FISHER NON-NEGATIVE MATRIX FACTORIZATION FOR LEARNING LOCAL FEATURES

Yuan Wang *Yunde Jia*

Computer Science Department
Beijing Institute of Technology
Beijing 100081, P.R.CHINA
yjiar@bit.edu.cn

Changbo Hu *Matthew Turk*

Computer Science Department
University of California
Santa Barbara, CA 93106, USA
{cbhu, mturk}@cs.ucsb.edu

ABSTRACT

In this paper, we propose a novel subspace method called Fisher non-negative matrix factorization (FNMF) for face recognition. FNMF is based on non-negative matrix factorization (NMF), which is a part-based image representation method proposed by Lee. NMF allows only additive combinations of non-negative basis components. The NMF bases are spatially global, whereas local bases would be preferred. Therefore, Stan et al proposed local non-negative matrix factorization (LNMF) to achieve a localized NMF representation by adding more constraints to enforce spatial locality; one of these constraints enforces the bases to be orthogonal to each other, just like the constraints which PCA imposes on its bases. However, LNMF does not encode discrimination information for a classification problem. In this paper, we impose Fisher constraints on the NMF algorithm, which results in the novel FNMF algorithm. Our experiments show that FNMF achieves better performance than LNMF.

1 INTRODUCTION

Face recognition is very challenging because of illumination, facial expression and pose variations. It has received extensive attention last two decades, not only because it has several potential applications in areas such as human computer interaction (HCI), biometrics and security, but also it is a typical pattern recognition problem whose solution would help in many other classification problems.

Subspace methods have demonstrated their success in numerous visual recognition tasks such as face recognition, face detection and tracking. These methods, such as Principle Component Analysis (PCA) [1,2], Fisher Linear Discriminant Analysis (FLDA) [3], Independent Component Analysis (ICA) [2,4] and non-negative matrix factorization (NMF) [6,7], learn to represent a face as a linear combination of basis images, but in different ways. The basis images of PCA are orthogonal and have a statistical interpretation as the directions of largest variance. FLDA seeks to find a linear transformation that can maximize the between-class scatter and minimize the within-class scatter. ICA is a linear non-orthogonal transform, which yields a representation in which unknown linear mixtures of multi-dimensional random variables are made as statistically independent as possible. NMF factorizes the image database into two matrix factors whose entries are all non-negative and produces a part-based representation of images because it allows only additive, not subtractive, combinations of basis components. For these reasons, the non-negativity constraints are compatible with the intuitive notion of combining parts to form a whole. Because a part-based representation can naturally deal with partial occlusion and some illumination problems, it has received much attention recently.

Local non-negative matrix factorization (LNMF) [5] has been proposed to achieve a more localized NMF algorithm with the aim of computing spatially localized bases from a face database by adding three constraints that modify the objective function in the NMF algorithm. One of these is an orthogonality constraint that is essentially the same as the constraint which PCA imposes on its bases.

This paper proposes a new variation to NMF. We take NMF as a framework for face recognition, and then add characteristics of FLDA. Specifically, we propose a novel subspace method called Fisher non-negative factorization (FNMF) and our algorithm can produce both additive and spatially localized basis images as LNMF. It also encodes discrimination information for face recognition.

2 RELATED WORK

There is psychological and physiological evidence for part-based representations in the brain. Lee and Seung proposed NMF [6,7] for learning parts of faces, and the non-negative constraint added to the matrix factorization is compatible with the intuitive notion of combining parts to form a whole face. However, the NMF algorithm can only get global, not spatially localized, parts from the training set. To improve the NMF algorithm, Local MNF (LNMF) [5] was proposed for learning spatially localized, part-based representations of visual patterns. The remainder of this section will introduce NMF and LNMF.

