

DISCRIMINANT NMFFACES FOR FRONTAL FACE VERIFICATION

Stefanos Zafeiriou, Anastasios Tefas and Ioannis Pitas

Dept. of Informatics, Aristotle University of Thessaloniki, Box 451, 54124 Thessaloniki, Greece
e-mail: {dralbert, tefas, pitas}@zeus.csd.auth.gr

ABSTRACT

In this paper, a novel supervised feature extraction method is presented. The method employs discriminant analysis in the features derived by *Non-negative Matrix Factorization* (NMF). In this way, a two phase discriminant feature extraction procedure is implemented, namely NMF plus *Linear Discriminant Analysis* (LDA). The introduced method has been applied to the problem of frontal face verification using the well known XM2VTS database, where a better performance than NMF, Eigenfaces and Fisherfaces has been achieved.

1. INTRODUCTION

Face recognition/verification has attracted the attention of researchers for more than two decades and is among the most popular research areas in the field of computer vision and pattern recognition.

The most popular among the techniques used for frontal face recognition/verification are the subspace methods. The subspace algorithms consider the entire image as a feature vector and their aim is to find projections (bases) that optimize some criterion defined over the feature vectors that correspond to different classes. Then the original high dimensional image space is projected into a low dimensional one. The classification is usually performed according to a simple distance measure in the final multidimensional space.

Various criteria have been employed in order to find the basis of the low dimensional spaces. Some of them have been defined in order to find projections that they best express the population (e.g. *Principal Component Analysis* (PCA) [1], NMF [2], *Local Non-negative Matrix Factorization* (LNMF) [3]) without using the information of how the data are separated to different classes. Another class of criteria is the one that deals directly with discrimination between classes (e.g. LDA [4]).

A subspace method that aims at finding a face representation by using basis images without using class infor-

mation is NMF [2]. The NMF algorithm, like PCA, represents a face as a linear combination of basis. The difference with PCA is that it does not allow negative elements in both the basis vectors and the weights of the linear combination. This constraint results to radically different basis than PCA. On the one hand the basis of PCA are eigenfaces, some of which resemble distorted versions of the entire face. On the other hand the basis of NMF are localized features that correspond better to the intuitive notions of face parts [2]. An extension of NMF that gives even more localized basis by imposing additional locality constraints is the so-called LNMF [3].

In this paper, we develop a technique for exploiting discriminant information in NMF. This technique uses the NMF basis images in order to discover a low dimensional space and search for discriminant projections in this space. This is similar to Fisherfaces [4, 5], where an initial PCA based dimensionality reduction step is used, before applying LDA in this new space for finding discriminant projections. Of course the motivations of Fisherfaces and the proposed NMF plus LDA are different. In Fisherfaces, first PCA is used in order to satisfy the invertibility of the within scatter matrix and afterwards LDA is used in this new space. In the proposed NMF plus LDA method LDA is used along with NMF in order to investigate whether there is any discriminant information in part-based decompositions, like NMF. The proposed method has been tested for frontal face verification using the XM2VTS database.

2. FRONTAL FACE VERIFICATION AND SUBSPACE TECHNIQUES

In this Section, we will briefly outline the problem of frontal face verification and the framework under which a subspace method can be used in order to solve this problem.

Let \mathcal{U} be a facial image database. Each facial image $\mathbf{x} \in \mathcal{U}$ is supposed to belong to one of the K facial (person) classes $\{\mathcal{U}_1, \mathcal{U}_2, \dots, \mathcal{U}_K\}$ with $\mathcal{U} = \bigcup_{i=1}^K \mathcal{U}_i$. For a face verification system that uses the database \mathcal{U} , a genuine (or client) claim is performed when a person t provides its facial image \mathbf{x} , claims that $\mathbf{x} \in \mathcal{U}_r$ and $t = r$. When a person t provides its facial image \mathbf{x} and claims that $\mathbf{x} \in \mathcal{U}_r$, with $t \neq$

This work is funded by the integrated project BioSec IST-2002-001766 (Biometric Security, <http://www.biosec.org>), under Information Society Technologies (IST) priority of the 6th Framework Programme of the European Community

r , an impostor claim occurs. The scope of a face verification system is to handle properly these claims by accepting the genuine claims and rejecting the impostor ones.

