Constructing Facial Identity Surfaces in a Nonlinear Discriminating Space

Yongmin Li, Shaogang Gong and Heather Liddell
Department of Computer Science
Queen Mary, University of London
London E1 4NS, UK
Email: {yongmin,sgg,heather}@dcs.qmul.ac.uk

Abstract

Recognising face with large pose variation is more challenging than that in a fixed view, e.g. frontal-view, due to the severe non-linearity caused by rotation in depth, self-shading and self-occlusion. To address this problem, a multi-view dynamic face model is designed to extract the shape-and-pose-free facial texture patterns from multi-view face images. Kernel Discriminant Analysis is developed to extract the significant non-linear discriminating features which maximise the between-class variance and minimise the within-class variance. By using the kernel technique, this process is equivalent to a Linear Discriminant Analysis in a high-dimensional feature space which can be solved conveniently. The identity surfaces are then constructed from these non-linear discriminating features. Face recognition can be performed dynamically from an image sequence by matching an object trajectory and model trajectories on the identity surfaces.

1 Introduction

Face recognition is emerging as an active research area in computer vision. Over the past decade, various approaches such as Eigenfaces [18], Elastic Graph model [10], Linear Object Classes [21], Active Shape Models (ASMs) [4] and Active Appearance Models (AAMs) [3] have been proposed to address this problem. It is important to point out that most of the previous work in face recognition is mainly concerned with frontal view or near frontal views. Due to the severe non-linearity caused by rotation in depth, self-occlusion, self-shading and illumination change, recognising faces with large pose variation is more challenging than that at a fixed view, e.g. frontal view.

Extracting the discriminating features, which maximise the between-class variance and minimise the within-class variance, is crucial to face recognition, especially when faces are undergoing large pose variation. Principal Component Analysis (PCA), also known as eigenface method, has been widely adopted in this research area [16, 18]. However, it is worth noting that the features extracted by PCA are actually "global" features for all face classes, thus they are not necessarily representative for discriminating one face class from others. Linear Discriminant Analysis (LDA), which seeks to find a linear transformation by maximising the between-class variance and minimising the within-class variance, proved to be a more suitable technique for classification [6, 17]. Although LDA can provide a significant discriminating improvement to the task of face recognition, it is a linear technique in nature. When severe non-linearity is involved, this method is intrinsically poor. Another shortcoming of LDA lies in the fact that the number of basis vectors is limited by the number of face classes, therefore it would be less representative when a small set of subjects is concerned. Kernel PCA (KPCA) has been developed to extract the non-linear principal components for pattern recognition problems [15, 14]. However, as with PCA, KPCA captures the *overall* variance of all patterns which are inadequate for discriminating purposes.

Another limitation of the previous studies is that the methodology adopted for recognition is largely based on matching static face images. Psychology and physiology research showed that the human vision system's ability to recognise animated faces is better than that on randomly ordered still face images (i.e. the same set of images, but displayed in random order without the temporal context of moving faces) [9, 2]. For computer vision systems, although some work has been reported [8, 5, 7], the problem of recognising the dynamics of human faces in a spatio-temporal context remains largely unresolved.

In this work, we present a comprehensive approach to address the three challenging problems in face recognition stated above. A multi-view dynamic face model is designed to extract the *shape-and-pose-free* facial texture patterns for accurate across-view registration. Kernel Discriminant Analysis (KDA), a kernel based method, is developed to compute the non-linear discriminating basis vectors. Finally face recognition is performed dynamically by matching an object trajectory tracked from an image sequence with model trajectories synthesised on *identity surfaces*.

2 Kernel Discriminant Analysis

As stated in the previous section, both PCA and LDA are limited to linear problems, and KPCA is designed to deal with the *overall* rather than the *discriminating* variance. In this work, Kernel Discriminant Analysis, a nonlinear discriminating approach based on the kernel technique [20, 15, 13, 1], is developed for extracting the nonlinear discriminating features.

