# Face Representation Based on Extended Non-negative Matrix Factorization

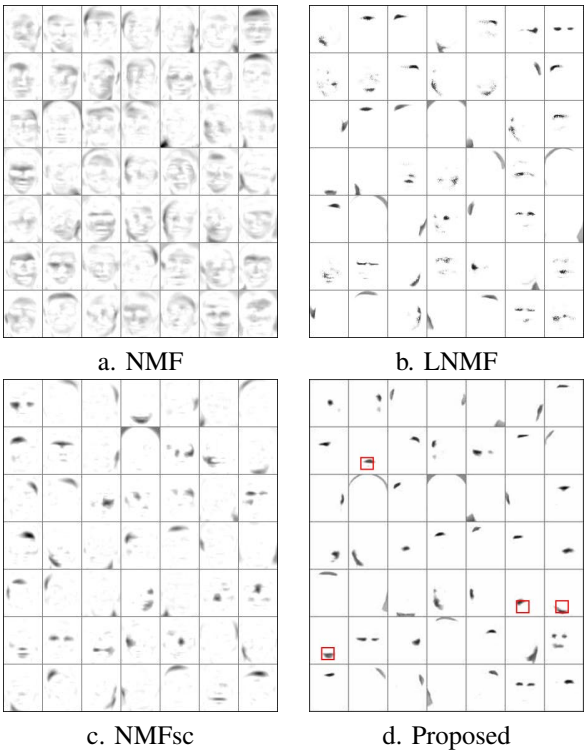
Ce Zhan, Wanqing Li, and Philip Ogunbona
School of Computer Science and Software Engineering
University of Wollongong, Australia
Email: {cz847, wanqing, philipo}@uow.edu.au

***Abstract*—In this paper, we improve Non-negative Matrix Factorization (NMF) in two ways to produce localized, part-based face representation. Firstly, NMF is extended by imposing orthogonality constraint on basis matrix while controlling the sparseness of coefficient matrix. Secondly, a new initialization method is proposed for the extended version of NMF to find spatially localized basis images. Furthermore, we define an indicator, referred to as locality $L$ to quantitatively evaluate the efficiency of a subspace projection-based method with respect to the capability of realizing the local part-based representation. Experiments based on benchmark face databases demonstrated the efficacy of the proposed method.**

## I. Introduction

Over the past few years, great efforts have been made to achieve local part-based face representations in facial analysis systems [1], [2], [3]. On the one hand, such representations are consistent with psychological and physiological evidence [4] of face perception in the human brain. On the other hand, comparing with holistic representations, local part-based representations are generally more robust to different variations such as partial occlusions and local distortions, since most of the variations in appearance affect only part of the face. Recently, rather than manually and intuitively defining the local features, more attentions is paid to learn the local representation from examples. One widely used tool for such learning is Nonnegative Matrix Factorization (NMF) [5]. NMF represent a face image as a non-negative linear combination of low rank basis images and imposes non-negativity constraints in learning the bases. As a result, the pixel values of resulting basis images are all non-negative and only additive combinations of the basis images are allowed. Therefore, NMF is considered to be compatible with the intuition notion of combining parts to form a whole in an accumulative means, the learned basis images are tend to be local facial parts.

Many factors affect NMF producing the local part-based representation. One of the key factors is the property of training dataset used to learn the basis images. Donoho et al. [6] theoretically proved that NMF does not necessarily decompose an object into parts and concluded three conditions (generative model, separability and complete factorial sampling) which training datasets should obey to guaranty NMF producing a unique, part-based decomposition. Theses conditions are not able to be satisfied in most real datasets, therefore, NMF unavoidably converges to local minima and produces different basis images corresponding to different initializations. Besides the conditions suggested by Donoho et al., other factors in real datasets especially misalignment of the object further affect the decomposition of NMF. As reported by Li et al. [7], when NMF is applied on ORL face database [8], in which faces are not well aligned, the learned basis images are holistic rather than local part-based (as can be seen in Figure 1a; we have reproduced the results).



a. NMF b. LNMF c. NMFsc d. Proposed

Fig. 1. Basis images learned from ORL database using different methods ($r = 49$).

