FACIAL FIDUCIAL POINTS DETECTION USING DISCRIMINATIVE FILTERING ON PRINCIPAL COMPONENTS

Waldir S. S. Júnior [1,2], *Gabriel M. Araújo* [1], *Eduardo A. B. da Silva* [1], *Siome K. Goldenstein* [3]

[1] PEE/COPPE/DEL/POLI, Universidade Federal do Rio de Janeiro,
Cx.P. 68504, Rio de Janeiro, RJ, CEP 21945-970, Brazil.
[2] DET/CETELI, Universidade Federal do Amazonas,
R. O. Octávio Jordão Ramos, 3000, Coroado I, Manaus, AM, CEP 69077-000, Brazil.
[3] IC, Universidade Estadual de Campinas, Cx.P. 6176, Campinas, SP, CEP 13084-971, Brazil.
e-mails: waldirjr@ufam.edu.br, gmatos@lps.ufrj.br, eduardo@lps.ufrj.br, siome@ic.unicamp.br.

ABSTRACT

The Discriminative Filtering technique performs pattern recognition using a two-dimensional filter. It has a closed-form design, based on the pattern and the statistics of the image set. Here, we investigate the use of Discriminative Filtering for detecting fiducial points in human faces. We show that designing discriminative filters for the principal components increases robustness. The method is assessed in a fiducial points detection framework using a Gentle AdaBoost classifier.

***Index Terms*—** Pattern Recognition, Discriminative Filtering, Principal Component Analysis, Fiducial Points, AdaBoost.

1. INTRODUCTION

In the last years, there has been some effort in developing methods for pattern recognition that use linear filtering, such as Expansion Matching (EXM) [1] and Discriminative Filtering [2, 3] in addition to methods that use shape models [4]. In discriminative filtering, the filters are designed using closed-form expressions. For example, in [2], detectors that recognize letters of the alphabet were implemented. However, these methods still lack robustness in case of both noise and small pattern mismatches.

In this work, we propose a new form of incorporating robustness into the Discriminative Filtering framework, through the Principal Components Analysis. This is done so that small changes in the target patterns do not impair the effectiveness of the detection algorithm. A different discriminative filter is designed, for each of the directions of highest energy, from the pattern set which we desire to detect/recognize. Specifically, the patterns of interest used in this work are salient characteristics (or landmarks) of human faces, such as corners of eyes, pupils, and corners of noses, among others, which are known in the literature as good fiducial points [5].

In order to evaluate the proposed method we use 11 fiducial points of 505 images from BioID [6] database. Simulation results show that it outperforms a classifier based on the early discriminative filters [2, 3] as well as the SVM [7] classifier, both in terms of false positive and true positive rates.

The remainder of this paper is organized as follows: in Section 2 a review of the Discriminative Filtering method and Principal Components Analysis is provided. The discriminative filter design for principal components is discussed in Section 3. Section 4 presents the experimental results. Finally, conclusions are drawn in Section 5.

2. BACKGROUND REVIEW

2.1. Discriminative Filtering

The Discriminative Filtering [3] is a method, based on two-dimensional filtering, which was developed for pattern recognition. The goal of the discriminative filter $\mathbf{\Theta}_{M\times M}$ is to maximize the energy of a coefficient $c(m,n)$ in a matrix $\mathbf{C}_{M\times M}$, obtained when the pattern $\mathbf{U}_{M\times M}$ (which we want to detect) is convolved with $\mathbf{\Theta}$. The metric used to measure the energy of $c(m,n)$ is called DSNR_2, (*Two-dimensional Discriminative Signal to Noise Ratio*), given by:

$$(\text{DSNR}_2)_{(m,n)} = \frac{c(m,n)^2}{\sum_{i=1}^{M}\sum_{j=1}^{M} c(i,j)^2 - c(m,n)^2}, \quad (1)$$

where $c(m,n)$ is the two-dimensional convolution between the filter $\mathbf{\Theta}$ and the pattern $\mathbf{U}$.

Recently, Mendonça and da Silva [8] showed how to achieve an analytical solution to $\mathbf{\Theta}$ by using an alternative interpretation of the referred process, based on the impulse restoration analogy [9]. The pattern $f(m,n)$, located in position

This research was supported by the FAPEAM agency.

(m_0, n_0) of image $g(m, n)$, can be written as follows:

$$g(m, n) = f(m - m_0, n - n_0) + b(m, n), \tag{2}$$

where $b(m, n)$ is the image content without the target pattern.

