Building Class Sensitive Models for Tracking Applications

Derek R. Magee and Roger D. Boyle
School of Computer Studies
University of Leeds, Leeds LS2 9JT, UK
drm@scs.leeds.ac.uk

Abstract

A method of generating complementary eigenspaces optimised for inter-class and intra-class separability respectively is presented. The objective of creating these spaces is to improve the efficiency of eigenspace search algorithms. The inter-class optimised space may also be used to improve classification and a quantitative evaluation of this against conventional Principal Component Analysis and Canonical analysis (based on Linear Discriminant Analysis) is presented. A qualitative comparison of the intra-class optimised space and spaces produced by Principal Component Analysis on single class data is also presented.

1 Introduction

Principal Component Analysis (PCA) is an efficient way of parameterising the variance within a multivariate data set such that the dimensionality may be reduced without greatly affecting approximation accuracy. This is done by finding the eigenvectors of a covariance matrix formed from the data set and forming an 'eigenspace' based on these. In many cases the variance of the data set along the least significant eigenvectors is negligible and thus these dimensions may be ignored and the dimensionality reduced. PCA is widely used in statistics and has been used to great effect in Machine Vision to simplify the variation in complex data sets such as faces [11], shape outlines [2] and human motion (gait) [5]. If physical modes of variation within a data set are linear and well decoupled then the parameterisation will map closely to these physical modes of variation. In real data sets physical modes of variation are not always linear or decoupled, however PCA will often produce a good linear approximation of the variation within the data set.

The fact that data variation parameterisation maps well to physical variation means a subset of the model parameterisation may be used as a classifier. The problem with the parameterisation produced by PCA is that both inter and intra-class variation are included. In cases where physical variation is non-linear and not well decoupled this problem is worsened by parameters including both inter and intra-class information. Canonical analysis is a variation on PCA in which class information is included such that the ratio of inter-class to intra-class variance is maximised. This is based around the assumption that the variation between class means contains no intra-class properties, in other words class means

BMVC 1999 doi:10.5244/C.13.59

have comparable intra-class properties. This is not necessarily a valid assumption, as will be shown later in this paper.

In this paper we present an alternative to canonical analysis in which two parameterisations are produced optimised for inter and intra-class variation respectively. This method, known as 'delta analysis', does not rely on assumptions about class means. It is demonstrated to give greater class separation than canonical analysis on three data sets: two sets of shape outlines based on the Point Distribution Models [2] (synthetic and real data) and on Eigenface images [11].

The two spaces produced by delta analysis may be used in conjunction to describe the variation in a data set with separate inter and intra-class components. This representation is useful in such applications as object tracking where the inter-class variation of an object is zero over time and as such need only be determined once. This improves the efficiency of tracking as only intra-class properties need be determined at each time point.

2 Background

2.1 The Point Distribution Model and the Vector Distribution Model

The Point Distribution Model (PDM) [2] uses PCA to represent the variation within a class of shapes by modelling the positions of various 'landmark points'. The model is built from a set of training shapes which are normalised by position, scale and rotation using an iterative process. The x and y values of each set of normalised landmark points are formed into a vector and PCA is performed on them. By altering each component individually in the Eigenspace produced, a set of linear 'modes of variation' are produced. In many cases these are close to the physical ways in which the class of shapes varies. When this is the case these components may be used to classify items within the data set. An alternative to the PDM is the Vector Distribution Model [7] which uses relative vectors rather than points but is in other ways identical to the PDM.

2.2 Canonical analysis and the Canonical Space Transformation

Canonical analysis or Linear Discriminant Analysis [6] is a statistical technique used on multivariate data sets which contain multiple data classes. The objective of this technique is to separate out inter-class variation from intra-class variation by creating a new space in a similar way to PCA. For this method two covariance matrices are required: an 'intra-class' matrix (S_w) and an 'inter-class' matrix (S_b). These are formed as in equations 1 and 2 respectively. Once the covariance matrices have been formed the generalised eigen equation given in equation 3 is solved to give a set of eigenvectors which describe a new space. It should be noted that this space is not necessarily optimised to allow dimensionality reduction as with PCA, however the first components describe inter-class variation and can as such be used as a data classifier. This has been used recently for face recognition [4, 10], image retrieval systems [10] and gait classification [5].