Let the facial image database \mathcal{U} be comprised by L facial images $\mathbf{x}_j \in \mathfrak{R}_+^F$, where $\mathfrak{R}_+ = [0, +\infty)$ and let the cardinality of each facial class \mathcal{U}_r to be N_r . A linear subspace transformation of the original F -dimensional space onto a M -dimensional subspace (usually $M \ll F$) is a matrix $\mathbf{W} \in \mathfrak{R}^{M \times F}$ estimated using the database \mathcal{U} . The new feature vector $\hat{\mathbf{x}} \in \mathfrak{R}^M$ is given by:

$$\hat{\mathbf{x}} = \mathbf{W}\mathbf{x}. \quad (1)$$

The rows of the matrix \mathbf{W} contain the basis of the lower dimension feature space.

After the projection given by (1), a distance metric is chosen in order to measure the similarity of a test facial image to a certain class. This similarity measure can be the L_1 norm, the L_2 norm, the normalized correlation or the Mahalanobis distance [6]. In case of face verification, the algorithm should also learn a threshold on the similarity measure in order to accept or reject a client/impostor claim.

3. THE NMF ALGORITHM

In order to apply NMF, the matrix $\mathbf{X} \in \mathfrak{R}_+^{F \times L} = [x_{i,j}]$ should be constructed, where $x_{i,j}$ is the i -th element of the j -th image. In other words the j -th column of \mathbf{X} is the \mathbf{u}_j facial image. NMF aims to find two matrices $\mathbf{Z} \in \mathfrak{R}_+^{F \times M} = [z_{i,k}]$ and $\mathbf{H} \in \mathfrak{R}_+^{M \times L} = [h_{k,j}]$ such that,

$$\mathbf{X} \approx \mathbf{Z}\mathbf{H}. \quad (2)$$

The facial image \mathbf{u}_j after the NMF decomposition can be written as $\mathbf{u}_j \approx \mathbf{Z}\mathbf{h}_j$, where \mathbf{h}_j is the j -th column of \mathbf{H} . Thus, the lines of the matrix \mathbf{Z} can be considered as basis images and the vector \mathbf{h}_j as the corresponding weight vector. The \mathbf{h}_j vectors can also be considered as the projected vectors of a lower dimensional feature space.

The NMF imposes non-negative constraints in both the elements of $z_{i,k}$ and of $h_{k,j}$. Thus, only non-subtractive combinations are allowed. This is believed to correspond better to the intuitive notion of combining facial parts in order to create a complete face.

One of the algorithms initially proposed for finding the matrices \mathbf{Z} and \mathbf{H} used the following metric:

$$D_N(\mathbf{X}||\mathbf{Z}\mathbf{H}) = \sum_{i,j} (x_{i,j} \ln(\frac{x_{i,j}}{\sum_l z_{i,l}h_{l,j}}) + \sum_k z_{i,k}h_{k,j} - x_{i,j}) \quad (3)$$

as the measure of the cost for factoring \mathbf{X} into $\mathbf{Z}\mathbf{H}$ [7].

The NMF factorization is the outcome of optimization :

$$\begin{aligned} \min_{\mathbf{Z}, \mathbf{H}} D_N(\mathbf{X}||\mathbf{Z}\mathbf{H}) \text{ subject to} \quad (4) \\ z_{i,k} \geq 0, h_{k,j} \geq 0, \sum_i z_{i,j} = 1, \forall j \end{aligned}$$

By using an auxiliary function and the Expectation Maximization algorithm [7], the following update rules for $h_{k,j}$ and $z_{i,k}$ guarantee a non increasing behavior of (3). The update rule for the t -th iteration for $h_{k,j}$ is given by:

$$h_{k,j}^{(t)} = h_{k,j}^{(t-1)} \frac{\sum_i z_{i,k}^{(t-1)} \frac{x_{i,j}}{\sum_l z_{i,l}^{(t-1)} h_{l,j}^{(t-1)}}}{\sum_i z_{i,k}^{(t-1)}} \quad (5)$$

whereas for the $z_{i,k}$ the update rule is given by:

$$z_{i,k}^{(t)} = z_{i,k}^{(t-1)} \frac{\sum_j h_{k,j}^{(t)} \frac{x_{i,j}}{\sum_l z_{i,l}^{(t-1)} h_{l,j}^{(t)}}}{\sum_j h_{k,j}^{(t)}}. \quad (6)$$