Note that, in equation (2), $g(m, n)$ is obtained by convolving $f(m, n)$ with an impulse in (m_0, n_0), adding a signal $b(m, n)$ to the result. With that model in mind, finding the position (m_0, n_0) is equivalent to a restoration problem where the observed image is an impulse distorted by $f(m, n)$ and corrupted by $b(m, n)$, which can be regarded as noise. Using a matrix notation, we have:

$$\mathbf{g} = \mathbf{F}\boldsymbol{\delta} + \mathbf{b}, \tag{3}$$

where $\mathbf{F}$ is a circulant matrix whose first column is given by the pattern.

Thus, the optimal filter is the one that restores an impulse $\boldsymbol{\delta}$ distorted by the target pattern and corrupted by noise. The estimation process for $\boldsymbol{\delta}$, resulting in $\widehat{\boldsymbol{\delta}}$, is carried on by using the *Linear Minimum Mean Square Error* (LMMSE) [10] estimate. The best linear approximation, which minimizes the mean squared error $E[\|\boldsymbol{\delta} - \widehat{\boldsymbol{\delta}}\|^2]$, is given by:

$$\widehat{\boldsymbol{\delta}} = \mathbf{F}^T \left[\mathbf{F}\mathbf{F}^T + \mathbf{C_b}\right]^{-1} \mathbf{g} = \mathbf{A}\mathbf{g}, \tag{4}$$

where $\mathbf{C_b}$ is the covariance matrix of the noise. The discriminative filter $\boldsymbol{\Theta}$ can be obtained from $\mathbf{A}$ by inspection [2]. The superscript ^{T} denotes the Hermitian of the matrix.

2.2. Principal Components Analysis

The Principal Component Analysis (PCA) is a technique initially developed for dimensionality reduction problems. However, it has also found widespread use on pattern recognition. Examples of its application are the Eigenfaces [11] and Eigenfeatures [12] approaches developed for face recognition and features detection, respectively.

The mathematical formulation of PCA can be obtained as follows: supposing that $\mathcal{U}$ is a random vector, the optimal base $\boldsymbol{\Phi}$, composed by N orthogonal principal directions, can be obtained through diagonalization of the covariance matrix $\boldsymbol{\Sigma}_{\mathcal{U}}$, using the solution computed with the eigenvalues approach [12]. More specifically,

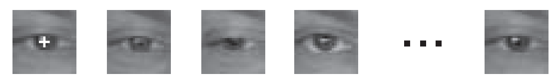
$$\boldsymbol{\Lambda} = \boldsymbol{\Phi}^T \boldsymbol{\Sigma}_{\mathcal{U}} \boldsymbol{\Phi}, \tag{5}$$

where $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_N]$ is the matrix of ortogonal eigenvectors of $\boldsymbol{\Sigma}_{\mathcal{U}}$ and $\boldsymbol{\Lambda}$ is a diagonal matrix, whose diagonal elements are the eigenvalues $\lambda_1 \geq \ldots \geq \lambda_N$ of $\boldsymbol{\Sigma}_{\mathcal{U}}$.

3. DISCRIMINATIVE FILTERING USING PRINCIPAL COMPONENT ANALYSIS

Here we propose increase the robustness of discriminative filtering by designing discriminative filters for the principal components of highest energy. Small changes in the target pattern do not modify a great deal the energy of the principal components. Since the discriminative filters are linear, one expects that such small changes produce only small perturbations at the filters' outputs.

A different discriminative filter is designed for each of the directions of highest energy, considering the common features in the pattern set we wish to detect. The patterns of interest used in this work are salient features of human faces, like corners of eyes, noses, and center of eyes, which are referred to the literature as good fiducial points [5] (see figure 1).



$\mathbf{U}_1$ $\mathbf{U}_2$ $\mathbf{U}_3$ $\mathbf{U}_4$ … $\mathbf{U}_M$

Fig. 1. Set formed by M blocks $\mathbf{U}$ whose centers are fiducial points of the human eye center. Note, in the first image, the mark on the fiducial point.

Suppose that the realizations of the random variable $\mathcal{U}$ are equal to $\mathbf{U}_1, \ldots, \mathbf{U}_M$. Given that, the implementation of discriminative filters $\boldsymbol{\Theta}_{\boldsymbol{\phi}_1}, \ldots, \boldsymbol{\Theta}_{\boldsymbol{\phi}_N}$ is carried on by using N principal directions, which are $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_N]$ (see equation (5)), with eigenvalues $\lambda_1, \ldots, \lambda_N$ associated to them.