A database of *m* face images, each of which contains *n* non-negative pixel values, is represented by an $n \times m$ matrix V, where each column denotes one of the *m* facial images. Basis images computed from the database are denoted by an $n \times r$ matrix W, where *r* is the number of basis images. To reduce the dimensionality of V, r should be less than m. Hence the factorization is

$$V \approx WH \tag{1}$$

where H consists of the coefficients by which a face is represented with a linear combination of basis images.

Many matrix factorizations allow the entries of W and H to be of arbitrary sign. Therefore, the basis images of this kind do not have an obvious visual interpretation because there are complex cancellations between positive and negative numbers when the basis images are used in linear combinations. However, a matrix factorization with a non-negative constraint can produce basis images that have an intuitive meaning, since the entries of W and H are all non-negative.

2.1 NMF

NMF enforces the non-negative constraints on W and H. Thus the basis images can be combined to form a whole face in an intuitive, additive fashion. NMF uses the divergence of V from its approximation $Y = WH$ as the measure of cost for factorizing V into WH. The divergence function, used as an objective function in NMF, is defined as

$$D(V\|Y) = \sum_{i,j}\left(v_{ij}\log\frac{v_{ij}}{y_{ij}} - v_{ij} + y_{ij}\right) \tag{2}$$

NMF factorization is a solution to the following optimization problem:

$$\min_{B,H} \quad D(V\|WH)$$

$$\text{s.t.} \quad W, H \geq 0, \sum_i b_{ij} = 1 \quad \forall j$$

where $W, H \geq 0$ indicates that all elements of W and H are to be non-negative; b_j are the basis images. This optimization can be done by using the following multiplicative update rules:

$$W_{ia} \leftarrow W_{ia}\sum_{\mu}\frac{V_{i\mu}}{(WH)_{i\mu}}H_{a\mu} \tag{3a}$$

$$W_{ia} \leftarrow \frac{W_{ia}}{\sum_j W_{ja}} \tag{3b}$$

$$H_{a\mu} \leftarrow H_{au}\sum_i W_{ia}\frac{V_{i\mu}}{W_{i\mu}} \tag{3c}$$

2.2 LNMF

Local NMF [5] aims to improve the locality of the learned features by imposing additional constraints. It incorporates the following three constraints into the original NMF formulation.

- LNMF attempts to minimize the number of basis components required to represent V. This implies that a basis component should not be further decomposed into more components.
- To minimize redundancy between different bases, LNMF attempts to make different bases as orthogonal as possible.
- Only bases containing the most important information should be retained. LNMF attempts to maximize the total "activity" on each component, i.e., the total squared projection coefficients summed over all training images.

LNMF incorporates the above constraints into the original NMF formulation and defines the following constrained divergence as the objective function:

$$D(V\|WH) = \sum_{i,j}\left(v_{ij}\log\frac{v_{ij}}{y_{ij}} - v_{ij} + y_{ij}\right) + \alpha\sum_{ij}u_{ij} - \beta\sum_i q_{ii} \tag{4}$$

where $\alpha, \beta > 0$ are constants, $(W^TW) = U = [u_{ij}]$, and $(HH^T = Q = [q_{ij}])$. This optimization can be done by using a multiplicative update rules which was presented by Common [4].

3 FISHER NMF

3.1 Algorithm

Of the three constraints that LNMF imposes on NMF, one is similar to PCA in that it constrains the bases to be orthogonal to each other. To achieve good recognition results and also get intuitive bases, we propose a novel subspace method using Fisher linear discriminant analysis (FLDA), called Fisher non-negative matrix factorization (FNMF).

FLDA has been successfully applied to the problem of face recognition. The main idea of FNMF is to add the Fisher constraint to the original NMF formulation. Because the columns of the encoding matrix *H* have a one-to-one correspondence with the columns of the original matrix *V*, we seek to maximize the between-class scatter and minimize the within-class scatter of *H*.