Since $\mathbf{x}_j \approx \mathbf{Z}\mathbf{h}_j$, a natural way to compute the projection of \mathbf{x}_j to a lower dimensional feature space using NMF is $\hat{\mathbf{x}}_j = \mathbf{Z}^\dagger \mathbf{x}_j$. The pseudo-inverse \mathbf{Z}^\dagger can be calculated using singular value decomposition methods [8, 9]. In order to proceed to the dimensionality reduction, it has been also claimed that \mathbf{Z}^T can be used as an alternative [10], due to the fact that the calculation of \mathbf{Z}^\dagger may suffer from numerical instability.

In any case, we can not use the coefficient matrix of \mathbf{H} computed directly from (5) (which gives us its values in the training phase), since we do not have any expression for calculating this representation for the test images.

4. NMF PLUS LDA

The previously presented method, NMF, does not use the information about how the various facial images are separated into different facial classes. The most straightforward way in order to exploit discriminant information in NMF is to try to discover discriminant projections for the facial image vectors after the projection to the basis image matrix \mathbf{Z}^\dagger . Let the matrix \mathbf{X} that contains all the facial images of the database \mathcal{U} , be organized as follows. The j -th column of the database \mathbf{X} is the ρ -th image of the r -th class. Thus, $j = \sum_{i=1}^{r-1} N_i + \rho$.

The vector \mathbf{h}_j that correspond to the j th column of the matrix \mathbf{H} , is the coefficient vector for the ρ th facial image of the r th class and will be denoted as $\boldsymbol{\eta}_\rho^{(r)} = [\eta_{\rho,1}^{(r)} \dots \eta_{\rho,M}^{(r)}]^T$. The mean vector of the vectors $\boldsymbol{\eta}_\rho^{(r)}$ for the class r is denoted as $\boldsymbol{\mu}^{(r)} = [\mu_1^{(r)} \dots \mu_M^{(r)}]^T$ and the mean of all classes as $\boldsymbol{\mu} = [\mu_1 \dots \mu_M]^T$. Then, the within scatter for the coefficient vectors \mathbf{h}_j is defined as:

$$\mathbf{S}_w = \sum_{r=1}^K \sum_{\rho=1}^{N_r} (\boldsymbol{\eta}_\rho^{(r)} - \boldsymbol{\mu}^{(r)})(\boldsymbol{\eta}_\rho^{(r)} - \boldsymbol{\mu}^{(r)})^T \quad (7)$$

whereas the between scatter matrix is defined as:

$$\mathbf{S}_b = \sum_{r=1}^K N_r (\boldsymbol{\mu}^{(r)} - \boldsymbol{\mu})(\boldsymbol{\mu}^{(r)} - \boldsymbol{\mu})^T. \quad (8)$$

The matrix \mathbf{S}_w defines the scatter of sample vector coefficients around their class mean. The dispersion of samples that belong to the same class around their corresponding mean should be as small as possible. A convenient metric for the dispersion of the samples is the trace of \mathbf{S}_w . The matrix, \mathbf{S}_b denotes the between-class scatter matrix and defines the scatter of the mean vectors of all classes around the global mean $\boldsymbol{\mu}$. Each class formed by the samples that belong to the same class must be as far as possible from the other classes. Therefore, the trace of \mathbf{S}_b should be as large as possible. By taking into consideration the previous remarks, the well known Fisher discriminant criterion is constructed as:

$$J(\boldsymbol{\Psi}) = \frac{\text{tr}[\boldsymbol{\Psi}^T \mathbf{S}_b \boldsymbol{\Psi}]}{\text{tr}[\boldsymbol{\Psi}^T \mathbf{S}_w \boldsymbol{\Psi}]} \quad (9)$$

where $\text{tr}[\mathbf{R}]$ is the trace of the matrix \mathbf{R} . The maximization of J yields a set of discriminant projections that is given by the columns of the matrix $\boldsymbol{\Psi} = [\psi_1 \dots \psi_{K-1}]$. If \mathbf{S}_w is invertible then the projection matrix $\boldsymbol{\Psi}$ is given by the generalized eigenvectors of $\mathbf{S}_w^{-1} \mathbf{S}_b$.

