Single Model Face Database Retrieval by Directional Corner Points

Author

Gao, YS

Published

2004

Conference Title

10TH INTERNATIONAL MULTIMEDIA MODELLING CONFERENCE, PROCEEDINGS

DOI

10.1109/MULMM.2004.1264961

Single Model Face Database Retrieval by Directional Corner Points

Yongsheng Gao
School of Microelectronic Engineering, Griffith University, Australia
Email: yongsheng.gao@griffith.edu.au

Abstract

This paper presents a new face image retrieval approach using directional corner points (DCPs). Though much work is being done on face similarity matching techniques, little attention is given to the design of face matching scheme suitable for visual retrieval in single model databases where accuracy, robustness to scale and environmental changes, and computational efficiency are three important issues to be concerned. This research demonstrates that the proposed DCP approach provides a new solution, which is both robust to scale and environmental changes, and efficient in computation, for retrieving human faces in single model databases.

1. Introduction

In recent years, image retrieval by visual content from database has become a major research area due to the ever-increasing rate at which images are generated in many information systems [1]. Visual retrieval of face images, such as album search that helps police officers identify a suspect by searching a country or state photo database, is of particular interest in law enforcement applications. Bach et al. [2] first attempted this problem using an interactive manner. Automatic similarity retrieval in a face database presents a significant challenge to image database retrieval researchers due to the small inter-class variation in the face database. Human faces are very similar in structure with minor differences from person to person. Furthermore, scale variations, appearance changes due to environmental changes (i.e., lighting condition changes) and subject actions (i.e., head pose variations and facial expressions) increase the intra-class variation. These factors further complicate the face retrieval task as one of the most difficult problems in image retrieval. Though much work is being done on face similarity matching techniques [3],[4],[5],[6], [7] little attention is given to the design of face matching scheme suitable for visual retrieval in single model databases where accuracy, robustness to scale and environmental changes, and computational efficiency are three important issues to be concerned. For image retrieval, the method employed needs to satisfy following conditions: (1) It must be suitable for retrieval from single model databases because the face databases in law enforcement contain only one photo per person. (2) It should be robust to scale and environmental changes. Face detection and localization is a prior stage for automatic recognition of faces. It is inevitable that a face is localized correctly with minor scale variation. Thus it is an important requirement that a face matching algorithm is robust to scale variations. The robustness to environmental changes (i.e., lighting condition changes) is another critical issue for face image retrieval, such as photo album searching. When collecting the image data, subjects are usually required to be cooperative to have a fronto-parallel head pose with neutral expression such that the effect of appearance changes due to subject actions is minimized. However, the query image and the model image of each person are usually taken at different time and in different places. Their lighting conditions are different and unknown. Moreover, the model images from different subjects are taken under different lighting conditions. It is impossible to get a query image under the same lighting condition as when all the model images in the photo database are taken. Hence, the environmental changes could thus be the key remaining problem in this situation. (3) It must be fast in computation.

This paper proposes a new face description and similarity measuring technique for visual similarity retrieval in single model face databases, which is robust to scale and environmental changes, and efficient in computation. The proposed method employs directional corner point (DCP) matching in which directional information showing connectivity to its neighbors is utilized in the point correspondence. Unlike neural network approaches that multiple faces per person are required to train the system to optimal setting, the DCP method is suitable for single model face database retrieval and fast in computation with encouraging accuracy. In the following, Section 2 presents a lighting-insensitive face descriptor, which incorporates structural information with spatial features. In Section 3, an efficient warping algorithm for visual similarity retrieval is proposed. Particular attention is given to the discriminative power to distinguish similar objects. In Section 4, the proposed system is extensively examined with various experimental conditions, and compared with a benchmark system.

Proceedings of the 10th International Multimedia Modelling Conference (MMM'04)
0-7695-2084-7/04 $ 20.00 © 2004 IEEE

Finally, the paper concludes in Section 5.

Figure 1. An example of face DCPs superimposed on the face image.

