

Face Recognition from Multi-Pose Image Sequence

Zoran Biuk, Sven Loncaric

Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia

zbiuk@hotmail.com, sven.loncaric@fer.hr

Abstract

A novel approach to face recognition based on multi-pose image sequence is presented in this paper. In this approach faces are represented by their pattern vectors (projections to eigenfaces) in eigenspace. Instead of recognizing a face from a single view, a sequence of images showing face movement (from left to the right profile) is used for recognition. Pattern vectors corresponding to multiple poses build a trajectory in eigenspace where each trajectory belongs to one face sequence (profile to profile). In the training phase, sequences of poses construct prototypes trajectories, and in recognition phase, an unknown face trajectory is taken into comparison with prototypes. New models of matching are presented and analyzed as well as influence of some parameters to recognition ratio.

1. Introduction

During the past decade, the issue of face recognition has been extensively addressed. Techniques for computer vision-based face recognition can be largely divided into three categories: 3D model-based [15], 2D geometric feature-based [16], and 2D appearance-based matching [1,11]. A face model must exhibit invariance under changes in viewing conditions if robust recognition is to be performed. Invariance to changes in illumination, scale, translations and small rotations in the image plane can be achieved through a process of normalization of face images, but changes in face pose (large rotation in depth) cannot be easily “normalized”. 3D models and 2D geometric feature cannot be extracted robustly under large rotation in depth. Furthermore, it is very difficult to find face features relevant at different poses. Therefore, appearance-based model is more promising under changing viewing conditions.

In most of previous work, basic methodology adopted for recognition is largely based on matching of static face image patterns in a given feature space. More recently, there has been some work on face recognition using video sequences [4]. However, the issue of recognizing faces from sequences of rotating head is largely unresolved.

Therefore, we had analyzed a method of *Yongmin and Gong* and simulated the effect of variation of some parameters to recognition ratio (using a program written in MATLAB), as well as have proposed improvement of basic matching criteria from [8].

2. The eigenface method

Consider the set of all possible images, those representing a face, make up only a small fraction of it. We decided to represent images as very long vectors, instead of the matrix representation. This makes up entire image space where each face is a point. Since faces possess similar structure (eye, nose and mouth position, etc.) vectors representing them will be correlated and faces will group at certain location in the Image space. The idea behind eigenimages (in our case eigenfaces) is to find a lower dimensional space in which lower-dimensional vectors well describe faces.

In order to efficiently describe this cluster of images, we have to choose the set of directions in the images space along which the variance of the cluster is maximum. This is achieved through standard procedure of Principal Component Analysis (PCA), or K-L Transform. A direction defined in terms of the coordinate of its extremity is in the image space actually an image.

Transforming coordinates amounts to projection onto new coordinates and expressing an image as a linear combination of base images. The identified directions from the KLT thus are images, or more precisely eigen-images, and in our case we will call them eigenfaces because we are describing faces. Recognizing similar faces, is equivalent to identifying which is the closest point to the novel face, in the newly defined face-space.

2.1. Mathematical formulation

The notion of direction of variance in a high dimensional space can be extracted from the covariance matrix of the data points. The eigen-vectors of the covariance matrix points in directions of maximum variance of the data and mean square error between original and

transformed image is minimized by selecting the eigenvectors associated with largest eigenvalue.

Consider an face image G_i among collection of M images, define \mathbf{y} as the average image:

$$\mathbf{y} = 1/M \sum_i \Gamma_i, \quad i = 1, \dots, M \quad (1)$$

Every image differ from the mean by a vector $F_i = G_i - \mathbf{y}$. The covariance matrix of the data is thus defined as:

$$C = 1/M \sum_i \Phi_i \Phi_i^T = AA^T, \quad i = 1, \dots, M \quad (2)$$

where $A = [F_1 F_2 \dots F_M]$ and C has dimension $wh \times wh$ where w is the width of the image and h is the height. The size of this matrix is enormous and computing the wh eigenvectors of C is computationally hard. However, since we only sum up a finite number of image vectors M ($M \ll wh$) the rank of this matrix can not exceed $M-1$. Now, if we consider V_i to be the eigenvectors of matrix $A^T A$ whilst $A^T A$ is only a $M \times M$ matrix, i.e.