In this paper, we improve NMF in two ways to produce localized, part-based representation. Firstly, NMF is extended by imposing orthogonality constraint on basis matrix while controlling the sparseness of coefficient matrix. Secondly, a new initialization method is proposed for the extended version of NMF to find spatially localized basis images. Furthermore, we define an indicator, referred to as locality $L$ to quantitatively evaluate the efficiency of a subspace projection-based

method with respect to the capability of realizing the local part-based representation. The rest of the paper is organized as follows: In Section II a brief introduction is given on non-negative matrix factorization (NMF) and its major variants. Details of the proposed extended NMF (ENMF) are described in Section III. Section IV presents the experimental results and conclusions are drawn in Section V.

## II. Related Works

### A. NMF

Given a non-negative data matrix $V = (v_{ij})_{m\times n}$, NMF finds non-negative matrices $W = (w_{ij})_{m\times r}$, and $H = (h_{ij})_{r\times n}$, such that $V \approx WH$. The rank $r$ of the factorization is generally chosen to satisfy $(n + m)r < mn$, so that the product $WH$ can be regarded as a compressed form of the data in $V$. Let $V$ represents a face database, each column of $V$ contains $m$ pixel values of one of the $n$ face images in the database. Then, each face in $V$ can be represented by a linear combination of $r$ columns of $W$, the columns are called basis vectors (images). Each column of $H$ is called a coefficient vector, that is in one-to-one correspondence with a face in $V$ and describes how strongly each basis is present in the face.

NMF can be taken as an optimization problem, where $W$ and $H$ are chosen to minimize the reconstruction error between $V$ and $WH$. Various error functions (objective functions) have been proposed, one widely used is the Euclidean distance function:

$$E(W, H) = \|V - WH\|^2 = \sum_{i,j}(V_{ij} - (WH_{ij}))^2 \qquad (1)$$

Although the minimization problem is convex with respect to $W$ and $H$ separately, it is not convex in both simultaneously. Paatero and Tapper [9] proposed a gradient decent method for the optimization, Lee and Seung [10] devised a multiplicative algorithm to search a local optimum.

### B. Variants of NMF

To improve NMF in learning local part-based representation, Li et al. proposed a local NMF method (LNMF) [7], that adds three additional constraints on NMF to enforce maximum sparsity in $H$ and maximum orthogonality in $W$. Figure 1b shows the basis images learned from ORL database using LNMF. Compared with NMF, we see that features gained by LNMF are more localized. However, some of the basis images are still global. This is mainly due to the introduction of maximum sparsity constraint on coefficient matrix. A high sparseness in columns of $H$ forces each coefficient to represent more of the image, and then the basis images tend to be global. Consider the extreme case when only one element in each column of $H$ is allowed to be nonzero, then the NMF reduces to vector quantization (VQ), and all the basis images become holistic prototypical faces. At the same time, high sparseness in rows of $H$ causes each learned basis present in a very small fraction of the training images, thus the learned bases are tend to be redundant, which is conflict with the intention of the orthogonality constraint. Furthermore, since more constraints are imposed, the convergence of LNMF is very slow.

As an effect of part-based decomposition, NMF usually produces sparse representation. $W$ is sparse since the learned bases tend to be non-global. $H$ is often sparse because any given sample does not contain of all the available parts (bases). Hoyer [11] proposed a method called NMF with sparseness constraints (NMFsc) to explicitly control the sparseness of $W$ and $H$. A particular sparseness for columns of $W$ requires each learned basis image has a certain fraction of pixels with values greater than zero, however, although for a very high sparseness, the small fraction of nonzero pixels are not necessarily locally distributed in the basis image. We show the basis images learned from ORL database using NMFsc in Figure 1c, where $S_w$ is set to 0.75 and $S_h$ is unconstrained as the best result achieved in [11]. As can be seen from the figure, by only directly controlling the sparseness of the representation, NMFsc does not give a better part-based representation than LNMF.

