathbf{X}_i as seen from the basis views, by computing their landmark points $(x_i'(j), y_i'(j))$ & $(x_i''(j), y_i''(j))$, via the fundamental constraint equations:

$$\begin{bmatrix} x_i(j) & y_i(j) & 1 \end{bmatrix} \mathbf{F}_{i1} \begin{bmatrix} x_i'(j) \\ y_i'(j) \\ 1 \end{bmatrix} = 0 \quad (23)$$

$$\begin{bmatrix} x_i(j) & y_i(j) & 1 \end{bmatrix} \mathbf{F}_{i2} \begin{bmatrix} x_i''(j) \\ y_i''(j) \\ 1 \end{bmatrix} = 0 \quad (24)$$



In each case we have one equation and two unknowns and hence we solve each equation as a linear optimization problem using the coordinates of the current landmark points as initial estimates in a Gauss-Newton method with line-search technique.

4.2 The Extended Procrustes Alignment (EPA) Algorithm

The aim of the EPA algorithm is to align a given set of images to the mean shape in each of the basis views ($\bar{\mathbf{X}}'$ & $\bar{\mathbf{X}}''$). However, the two mean shapes and even the basis views are initially unknown. Therefore, we commence by selecting *two* reference images (\mathbf{X}_{r1} & \mathbf{X}_{r2}), from the given set, whose views (or poses) define the basis views. Although we do not have a set of images of a rigid object, we use the combined error (ε^2) in Equation (22) to select our two reference images, as described in Section 3.3. Experiments show that this yields better results than selecting the two reference images at random.

Once we have the two reference images that define the basis views, we proceed by aligning the remaining images to them, as described above (see Section 4.1). This alignment results in two images of each training shape, as seen from each of the basis views. Thus, we have two sets of aligned images with no pose variations, and we can utilize the GPA algorithm to extract the mean shape of each set. We use these mean shapes as our current estimates of $\bar{\mathbf{X}}'$ & $\bar{\mathbf{X}}''$, which can be updated by iteratively aligning all the original images to the current estimates and re-computing the two mean shapes via GPA.

Each time we use the GPA algorithm to extract the mean shape of a set of aligned images, we use the relevant reference image (\mathbf{X}_{r1} or \mathbf{X}_{r2}) as an external reference in order to ensure that the mean shapes remain in the basis views and that the system converges.

Thus, given a set of images of a deformable object in different poses and in different shape-states, this process may be described as a simple three-step algorithm, as follows:

The EPA Algorithm

1. Pick the two best examples (Section 3.3) as the two reference images \mathbf{X}_{r1} & \mathbf{X}_{r2} and align all the remaining examples to them, via the relevant CATTs (Section 4.1).
2. Compute the two mean shapes, $\bar{\mathbf{X}}'$ & $\bar{\mathbf{X}}''$, of the two sets of aligned images, via GPA, using \mathbf{X}_{r1} & \mathbf{X}_{r2} as external references.
3. Align all the original examples to $\bar{\mathbf{X}}'$ & $\bar{\mathbf{X}}''$ (Section 4.1).

We repeat steps 2 & 3 until the difference between consecutive estimates of the two means is smaller than some given tolerance.

5 The Integrated Shape and Pose Model (ISPM)

The Integrated Shape and Pose Model (ISPM) is built from a labelled set of training images, where key landmark points are marked in each example image. The EPA algorithm (Section 4.2) is used to align the training images, in order to remove the extrinsic shape variations. This algorithm results in a set of aligned images and their mean shape, in each of the two reference poses which define the basis views. Since we have a correspondence between the two sets of aligned images, we can build a coupled-view FSM as described in Section 2.1.



Figure 1: Some example images from the training set used to build the ISPM. From left to right we see an angry, happy, neutral, sad and surprised face having a pose angle of approximately $+30^\circ$, $+15^\circ$, 0° , -15° and -30° respectively

The ISPM is then formed by linking the coupled-view FSM with the reformulated version of LCV (Section 3.2) by using the two coupled views as the basis views. The shape of the object, as seen from the two basis views, can be manipulated by the coupled-view FSM and an object of that shape can then be synthesized in a new view via the reformulated LCV model.