2. Directional corner points

In this study, a new face feature descriptor, directional corner points (DCPs), is proposed to integrate the structural connectivity information with spatial features of a face image. After thinning of the edge curves, a corner detection process is applied to generate the DCPs of a face. A DCP $P(x, y, \theta_1, \theta_2)$ is represented by its Cartesian coordinates (x, y) and two directional attributes θ_1 and θ_2. θ_1 is the angular value of "horn 1" that points to its anterior neighboring corner point. Similarly, θ_2 is the angular value of "horn 2" that points to its posterior neighboring corner point. θ_1 and θ_2 range from 0° to 360°. If a DCP is a start point of a curve, a null is assigned to θ_1. If a DCP is an end point of a curve, a null is assigned to θ_2. An example of human face DCPs is illustrated in Figure 1. A DCP is either a corner point with two "horns" pointing to its two neighboring DCPs or a start/end point of the edge curve with a single "horn" pointing to its neighboring DCP. These "horns" provide isolated feature points with additional structural information about the connectivity to their neighbors. The DCP descriptor, using sparse points, further reduces the storage demand of an edge map and thus improves the computational efficiency to meet the high-speed requirement in visual retrieval of face databases. On the other hand, the structural attributes on the points enhance the discriminative power of the descriptor to improve the retrieval accuracy. The DCP descriptor is expected to be less sensitive to illumination changes due to the fact that it is a feature derived from low-level illumination insensitive edge map representation.

3. DCP matching

Based on above image coding, a face is represented by a DCP descriptor that is a set of directional corner points in a four-dimensional feature space of location and directions. The face retrieval process locates the face in the query image, generates the DCP descriptor of the query face and calculates the differences between the query DCP descriptor and the model descriptors in the database. The model in the database with minimum difference is considered as the correct return. Here we propose a warping process to perform the matching between two DCP descriptors. The difference between two faces is measured by a cost function of global warping.

Let $A(x^A, y^A, \theta_1^A, \theta_2^A)$ and $B(x^B, y^B, \theta_1^B, \theta_2^B)$ be two DCPs. A 3-step warping process, which consists of translation, rotation and open/close operations, is used to warp *A* to *B*.

(1) Translation operation: A translation operation from *A* to *B*, denoted as $T(A \to B)$, moves *A* to the location of *B* such that $x^A = x^B$ and $y^A = y^B$. The cost function for a translation operation from *A* to *B* is defined as

$$C[T(A \to B)] = \sqrt{(x^A - x^B)^2 + (y^A - y^B)^2} \tag{1}$$

(2) Rotation operation: A rotation operation from *A* to *B*, denoted as $R(A \to B)$, rotates *A* anticlockwise by α degree until the right "horn" of *A* coincides with the right "horn" of *B*. The "horn" of a double-horn DCP is defined as the right "horn" if one can rotate that "horn" anticlockwise to the other with an angle less than 180 degree. The cost function for a rotation operation from *A* to *B* is defined as

$$C[R(A \to B)] = \begin{cases} \alpha & \text{if} \quad \alpha \le 180^\circ \\ 360^\circ - \alpha & \text{if} \quad 180^\circ < \alpha \le 360^\circ \end{cases} \tag{2}$$

(3) Open/close operation: An open (or close) operation from *A* to *B*, denoted as $O/C(A \to B)$, opens (or closes) the two "horns" of *A* by β degree until the two "horns" of *A* coincides with the corresponding "horns" of *B* if the intersecting angle between the two "horns" of *A* is smaller (or greater) than that of *B*. The cost function for an open/close operation from *A* to *B* is defined as

$$C[O/C(A \to B)] = \beta \tag{3}$$

Let $A \to B$ denote a warping from *A* to *B*. The cost function for warping *A* to *B* is defined as a combined cost of above three operations in equation (4).

$$C(A \to B) = \sqrt{C^2[T(A \to B)] + f^2\{C[R(A \to B)] + C[O/C(A \to B)]\}} \tag{4}$$

$f(\cdot)$ is a non-linear function to map the angle to a scalar. It is desirable to ignore small angle variation, which is most likely segmentation error or intra-class variation, but penalize heavily on large deviation, which is most likely inter-class difference. In this study, a quadratic function $f(x)=\frac{x^2}{W}$ is used, where W is the weight to be determined experimentally in Section 4.