$$S_w = \frac{1}{n_t} \sum_{i=1}^{n_c} \sum_{j=1}^{n_i} (\mathbf{y_{i,j}} - \mu_\mathbf{i})(\mathbf{y_{i,j}} - \mu_\mathbf{i})^T \tag{1}$$

$$S_b = \frac{1}{n_t}\sum_{i=1}^{n_c}(\mu_{\mathbf{i}} - \mu_{\mathbf{y}})(\mu_{\mathbf{i}} - \mu_{\mathbf{y}})^T \tag{2}$$

$$S_b E = \lambda S_w E \tag{3}$$

Where:

S_w = Intra-class Covariance Matrix	n_i = Number of data items in class i
S_b = Inter-class Covariance Matrix	$\mathbf{y_{i,j}}$ = Data item j of class i
n_t = Total number of data items	$\mu_{\mathbf{i}}$ = Mean vector for class i
n_c = Number of classes	$\mu_{\mathbf{y}}$ = Mean vector for entire data set

It can be seen from equation 2 that the notion of 'inter-class variance' is based on the variance of the means. This means that if the means of the data sets do not have comparable intra-class characteristics, intra-class variation will be included in the optimisation (to some extent). In this situation the result of canonical analysis is a compromise as within class variation is both maximised and minimised simultaneously.

Swets and Weng [10] combine canonical analysis with PCA to give data reduction and improved classification. In this scheme canonical analysis is performed on the projections of the raw data in an eigenspace created by PCA. In this paper we shall refer to this scheme as the combined eigen-canonical transform.

2.3 Eigenfaces and High Dimensionality PCA

PCA has been applied to greylevels in face and other images to obtain a compact parameterisation of these complex data sets [9, 11]. Turk and Pentland [11] highlight the practical problem of finding the eigenvectors of a high dimensional problem and present a method for reducing the problem to the order of the number of samples (which is usually much lower than the number of pixels per sample in these cases). In this method the number of eigenvectors found is equal to the number of data items rather than the dimensionality, however this is not a problem as the data set can be described exactly using these eigenvectors. If the number of data items is large this may be reduced using clustering algorithms. Standard cannonical analysis does not apply well to raw face data as the intra-class covariance matrix is often singular due to the sample size being much smaller than the dimensionality. Belhumeur *et al.* [1] use a version of the combined eigen-canonical transform known as Fisherfaces to include class separation in their eigenface representation.

3 Creating Complementary Eigenspaces Using Delta Analysis

In section 2.2 it was described how canonical analysis can produce an eigenspace that is optimised for intra-class variability. For data sets where the mean vectors of classes within a data set do not have the same within class characteristics it can be seen that within class variation is included in the more significant eigenvectors. Delta analysis is an alternative to canonical analysis that does not use class mean information, instead modelling within class and between class variation explicitly using deltas between data items.

3.1 Creating an Intra-class Optimised Eigenspace

An intra-class optimised eigenspace is created by forming a covariance matrix from the deltas between each pair of data items as in equation 4. PCA is performed on this matrix to create a 'delta eigenspace'. For each pair there are two deltas ($\overline{AB}$ and $\overline{BA}$ for pair *A* and *B*) which are equal and opposite to each other. In this way the mean of the deltas is zero and the relationship between the data and its projection in eigenspace simplifies to that given in equation 5. This scheme assumes only that the intra-class variation between classes is comparable.

$$S_w = \frac{1}{n_t} \sum_{i=1}^{n_c} \sum_{j=1}^{n_i} \sum_{k=0, k \neq j}^{n_i} (\mathbf{y_{i,j}} - \mathbf{y_{i,k}})(\mathbf{y_{i,j}} - \mathbf{y_{i,k}})^T \quad (4)$$

Where:

$\mathbf{y_{x,y}}$ = Member y of class x	n_c = The total number of classes
n_t = The total number of data items	n_i = The number of members in class i

$$\mathbf{d} = \mathbf{y}E \quad (5)$$

Where:

$\mathbf{d}$ = A delta vector	E = A matrix of Eigenvectors
$\mathbf{y}$ = Vector in 'Delta Eigenspace'	(One Eigenvector per row)

It would not be correct to project raw data vectors into this 'Delta Eigenspace', however; as the deltas are relative and have mean zero they may be considered with respect to any point in data space. A new 'Data Projection space' with identical axes to the 'Delta Eigenspace' may thus be formed and this will be optimised for within class variation. The variances of the components in this 'Data Projection space' are not equal to the eigenvalues produced by PCA but may be calculated by projecting the data into this space.