$$(A^T A)V_i = \lambda_i V_i \quad (3)$$

where λ_i are the eigenvalues, then

$$A(A^T A)V_i = A(\lambda_i V_i) \quad (4)$$

which means

$$(AA^T)(AV_i) = \lambda_i (AV_i) \quad (5)$$

and AV_i are the eigenvectors of $C = AA^T$. Therefore, the eigenvectors of C are given by:

$$U_i = AV_i = \sum_k v_k^i \Phi_k, \quad k = 1, \dots, M \quad (6)$$

where $(i = 1, \dots, M-1)$ and v_k^i is the k th element of V_i . For a given set of eigenvectors U_k , a face can be projected onto the eigenvectors by:

$$p_k = U_k^T (\Gamma - \mathbf{y}) \quad k = 1, \dots, M' \quad (7)$$

where \mathbf{y} is the average face image given by equation (1), $M' \leq M-1$ (with $M' < M-1$, the representation is approximate). Now we have *pattern vector* of face G , defined as $P = [p_1, \dots, p_{M'}]$. A face image can be represented by its pattern vector and first M' eigenfaces of the face space (given by M' most significant eigenvectors of covariance matrix A). To recognize a novel face, it is needed to calculate distance between its pattern vector and P of known face.

3. Dynamics of Faces

So far we have assumed that face images are taken from a very similar view. However, mathematical formulation can be extended to multi views sequence where face is rotating from profile to profile (pose change is between -90° to 90°).



Figure 1. Face images taken from different view angels (profile to profile).

In the described method, two face images of the same person but with large view difference are more likely to be associated with two different persons. This is because that eigenfaces do not explicitly register any three-dimensional facial structure and images are not differentiated if they are taken from different views. A variance from different persons image are not significantly isolated from the variance from different views.

But if we consider the pose information, the problem becomes different. We can form trajectory of patterns for same person. Each point of trajectory corresponds to one view angle of the same person. It is noticeable that, even for very low dimensional features space (only the first 3 PCA features used), the identities of different persons captured by their trajectories are separable. Therefore, subjects can be discriminated.

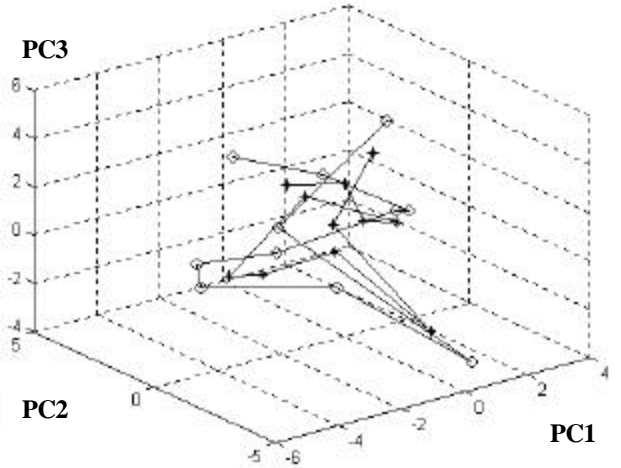


Figure 2. Pattern trajectories for 2 persons represented with first 3 principal components where yaw changes from -90° to 90° in 11 view angels (poses).

However, trajectories of same persons lie very close to each other in features space spanned by first 3 PCA components, as shown in Figure 3. Problem of face recognition can be simply matching the face trajectories.

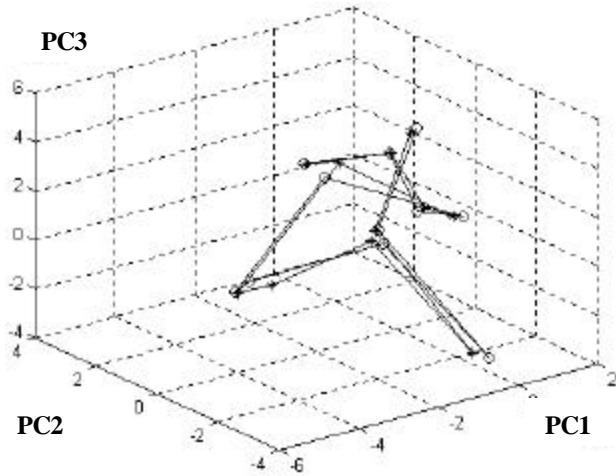


Figure 3. Pattern trajectories for 2 sequences of the same person represented with first 3 principal components.