The underlying principle of KDA can be described as follows: For a set of training patterns $\{\boldsymbol{x}\}$ which are categorised into C classes, $\boldsymbol{\phi}$ is defined as a non-linear map from the input space to a high-dimensional feature space. Then by performing LDA in the feature space, one can obtain a non-linear representation in the original input space. However, the computation in the high-dimensional feature space may be problematic or even impossible. By employing a kernel function

$$k(\boldsymbol{x},\boldsymbol{y}) = (\boldsymbol{\phi}(\mathbf{x}) \cdot \boldsymbol{\phi}(\mathbf{y})) \tag{1}$$

the inner product of two vectors $\boldsymbol{\phi}(\mathbf{x})$ and $\boldsymbol{\phi}(\mathbf{y})$ in the feature space can be calculated directly in the input space.

The problem can be finally formulated as an eigen-decomposition problem

$$\boldsymbol{A}\boldsymbol{\alpha} = \lambda\boldsymbol{\alpha} \tag{2}$$

The $N \times N$ matrix $\boldsymbol{A}$ is defined as

$$\boldsymbol{A} = \left(\sum_{c=1}^{C} \frac{1}{N_c} \boldsymbol{K}_c \boldsymbol{K}_c^{\mathrm{T}}\right)^{-1} \left(\sum_{c=1}^{C} \frac{1}{N_c^2} \boldsymbol{K}_c \boldsymbol{1}_{N_c} \boldsymbol{K}_c^{\mathrm{T}}\right) \tag{3}$$

where N is the number of all training patterns, N_c is the number of patterns in class c, $(\boldsymbol{K}_c)_{ij} := k(\boldsymbol{x}_i \cdot \boldsymbol{x}_j)$ is an $N \times N_c$ kernel matrix, and $(\boldsymbol{1}_{N_c})_{ij} := 1$ is an $N_c \times N_c$ matrix. More details of the underlying algorithm are available in [11].

For a new pattern $\boldsymbol{x}$, one can calculate its projection onto a KDA basis vector $\boldsymbol{v}$ in the high-dimensional feature space by

$$(\boldsymbol{\phi}(x) \cdot \boldsymbol{v}) = \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{k}_x \tag{4}$$

where $\boldsymbol{k}_x = (k(\boldsymbol{x},\boldsymbol{x}_1), k(\boldsymbol{x},\boldsymbol{x}_2), ..., k(\boldsymbol{x},\boldsymbol{x}_N))^{\mathrm{T}}$. Constructing the eigen matrix $\boldsymbol{U} = [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, ..., \boldsymbol{\alpha}_M]$ from the first M significant eigenvectors of $\boldsymbol{A}$, the projection of $\boldsymbol{x}$ in the M-dimensional KDA space is given by

$$\boldsymbol{y} = \boldsymbol{U}^{\mathrm{T}} \boldsymbol{k}_x \tag{5}$$

We use a "toy" problem to illustrate the characteristics of KDA in Figure 1. Two classes of patterns, denoted by circles and crosses respectively, have a significant non-linear distribution. We try to separate them with a *one dimensional* decision boundary of PCA, LDA, KPCA or KDA. Gaussian kernel is used in KPCA and KDA. The upper row shows the patterns and the discriminating curves computed by the four different methods. The lower row illustrates the intensity values of the one-dimensional features computed from PCA, LDA, KPCA and KDA. It can be seen clearly that PCA and LDA are incapable of providing correct classification because of their linear nature. Neither does KPCA do so since it is designed to extract the overall rather than the discriminating variance although it is nonlinear in principle. KDA gives the correct classification boundary: the discriminating curve accurately separates the two classes of patterns, and the feature intensity correctly reflects the actual pattern distribution.