## III. Extended nonnegative matrix factorization

### A. ENMF

The proposed extended NMF (ENMF) impose orthogonality constraint on basis matrix $W$ while controlling the sparseness of coefficient matrix $H$. To reduce the overlapping between basis images, different bases should be as orthogonal as possible so as to minimize the redundancy. Denote $U = W^TW$, the orthogonality constraint can be imposed by minimizing $\sum_{i,j,i\neq j} U_{i,j}$. As analyzed in Section II, high sparsity in the coefficient matrix makes sure that a bases cannot be further decomposed into more components, while at the same time, leads basis images tend to be global. Therefore, we chose to explicitly control the sparseness level of $H$, so that a compromise can be made between localization and overlapping. The objective function of the ENMF is defined as:

$$E(W, H) = \frac{1}{2}\sum_{i,j}(V_{ij} - (WH)_{ij})^2 + \alpha \sum_{i,j,i\neq j} U_{i,j} \qquad (2)$$

where $U = W^TW$, $\alpha$ is a small positive constant. Then the ENMF is defined as following optimization problem:

$$\min_{W,H} E(W, H) \quad s.t.\ W, H \geq 0, \sum_i W_{ij} = 1\ \ \forall j \qquad (3)$$
$$sparseness(h_j) = S_h, \forall j$$

where $h_j$ is the $j$-th row of $H$; $S_h$ are the desired sparsenesses of $H$; the sparseness is measured based on the relationship between the $L_1$ norm and the $L_2$ norm [11]:

$$\text{sparseness}(h_j) = \frac{\sqrt{n} - \|h_j\|_1 / \|h_j\|_2}{\sqrt{n} - 1} \qquad (4)$$

where $n$ is the dimensionality of row vector $h_j$. This measure quantifies how much energy of the vector is packed into a few components. This function evaluates to 1 if and only if

$h_j$ contains a single nonzero component. Its value is 0 if and only if all components are non-zero and equal.

A local solution to (2) can be found by using the following two step update rules:

1)

$$W_{ia} \leftarrow W_{ia} \frac{(VH^T)_{ia}}{(WHH^T)_{ia} + \alpha \sum_i W_{ia}} \tag{5}$$

2)

$$H_{a\mu} \leftarrow H_{a\mu} - \mu_{a\mu}[W^T(WH - V)]_{a\mu} \tag{6}$$

Then project each row of $H$ to be non-negative, have unit $L_2$ norm, and $L_1$ norm set to achieve desired sparseness $S_h$. (For the projection method, please refer to [11].) $\mu_{a\mu}$ is the step size, and allowed to change at every iteration. we initially set $\mu_{a\mu}$ to 1, then multiply it by one-half at each subsequent iteration.

### B. Initialization

As introduced in Section II each column of training data matrix $V$ represents one image in the database, while each row of $V$ actually reflects the variations of one pixel across the time. The main idea of the proposed initialization method is to group pixels in images into parts before the decomposition based on the row vectors of $V$, so that the emphasis is placed on local part-based representation.

Let $R = \{1, \ldots, m\}$ denote the set of indices of row vectors in $V$, $R_j$ denote a subset of $R$. Suppose all of the row vectors in $V$ can be grouped into $k$ clusters, each cluster $C_i$ associates with a submatrix $V_i$ $(i = 1, \ldots, k)$, $V_i = V[R_i; :]$ that consists of the corresponding row vectors of $V$. Based on rank-one approximation, we have:

$$V_i \approx u_i \sigma_i p_i^T \tag{7}$$

where $u_i$ and $p_i$ are left and right singular vectors associated with the largest singular value of $V_i$. Then the basis matrix $W$ and coefficient matrix $H$ can be initialized as:

$$W[R_i; i] = u_i \tag{8}$$

$$W[\bar{R}_i; i] = \epsilon \tag{9}$$

$$H[i; :] = \sigma_i p_i^T \tag{10}$$

where $W[\bar{R}_i; i]$ denotes the rest of elements in $ith$ column of $W$ excluding elements with the indices in $R_i$. $\epsilon$ is a small positive value replacing 0 to prevent parameters from being locked under the multiplicative update rules.

Figure 1d shows an example of the bases learned from ORL database using the proposed ENMF, $S_h$ is set to 0.1, $\alpha$ is set to 1 and $\epsilon$ is set to 0.001, K-Means is employed to group the row vectors in training data matrix $V$ during initialization. As can be seen from the figure, more localized, less overlapped basis images are obtained, and limited bases contribute to each specific local facial area. At the same time, the initialization makes ENMF converge much faster than LNMF.