3.2 Creating an Inter-class Optimised Eigenspace

Creating an optimal inter-class space is harder than creating an optimal intra-class space since intra-class variation is included in the inter-class deltas. One approach to isolating inter-class information is to form the product of an inter-class covariance matrix and the inverse of an intra-class covariance matrix. This is similar to canonical analysis described in section 2.2, but does not always produce real eigenvalues (this can be a problem with regular cannonical analysis also [1]).

Instead we use the variance of the intra-class deltas to normalise the inter-class deltas by dividing each inter-class delta component by the equivalent intra-class variance. This approach does not include intra-class covariance information and thus relies on inter-class modes of variation being fairly well de-coupled from within class modes of variation but will always produce a real solution. Equation 6 describes the exact formulation of the covariance matrix.

$$S_b = \frac{1}{n_t} \sum_{i=1}^{n_c} \sum_{j=1, j \neq i}^{n_c} \sum_{k=1}^{n_i} \sum_{l=1}^{n_j} \delta_{i,j,k,l} \delta_{i,j,k,l}^T \quad (6)$$

Where:

$\delta_{i,j,k,l} = (\mathbf{y}_{i,j} - \mathbf{y}_{k,l})\mathbf{v}_w$	n_t = Total number of data items
$\mathbf{y}_{x,y}$ = Data Vector y of class x	n_c = Number of classes
$\mathbf{v}_w = (\frac{1}{v_1}, \frac{1}{v_2}, ..., \frac{1}{v_n})$	n_i = Number of members in class i
(The Inverse Intra-class Variance Vector)	n_j = Number of members in class j

A between class 'Data Projection Space' is formed from the eigenvectors of this covariance matrix in exactly the same way as the within class space (see section 3.1).

3.3 Combining Intra-Class and Inter-Class Spaces in Practical Applications

If we examine the example variance graphs in figure 3.1, which are taken from the cow outline example in section 5, we see at first glance that the space produced using conventional PCA appears much better optimised for dimensionality reduction than the two spaces produced by delta analysis. However, if we project the first few vectors of these eigenspaces into PCA eigenspace (see figure 3.2) it is interesting to note that, for this data set, most of the energy in these vectors lies in the first few components of this eigenspace and all of the first few eigen components have power in at least one of the delta space vectors.

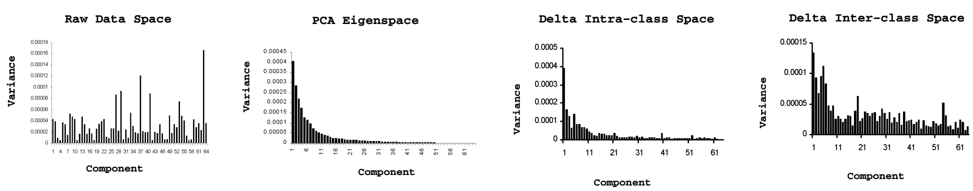

Figure 3.1: Variances of Components in Various Spaces

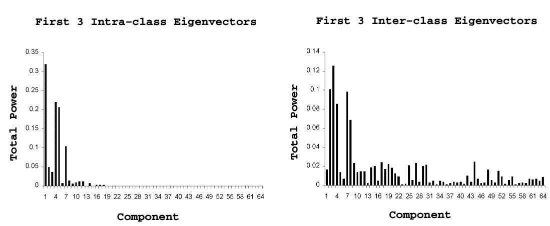

Figure 3.2: Power of Inter and Intra-class Eigenvectors Projected into PCA Eigenspace

It is not hard to see that we can thus approximate the first few eigenvectors in PCA eigenspace and thus the entire data set by the first few eigenvectors in within and between class 'Data Projection Space' as in equation 7.

$$\mathbf{x} \approx \bar{\mathbf{x}} + \sum_{n=1}^{n_{wc}} w_n \mathbf{e}_{wc_n} + \sum_{n=1}^{n_{bc}} b_n \mathbf{e}_{bc_n} \tag{7}$$

Where:

$\mathbf{x}$ = Data vector to be approximated	n_{wc} = Dimensionality of Truncated Within Class Space
$\bar{\mathbf{x}}$ = The mean vector	n_{bc} = Dimensionality of Truncated Between Class Space
$\mathbf{e}_{wc_n}$ = Within Class Eigenvector n	w_n = Component Value for Within Class Eigenvector n
$\mathbf{e}_{bc_n}$ = Between Class Eigenvector n	b_n = Component Value for Between Class Eigenvector n