To extract the shape parameters from a given image, we first align it to the two basis views via the corresponding CATT. The individual shape parameters, \mathbf{b}_1 & \mathbf{b}_2 , of the two resulting aligned images are then extracted via Equation (6). \mathbf{b}_1 & \mathbf{b}_2 are then combined as in Equation (7) and the coupled-view FSM parameters \mathbf{c} are extracted via:

$$\mathbf{c} = \mathbf{P}_b^T (\mathbf{j} - \bar{\mathbf{j}}) \quad (25)$$

To generate a new view of a desired shape, the input to the system is the combined shape vector, \mathbf{c} , of the coupled-view FSM. The individual FSM parameters, \mathbf{b}_1 & \mathbf{b}_2 , can be generated via Equation (8) and used to re-construct the desired shape as seen from the two basis views, via Equation (4). The desired shape is then synthesized in a new view via the reformulated LCV model, by use of a *valid* CATT (i.e which satisfies the constraints in Equations (16) & (17)).

6 Results

We used a set of training images, in which key landmark points were manually located, to build both an ISPM and a conventional FSM [4]. An FSM is able to cope with pose variations of up to 45° [3]. Therefore, for a valid comparison, we used images of a face which was allowed to rotate about the vertical axis with a maximum pose variation of $\sim 30^\circ$. Five different expressions (Neutral, Angry, Happy, Sad & Surprised) were randomly sampled in the different poses. We used a set of 39 images, consisting of three random expressions sampled at $\sim 5^\circ$ intervals from $\sim -30^\circ$ to $\sim +30^\circ$ where 0° corresponds to the frontal view. Some examples of the training images are given in Figure 1. The resulting two most significant modes of the ISPM and FSM are shown in Figure 2.

Over-laying the images of the first mode of each model (Figure 3 (a)) shows that the ISPM displays significantly less pose variation than the FSM. The eigenvalues of the models (Figure 3 (b)) reveal that the ISPM has significantly less variance than the FSM. We would expect to see this, owing to the reduction of pose variations modelled. These

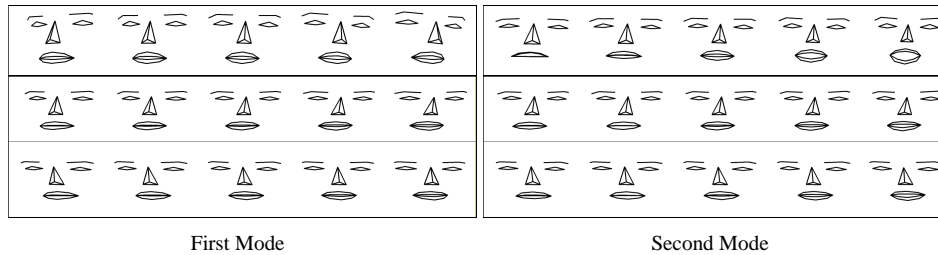


Figure 2: First & second modes of the FSM (top row) & the ISMP coupled-models (bottom two rows)

preliminary results indicate that the EPA is, to a certain extent, capable of achieving pose-alignment in a 2D context, via the CATT. Figure 3 (b) also shows that there is a correlation between the coupled models of the ISPM, though the correlation requires improvement.

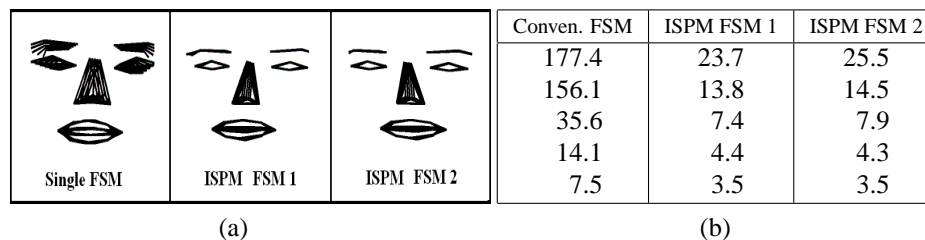


Figure 3: (a) All the images of the first mode of each model, as seen in Figure 2, over-laid (b) The eigenvalues (variance) of the first five modes of each model

7 Conclusions and Future work

We have demonstrated that the shape and pose of a flexible object, such as a human face, can be modelled independently by integrating a coupled-view FSM with a reformulated LCV technique. We use the Extended Procrustes Alignment (EPA) algorithm to align the training images to remove all extrinsic degrees of freedom. The resulting two sets of aligned images are then used to build two FSMs which are combined to form a coupled-view FSM, which models the pose-free, intrinsic shape variations of the imaged object. The coupled-view FSM and reformulated LCV model are linked via the basis views to form the ISPM. An FSM, as constructed by Cootes et al [4] was used for comparison.