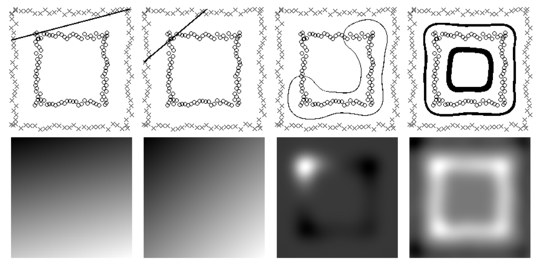
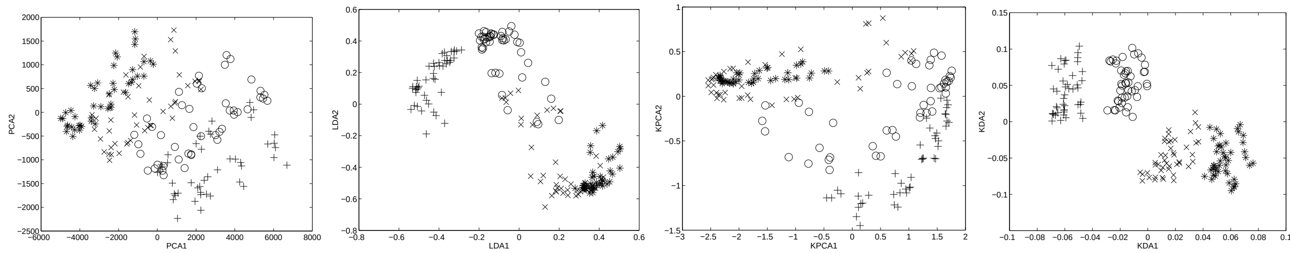


Figure 1. Solving a nonlinear classification problem with, from left to right, PCA, LDA, KPCA and KDA.

3 Multi-View Dynamic Face Model

Due to the severe non-linearity caused by rotation in depth, self-occlusion, self-shading and illumination change, modelling the appearance of faces across multiple views is much more challenging than that from a fixed, e.g. frontal, view. Another significant difficulty for multi-view face recognition comes from the fact that the appearances of different people from the same view are often more similar than those of the same person from different views.

A multi-view dynamic face model, which consists of a sparse 3D Point Distribution Model (PDM) [4], a *shape-and-pose-free* texture model, and an affine geometrical model, is developed in this work. The 3D shape vector of a face is estimated from a set of 2D face images in different views, i.e. given a set of 2D face images with known pose and 2D positions of the landmarks, the 3D shape vector can be estimated using linear regression. To decouple the covariance between shape and texture, a face image fitted by the shape model is warped to the mean shape at frontal view (with $0°$ in both tilt and yaw), obtaining a *shape-and-pose-free* texture pattern. This is implemented by forming a triangulation from the landmarks and employing a piece-wise



(a) PCA (b) LDA (c) KPCA (d) KDA

Figure 2. Distribution of multi-view face patterns in PCA, LDA, KPCA and KDA spaces.

affine transformation between each triangle pair. When part of a face is invisible in an image due to rotation in depth, the facial texture is recovered from the visible side of face using the bilateral symmetry of faces. By warping to the mean shape, one obtains the shape-free texture of the given face image. Furthermore, by warping to the frontal view, a pose-free texture representation is achieved. We applied PCA to the 3D shape patterns and *shape-and-pose-free* texture patterns respectively to obtain a low dimensional statistical model.

Based on the analysis above, a face pattern can be represented in the following way. First, the 3D shape model is fitted to a given image or an image sequence containing faces. Then the face texture is warped onto the mean shape of the 3D PDM model in frontal view. Finally, by adding parameters controlling pose, shift and scale, the complete parameter set of the dynamic model for a given face pattern is $\mathbf{c} = (\mathbf{s}, \mathbf{t}, \alpha, \beta, dx, dy, r)^{\mathtt{T}}$ where $\mathbf{s}$ is the shape parameter, $\mathbf{t}$ is the texture parameter, (α, β) is pose in tilt and yaw, (dx, dy) is the translation of the centroid of the face, and r is its scale. More details of model construction and fitting are described in [12].

Once the model is constructed, it can be automatically fitted on new images or video sequences containing faces. The *shape-and-pose-free* texture patterns obtained from model fitting are adopted for face recognition. In our experiments, we also tried to use the shape patterns for recognition, however, the performance was not as good as that of using textures.

4 Extracting the Non-linear Discriminating Features of Multi-view Face Patterns

There are mainly two kinds of variance involved for multi-view face recognition, variance from identities (between-class variance) and variance from other sources such as pose, illumination and expression changes (within-class variance). The task of face recognition is to emphasise the former and suppress the latter. Although the within-class variance has been reduced by forming the *shape-and-pose-free* facial texture patterns, the underlying discriminating features for different face classes have not been represented explicitly. Therefore such a representation in itself may not be efficient for recognition.