The values of n_{wc} and n_{bc} will obviously depend on the particular data set but this example suggests significant dimensionality reduction may be possible. The disadvantage of this scheme is that within class eigenvectors are not orthogonal to between class eigenvectors and thus there is not necessarily a unique solution when X is known and we wish to find values for w_n and b_n. A least squares minimisation method may be used to find an optimal solution however it may be computationally more efficient to firstly find an optimal solution for inter-class parameters ($b_{1\rightarrow n}$) only and then complete the solution by finding the optimal intra-class parameters ($w_{1\rightarrow n}$). This scheme applies well to object tracking where inter-class parameters are constant over time and need only be found once. An evaluation of the approximation accuracy of this scheme compared to PCA is presented in section 4.3. Edwards *et al.* [3] present an alternative scheme for combining different types of inter-class variation.

4 Application of Method to Data

Delta analysis has been tried on two sets of point data, a synthetic set (Triangle people), and a set derrived from live data (cow outlines) and a set of intensity images (AT&T Database of faces [8] - 10 examples of 40 people). Triangle people are very simple figures made from card triangles and a drawing pin to pivot the ‘legs’. 20 examples of 10 individuals were created using an 8 point approximation (see Figure 4.1a). 118 cow outlines were obtained by associating points with landmarks on cows from video sequences (see figure 4.1b). Data sets were divided into equal sized ‘training’ and ‘test’ sets for classification experiments. Vector distribution models (see section 2.1) were built from the point data sets and delta analysis performed on the raw vector data. Delta analysis was applied to the face data using a simple nearest neighbour clustering algorithm on the deltas and the PCA method described in section 2.3 (reducing covariance dimensionality from 10304 to 200).

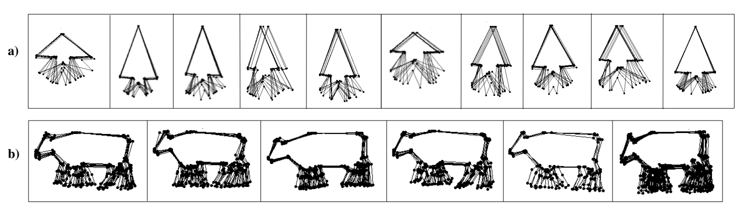

Figure 4.1: (a) ‘Triangle people’ data (b) Cow data

4.1 Evaluation of Inter-class Separation

Delta analysis was performed on the two sets of point data and the resultant space was used as a nearest neighbour classifier. This was compared to nearest neighbour classifiers in canonical space, combined eigen-canonical space, PCA eigenspace and raw vector space. Delta analysis was also performed on PCA data in a similar way to the combined eigen-canonical transform [10]. It is debatable whether Euclidean or Mahalanobis distance (a distance metric normalised by variance) should be used in spaces where the variance of components is known so both were tried where possible for evaluation. Figures 4.2 to 4.5 show the results of these tests. The final row in the tables, ‘Information content’, gives the sum of the PCA eigenvalues for components used divided by the sum of all eigenvalues to give an idea of the information content of the reduced space.

Space Used	No. Components Used							
	1	2	3	4	...	8	...	16
BC Delta Space (Mahalanobis)	88.5	97	99	96	...	85.5	...	85.5
BC Delta Space (Euclidean)	88.5	99.5	99.5	99.5	...	90	...	94.5
PCA (Mahalanobis)	86.5	52	78	79	...	95	...	98
PCA (Euclidean)	86.5	81	96	95	...	95.5	...	95.5
Eigen-Canonical (Max. Dimensionality)	27.5	57	60.5	73.5	...	42.5	...	44.5
Eigen-Canonical (Half Dimensionality)	46	62	71	76.5	...	33.5	...	-
Regular Canonical	71.5	93	91	89	....	40	...	24.5
PCA+Delta BC Space (Mahalanobis)	28.5	66.5	86.5	92	...	95.5	...	95.5
PCA+Delta BC Space (Euclidean)	28.5	65	87	92	...	98	...	98
Raw Vector	-	-	-	-	...	-	...	94.5
Information content	0.59	0.77	0.89	0.98	...	0.99	...	1

Figure 4.2: Classification Accuracy (%) For Various Spaces (Triangle People)