The use of the principal directions provides an interesting method to determine the covariance matrix $\mathbf{C}_{\mathbf{b}i}$ associated to the discriminative filter for the component i (see equation (4)). Supposing that we intend to detect the principal direction $\boldsymbol{\phi}_i$ the patterns that are not of interest lie in the subspace orthogonal to $\boldsymbol{\phi}_i$. Therefore, the noise associated to component i can be regarded as the projection of typical patterns on the subspace orthogonal to $\boldsymbol{\phi}_i$. It thus has energy λ_j in direction $\boldsymbol{\phi}_j$, $j \neq i$. Thus, the covariance matrix $\mathbf{C}_{\mathbf{b}i}$ associated with direction $\boldsymbol{\phi}_i$ can be expressed as:

$$\mathbf{C}_{\mathbf{b}i} = \sum_{\substack{j=1 \\ j \neq i}}^{N} \lambda_j \mathbf{F}_{\boldsymbol{\phi}_j} \mathbf{F}_{\boldsymbol{\phi}_j}^T, \tag{6}$$

where $\mathbf{F}_{\boldsymbol{\phi}_j}$ is the circulant matrix whose first column is given by the direction $\boldsymbol{\phi}_j$.

In this case, equation (4) is changed to:

$$\mathbf{A}_{\boldsymbol{\phi}_i} = p_i \mathbf{F}_{\boldsymbol{\phi}_i}^T \left[p_i \mathbf{F}_{\boldsymbol{\phi}_i} \mathbf{F}_{\boldsymbol{\phi}_i}^T + (1 - p_i)\mathbf{C}_{\mathbf{b}i}\right]^{-1}, \tag{7}$$

where p_i is the *a priori* probability of $\boldsymbol{\phi}_i$ and $\boldsymbol{\Phi}$ is the base that diagonalizes the covariance matrix $\boldsymbol{\Sigma}_{\mathcal{U}}$ of the random variable associated to the target patterns. A supposition the leads to goods results is to consider all p_i to be equal.

Finally, by replacing (6) in equation (7) we have the full form of matrices $\mathbf{A}_{\boldsymbol{\phi}_i}$:

$$\mathbf{A}_{\boldsymbol{\phi}_i} = p_i \mathbf{F}_{\boldsymbol{\phi}_i}^T \left[p_i \mathbf{F}_{\boldsymbol{\phi}_i} \mathbf{F}_{\boldsymbol{\phi}_i}^T + (1 - p_i) \sum_{\substack{j=1 \\ j \neq i}}^{N} \lambda_j \mathbf{F}_{\boldsymbol{\phi}_j} \mathbf{F}_{\boldsymbol{\phi}_j}^T\right]^{-1}. \tag{8}$$

Again, the discriminative filter $\mathbf{\Theta}_{\phi_i}$ can be obtained from $\mathbf{A}_{\phi_i}$ by inspection [2].

4. EXPERIMENTS AND RESULTS

4.1. Pre-processing

Before the detection algorithm is used, we apply four pre-processing steps to each image, as shown in figure 2. In the first one, we apply the Viola-Jones face detection algorithm [13] to each input image. After that, the face is scaled to a predefined resolution, and then we perform illumination correction [14]. In the final step, we use a Gaussian prior model for fiducial points. A candidate to a fiducial point with label q in the image, with coordinates equal to $\mathbf{y}$, is considered to be inside the ellipse defined by the Mahalanobis distance to the average of all fiducial points in the training set, given by

$$\max_{\text{label}(\mathbf{w})=q} \{1.05|\mathbf{w}-\boldsymbol{\mu}|^2\} \geq (\mathbf{y}-\boldsymbol{\mu})^t \, \mathbf{\Sigma}_{\mathcal{X}}^{-1} \, (\mathbf{y}-\boldsymbol{\mu}) \,, \quad (9)$$

where $\mathcal{X}$ is the random vector whose realization is equal to the fiducial points, $\boldsymbol{\mu}$ is the vector mean and $\mathbf{\Sigma}_{\mathcal{X}}$ is the covariance matrix of $\mathcal{X}$.

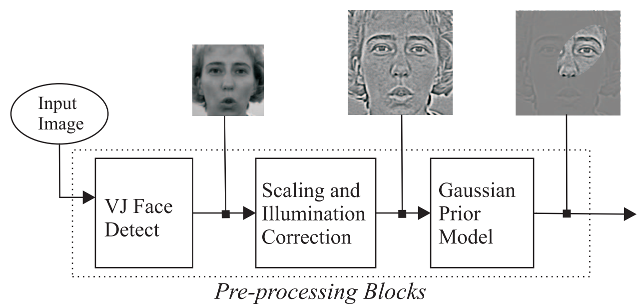


Fig. 2. The pre-processing steps.