For a given sequence containing faces to be recognized, one can obtain a trajectory by projecting the face patterns into PCA space. On the other side, using the known prototype patterns, it is easy to construct trajectory for all known faces. Therefore, the recognition problem can be solved by matching the novel trajectory to a set of prototype trajectories. In [8] it is defined distance between the novel trajectory and a model trajectory m at each frame t as:

$$d_m = \sum_i w_i d_{mi}, \quad i = 1, \dots, t \quad (8)$$

where d_{mi} is the distance in feature space between model point and object point at frame i , and w_i is the weight on this distance. To achieve more robust and accurate recognition, it is needed to perform accumulated trajectories. In other word, accumulation of positive identity information will overwhelm any misidentification over time if recognition is performed on accumulated evidence. But it may not be consistently the best match in every frame over time. This idea is illustrated in Figure 4,5. Distance can be measured as Euclidean distance in vectors space or some other like city-block distance (also called D_4 distance), etc. Difference between two lowest accumulated distances, as shown in Figure 5, can be very small. Therefore, it is important to take into account most suitable distance measure or extend matching criteria to perform robust and most accurate recognition. In this paper, we recommend an extend matching criteria for better recognition ratio and present exhaustive analysis of eigenfaces number and views number to recognition results. Because of more views used for recognition (11 in our case) and robustness of eigenfaces method, high accurate result is expected.

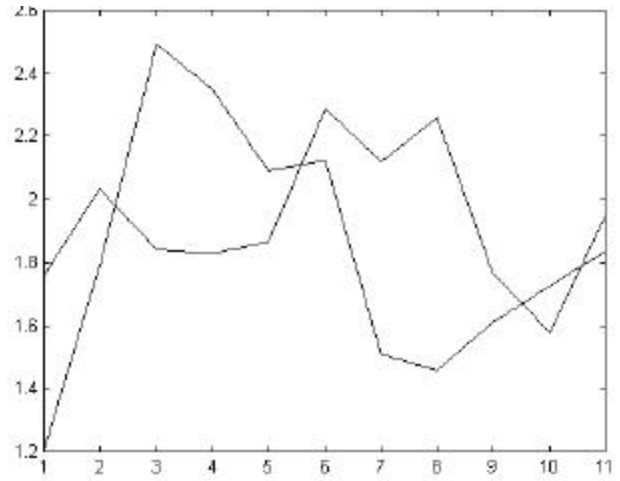


Figure 4. Distance measured independently in each of 11 frames between object pattern and 2 model patterns.

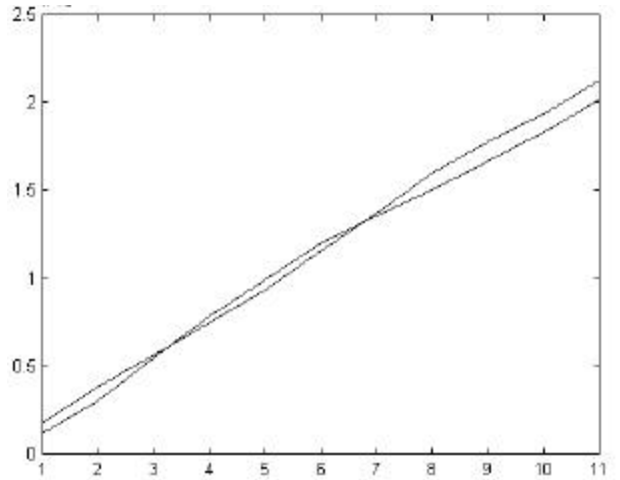


Figure 5. Accumulated distance after each of 11 frames between object pattern and 2 model patterns,. Subject associated to lower line (smaller distance) after 11 frames is true subject.

4. Experiments

We used face images taken from 28 subjects, with changes from -90° to 90° in yaw. Each subject is captured in 2-4 sequences of 11 frames (one frontal image, five from left side and five from right side). One of the captured sequences of each person is used for training (model subjects) and others for recognition. All images were scaled to size 112 x 92 pixels. Simulation program is written in MATLAB. Face images for recognition are captured under real conditions, including very different face expression, glasses, hair, rotation in image plane, etc. to

demonstrate robustness of the eigenfaces method and matching criteria.



Figure 6. Real conditions effect, (above) sequence used for training, (down) sequence used for recognition – without glasses, different face expression, different pose angle in analogous frames, hand in last frame, etc.

First experiment is made using all of 11 frames of each subject and with varying dimension of features space. It is expected to achieve better recognition ratio when more eigenfaces are used.

Table 1. Recognition results of using first n eigenfaces

Eigenfaces (n)	Used frames	Recognition rate (%)
1	11	85.7
2	11	92.8
3	11	96.4
4	11	100
5	11	100
6	11	100
27	11	100

It is evident from Table 1, that even with only two eigenfaces used for representation, significant recognition ratio is achieved (only one subject of 28 did not recognized successfully). That is because of all of 11 frames are used. Note, that just four eigenimages are enough for reliable recognition in this case. However, for higher number of subjects to be recognized, better approximation is needed (more eigenfaces for representation).