We define the following constrained divergence as the new objective function for FNMF:

$$D(V\|WH) = \sum_{i,j}\left(v_{ij}\log\frac{v_{ij}}{y_{ij}} - v_{ij} + y_{ij}\right) + \alpha S_W - \alpha S_B \quad (5)$$

where $\alpha > 0$ is a constant, S_W is the within-class scatter of the encoding matrix *H*, and S_B is the between-class scatter of *H*. Let n_i denote the number of vectors in the *ith* class and C the number of class. We define S_W and S_B as follows:

$$S_W = \frac{1}{C}\sum_{i=1}^{C}\frac{1}{n_i}\sum_{j=1}^{n_i}(h_j - u_i)(h_j - u_i)^T \quad (6)$$

$$S_B = \frac{1}{C(C-1)}\sum_{i=1}^{C}\sum_{j=1}^{C}(u_i - u_j)(u_i - u_j)^T \quad (7)$$

where $u_i = \frac{1}{n_i}\sum_{j=1}^{n_i} h_j$ denotes the mean value of class *i* in *H*.

The following update rules implement a local solution to the above constrained minimization.

$$h_{kl} \leftarrow -b + \sqrt{b^2 + 4\left(\sum_i v_{il}\frac{w_{ik}h'_{kl}}{\sum_k w_{ik}h'_{kl}}\right)\left(\frac{2}{n_i C} - \frac{4}{n_i^2(C-1)}\right)} \quad (8)$$

$$w_{kl} \leftarrow \frac{w_{kl}\sum_j v_{kj}\frac{h_{lj}}{\sum_k w_{kl}h_{lj}}}{\sum_j h_{lj}} \quad (9)$$

$$w_{kl} \leftarrow \frac{w_{kl}}{\sum_k w_{kl}} \quad (10)$$

where

$$b = \frac{4}{n_i C(C-1)}\sum_j (u_{kj} - (u_{ki} - \frac{h'_{kl}}{n_i})) - \frac{2}{n_i C}u_{ki} + 1.$$

3.2 Convergence proof

Our update rules are based on a technique which minimizes an objective function $L(X)$ by using an auxiliary function. $G(X, X')$ is defined as an auxiliary function for $L(X)$ if $G(X, X') \geq L(X)$ and $G(X, X) = L(X)$ are satisfied. If G is an auxiliary function, then $L(X)$ is non-increasing when X is updated by

$$X^{(t+1)} = \arg\min_X G(X, X^{(t)}) \quad (11)$$

because

$$L(X^{(t+1)}) \leq G(X^{(t+1)}, X^{(t)}) \leq G(X^{(t)}, X^{(t)}) = L(X^{(t)}).$$

H is updated by minimizing $L(H) = D(V\|WH)$ with W fixed.

We construct an auxiliary function for $L(H)$ as such

$$G(H, H') = \sum_{i,j} v_{ij}\log v_{ij} - \sum_{i,j,k} x_{ij}\frac{w_{ik}h'_{kj}}{\sum_k w_{ik}h'_{kj}}\left(\log(w_{ik}h_{kj}) - \log\frac{w_{ik}h'_{kj}}{\sum_k w_{ik}h'_{kj}}\right) + \sum_{i,j} y_{ij} - \sum_{i,j} v_{ij} + \alpha S_W - \alpha S_B \quad (12)$$

$G(H, H') = L(H)$ is easily verified, so we will just prove $G(H, H') \geq L(H)$ as follows. Because $\log(\sum_k w_{ik}h_{kj})$ is a convex function, the following holds for all i, j and $\sum_k \sigma_{ijk} = 1$:

$$-\log\left(\sum_k w_{ik}h_{kj}\right) \le -\sum_k \sigma_{ijk} \log \frac{w_{ik}h_{kj}}{\gamma\sigma_{ijk}},$$

where $\sigma_{ijk} = \dfrac{w_{ik}h'_{kj}}{\sum_k w_{ik}h'_{kj}}$.