There is not upper limit for how many basis someone can construct using NMF decomposition in (6) and unless we create a limited number of basis by NMF the matrix \mathbf{S}_w is singular. That is, there always exist vectors ϕ_i that satisfy $\phi_i^T \mathbf{S}_w \phi_i = 0$. These vectors turn out to be very effective if they satisfy $\phi_i^T \mathbf{S}_b \phi_i > 0$ at the same time [11, 12]-[14]. In that case the Fisher discriminant criterion degenerates into the following between-class scatter criterion:

$$J_b(\boldsymbol{\Phi}) = \text{tr}[\boldsymbol{\Phi}^T \mathbf{S}_b \boldsymbol{\Phi}] \quad (\boldsymbol{\Phi} = [\dots \phi_i \dots], \|\phi_i\| = 1). \quad (10)$$

We will use the main results of [14] in order to extract discriminant features using an arbitrary number of NMF basis. The discriminant features are then extracted by the minimization of the criterions (9) and (10). The discriminant projections that are derived by the (9) will be called *regular discriminant projections* (or *regular NMFfaces*) while the ones created by (10) will be called *irregular discriminant projections* (or *irregular NMFfaces*).

Let the total scatter matrix of the feature vectors \mathbf{h}_j be defined as:

$$\mathbf{S}_t = \mathbf{S}_w + \mathbf{S}_b \quad (11)$$

it is easy to prove that the matrix \mathbf{S}_t is a compact and self-adjoint operator in \mathbb{R}^M [14]. Thus, its eigenvector system forms an orthonormal basis for \mathbb{R}^M [14].

Let \mathcal{O} and \mathcal{O}^\perp be the two complementary spaces spanned by the orthonormal eigenvectors that correspond to no-zero and to zero eigenvalues of \mathbf{S}_t , respectively. It is easy to prove, using the theory developed in [14], that \mathcal{O}^\perp does not contain any discriminant information in respect to the criterion (9) and (10). The isomorphic mapping in order to move from the feature space of the vectors \mathbf{h}_j to \mathcal{O} is the

matrix $\boldsymbol{\Pi}$ whose columns are the orthonormal eigenvectors of \mathbf{S}_t that correspond to its non-zero eigenvalues. In order to find the non-zero eigenvectors of \mathbf{S}_t efficiently, we can use algorithms like [15].

Let $\hat{\mathbf{S}}_w$ and $\hat{\mathbf{S}}_b$ be the within scatter and the between scatter matrices in the space \mathcal{O} . These matrices are given by $\hat{\mathbf{S}}_w = \boldsymbol{\Pi}^T \mathbf{S}_w \boldsymbol{\Pi}$ and by $\hat{\mathbf{S}}_b = \boldsymbol{\Pi}^T \mathbf{S}_b \boldsymbol{\Pi}$. In the space \mathcal{O} the matrix $\hat{\mathbf{S}}_w$ is still singular. Let $\boldsymbol{\Xi}_1$ and $\boldsymbol{\Xi}_2$ be the orthonormal eigenvectors that correspond to non-zero and to zero eigenvalues of the matrix $\hat{\mathbf{S}}_w$, respectively.