## IV. Experimental results

### A. Measuring the locality

In Figure 1, we have shown that the proposed ENMF is able to produce more localized and less overlapped basis images. To further quantitatively evaluate the efficiency of ENMF with respect to the capability of realizing the local part-based representation, we define an indicator, referred to as locality $L$, for a representation learnt using a subspace projection-based method. Specifically, after basis images are learned using a subspace projection-based method, let $r$ denote the total number of basis images (rank), $e_i$ denote the number of basis images related to a local region $g_i$ (Figure 1d shows an example of ENMF bases related to a "mouth" area), assume the whole image is divided into $m$ sub-regions $g_1, \ldots, g_m$, then $L$ is defined as:

$$L = \frac{1}{m} \sum_{i=1}^{m} e_i / r \tag{11}$$

$L$ is an important parameter for a representation learned with subspace projection-based methods. When the basis images are holistic, the value of $L$ is close to 1, almost all the bases contribute to a local region. While local part-based basis images would obtain a small value of $L$, only few bases are related to each specific local region. With a small value of $L$, local variations such as partial occlusions and local distortions only affect a small part of the coefficients used to represent an image, thus a more robust representation is achieved than those with larger value of $L$.

We calculate the locality $L$ for NMF, LNMF, NMFsc and ENMF based on two benchmark databases ORL [8] and FERET [12] face databases. The ORL face database [8] mainly addresses the pose variation especially out-of-plane rotations of faces, while the FERET database only focused frontal view of faces. Only Images in the gallery subset of FERET database are used in the experiment, the images are transformed to simulate three kinds of possible in-plane misalignment: random translation within $[+5, -5]$ pixels in vertical or horizontal; random rotation within $[+15°, -15°]$; random scaling within $[0.9, 1.1]$. After the transformation, all images are cropped to $64 \times 64$ face images. As for ORL database, face area of the images are first detected and then resized to $64 \times 64$. Sample face images used in the experiments are shown in Figure 2.

To calculate $L$ for each obtained representation, the face image is equally divided into non-overlapping $8 \times 8$ sub-regions. For NMFsc, we follow Hoyer's [11] work and set $S_w$ to 0.75 and keep $S_h$ unconstrained. For ENMF, $S_h$ is set to 0.1 $\alpha$ is set to 1 and $\epsilon$ is set to 0.001. K-Means is employed to group the row vectors in training data matrix $V$ during initialization. The traditional NMF, LNMF, NMFsc and ENMF are applied on the ORL database and transformed FERET database with different total number of bases $r$. Table 1 and Table 2 show the calculated $L$ for all the obtained representations learned from ORL database and FERET database respectively. As can be seen from the tables, the proposed ENMF achieves the

Transformed FERET

ORL

Fig. 2. Sample face images used in the experiments

| Method | $r = 36$ | $r = 49$ | $r = 64$ | $r = 81$ | $r = 100$ |
|---|---|---|---|---|---|
| NMF | 0.973 | 0.976 | 0.976 | 0.976 | 0.979 |
| NMFsc | 0.423 | 0.571 | 0.727 | 0.757 | 0.801 |
| LNMF | 0.258 | 0.238 | 0.199 | 0.195 | 0.187 |
| ENMF | 0.194 | 0.172 | 0.127 | 0.125 | 0.121 |

TABLE I
THE LOCALITY $L$ FOR DIFFERENT REPRESENTATIONS LEARNED FROM FERET DATABASE

smallest $L$ in all the cases, especially for ORL database that addresses out of plane rotations of the faces.

### *B. Face recognition*

In this experiment, NMF, LNMF, NMFsc and ENMF representations are comparatively evaluated for face recognition based on ORL database. The following simple recognition scheme is used:

1) Feature extraction. After the subspace is learned from training images, let $\bar{t}$ be the mean of training images. Each training face image $t_i$ is projected into the learned subspace as a feature vector $f_i = W^{-1}(t_i - \bar{t})$ which is then used as a prototype feature point. A query face image $q$ to be classified is represented by its projection in the subspace as $f_q = W^{-1}(q - \bar{t})$.
2) Nearest neighbor classification. The Euclidean distance between the query and each prototype, $d(f_q, f_i)$,is calculated. The query is classified to the class to which the closest prototype belongs.