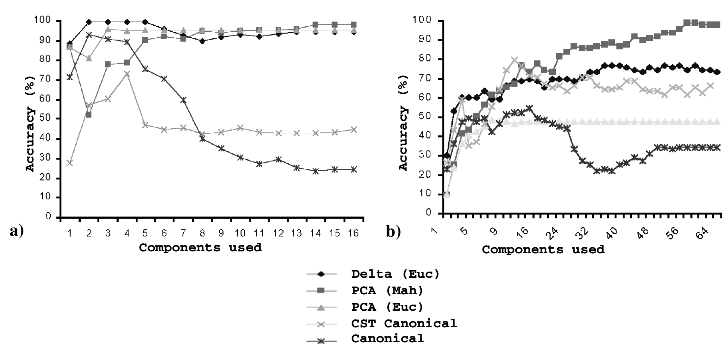

Figure 4.3: Classification Accuracy (%) For Various Spaces a) Triangle People b) Cows

Space Used	No. Components Used							
	1	2	3	4	...	32	...	64
BC Delta Space (Mahalanobis)	30.3	53.5	58.6	59.6	...	72.7	...	73.7
BC Delta Space (Euclidian)	30.3	53.5	60.6	60.6	...	73.7	...	73.7
PCA (Mahalanobis)	11.1	25.3	41.4	43.4	....	85.86	...	97.8
PCA (Euclidean)	10.1	24.2	36.4	40.4	...	47.5	...	47.5
Eigen-Canonical (Max. Dimensionality)	26.3	43.4	59.6	35.4	...	66.7	...	96.0
Eigen-Canonical (Half Dimensionality)	28.3	41.4	50.5	45.5	...	55.6	...	-
Regular Canonical	23.2	36.4	47.5	49.5	...	22.2	...	34.3
PCA+Delta BC Space (Mahalanobis)	10.1	10.1	10.1	11.1	...	46.4	...	57.5
PCA+Delta BC Space (Euclidean)	11.1	13.1	14.1	13.1	...	53.5	...	60.6
Raw Vector	-	-	-	-	...	-	...	73.7
Information content	0.20	0.33	0.44	0.52	...	0.97	...	1

Figure 4.4: Classification Accuracy (%) For Various Spaces (Cows)

Space Used	No. Components Used				Max. accuracy
	1	2	3	4	
Approx. Delta BC Space (Mahalanobis)	21	42	52.5	60.5	92.5[115]
PCA (Mahalanobis)	15	36.75	49.75	57	94 [79]
PCA (Euclidean)	15	36.3	51.8	58.5	90.3 [71]
Eigen-Canonical (Dimensionality = 100)	11	29	38.8	50	74.8 [20]
PCA+Delta BC Space (Mahalanobis)	2.3	1.5	3	4	93 [197]
PCA+BC Delta Space (Euclidian)	2.3	1.5	2.8	3.5	91.3 [188]
Raw Data	-	-	-	-	90.3 [10304]

Note: Value in square brackets is no. of components used.

Figure 4.5: Nearest Neighbour Classification Accuracy (%) For Various Spaces (Face Data)

4.2 Evaluation of Intra-class Separation

The intra-class space produced by delta analysis was compared to spaces produced using PCA on single class data. The 'modes of variation' in these spaces were observed to be qualitatively similar. To evaluate this we projected the first three eigenvectors of the intra-class space into these single class PCA eigenspaces. The results, shown in figure 4.6, demonstrate that these eigenvectors lie almost exclusively within the subspace defined by the first four PCA eigenvectors and (for the triangle people especially) are composed mostly of a single PCA eigenvector.

a) b)

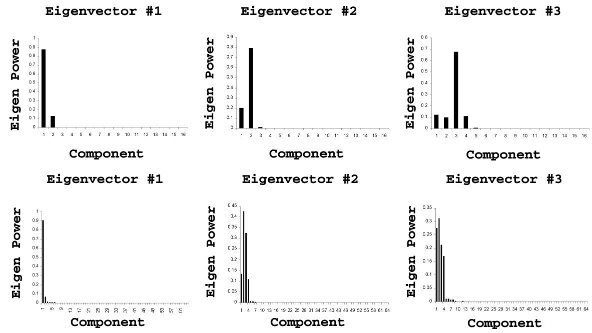

Figure 4.6: Mean of Projections of Eigenvectors into Single Class Spaces a) Triangle People b) Cows

4.3 Evaluation of Combined Representation

We approximated both real and synthetic data using truncated delta and standard PCA spaces to evaluate the trade off between data compression and accuracy in these two schemes. The mean vector error was calculated and normalised by the average vector size over all data sets. Delta space approximations were performed with equal numbers of inter and intra class components (i.e. 2 components = 1 intra and 1 inter class component).