In the space spanned by the vectors contained in $\boldsymbol{\Xi}_1$ the discriminant projections are given by the columns of the matrix $\boldsymbol{\Theta}_1$ that are the eigenvectors of $\hat{\mathbf{S}}_w^{-1} \hat{\mathbf{S}}_b$, where $\hat{\mathbf{S}}_w = \boldsymbol{\Xi}_1^T \hat{\mathbf{S}}_w \boldsymbol{\Xi}_1$ and $\hat{\mathbf{S}}_b = \boldsymbol{\Xi}_1^T \hat{\mathbf{S}}_b \boldsymbol{\Xi}_1$. In the space that is spanned by the columns of $\boldsymbol{\Xi}_2$ it can be easily proven that $\hat{\mathbf{S}}_b = \boldsymbol{\Xi}_2^T \hat{\mathbf{S}}_b \boldsymbol{\Xi}_2$ is not singular [14]. Thus, the discriminant projections in this space are given by the matrix $\boldsymbol{\Theta}_2$ that has as columns the orthonormal eigenvectors of $\hat{\mathbf{S}}_b$.

The linear transform that extracts the regular discriminant features using NMF is:

$$\hat{\boldsymbol{\Phi}}_1 = \boldsymbol{\Theta}_1^T \boldsymbol{\Xi}_1^T \boldsymbol{\Pi}^T \mathbf{Z}^\dagger, \quad (12)$$

whereas, the linear transform that extracts the irregular discriminant features using NMF is:

$$\hat{\boldsymbol{\Phi}}_2 = \boldsymbol{\Theta}_2^T \boldsymbol{\Xi}_2^T \boldsymbol{\Pi}^T \mathbf{Z}^\dagger \quad (13)$$

where \mathbf{Z} is the decomposition of NMF given by (6). The total number of discriminant projections derived by this procedure is $2(K - 1)$.

5. EXPERIMENTAL RESULTS

The experiments were conducted in the XM2VTS database using the protocol described in [16]. The images were aligned semi-automatically according to the eyes position of each facial image using the eye coordinates. The facial images were down-scaled to 64×64 resolution. Histogram equalization was used for normalizing the facial images.

5.1. Training Procedure

The XM2VTS database provides two experiment setups namely, Configuration I and Configuration II [16]. Each configuration is divided in three different sets the training set, the evaluation set and the test set. The training set is used to create client and impostor models for each person. The evaluation is used to learn the thresholds.

The training set of the Configuration I contains 200 persons with 3 images per person. The evaluation set contains 3 images per client for genuine claims and 25 evaluation impostors with 8 images per impostor. Thus, evaluation set gives a total of $3 \times 200 = 600$ client claims

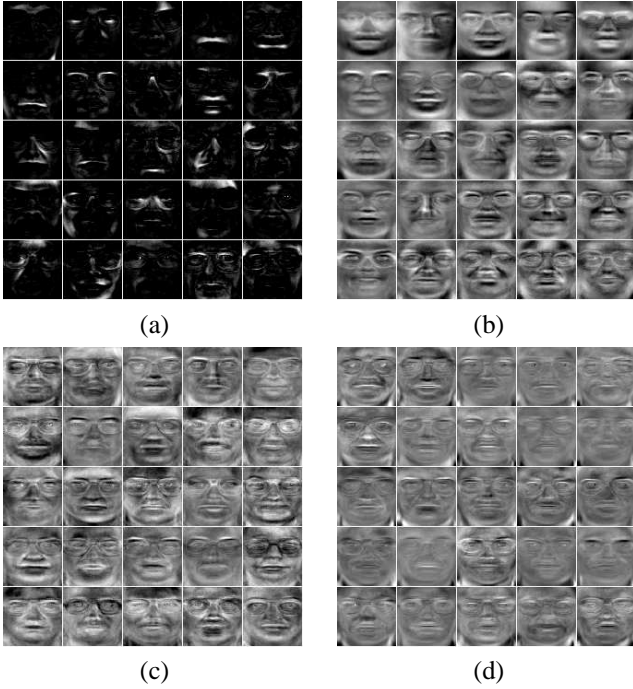


Fig. 1. A set of 25 basis images for (a) NMF (b) EigenFaces, (c) FisherFaces (d) the proposed NMF-Faces (regular).

and $25 \times 8 \times 200 = 40,000$ impostor claims. The test set has 2 images per client and 70 impostors with 8 images per impostor and gives $2 \times 200 = 400$ client claims and $70 \times 8 \times 200 = 112,000$ impostor claims. The maximum number of Eigenfaces [15] given by the training set is 599. The number of classes is 200 and, thus, the number of Fisherfaces [4] is 199. For NMF plus LDA, 1000 basis images have been created initially using NMF and after the regular and irregular discriminant information has been found according to (12) and (13) that gives a total of 398 projections (199 regular NMFfaces and 199 irregular NMFfaces).