We illustrate this situation as in Figure 2. The multi-view face patterns of different face classes are first warped to the *shape-and-pose-free* form, then they are projected and displayed in the first two significant dimensions of PCA, LDA, KPCA and KDA. For the sake of conciseness, only patterns from four face classed are shown here. It is noted that, with PCA and KPCA, the variance from different face classes is not efficiently separated from that of pose change, or more precisely, the former is even overshadowed by the latter. Although the patterns are more separable using LDA, the performance is not as good as KDA since the non-linearity is not appropriately addressed due to the linear limitation of LDA. In this work, we adopt the KDA vectors of facial texture patterns to represent faces.

5 Recognising Multi-view Faces Using Identity Surfaces

The traditional techniques for face recognition include computing the Euclidean or Mahalanobis distance to a face template and estimating the density of patterns using multi-modal models. However, the problem of *multi-view* face recognition can be solved more efficiently if the pose information is available. Based on this idea, we propose an approach to multi-view face recognition by constructing *identity surfaces* in a discriminating feature space.

As shown in Figure 3, each subject to be recognised is represented by a unique hyper surface based on pose information. In other words, the two basis coordinates stand for the head pose: tilt and yaw, and the other coordinates are used to represent the discriminating feature patterns of faces. For each pair of tilt and yaw, there is one unique "point" for a face class. The distribution of all these "points" of a same face class forms a hyper surface in this feature space. We call this surface an *identity surface*.

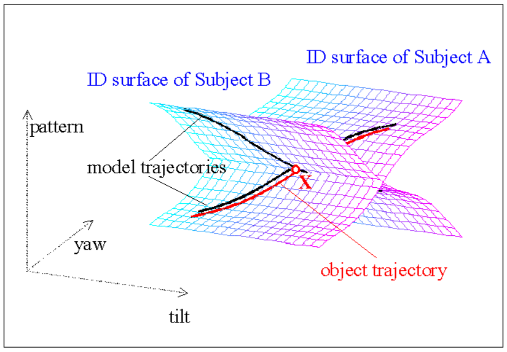


Figure 3. Identity surfaces.

5.1 Synthesising Identity Surfaces

We propose to synthesise the *identity surface* of a subject from a small sample of face patterns which sparsely cover the view sphere. The basic idea is to approximate the *identity surface* using a set of N_p planes separated by a number of N_v predefined views. The problem can be formally defined as follows:

Suppose x, y are tilt and yaw respectively, $\mathbf{z}$ is the discriminating feature vector of a face pattern, e.g. the KDA vector. $(x_{01}, y_{01}), (x_{02}, y_{02}), ..., (x_{0N_v}, y_{0N_v})$ are predefined views which separate the view plane into N_p pieces. On each of these N_p pieces, the *identity surface* is approximated by a plane

$$\mathbf{z} = \mathbf{a}x + \mathbf{b}y + \mathbf{c} \tag{6}$$

Suppose the M_i sample patterns covered by the ith plane are $(x_{i1}, y_{i1}, \mathbf{z}_{i1}), (x_{i2}, y_{i2}, \mathbf{z}_{i2}), ..., (x_{iM_i}, y_{iM_i}, \mathbf{z}_{iM_i})$, then one minimises

$$\mathcal{Q} = \sum_{i}^{N_p} \sum_{m}^{M_i} \|\mathbf{a}_i x_{im} + \mathbf{b}_i y_{im} + \mathbf{c}_i - \mathbf{z}_{im}\|^2 \tag{7}$$

$$\begin{aligned} \text{subject to} \quad : \quad & \mathbf{a}_i x_{0k} + \mathbf{b}_i y_{0k} + \mathbf{c}_i = \mathbf{a}_j x_{0k} + \mathbf{b}_j y_{0k} + \mathbf{c}_j \\ & k = 0, 1, ..., N_v, \\ & \text{planes } i, j \text{ intersect at } (x_{0k}, y_{0k}). \end{aligned} \tag{8}$$

This is a quadratic optimisation problem which can be solved using the interior point method [19].