In the next experiment we varied the number of frames and investigated its influence to recognition results. We numerated frames from 1 to 11, so frontal pose is 6th pose, left profile is first and right profile is 11th frame.

Table 2. Recognition results of using only some frames and 5 eigenfaces

Eigenfaces (n)	Used frames	Recognition rate (%)
5	6 - 6	67.8
5	5 - 7	78.5
5	4 - 8	89.2
5	3 - 9	92.8
5	2 - 10	96.4
5	1 - 11	100

Table 3. Recognition results of using only some frames and 5 eigenfaces

Eigenfaces (n)	Used frames	Recognition rate (%)
3	6 - 6	57.1
3	5 - 7	64.2
3	4 - 8	75.0
3	3 - 9	82.1
3	2 - 10	85.7
3	1 - 11	96.4

As we have expected, the more frames are used the better results are achieved. It is also noticeable from Tables 2,3 that first three or five eigenfaces do not achieve enough rates if only some frames are used for recognition. Recognition rate is very sensitive with respect to the available poses when it is decreased below 11. For 3 eigenfaces approximation, thus, 11 or more frames are desirable. It is also evident from Tables 2,3 that poses close to left and right profiles are less significant for recognition than frontal and half profile poses.

Matching criteria used in all above experiments were accumulated Euclidean distance along entire trajectory. Same test were repeated with city-block distance measure. However, results did not changed significantly.

One can try to measure distance among trajectories in the points between two frames or take it into account with distance in true pose, but this provide worse results, especially for lower order of approximation (1-3 eigenfaces). To understand this effect it is important to consider Figure 2. For 11 or similar number of frames, trajectories are not smooth curves but it “jumps” in eigenspace and midpoints give inaccurate pose reconstruction. This effect is more significant when projection to lower order of eigenspace is made.

4.1. Improving matching criteria

One of the aims of this recognition method is to provide high recognition ratio using as small as possible eigenfaces and frames to minimize processing. It is obviously from experimental results that space and pose reduction yield worse results. To compensate that problem we tried to change matching criteria. Instead of accumulating distance between trajectories, we considered each frame separately. The idea is to make recognition at each pose independently and after processing all of poses choose subject with maximum score across the sequence. After reducing space dimensionality (for example to only one eigenface) considerable errors are introduced and separate recognition do not provide correct results. At some poses right subject is not positioned on the first place but, maybe, on second or third. Therefore, it is

needed to take into account all three subjects, best positioned at each frame. We have to take it with different weight according to its position. Using this measurement, considerable improvement is provided at low dimensional eigenspaces. For example, only two subjects are unrecognized when only one eigenface is applied (used weight for scoring first three positions are 1, 0.9 and 0.8). Standard distance measurement yields four unrecognized subjects. In case of higher dimensional space, this method is much worse than accumulated distance between trajectories. At very low dimensional space, errors in accumulated distance measurement are compensated with nonlinear scoring criteria.

It is true that measuring the distance between the object and model trajectories is only a simple implementation of the approach. More elaborated methods, can be more effective. Instead of using only positions in feature space we can take into account, also, an information about trajectory gradient at each point (pose) and combine it with positions (distance between patterns). Accumulated gradient difference at each frame t is defined as:

$$G_t = \sum_i \Delta G_i, \quad i = 1, \dots, t \quad (9)$$

where ΔG_i is gradient difference between object trajectory and model trajectory at frame i . After processing all frames, amplitude of accumulated gradient difference is taken into account with accumulated distance between trajectories. Combined distance between object and model trajectory we define as:

$$D = (G^2 + d^2)^{1/2} \quad (10)$$

where G is accumulated gradient difference over all frames between model and object trajectory, d is accumulated distance between trajectories defined by (8). Trajectory with smallest cumulative distance D to object trajectory is associated to person to be recognized.

In Table 4. are summarized recognition results provided using our combined distance and compared with standard trajectory distance. According to high recognition rates with both criteria when all frames and many eigenfaces are used, only interesting cases are presented (lower space dimensions and partial frame sequence). It is evident that combined distance yield better results, especially for 1 and 2 eigenfaces, or if more eigenfaces are used with partial sequence. Matching criteria based on (8) is not enough robust alone when worse feature approximation is applied. However, cumulative gradient “carry” additional information about trajectory flow and in combination with cumulative distance provide better recognition ratio and robustness.