4.2. Training and Detection

The training procedure was performed with an image set whose elements were labeled with fiducial points. A discriminative filter was designed for each of the principal directions with highest energies (see Section 3). They generate as output a vector whose coordinates are the DSNRs (equation (1) and figure 3) associated with each principal direction. The decision between positives (fiducial points) and negatives (other patterns) is obtained through a Gentle AdaBoost [15] classifier using the GML AdaBoost Matlab toolbox [16]. We employ a two-stage cascade scheme, were the true positives of the first stage are used as a training set to design the second stage Gentle AdaBoost classifier.

For detecting patterns of interest, we use the filters and the classifier designed in the training procedure, which identify pixels as fiducial points (positives) or other patterns (negatives). All filters are circularly convolved with a sliding window, which has the same dimensions of the filter, going through the candidate blocks in the image (see figure 3 for details).

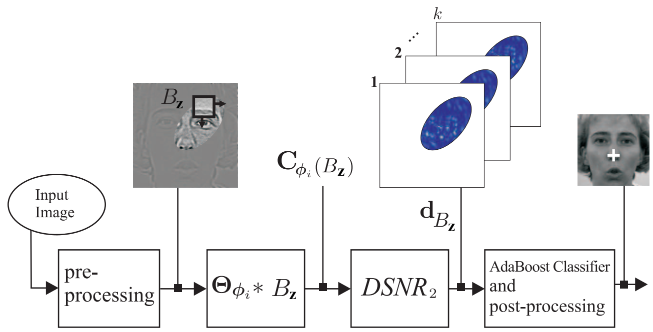


Fig. 3. The image, after pre-processing, will be processed through a sliding window $B_{\mathbf{z}}$, where $\mathbf{z}$ is the coordinate corresponding to the center of the block. $\mathbf{C}_{\phi_i}(B_{\mathbf{z}})$ is obtained by filtering $B_{\mathbf{z}}$ with each $\mathbf{\Theta}_{\phi_i}$, where $i = \{1, \ldots, k\}$, k being the number of principal components. The output of the box $DSNR_2$ provides the $DSNR_2$ (equation (1)) for each $\mathbf{C}_{\phi_i}(B_{\mathbf{z}})$. So, each block $B_{\mathbf{z}}$ has a $DSNR_2$ vector associated called $\mathbf{d}_{B_{\mathbf{z}}}$. The classifier generated with the AdaBoost algorithm will use $\mathbf{d}_{B_{\mathbf{z}}}$ to classify $B_{\mathbf{z}}$. Finally, we perform a post-processing step consisting of scaling to original image size and a simple clustering algorithm to group the close labels.

4.3. Simulation Results

In order to evaluate the proposed method we use 11 fiducial points of 505 images from BioID [6] database. The tests were made using cross-validation of the 505 images with 7-folds. The True Positive (TP) and False Positive (FP) rates were computed. We consider that a candidate is a true positive when its distance to the original labeling (annotated manually) is smaller than 10% of the face intra-ocular (between the pupils) distance.

In table 1 we show the results of the proposed method (FD-PCA) for each of the fiducial points. We include, for comparison purposes, also results of the discriminative filters without the PCA step (FD-ori) as well as the ones of a two-stage cascade of the Support Vector Machine (SVM) [7] method. Note that the differences among the various methods correspond only to the two middle boxes in figure 3. From this table, we can see that the proposed method outperforms FD-ori for all fiducial points. In addition, our approach outperforms the SVM-based for the fiducial points 0, 2, 3, 4, 5, 7 and 9 while showing competitive results for the fiducial points 1, 3, 6 and 8. Figure 4 illustrates the fiducial points found by the proposed method on two different faces. Note that it succeeded in finding all the 11 fiducial points, with one false positive at the center point on the outer edge of upper lip, that has been misclassified as the center of the nose in both images.

Table 1. Average performance, using cross-validation with 7-folds, to each fiducial point of 505 images from BioID database. The picture labeling show all the fiducial points considered. TP is the percentage of true positives and FP the percentage of false positives.