5.2 Dynamic Face Recognition by Trajectory Matching

For an unknown face pattern $(x, y, \mathbf{z_0})$ where $\mathbf{z_0}$ is the KDA vector and x, y are the pose in tilt and yaw, one can classify this pattern into one of the known face classes by computing the distance to each of the *identity surfaces* as the Euclidean distance between $\mathbf{z_0}$ and the corresponding point on the *identity surface* $\mathbf{z}$

$$d = \|\mathbf{z_0} - \mathbf{z}\| \tag{9}$$

where $\mathbf{z}$ is given by (6).

As shown in Figure 3, when a face is tracked continuously in an image sequence using the multi-view dynamic face model described in Section 3, an object trajectory is obtained by projecting the face patterns into the KDA feature space. On the other hand, according to the pose information of the face patterns, one can build the model trajectory on the *identity surface* of each subject using the same pose information and temporal order of the object trajectory. These two kinds of trajectories, i.e. object and model trajectories, encode the spatio-temporal information of the tracked face. And finally, the recognition problem can be solved by matching the object trajectory to a set of model trajectories. A preliminary realisation of trajectory matching is implemented by computing the trajectory distances up to time slice t

$$d_m = \sum_{i=1}^{t} w_i d_{mi} \tag{10}$$

where d_{mi}, the pattern distance between the face pattern captured in the ith frame and the *identity surface* of the mth subject, is computed from (9), and w_i is the weight on this distance. Finally, the optimal m with minimum d_m is chosen as the recognition result.

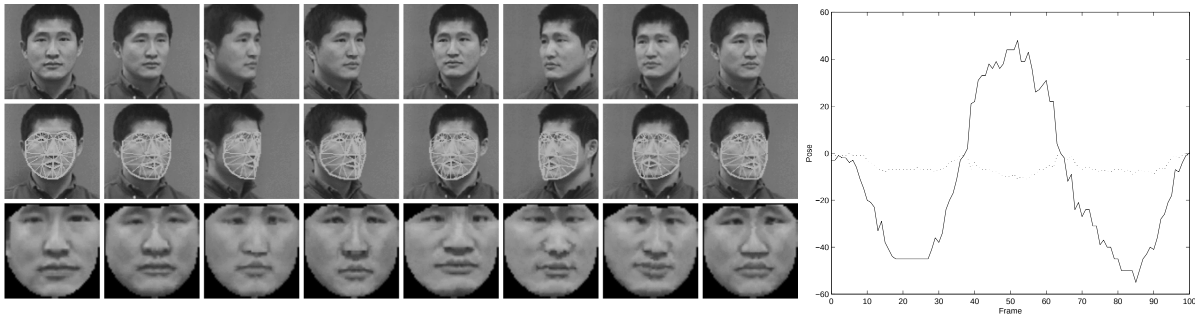
6 Experiments

We demonstrate the performance of this approach on a small scale multi-view face recognition problem. Twelve sequences, one of each subject, were used as training sequences. The sequence length varies from 40 to 140 frames. We randomly selected 180 images (15 images of each subject) to train KDA, where Gaussian kernel was adopted. Recognition was then performed on new test sequences of these subjects.

Figure 4 shows the results on one of the test sequences. The dimension of the KDA vectors is set to 10 in this experiment. It is noted that a more robust performance is achieved when recognition is carried out using the trajectory distances which include the accumulated evidence over time, although the pattern distances in each individual frame already provides good recognition accuracy on a frame by frame basis.