By substituting $A(x^A, y^A, \theta_1^A, \theta_2^A)$ and $B(x^B, y^B, \theta_1^B, \theta_2^B)$ into equation (4), we have

$$C(A \to B) = \sqrt{\begin{array}{l}(x^A - x^B)^2 + (y^A - y^B)^2 + \\ f^2\{\min[\Delta(\theta_1^A,\theta_1^B) + \Delta(\theta_2^A,\theta_2^B), \Delta(\theta_1^A,\theta_2^B) + \Delta(\theta_2^A,\theta_1^B)]\}\end{array}} \quad (5)$$

where

$$\Delta(\theta_i^A,\theta_j^B) = \begin{cases} \left|\theta_i^A - \theta_j^B\right| & \text{if } \left|\theta_i^A - \theta_j^B\right| \le 180^\circ \\ 360^\circ - \left|\theta_i^A - \theta_j^B\right| & \text{if } 180^\circ < \left|\theta_i^A - \theta_j^B\right| \le 360^\circ \end{cases} \quad (6)$$

$i, j = 1,2$

and $\theta_l^k\Big|_{l=1,2}^{k=A,B} \in [0^\circ, 360^\circ)$.

For warping between a single-horn DCP and a double-horn DCP, one of the four $\theta_l^k\Big|_{l=1,2}^{k=A,B}$ is null which represents an indefinite direction. Thus $\Delta(null, \theta_j^B)\Big|_{j=1,2}$ or $\Delta(\theta_i^A, null)\Big|_{i=1,2}$ is required for the calculation of equation (5). Since $\Delta(\theta_i^A, \theta_j^B)\Big|_{i,j=1,2}$ ranges from 0° to 180° as defined in equation (6), the maximum value, 180°, is assigned to $\Delta(null, \theta_j^B)\Big|_{j=1,2}$ or $\Delta(\theta_i^A, null)\Big|_{i=1,2}$ to penalize warping a single-horn DCP to a double-horn DCP, or vice versa. It is desirable to prohibit a warping between two DCPs of different types. The cost function for warping between a single-horn DCP and a double-horn DCP is the same as equation (5).

For warping between two single-horn DCPs, a rotation operation from A to B rotates A anticlockwise by α degree until the single "horn" of A coincides with the single "horn" of B. The cost function for a rotation operation between two single "horn" DCPs is of the same form as equation (2). It can be rewritten as equation (7) by substituting $A(x^A, y^A, \theta_1^A, \theta_2^A)$ and $B(x^B, y^B, \theta_1^B, \theta_2^B)$ into equation (2).

$$C[R(A \to B)] = \Delta(\theta_{nonull}^A, \theta_{nonull}^B) \quad (7)$$

where θ_{nonull}^A and θ_{nonull}^B are of non-null values which indicate the directions of the single-horns of A and B, respectively.

The cost function for an open/close operation in this case is defined the same as $C[R(A \to B)]$ such that the value of $C[R(A \to B)] + C[O/C(A \to B)]$ is within the same range of $[0^\circ, 360^\circ)$ in all of the three warping cases (i.e., double-horn to double-horn warping, single-horn to double-horn warping or vice versa, and single-horn to single-horn warping). Thus the cost function for warping a single-horn DCP (A) to another single-horn DCP (B) is defined as

$$C(A \to B) = \sqrt{(x^A - x^B)^2 + (y^A - y^B)^2 + f^2\{2 \times \Delta(\theta_{nonull}^A, \theta_{nonull}^B)\}} \quad (8)$$

In that a DCP is modeled as a point in the four-dimensional (4-D) feature space of location and directions, the representation of a generic face results to be a set of points in this space. The dissimilarity between two faces can be characterized by the warping cost between the two sets in the 4-D space, as shown in Figure 2.

Given two finite DCP sets $Q(A_1, A_2, \cdots, A_p)$ representing a query face and $M(B_1, B_2, \cdots, B_q)$ representing a model in the face database. p and q are the numbers of DCPs in Q and M. A DCP set to set warping process is proposed to establish every DCP correspondence between the two DCP sets by minimizing the global warping cost. For each DCP A_i in Q, its corresponding DCP B_j in M is identified as the one with minimum warping cost from A_i to B_j among all $B_j \in M$. The cost for establishing the paring for A_i can be calculated by

$$\min_{B_j \in M} C(A_i \to B_j) \quad (9)$$

The cost for warping the whole set Q to set M (i.e., establishing paring for all DCPs in Q), denoted as $Q \Rightarrow M$, is defined as

$$C(Q \Rightarrow M) = \frac{1}{p} \sum_{A_i \in Q} \min_{B_j \in M} C(A_i \to B_j) \quad (10)$$

Finally, the dissimilarity between Q and M is defined as the maximum value of the two minimum costs to establish bilateral correspondences from Q to M and vice versa.

$$D(Q, M) = \max[C(Q \Rightarrow M), C(M \Rightarrow Q)] \quad (11)$$

Figure 2. An example pair of DCP sets (in red and green respectively) from different face images of the same person as in Figure 1.