| Method | $r = 36$ | $r = 49$ | $r = 64$ | $r = 81$ | $r = 100$ |
|---|---|---|---|---|---|
| NMF | 1 | 1 | 1 | 1 | 1 |
| NMFsc | 0.667 | 0.836 | 0.946 | 0.956 | 0.969 |
| LNMF | 0.422 | 0.396 | 0.345 | 0.341 | 0.329 |
| ENMF | 0.275 | 0.232 | 0.197 | 0.193 | 0.186 |

TABLE II
THE LOCALITY $L$ FOR DIFFERENT REPRESENTATIONS LEARNED FROM ORL DATABASE

| Method | $r = 36$ | $r = 49$ | $r = 64$ | $r = 81$ | $r = 100$ |
|---|---|---|---|---|---|
| NMF | 0.43 | 0.41 | 0.38 | 0.37 | 0.39 |
| NMFsc | 0.84 | 0.82 | 0.81 | 0.85 | 0.88 |
| LNMF | 0.91 | 0.91 | 0.92 | 0.94 | 0.93 |
| ENMF | 0.94 | 0.96 | 0.96 | 0.97 | 0.97 |

TABLE III
THE RECOGNITION RATES FOR DIFFERENT METHODS

Different from the first experiment, all images from ORL database are directly used without normalization. The set of 10 images for each person is randomly partitioned into a training subset of 5 images and a test set of the other 5. The training set is then used to learn the subspaces (basis images), and the test set for evaluation. All the compared methods take the same training and test data. For NMFsc and ENMF, the same parameters are used as in the first experiment. Table IV-B shows the recognition rates for the compared representations with different total number of bases $r$. As can be seen from the table, the performance of traditional NMF is surprisingly poor, ENMF achieves the best recognition accuracy.

## V. CONCLUSION

In this paper, by imposing orthogonality constraint on basis matrix while controlling the sparseness of coefficient matrix, the NMF is extended for producing a localized, part-based representation for face images. Since the proposed ENMF is a local minimizer, gives different representations from different initial conditions. We thus also propose a new initialization method to place the emphasis of decomposition on local part-based representation. Furthermore, we define an indicator, referred to as locality $L$ to quantitatively evaluate the efficiency of a subspace projection-based method with respect to the capability of realizing the local part-based representation. Experimental results have shown that the proposed ENMF derives more localized basis images for face representation than NMF and its major extensions, and not sensitive to training data, especially robust to misalignment of faces.

## REFERENCES


[1] S. Z. Li and A. K. Jain, *Handbook of Face Recognition*, S. Z. Li and A. K. Jain, Eds. Springer, 2004.

[2] Y. Fu, G. Guo, and T. S. Huang, "Age synthesis and estimation via faces: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 1955–1976, 2010.

[3] M. Pantic and L. Rothkrantz, "Automatic analysis of facial expressions: the state of the art," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1424–1445, 2000.

[4] S. E. Palmer, "Hierarchical structure in perceptual representation," *Cognitive Psychology*, vol. 9, pp. 441–474, 1977.

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

[6] D. Donoho and V. Stodden, "When does non-negative matrix factorization give a correct decomposition into parts?" in *Proceedings of NIPS'2003*, 2003.

[7] S. Z. Li, X. Hou, H. Zhang, and Q. Cheng, "Learning spatially localized, parts-based representation," in *Proceedings of CVPR'01*, vol. 1, 2001, pp. 207–212.

[8] F. Samaria and A. Harter, "Parameterisation of a stochastic model for human face identification," *Proceedings of 2nd IEEE Workshop on Applications of Computer Vision*, pp. 138–142, 1994.

[9] P. Paatero and U. Tapper, "Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values," *Environmetrics*, vol. 5, no. 2, pp. 1180–4009, 1994.

[10] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proceedings of NIPS'2000*, 2000, pp. 556–562.

[11] P. O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *The Journal of Machine Learning Research*, vol. 5, no. 5, pp. 1457–1469, 2004.

[12] P. Phillips, H. Wechsler, J. Huang, and P. Rauss, "The feret database and evaluation procedure for face recognition algorithms," *Image and Vision Computing*, vol. 16, pp. 295–306, 1998.