Table 4. Comparison of standard (Euclidean) cumulative distance criteria and combined criteria.

Eigenfaces	Used frames	Combined	Distance
1	5 – 7	60.7	53.5
1	4 – 8	75.0	75.0
1	3 – 9	78.5	75.0
1	2 – 10	78.5	78.5
1	1 - 11	85.7	85.7
2	5 – 7	71.4	64.2
2	4 – 8	82.1	82.1
2	3 – 9	82.1	82.1
2	2 – 10	85.7	85.7
2	1 - 11	96.4	92.8
3	5 – 7	78.5	64.2
3	4 – 8	75.0	75.0
3	3 – 9	82.1	82.1
3	2 – 10	89.2	85.7
3	1 - 11	96.4	96.4
5	5 – 7	82.1	78.5
5	4 – 8	89.2	89.2
5	3 – 9	92.8	92.8
5	2 – 10	96.4	96.4
5	1 - 11	100	100

5. Conclusions

In this work, we presented an approach for dynamic recognition of faces across multi-pose sequence by matching trajectories in PCA feature space. For a given sequence containing faces to be recognized, one can obtain a trajectory by projecting face patterns into PCA space. For prototype subjects, it is easy to construct identity model trajectories using prototype patterns. Thus, a face recognition problem can be solved by matching the object trajectory to a set of identity model trajectories.

We wrote a program to simulate this method and influence of PCA space dimensionality and pose number to recognition ratio as well as different distance measure. For eleven or more frames and four or more eigenspaces, this method provide excellent results (with our set of 28 subjects), but recognition ratio is decreased when either parameter is reduced. Therefore we proposed matching method with scoring at each trajectory point and then decision making on the base of maximum score. This matching criteria yields better results in case of small number of PCA components (1-2 eigenfaces). Furthermore, we proposed a matching method based on cumulative

trajectories distance combined with trajectory gradient and have obtained better recognition results.

References

- [1] Turk, M., and Pentland, A., "Eigenfaces for recognition", *Journal of Cognitive Neuroscience*, 1991., pp. 3(1):71-86.
- [2] Murase, H., and Nayar, S. K., "Learning and Recognition of 3D Objects from Appearance" *In IEEE 2nd Qualitative Vision Workshop*, New York, June 1993.
- [3] S. Gong, E.-J. Ong, and S. McKenna, "Learning to associate faces across views in vector space of similarities to prototypes", *In British Machine Vision Conference*, Southampton, England, 1998., pp 54-64.
- [4] 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*, Hamburg, Germany, 1994., pp 96-112.
- [5] B. Moghaddam and A. Pentland, "Face recognition using view-based and modular eigenspaces", *In Automatic Systems for the Identification and Inspection of Humans, SPIE*, 1994., volume 2277.
- [6] O. Yamaguchi, K. Fukui, and K. Maeda. "Face recognition using temporal image sequence", *In IEEE International Conference on Automatic Face & Gesture Recognition*, Nara, Japan, 1998., pp318-323.
- [7] K. Fukunaga, "Introduction to statistical pattern recognition", *Academic Press*, 1972.
- [8] Y. Li, S. Gong, and H. Liddell, "Exploiting the Dynamics of Faces in Spatial-temporal Context", *In Proc. International Conference on Control, Automation, Robotics and Vision*, Singapore, December 2000.
- [9] K. Sung and T. Poggio, "Example-based learning for view-based human face detection", *Technical report, Massachusetts Institute of Technology*, 1994. A.I.MEMO 1521.
- [10] Y. Li, S. Gong, and H. Liddell, Multi-view face detection using support vector machines and eigenspace modelling. *In Proc. International Conference on Knowledge-based Intelligent Engineering Systems and Allied Technologies*, Brighton, UK, September 2000.pp. 241-245.
- [11] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces", *PAMI*, 1990.,12(1).
- [12] G. Robertson and I. Craw, "Testing face recognition systems", *In BMVC*, Guildford, Surrey, England, 1993.
- [13] R. Brunelli and T. Poggio, "Face recognition: Features vs. templates", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, October 1993., 15(10).
- [14] D. Swets and J. Weng, "Using discriminant eigenfeatures for image retrieval", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, August 1996., PAMI-18(8):831-836.
- [15] V. Bruce, A. Coombes, and R. Richards, "Describing the shapes of faces using surface primitives", *Image and Vision Computing*, 11,1993.
- [16] I. Craw, E. Ellis, and J. Lishman, "Automatic extraction of face features", *Pattern Recognition Letters*, 1987.