So

$$-\log\left(\sum_k w_{ik}h_{kj}\right) \le$$

$$-\sum_k \frac{w_{ik}h'_{kj}}{\sum_k w_{ik}h'_{kj}}\left(\log w_{ik}h_{kj} - \log \frac{w_{ik}h'_{kj}}{\sum_k w_{ik}h'_{kj}}\right)$$

which is $G(H,H') \ge L(H)$.

In order to minimize $L(H)$ w.r.t. H, we can update H using

$$H^{(t+1)} = \arg\min_H G(H, H^{(t)})$$

H can be found by letting $\dfrac{\partial G(H,H')}{\partial h_{kl}} = 0$ for all $k,l,$ since

$$\frac{\partial G(H,H')}{\partial h_{kl}} = -\sum_i v_{il} \frac{w_{ik}h'_{kl}}{\sum_k w_{ik}h'_{kl}} \frac{1}{h_{kl}} +$$

$$\sum_i w_{ik} + \alpha \frac{2}{n_i C}(h_{kl} - u_{ki}) - \tag{13}$$

$$\alpha \frac{4}{n_i C(C-1)} \sum_j (u_{kj} - (\frac{h_{kl}}{n_i} + u_{ki} - \frac{h'_{kl}}{n_i}))$$

In fact, α is just a constant, so we can define $\alpha = 1$. n_i corresponds to the number of face vectors in the class to which h_{kl} belongs, and C is the number of face classes. We find that

$$h_{kl} = \frac{-b + \sqrt{b^2 + 4\left(\sum_i v_{il} \frac{w_{ik}h'_{kl}}{\sum_k w_{ik}h'_{kl}}\right)\left(\frac{2}{n_i C} - \frac{4}{n_i^2(C-1)}\right)}}{2\left(\frac{2}{n_i C} - \frac{4}{n_i^2(C-1)}\right)} \tag{14}$$

where

$$b = \frac{4}{n_i C(C-1)} \sum_j (u_{kj} - (u_{ki} - \frac{h'_{kl}}{n_i})) - \frac{2}{n_i C} u_{ki} + 1 \quad .$$

$\left(\dfrac{2}{n_i C} - \dfrac{4}{n_i^2(C-1)}\right)$ is just a positive constant when $n_i \ge 2$ (we can easily ensure $n_i \ge 2$) and has little effect on the update rules, so we can replace (14) by (8).

Just like updating H, we can update W by minimizing $L(W) = D(V\|WH)$ with H fixed. The auxiliary function for $L(W)$ is

$$G(W,W') = \sum_{i,j} v_{ij} \log v_{ij} -$$

$$\sum_{i,j,k} x_{ij} \frac{w'_{ik}h_{kj}}{\sum_k w'_{ik}h_{kj}}\left(\log(w_{ik}h_{kj}) - \log \frac{w'_{ik}h_{kj}}{\sum_k w'_{ik}h_{kj}}\right) + \tag{15}$$

$$\sum_{i,j} y_{ij} - \sum_{i,j} v_{ij} + \alpha S_W - \alpha S_B$$

We can prove $G(W,W) = L(W)$ and $G(W,W') \ge L(W)$ in the same way as proving $G(H,H') = L(H)$ and $G(H,H') \ge L(H)$. By letting $\dfrac{\partial G(W,W')}{\partial w_{kl}} = 0$, we find

$$w_{kl} = \frac{w'_{kl} \sum_j v_{kj} \dfrac{h_{lj}}{\sum_k w'_{kl} h_{lj}}}{\sum_j h_{lj}}$$

According to the above analysis, we conclude that the three step update rules lead to a sequence of non-increasing values of $D(V\|WH)$, and hence a local minimum.

4 EXPERIMENTS

Our experiments were performed on two benchmarks: the ORL database and a dataset from the FERET database. The nearest neighbor (NN) classifier was used for all face recognition experiments. On each benchmark, we reduced the face images from 112×92 to 28×23 for efficiency.