a)

Components Used	2	4	6	8	16
PCA Eigenspace	11.4[11.3]	3.8[3.6]	2.4[2.2]	1.5[1.2]	0[0]
Delta Eigenspaces	14.9[13.0]	10.3[10.2]	8.3[9.3]	7.8[9.1]	0[0]

b)

Components Used	2	4	6	8	16	32	64
PCA Eigenspace	18.3[16.7]	15.4[14.2]	13.9[11.8]	12.4[10.2]	8.6[6.9]	4.4[3.6]	0[0]
Delta Eigenspaces	19.0[18.1]	17.5[16.4]	16.4[14.9]	15.6[14.2]	12.5[11.0]	8.5[7.4]	0[0]

c)

Components Used	10	20	50	100	200
PCA Eigenspace	16.6 [14.4]	13.9 [12.5]	10.1 [9.3]	6.7 [6.1]	0 [0]
Delta Eigenspaces	18.9 [15.9]	16.5 [14.2]	13.1 [11.5]	9.7 [8.8]	5.8 [5.6]

Figure 4.7: Mean and Standard Deviation of Error Rates (%) For Different Approximations a) Triangle People b) Cows c) Faces

Tables 4.8a-c show that the approximation accuracy for a delta space approximation is comparable but marginally worse than the equivalent eigenspace approximation with the same number of components. This is as would be expected as the eigenspace representation is optimised for dimensionality reduction. This reduction in dimensionality reduction must however be traded against the advantages of class separation.

5 Conclusions

It has been shown that delta analysis can outperform PCA, canonical analysis and combined PCA and canonical analysis (CST) in simultaneously producing optimal inter-class separation and dimensionality reduction. The intra-class optimised spaces produced by delta analysis are qualitatively similar to spaces produced by performing PCA on single class data sets assuming that intra-class variation is comparable across data sets. It has been shown in section 3.3 that subsets of the two parameterisations produced may be combined to form a powerful approximation to the data which may be used in shape search.

It has also been shown that canonical analysis and delta analysis do not always perform as well on the modified data set produced using PCA as on the raw data. In the eigenface data these methods applied to PCA data actually perform worse than PCA alone for the initial modes. The reason for these observations is that these methods work best when physical modes of variation are well de-coupled in the data representation. The eigenspace transform performed by PCA can, under certain circumstances, make the task of class separation harder by producing mixed modes which are combinations of two or more physical modes of variation. This is not to say this two stage approach is not valid as it may reduce the computational cost of creating and working with these inter and intra-class optimised spaces which is important in many applications.

The delta analysis approach is applicable to a wide variety of tasks including object recognition using outlines or greylevels and deformable object tracking as well as non machine vision applications where a multivariate classifier is required.

References

[1] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:711–720, 1997.

[2] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J.Graham. Training models of shape from sets of examples. In *Proc. British Machine Vision Conference*, pages 9–18, 1992.

[3] G. Edwards, A. Lanitis, C. Taylor, and T. Cootes. Statistical models of face images - improving specifity. *Image and Vision Computing*, 1997.

[4] K. Etemad and R. Chellappa. Discriminant analysis for recognition of human face images. In *First International Conference on Audio and Video Based Biometric Person Authentication*, pages 127–142, 1997.

[5] P. Huang, C. Harris, and M. Nixon. Recognising humans by gait via parametric canonical space. In *ICSC Symposium on Engineering of Intelligent Systems*, 1998.

[6] W.J. Krzanowski. *Kendall's library of statistics: Multivariate analysis*. Edward Arnold, 1994.

[7] D. Magee and R. Boyle. Building shape models from image sequences using piecewise linear approximation. In *Proc. British Machine Vision Conference*, pages 398–408, 1998.

[8] Ferdinando Samaria and Andy Harter. Parameterisation of a stochastic model for human face identification. In *Proceedings of 2nd IEEE Workshop on Applications of Computer Vision*, 1994.

[9] L. Sirovich and M. Kirby. A low-dimensional procedure for the characterisation of human faces. *Journal of the Optical Society of America*, 4(3):519–524, 1987.

[10] D. Swets and J. Weng. Using discriminant eigenfeatures for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18:831–836, 1996.

[11] Matthew Turk and Alex Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3:71–86, 1991.