By a visual inspection of the images of Figure 1, it can be seen that Eigenfaces, Fisherfaces and regular NMFfaces (it also holds for the irregular) resemble degraded versions of faces.

5.2. Experimental Results in Configuration I

The facial images have been then projected using these basis into a low dimensional feature space and the normalized correlation was used in order to define the similarity measure between two faces as:

$$D(\mathbf{x}_r, \mathbf{x}_t) = \frac{\hat{\mathbf{x}}_r^T \hat{\mathbf{x}}_t}{\|\hat{\mathbf{x}}_r\| \|\hat{\mathbf{x}}_t\|} \quad (14)$$

where \mathbf{x}_r and \mathbf{x}_t are the reference and the test facial image, respectively while $\hat{\mathbf{x}}_r$ and $\hat{\mathbf{x}}_t$ are their projections to one of the subspace.

In case of NMF plus LDA two different discriminant projection are found by (12) and (13). Thus, two different similarity values are created by $D_g(\mathbf{x}_r, \mathbf{x}_t) = \frac{(\Psi_1 \mathbf{x}_r)(\Psi_1 \mathbf{x}_t)^T}{\|\Psi_1 \mathbf{x}_r\| \|\Psi_1 \mathbf{x}_t\|}$ and by $D_u(\mathbf{x}_r, \mathbf{x}_t) = \frac{(\Psi_2 \mathbf{x}_r)(\Psi_2 \mathbf{x}_t)^T}{\|\Psi_2 \mathbf{x}_r\| \|\Psi_2 \mathbf{x}_t\|}$ for the regular and the irregular discriminant information, respectively. In [14] it has been proposed to use a simple fusion technique by weighting the irregular score with some empirical coefficient. Instead of using the empirical parameter we used the evaluation set of the Configuration I in order to learn a discriminant weighting vector \mathbf{w} using also LDA. The final similarity measure between the facial image vectors \mathbf{x}_r and \mathbf{x}_t is given by:

$$D_t(\mathbf{x}_r, \mathbf{x}_t) = \mathbf{w}^T [D_g(\mathbf{x}_r, \mathbf{x}_t) \ D_u(\mathbf{x}_r, \mathbf{x}_t)]^T. \quad (15)$$

The similarity measures for each person, calculated in both evaluation and training set form the distance vector $\mathbf{d}(r)$. The elements of the vector $\mathbf{d}(r)$ are sorted in descending order and are used for the person specific thresholds on the distance measure. Let $T_Q(r)$ denote the Q -th order statistic of the vector of distances, $\mathbf{d}(r)$. The threshold of the person r is chosen to be equal to $T_Q(r)$. Let \mathbf{x}_r^1 , \mathbf{x}_r^2 and \mathbf{x}_r^3 be the 3 instances of the person r in the training set. A claim of a person (with a facial image \mathbf{x}_t) to the identity r is considered valid if $\max_j \{D(\mathbf{x}_r^j, \mathbf{x}_t)\} < T_Q(r)$. Obviously when varying Q , different pairs of *False Acceptance Rate* (FAR) and *False Rejection Rate* (FRR) can be created and that way a ROC curve is produced and the *Equal Error Rate* (EER) can be measured [16, 17].

The performance of the methods that project to face basis like Eigenfaces, Fisherfaces and NMFfaces (regular and irregular) for various feature dimensions is illustrated in Figure 2. The best EER achieved was 0.8% when 80 regular and 80 irregular projections have been kept. The best EER for Fisherfaces has been 1.6% and for Eigenfaces 4.3%. The best EER achieved for NMF was more than 8%, thus, a separate curve showing the performance of NMF for various feature dimensions was not included in Figure 2.

6. CONCLUSIONS

We proposed a supervised feature extraction technique in order to improve the classification performance of NMF. The discriminant technique gives basis images that are holistic and is comprised of two different phases, namely NMF and LDA thus producing the so-called NMFfaces. The new subspace technique has been applied to frontal face verification, where, it was verified that the proposed NMFfaces outperform the well-known Fisherfaces and Eigenfaces.