4. EXPERIMENTAL RESULTS

A complete system performance examination that covers all aspects of face retrieval was conducted. The following issues for face database retrieval are investigated.

1. Retrieval accuracy under controlled condition.
2. Sensitivity to scale variations.
3. Sensitivity to environmental changes. When collecting the image data in law enforcement, the subjects are usually asked to have a neutral expression and fronto-parallel face images are taken such that the effect of subject actions is minimized. However the lighting conditions between the query image and the models are usually different because they are taken at different time and in different places.
4. Computational speed.

The system was compared with the eigenface method [5], one of the best face image retrieval approaches. In this study, two well-known and publicly available face databases were tested. The AR face database [8], [9] from Purdue University was used to evaluate the system performances under controlled condition, scale variations and lighting condition changes. The database consists of over 3200 color images of the frontal view faces of 126 people (70 men and 56 women). There are 26 different images per person, recorded in two different sessions with a two-week time interval, each session consisting of 13 images. However, some images were found lost or corrupted after downloading through Internet. One hundred and twelve sets of images (61 men and 51 women) are complete and can be used. No restrictions on wear (clothes, glasses, etc.), make-up, hairstyle etc. were imposed to the participants. Since the number of face images per person in the AR database is larger than in many other public available face databases and only one image is used for training purpose, we are able to test the proposed approach under a large variety of conditions. The database from the University of Bern [10] was used to examine the system performances under controlled condition. The database contains frontal views of 30 people with different head pose variations (Two fronto-parallel pose, two looking to the right, two looking to the left, two looking downwards and two looking upwards). In all the experiments, a preprocessing to locate the faces was applied. Original images were normalized (in scale and orientation) such that the two eyes were aligned roughly at the same position with a distance of 80 pixels. Then the facial areas were cropped into the final images for query and retrieval.

4.1 *W* determination

The effect of W in equation (5) was investigated using the AR database. All the neutral expression faces under background lighting condition taken in the first session were used to construct the model database. The neutral expression faces under background lighting condition taken in the second session were used as query images. The retrieval rate is plotted against the W in Figure 3. W was found easily tuned since the retrieval rate remained higher than 93.75% when W ranged from 500 to 1200 (Figure 3). The algorithm could only achieve a 2.68% retrieval rate when W=1. It improved quickly with the increase of *W* and reached the optimal value of 94.64% when *W* was 800 and remained unchanged till 950. For all the other experiments in this study, *W* was set as 900.

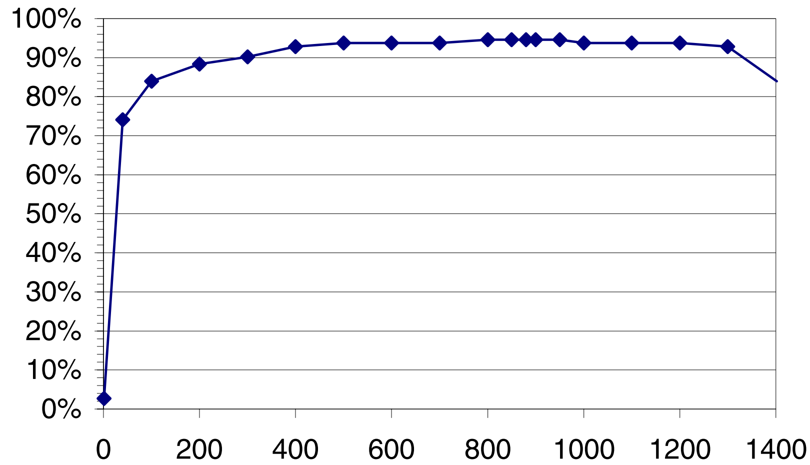

Figure 3. The selection of *W*.