4.1 Cambridge ORL database

We used the ORL face database composed of 400 images of size 112×92. There are 40 persons, 10 images each person. The images were taken at different times, lighting and facial expressions. The faces are in an upright position in frontal view, with a slight left-right rotation. Figure 1 shows some samples images from the database.

Each set of 10 images for a person was randomly partitioned into a training set of five images and a test set of the other five images. The training set was then used to train the FNMF, LNMF and NMF algorithms, and the test set was used to evaluate face recognition. Both methods used the same training and test data.

Figure 1. ORL face samples

4.1.1 Learning basis components

We used NMF, LNMF and FNMF to learn the basis images of the training set from the ORL database by the update rules described by equations (8), (9), and (10).

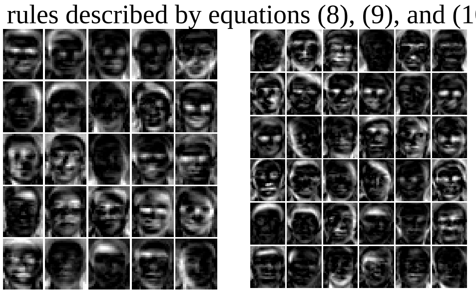

Figure 2. Basis images of NMF

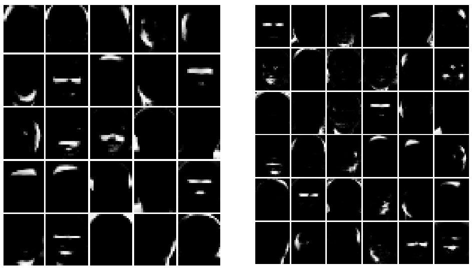

Figure 3. Basis images of LMF

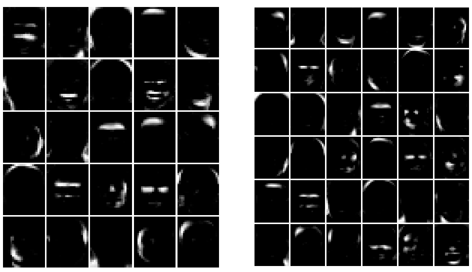

Figure 4. Basis images of FNMF

Figure 2, Figure 3 and Figure 4 show, respectively, bases of NMF, LNMF and FNMF, which were all learned from the training faces. The numbers of bases are 25 and 36. The images show that the bases trained by NMF were additive, but not spatially localized, for representation of faces. At the same time, the bases trained from FNMF and LNMF are both additive and spatially localized for representing faces.

4.1.2 Face recognition on the ORL database

In this experiment, FNMF, LNMF and NMF were compared for face recognition on the ORL database. Figure 5 shows the recognition results for the three methods. The horizontal axis represents the square root of the number of bases.

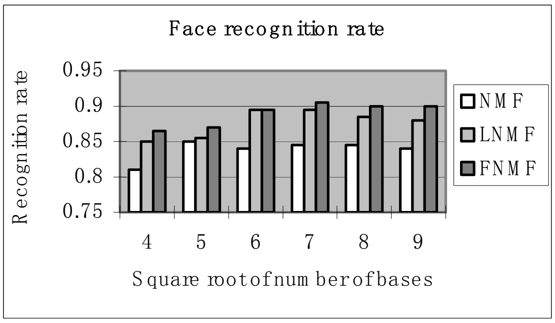

Figure 5. Face recognition on ORL database

This experiment shows that FNMF gives higher recognition rates than NMF and LNMF on the ORL database. The recognition rate of NMF was the lowest.

4.2 FERET dataset

There are 70 persons in this dataset of the FERET [10] face database. Each person has six different frontal-view images. There are three different illuminations and two different facial expressions for each illumination. Figure 6 shows some samples.