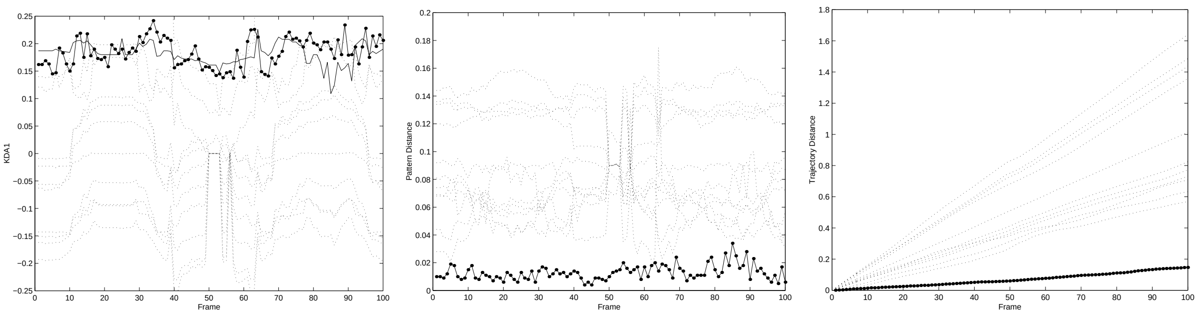
To compare with KDA, we applied the PCA, KPCA, and LDA techniques using the same set of face patterns. To make the results of different representations comparable, we define the following criterion

$$d' = \frac{1}{N} \sum_{i=1}^{N} \frac{C \cdot d_{i0}}{\sum_{j=1}^{C} d_{ij}} \tag{11}$$



(a) Sample frames, fitted 3D shape and the *shape-and-pose-free* texture patterns.

(b) Pose in tilt (dotted) and yaw (solid).



(c) Object and model trajectories.

(d) Pattern distances.

(e) Trajectory distances.

Figure 4. Video-based multi-view face recognition. (c) shows the object trajectory (solid line with dots) and model trajectories in the first KDA dimension, among which the model trajectory from the ground-truth face class is highlighted with solid line. It is noted from (d) and (e) that the pattern distances can give an accurate recognition result; however, the trajectory distances provide a more robust performance, especially its accumulated effects (i.e. discriminating ability) over time.

where C is the number of face classes, N is the total number of test face patterns, d_{ij} is the pattern distance between the ith test pattern and the jth face class, and d_{i0} is the pattern distance between the ith test pattern and the ground-truth face class.

Criterion d' can be interpreted as a summation of normalised pattern distances to their ground-truth face class. The smaller the d', the more reliable the classification performance. Figure 5 shows the values of d' for different representations, PCA, KPCA, LDA and KDA, with respect to the dimension of the feature spaces. The results indicate that KDA gives the most reliable classification performance.

The recognition accuracies with respect to the dimension of feature spaces are shown in Figure 6. It is interesting to note that the KDA features are very efficient. A 93.9% recognition accuracy was achieved when the dimension of the KDA vector was only 2. However, it is also noted that, for the small scale problem (12 subjects), PCA, LDA and KDA perform equally when more than 6 dimensional features are adopted. We will investigate how this approach generalises to large scale problems in future work.

7 Conclusions

In this paper, we have presented a comprehensive approach to multi-view dynamic face recognition. This approach is designed to addressed three challenging problems: modelling faces across multi-views, extracting non-linear discriminating features, and recognising faces dynamically in a spatio-temporal context.

Recognising faces with large pose variation involves a severe non-linearity caused by rotation in depth, self-occlusion, self-shading, and illumination change. To model the across-view faces, we developed a dynamic face model, which includes a 3D PDM, a *shape-and-pose-free* texture model, and an affine geometrical model. By representing faces with the *shape-and-pose-free* texture patterns, the variance from pose change is suppressed.

PCA, LDA and KPCA have been widely used in face recognition. But PCA and LDA are limited to the linear applications while KPCA seeks to capture the *overall* rather than the *discriminating* variance of patterns even though it is non-linear. To efficiently extract the discriminating features of multi-class patterns with severe non-linearity, KDA, which implicitly performs LDA in a non-linear feature space through a kernel function, is developed in this work. When applying KDA to the *shape-and-pose-free* texture patterns, the variance from pose change is further reduced while the between-class variance is emphasised.

Instead of matching templates or estimating multi-modal density, the *identity surfaces* of face classes are constructed in a discriminating feature space. Recognition is then performed dynamically by matching an object trajectory tracked from an image sequence with a set of model trajectories synthesised on the *identity surfaces*. Experimental results showed that this approach provides robust and accurate recognition.