4.2 Face retrieval under controlled condition

The face images under controlled condition in the database [10] were also used to evaluate the performance of the proposed approach. The retrieval rates are summarized in Table 1. The DCP experiments were conducted under three retrieval conditions, namely, top 1, top 5 and top 10 retrievals. In the top 1 retrieval, a correct retrieval was counted when the best returned face in the model database was from the same person of the query face. In the top 5 or the top 10 retrieval, a correct retrieval was counted when the face image from the same person of the query face was among the best 5 or 10 returned faces in the model database respectively.

4.3 Sensitivity to scale variations

A sensitivity analysis to scale variations was conducted using the AR database. The scale variations were generated by applying a random scaling factor, which was uniformly distributed within [1-10%, 1+10%], to the query images. Four faces with different scales were generated from each query image. Thus, we had 448 query faces for the retrieval test with scale variations ranging from –10% to +10%. The scales of the models were not changed. The experimental results are tabulated in Table 2. The results show that the proposed DCP approach outperformed the eigenface approach by 23.4%, which means that it is much more robust to scale variations than the eigenface method. This is a very attractive property that can alleviate the difficulty of precisely locating faces in the prior face detection stage.

Table 1. Retrieval accuracies of the proposed DCP and the eigenface (20-eigenvectors) approaches.

	Approach	Retrieval rate
Bern database [10]	Eigenface	100%
	DCP	100%
AR database [9]	Eigenface	55.4%
	DCP (top 1 retrieval)	94.64%
	DCP (top 5 retrieval)	99.11%
	DCP (top 10 retrieval)	100%

Table 2. Retrieval accuracies with scale variations.

	Top 1 retrieval	Top 5 retrieval
Eigenface (112-eigenvectors)	44.9%	68.8%
DCP method	68.30%	73.22%

Table 3. Retrieval accuracies with lighting condition changes.

Query faces	Eigenface		DCP method			
			Top 1 retrieval	Decrease	Top 5 retrieval	Decrease
Normal condition (Benchmark)	20-eigenvectors	55.36%	94.64%	--	99.11%	--
	112-eigenvectors	78.57%				
Left light on	20-eigenvectors	6.25%	87.50%	7.14%	96.43%	2.68%
	112-eigenvectors	9.82%				
	112-eigenvectors w/o 1st 3 components	26.79%				
Right light on	20-eigenvectors	4.46%	89.29%	5.35%	96.43	2.68%
	112-eigenvectors	7.14%				
	112-eigenvectors w/o 1st 3 components	49.11%				
Both lights on	20-eigenvectors	1.79%	61.61%	33.03%	81.25	17.86%
	112-eigenvectors	2.68%				
	112-eigenvectors w/o 1st 3 components	64.29%				

4.4 Sensitivity to environmental changes

The robustness to environmental changes (i.e., lighting condition changes) is one of the critical issues for face image retrieval systems, such as photo album searching systems. When collecting the image data, subjects are usually required to be cooperative to have a neutral expression and fronto-parallel view faces are taken such that the effect of subject actions (i.e., head pose variations and facial expressions) is minimized. Usually, passport/IC photos are used to construct a model face database. Thus the environmental changes could be the only remaining appearance variability problem in face retrieval. In general, the query image and the model image of each person are taken at different time and in different places. Their lighting conditions are different and unknown. Moreover, the model images from different subjects are taken under different lighting conditions. It is impossible to get a query image under the same lighting condition as when all the model images in the photo database are taken. Hence, the retrieval algorithm has to be robust to the variability in appearance due to lighting condition changes.

The issue addressed in this subsection is whether the DCP representation is sufficient or how well it performs for retrieving faces under varying lighting conditions. The experiment was designed using face images taken under different lighting conditions from the AR database. The faces in neutral expression with background illumination were used as single models of the subjects. The images under three different lighting conditions were used as query images. The experimental results on query images with three different lighting conditions are illustrated in Table 3 together with the retrieval results under the controlled condition as comparison benchmarks. In all the three experiments, the proposed DCP method significantly outperformed the eigenface approach. For the eigenface method, it has been suggested that the first three principal components are the primary components responding sensibly to lighting variations. The system error rate can thus be reduced by discarding these three most significant principal components [11]. Though the accuracies of the eigenface approach increased without using the first 3 eigenvectors, the DCP approach still significantly outperformed it when one light was on. The variations of lighting condition did affect the system performance. Nevertheless, the DCP approach is much more tolerant to lighting condition changes than the eigenface method. The effect on retrieval rates when on light was one produced only 7.14% and 5.35% decreases in retrieval accuracy for the DCP approach. When both lights were on, the error rate became much higher than that of only one light on. This evidence shows that the DCP would still be affected by extreme lighting condition changes, such as over-illumination, though it is less sensitive to some extent. The over-illumination would cause strong specular reflection on the face skin (It is no longer a Lambertian surface). Therefore the shape information on faces would have been suppressed or lost, which would result in the increase of the error rate.