Figure 6. FERET dataset face samples

Each set of six images for a person was randomly partitioned into a training set of three images and a test set of the other three images. The training set was then used to train FNMF, LNMF and NMF, and the test set was used to evaluate face recognition. Both methods used the same training and test data.

Figure 7 compares the results of NMF, LNMF and FNMF on the FERET database. The horizontal axis represents the square root of the number of bases.

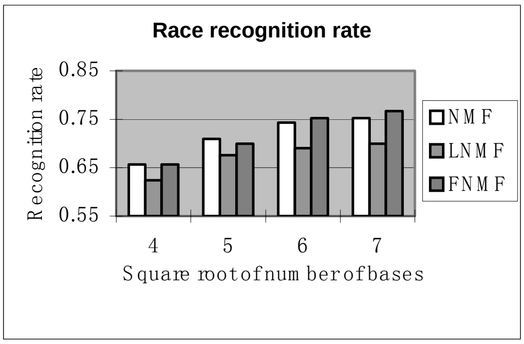

Figure 7. Face recognition on FERET database

The experiment shows that FNMF performed better than LNMF and slightly better than NMF for this data set.

4.3. Discussion

From the experimental results, we find that FNMF can obtain additive and spatially localized bases from a training set and achieve a higher recognition rate than LNMF and NMF. We also found that FNMF and LNMF perform well on the ORL database, which has little illumination variation, but LNMF is not as good as FNMF and NMF on the FERET dataset. It appears that Fisher Discrimination Analysis is better suited than PCA to deal with illumination variation, which confirms the experimental results of Belhumeur et al. [3].

Other constraints may be imposed on NMF, and in future work we will explore promising ones. For example, constraints that are part of ICA computation may be implemented to form ICA-NMF. We will experiment with this and other modifications to the basic NMF approach.

5 CONCLUSION

In this paper, we presented a new constrained non-negative matrix factorization algorithm, called Fisher non-negative matrix factorization (FNMF), for face recognition and learning part-based subspace representations. The main idea of FNMF is to take NMF as a framework and then add the Fisher constraint to the matrix factorization. We showed that using FNMF results in intuitive basis images and performs better than LNMF and NMF on face recognition.

6 REFERENCES

[1] M. Turk and A. Pentland. "Eigenfaces for recognition". *J.Cogn.Neurosci*. 3:71-86, 1991.

[2] K. Back, B.A. Draper, J.R. Beveridge and K. She. "PCA vs. ICA : A comparison on the FERET data set". *Int'l Conf. on Computer Vision, Pattern Recognition and Image Processing*, 2002.

[3] P.N. Belhumeur, J.P. Hespanha and D. Kriegman. "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection". *IEEE Trans. PAMI*, 19(7), pp. 711-720, 1997.

[4] P. Common. "Independent component analysis – a new concept?" *Signal Processing*, 36:287-314, 1994.

[5] S.Z. Li, X.W. Hou, and H.J. Zhang. "Learning Spatially Localized, Part-based Representation". *IEEE Conf. on Computer Vision and Pattern Recognition*, 2001.

[6] D.D. Lee and H.S. Seung. "Learning the parts of objects by non-negative matrix factorization". *Nature*, 401:788-791, 1999.

[7] D.D. Lee and H.S. Seung. "Algorithms for non-negative matrix factorization". In *Proceedings of Neural Information Processing Systems*, 2000.

[8] M.-H. Yang. "Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition Using Kernel Methods". *IEEE Conf. on Automatic Face and Gesture Recognition*, 2002.

[9] D.D. Lee and H.S. Seung. "Unsupervised learning by convex and conic coding". *Adv.Neural Info. Proc.Syst.* 9,515-521, 1997.

[10] P.J. Phillips, H. Moon, S. Rizvi, and P. Rauss. "The FERET evaluation methodology for face recognition algorithms". *IEEE Trans. PAMI*, Vol.22(10), pp. 1090-1104, 2000.