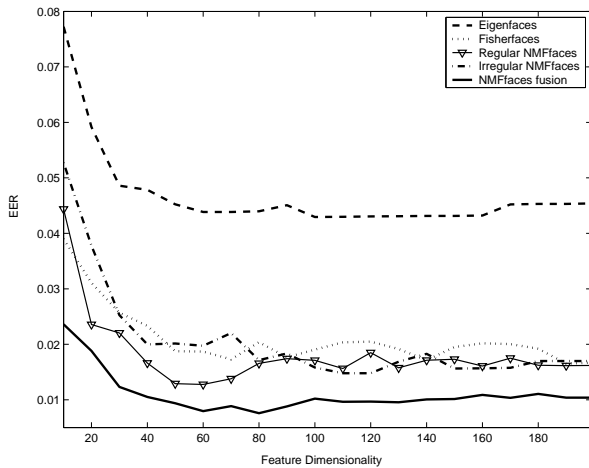


Fig. 2. EER for Configuration I plotted versus feature dimensionality for Eigenfaces, FisherFaces and the proposed NMFfaces (regular, irregular and fusion).

7. REFERENCES

- [1] M. Kirby and L. Sirovich, "Application of the karhunen-loeve procedure for the characterization of human faces.," *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp. 103–108, Jan. 1990.
- [2] D.D. Lee and H.S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.
- [3] S.Z. Li, X.W. Hou, and H.J. Zhang, "Learning spatially localized, parts-based representation," in *CVPR*, Kauai, HI, USA, December 8–14 2001, pp. 207–212.
- [4] 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, no. 7, pp. 711–720, July 1997.
- [5] D. L. Swets and J. Weng, "Using discriminant eigenfeatures for image retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 831–836, 1996.
- [6] L. Chengjun, "Gabor-based kernel PCA with fractional power polynomial models for face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 5, pp. 572 – 581, May 2004.
- [7] D.D. Lee and H.S. Seung, "Algorithms for non-negative matrix factorization," in *NIPS*, 2000, pp. 556–562.
- [8] G.H. Golub and C.F. VanLoan, *Matrix Computations*, third ed. John Hopkins Univ. Press, 1996.
- [9] J. Ye, R. Janardan, C. H. Park, and H. Park, "An optimization criterion for generalized discriminant analysis on undersampled problems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 8, pp. 982–994, 2004.
- [10] I. Buciu and I. Pitas, "A new sparse image representation algorithm applied to facial expression recognition," in *MLSP*, Sao Lus, Brazil, Sep. 29 - Oct. 1st 2004.
- [11] L. Juwei, K.N. Plataniotis, and A.N. Venetsanopoulos, "Face recognition using kernel direct discriminant analysis algorithms," *IEEE Transactions on Neural Networks*, vol. 14, no. 1, pp. 117–126, 2003.
- [12] L. Juwei, K.N. Plataniotis, and A.N. Venetsanopoulos, "Face recognition using lda-based algorithms," *IEEE Transactions on Neural Networks*, vol. 14, no. 1, pp. 195–200, 2003.
- [13] H. Cevikalp, M. Neamtu, M. Wilkes, and A. Barkana, "Discriminative common vectors for face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 1, pp. 4–13, 2005.
- [14] J. Yang, A.F. Frangi, J. Yang, D. Zhang, and Z. Jin, "KPCA plus LDA: A complete kernel fisher discriminant framework for feature extraction and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 2, pp. 230–244, 2005.
- [15] M. Turk and A. P. Pentland, "Eigenfaces for recognition.," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
- [16] K. Messer, J. Matas, J.V. Kittler, J. Luetin, and G. Maitre, "Xm2vtsdb: The extended m2vts database," in *AVBPA '99*, 1999, pp. 72–77.
- [17] A. Tefas, C. Kotropoulos, and I. Pitas, "Using support vector machines to enhance the performance of elastic graph matching for frontal face authentication," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 7, pp. 735–746, 2001.