	**FD-ori**		**SVM**		**FD-PCA (our)**	
	FP	TP	FP	TP	FP	TP
0	0.13	44.1	0.21	80.5	0.02	95.2
1	0.05	76.4	0.02	99.6	0.03	95.2
2	0.05	80.3	0.08	85.0	0.04	93.0
3	0.05	79.7	0.03	98.6	0.03	94.8
4	0.05	76.3	0.05	95.2	0.03	97.2
5	0.08	64.3	0.12	90.2	0.02	97.2
6	0.24	55.4	0.06	92.2	0.06	91.8
7	0.26	69.0	0.03	47.1	0.07	94.4
8	0.24	59.4	0.07	95.2	0.07	94.2
9	0.23	52.4	0.14	88.4	0.06	89.6
10	0.33	60.8	0.06	68.5	0.04	88.0

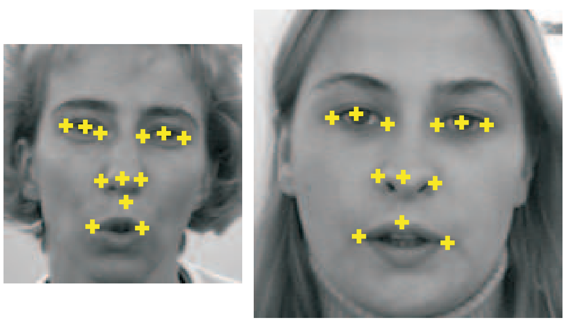


Fig. 4. Some examples of results in human faces.

5. CONCLUSIONS

In this work, we investigated the use of the Discriminative Filtering method to detect fiducial points in human faces. We developed a classifier based on discriminative filters designed for the principal components.

We used 11 fiducial points of 505 images from the BioID database to evaluate the performance of the proposed method. The results obtained show that the proposed approach outperforms a similar classifier based on discriminative filters designed without the PCA step, for all fiducial points. Our results also show that the performance of the proposed algorithm is in general superior to the one of the SVM classifier.

6. REFERENCES

[1] D. Nandy and J. Ben-Arie, "EXM eigen templates for detecting and classifying arbitrary junctions," in *Proceedings of the International Conference on Image Processing*, Kobe, Japan, October 1998, pp. 211–215.

[2] A.P. Mendonça and E.A.B. da Silva, "Multiple template detection using impulse restoration and discriminative filters," *IEE Electronics Letters*, vol. 39, no. 16, pp. 1172–1174, August 2003.

[3] A.P. Mendonça and E.A.B. da Silva, "Two-dimensional discriminative filters for image template detection," in *Proceedings of the International Conference on Image Processing*, Thessaloniki, Greece, September 2001.

[4] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models–their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995.

[5] A. Stefano, C. Paola, and L. Raffaella, "An efficient method to detect facial fiducial points for face recognition," in *Proceedings of the 17th International Conference on Pattern Recognition*, Cambridge, UK, August 2004, pp. 532–535.

[6] The BioID database, "*http://www.humanscan.de/support/downloads/facedb.php*," 2008.

[7] T. Joachims, C. Burges, B. Scholkopf, and A. Smola, Eds., *Advances in kernel methods: support vector learning*, chapter Making large-scale support vector machine learning practical, pp. 169–184, MIT press, Cambridge, MA, 1999.

[8] A.P. Mendonça and E.A.B. da Silva, "Closed-form solutions for discriminative filtering using impulse restoration techniques," *IEE Electronics Letters*, vol. 38, no. 22, pp. 1332–1333, October 2002.

[9] A. Abu Naser, N.P. Galatsanos, and M.N. Wernick, "Impulse restoration-based template-matching using the expectation-maximization algorithm," in *Proceedings of the International Conference on Image Processing*, Washington, DC, October 1997, pp. 158–161.

[10] A. Abu Naser, *Impulse Restoration-Based Template-Matching*, Ph.D. Dissertation, University of Illinois, Chicago, USA, 2000.

[11] M. Kirby and L. Sirovich, "Application of the karhunen-loeve procedure for the characterization of human faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, 1990.

[12] Z.L. Stan and K.J. Anil, *Handbook of Face Recognition*, Springer-Verlag, Secaucus, NJ, USA, 1ed edition, 2004.

[13] P. Viola and M. Jones, "Robust real-time object detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, July 2001.

[14] T. Xiaoy and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," in *Proceedings of the International Workshop on Analysis and Modeling of Faces and Gestures*, 2007, pp. 168–182.

[15] F. Jerome, H. Trevor, and R. Tibshirani, "Additive logistic regression: a statistical view of boosting," *Annals of Statistics*, vol. 28, pp. 2000, 1998.

[16] The GML AdaBoost Matlab toolbox, "*http://graphics.cs.msu.ru/en/science/research/machinelearning/adaboosttoolbox*," 2008.