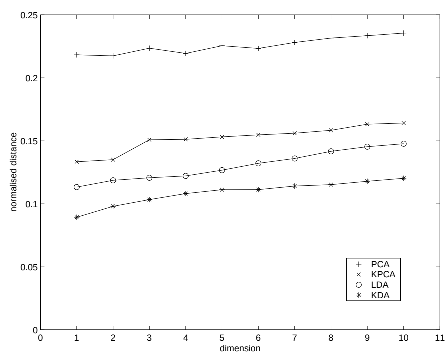


Figure 5. Recognition reliability.

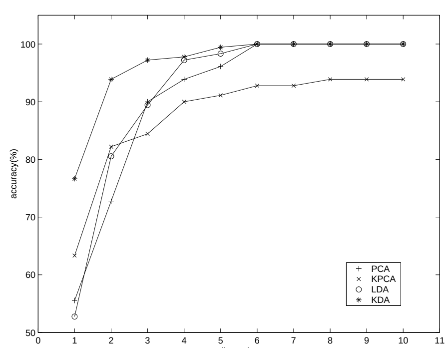


Figure 6. Recognition accuracy.

References

[1] G. Baudat and F. Anouar. Generalized discriminant analysis using a kernel approach. *Neural Computation*, 12:2385–2404, 2000.

[2] V. Bruce, A. Burton, and P. Hancock. Comparisons between human and computer recognition of faces. In *IEEE International Conference on Automatic Face & Gesture Recognition*, pages 408–413, Nara, Japan, 1998.

[3] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. In *European Conference on Computer Vision*, volume 2, pages 484–498, Freiburg, Germany, 1998.

[4] T. Cootes, C. Taylor, D. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[5] G. Edwards, C. Taylor, and T. Cootes. Learning to identify and track faces in sequences. In *IEEE International Conference on Automatic Face & Gesture Recognition*, pages 260–267, Nara, Japan, 1998.

[6] K. Fukunaga. *Introduction to statistical pattern recognitiion*. Academic Press, 1972.

[7] S. Gong, S. McKenna, and A. Psarrou. *Dynamic Vision: From Images to Face Recognition*. World Scientific Publishing and Imperial College Press, April 2000.

[8] S. Gong, A. Psarrou, I. Katsouli, and P. Palavouzis. Tracking and recognition of face sequences. In *European Workshop on Combined Real and Synthetic Image Processing for Broadcast and Video Production*, pages 96–112, Hamburg, Germany, 1994.

[9] B. Knight and A. Johnston. The role of movement in face recognition. *Visual Cognition*, 4:265–274, 1997.

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

[11] Y. Li, S. Gong, and H. Liddell. Constructing structures of facial identities using Kernel Discriminant Analysis. In *The 2nd International Workshop on Statistical and Computational Theories of Vision*, Vancouver, Canada, July 2001.

[12] Y. Li, S. Gong, and H. Liddell. Modelling faces dynamically across views and over time. In *IEEE International Conference on Computer Vision*, pages 554–559, Vancouver, Canada, July 2001.

[13] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. Muller. Fisher discriminant analysis with kernels. In *IEEE Neural Networks for Signal Processing Workshop*, pages 41–48, 1999.

[14] S. Romdhani, S. Gong, and A. Psarrou. On utilising template and feature-based correspondence in multi-view appearance models. In *European Conference on Computer Vision*, volume 1, pages 799–813, Dublin, Ireland, June 2000.

[15] B. Scholkopf, A. Smola, and K.-R. Muller. Kernel principal component analysis. In W. Gerstner, A. Germond, M. Hasler, and J.-D. Nicoud, editors, *Artificial Neural Networks – ICANN'97*, pages 583–588, Berlin, 1997. Springer Lecture Notes in Computer Science.

[16] L. Sirovich and M. Kirby. Low-dimensional procedure for the characterization of human faces. *Journal of Optical Society of America*, 4:519–524, 1987.

[17] D. Swets and J. Weng. Using discriminant eigenfeatures for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8):831–836, August 1996.

[18] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[19] R. Vanderbei. Loqo: An interior point code for quadratic programming. Technical report, Princeton University, 1994. Technical Report SOR 94-15.

[20] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer Verlag, New York, 1995.

[21] T. Vetter and T. Poggio. Linear object classes and image synthesis from a single example image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):733–742, 1997.