4.5 Computational speed

An experiment was conducted to evaluate the computational efficiency of the DCP matching using the AR face database. The average computational time for one match is 57ms. The experiment was conducted on a PC under Windows platform with 1.3GHz CPU and 512MB RAM. In a face retrieval system, searching is the most computationally expensive operation due to the large number of images available in the database. Therefore, it is a prerequisite of image retrieval systems to use efficient visual similarity matching algorithms. In most systems, face features are extracted/coded off-line from the original images, and stored in the face feature database. In querying process, the same features are extracted from the query face, and the features of the query image are compared with the features of each model image in the database. In practice, apart from adopting a fast face matching algorithm, pre-filtering operation [7] can be employed to further speed up the search by reducing the number of candidates.

5. Conclusion

This paper presents a new face image retrieval approach using directional corner points (DCPs). The proposed DCP is particularly designed to address the problem of image querying and retrieval from single model face databases where accuracy, robustness to scale and environmental changes, and computational efficiency are three important issues to be concerned. In order to meet the efficiency requirements of an image descriptor in high-dimensional 2-D image spaces, we extract directional corner points from edge maps to further reduce the storage demand and the computational expense of edge map matching, while keeping the advantage of insensitivity to illumination changes. On the other hand, the directional attributes are introduced to enhance the discriminative capability to cater for high accuracy requirement.

The algorithm has been evaluated using two well known and public available face databases of over 3000 images and compared with the eigenface approach, one of the best existing face retrieval techniques. It is a very encouraging finding that the proposed DCP approach performed superior to the eigenface approach in most of the comparison experiments. DCP correctly returned 100% and 94.64% of the queries on the face databases [10] and [9] respectively. The DCP approach significantly outperformed the eigenface method by 23.4% in retrieval rate by querying a face with scale variation. The effect on retrieval rates when one light was on degraded only by 7.14% and 5.35%. This study demonstrates that the proposed DCP method provides a new solution, which is both robust to scale and environmental changes, and efficient in computation, for retrieving human faces in single model databases.

References

[1] Y. A. Aslandogan and C. T. Yu, "Techniques and Systems for Image and Video Retrieval", IEEE Transactions on Knowledge and Data Engineering, Vol. 11, pp.56-63, 1999.

[2] J. R. Bach, S. Paul and R. Jain, "A Visual Information Management System for the Interactive Retrieval of Faces", IEEE Transactions on Knowledge and Data Engineering, Vol. 5, pp.619-628, 1993.

[3] S. Lawrence, C. L. Giles, A. C. Tsoi and A. D. Back, "Face Recognition: A Convolutional Neural-Network Approach", IEEE Trans. on Neural Networks, Vol. 8, pp. 98-113, 1997.

[4] R. Bruneli and T. Poggio, "Face Recognition: Features versus Templates", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 15, pp. 1042-1052, 1993.

[5] M. Turk and A. Pentland, "Eigenfaces for Recognition", Journal of Cognitive Neuroscience, Vol. 3, pp. 71-86, 1991.

[6] M. Lades, J. C. Vorbrüggen,, J. Buhmann, J. Lange, C. von der Malsburg, R. P. Würtz and M. Konen, "Distortion Invariant Object Recognition in the Dynamic Link Architecture", IEEE Transactions on Computers, Vol. 42, pp. 300-311, 1993.

[7] Y. Gao and M. K. H. Leung, "Face Recognition Using Line Edge Map", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No.6, pp.764-779, 2002.

[8] A. M. Martinez and R. Benavente, "The AR Face Database", CVC Technical Report #24, June 1998.

[9] Face database, Purdue University. http://rvl1.ecn.purdue.edu/~aleix/aleix_face_ DB.html

[10] Face database, the University of Bern. ftp://iamftp.unibe.ch/pub/Images/FaceImages/

[11] P. N. Belhumeur, J. P. Hespanha and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, pp. 711-720, 1997.

0-7695-2084-7/04 $ 20.00 © 2004 IEEE