We would like to extend the concept of the ISPM to modelling the appearance (grey-level and colour values) of images. Once appearance information is included, it should be possible to use the ISPM system for image searching in a manner similar to the active appearance models of Cootes et al [4].

We anticipate that the ISPM will then be useful in a variety of applications including: calculation of head pose, view-invariant face recognition, and head tracking; for example, to drive an animated avatar, or for video compression over the internet or very low band



width wireless links. The approach may also be of relevance to theories of human vision.

Acknowledgements

Benjamin Dias is funded by an *ORS award*¹, by financial support from *Unilever Research Colworth*² and by a *UCL Graduate Open Scholarship*³.

References

- [1] S. Avidan, T. Evgeniou, A. Shashua and T. Poggio. *Image-Based View Synthesis by Combining Trilinear Tensors and Learning Techniques*, in Proceedings of the ACM VRST 97, pp 103 - 110, September 1997.
- [2] L. Bretzner, *Multi-Scale Feature Tracking and Motion Estimation*, Dissertation, Computational Vision and Active Perception Laboratory (CVAP), pp 77 - 87, October, 1999.
- [3] T. F. Cootes, G. V. Wheeler, K. N. Walker and C. J. Taylor *Coupled-View Active Appearance Models*, Proc. British Machine Vision Conference, vol. 1, pp 52 - 61, 2000.
- [4] T. F. Cootes and C. J. Taylor, *Statistical Models of Appearance for Computer Vision*, http://www.wiau.man.ac.uk/bim/Models/app_model.ps.gz, October 2001.
- [5] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Third Edition, John Hopkins University Press, ISBN 0-8018-5414-8, 1996.
- [6] J. C. Gower, *Generalised Procrustes Analysis*, Psychometrika, 40, pp 33 - 51, 1975.
- [7] M. E. Hansard and B. F. Buxton, *Parametric View-Synthesis*, Proc. 6th European Conference on Computer Vision (I), pp 191 - 202, 2000.
- [8] R. Hartley and A. Zisserman, *Multiple View Geometry In Computer Vision*, Cambridge University Press, ISBN 0-521-62304-9, 2000.
- [9] D. M. Kennedy, B. F. Buxton and J. H. Gibly, *Application of the Total Least Squares Procedure to Linear View Interpolation*, BMVC, Vol. 1, pp 305 - 314, 1999.
- [10] I. Koufakis and B. F. Buxton, *Very Low Bit-Rate Face Video Compression using Linear Combination of 2D Face Views and Principal Component Analysis*, Image and Vision Computing 17, pp 1031 - 1051, 1998.
- [11] P. R. S. Mendonça and R. Cipolla, *Analysis and Computation of an Affine Trifocal Tensor*, Proceedings of BMVC, pp 125 - 133, September 1998.
- [12] G. Peters, *Theories of Three-Dimensional Object Perception - A Survey*, Part - I, Vol 1, pp 179 - 197, Transworld Research Network, 2000.
- [13] S. Pollard, M. Pilu, S. Hayes and A. Lorusso, *View Synthesis by Trinocular Edge Matching and Transfer*, Proc. BMVC, Vol. 2, pp 770 - 779, 1998.
- [14] A. Shashua, *Trilinear Tensor: The Fundamental Construct of Multiple-View Geometry and its Applications*, In Proc. International Workshop on Algebraic Frames For The Perception-Action Cycle (AFPAC97), Kiel, Germany, September 1997.
- [15] T. Thórhallsson and D. W. Murray, *The Tensors of Three Affine Views*, Proc. Int. Conf. on Computer Vision and Pattern Recognition, vol 1, pp 450 - 456, 1999.
- [16] S. Ullman, and R. Basri, *Recognition by linear combinations of models*, IEEE Transactions on PAMI, Vol. 13, No. 10, pp 992 - 1006, 1991.

¹Administered by the CVCP, Woburn House, 20 Tavistock Square, London WC1H 9HQ, UK.

²Unilever Research Colworth, Colworth House, Sharnbrook, Bedford MK44 1LQ, UK.

³Administered by University College London, Gower Street, London WC1E 6BT, UK.