21st International Conference on Pattern Recognition (ICPR 2012)
November 11-15, 2012. Tsukuba, Japan

Dual Subspace Nonnegative Matrix Factorization for Person-Invariant Facial Expression Recognition

Yi-Han Tu and Chiou-Ting Hsu
Department of Computer Science, National Tsing Hua University, Taiwan
s9962557@m99.nthu.edu.tw, and cthsu@cs.nthu.edu.tw

Abstract

Person-dependent appearance changes tend to increase difficulties in automatic facial expression recognition. Although one can use neutral face images to reduce the personal variations, acquisition of neutral face images may not always be possible in real cases. In order to remove the person-dependent influence from expressive images, we propose a dual subspace nonnegative matrix factorization (DSNMF) to decompose facial images into two parts: identity and expression parts. The identity part should characterize person-dependent variations, while the expression part should characterize person-invariant expression features. Our experimental results show that the proposed method significantly outperforms existing approaches on the CK+ and JAFFE expression databases.

1. Introduction

Facial expression analysis has received much attention in recent years. Many existing approaches tried to handle the environmental changes (e.g., pose, illumination) [1, 2] for expression recognition. However, automatic expression recognition is still very challenging because of different appearance changes among different individuals [3]. To reduce the personal variations in expression recognition, in [4, 5], the authors proposed to determine the person identity before conducting expression recognition. The overall performance, however, heavily relies on a robust face recognition system. On the other hand, few approaches [6, 7] tried to reduce the influence from personal variations. In [6, 7], the difference image, which is defined as the difference between a fully expressive image and a neural face image, has been proposed to remove personal appearance variations for expression recognition. Nevertheless, because neutral face images are not always available in real world applications, we need a better strategy to extract expression-related features by excluding person-dependent information from a fully expressive face image.

Facial images are highly structural and have been shown to reside on a low-dimensional manifold. Traditional facial representations, such as Eigenfaces (PCA) and Fisherfaces (LDA), tend to describe global facial structure. Since facial expression variations usually involve local appearance changes (e.g., eyes, mouth, etc.), such holistic representations usually fail to characterize specific local parts for expression analysis. Local facial components have also been shown to contain more discriminative information and outperform global features for face recognition [8, 9].

Therefore, we believe local or part-based representation is more appropriate for expression analysis. In recent years, non-negative matrix factorization (NMF) [10] has been shown to be more interpretable for facial image analysis and tend to generate part-based bases. Since the part-based NMF bases have the physical meaning of combining local parts to form a whole face, our proposed new facial representation will be developed based on NMF. Our goal is twofold: one is to extract expression-related features to characterize local appearance changes, and the other is to exclude person-dependent information from the representation.

In this paper, we propose a novel non-negative matrix factorization, called dual subspace nonnegative matrix factorization (DSNMF), to decompose facial images into identity and expression parts. The identity part should characterize person-dependent appearance variations; while the expression part is expected to characterize person-independent expression features with as little identity information as possible. The experimental results on CK+ [11] and JAFFE [12] expression databases show that our method is robust to human identity variation and greatly improves the

performance especially when no neutral face is available.

2. Dual subspace nonnegative matrix factorization (DSNMF)

Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in \mathbb{R}^{M\times N}$ denote the matrix containing the N vectorized facial images in the training database, where M is the number of pixels in each image. The goal of NMF is to decompose $\mathbf{X}$ into two non-negative matrices by

$$\mathbf{X} \approx \mathbf{WH}\,, \tag{1}$$

where $\mathbf{W} \in \mathbb{R}^{M\times d}$ is the non-negative basis matrix, $\mathbf{H} \in \mathbb{R}^{d\times N}$ is the non-negative coefficient matrix, and d is the number of basis images.

Since NMF is an unsupervised method, existing approaches [13, 14] have tried to include label information into NMF to improve the performance for classification problems. Nevertheless, while we conduct Graph-Preserving Sparse NMF (GSNMF) [13] and Projective Nonnegative Graph Embedding (PNGE) [14] for facial expression recognition, our experiments (as shown in Table 1) show that the graph embedding constraint alone is insufficient to extract expression-related features. On the other hand, although difference images (as shown Figure 1(c)) have been shown to be less sensitive to individual differences than the expressive images (Figure 1(a)), neural faces (Figure 1(b)) are rarely available in most of the expression recognition scenarios.

Therefore, we propose to decompose one expressive image into an identity part (with neutral expression) and an expression-related part under the non-negative constraint. Let $\mathbf{W} = [\mathbf{W}_\mathrm{I}\ \mathbf{W}_\mathrm{E}]$, where $\mathbf{W}_\mathrm{I} \in \mathbb{R}^{M\times(d-q)}$ and $\mathbf{W}_\mathrm{E} \in \mathbb{R}^{M\times q}$ denote the identity and expression bases, respectively. Once given an expressive image $\mathbf{x}_i$, we decompose the image by

$$\mathbf{x}_i \approx \mathbf{W}_\mathrm{I}\mathbf{h}_\mathrm{I} + \mathbf{W}_\mathrm{E}\mathbf{h}_\mathrm{E} = \mathbf{x}_i^\mathrm{I} + \mathbf{x}_i^\mathrm{E}\,, \tag{3}$$

where $\mathbf{h}_\mathrm{I}$ is the identity coefficient, $\mathbf{h}_\mathrm{E}$ is the expression coefficient, $\mathbf{x}_i^\mathrm{I}$ is the identity part, and $\mathbf{x}_i^\mathrm{E}$ is the expression part. Note that, the expression part $\mathbf{x}_i^\mathrm{E} \approx \mathbf{x}_i - \mathbf{x}_i^\mathrm{I}$ resembles the concept of difference images.

Given a training database, our goal is to decompose the whole data set by

$$\mathbf{X} \approx \mathbf{W}_\mathrm{I}\mathbf{H}_\mathrm{I} + \mathbf{W}_\mathrm{E}\mathbf{H}_\mathrm{E},\ \text{s.t.}\ \mathbf{W}_\mathrm{I}, \mathbf{W}_\mathrm{E}, \mathbf{H}_\mathrm{I}, \mathbf{H}_\mathrm{E} \geq 0\,, \tag{4}$$

where $\mathbf{H}_\mathrm{I} = [\mathbf{h}_\mathrm{I}^1, \mathbf{h}_\mathrm{I}^2, \dots, \mathbf{h}_\mathrm{I}^N] \in \mathbb{R}^{(d-q)\times N}$ and $\mathbf{H}_\mathrm{E} = [\mathbf{h}_\mathrm{E}^1, \mathbf{h}_\mathrm{E}^2, \dots, \mathbf{h}_\mathrm{E}^N] \in \mathbb{R}^{q\times N}$ are the coefficient matrices corresponding to $\mathbf{W}_\mathrm{I}$ and $\mathbf{W}_\mathrm{E}$, respectively. In Eq. (4), we expect that the expression part should describe only the expression-related variation, while the identity part should describe the personal variation depending only on their identity. We therefore include two additional constraints: (1) the variations between persons with the same expression should be minimized in the expression part; and (2) the variation between persons with the same identity should be minimized in the identity part. We formulate the above two constraints as follows:

$$\begin{cases} \min_{\mathbf{H}_\mathrm{E}} \sum_{i\neq j} \left\|\mathbf{h}_\mathrm{E}^i - \mathbf{h}_\mathrm{E}^j\right\|^2 \mathbf{S}_{ij}^\mathrm{E} = \min_{\mathbf{H}_\mathrm{E}} \mathrm{Tr}\left(\mathbf{H}_\mathrm{E}\mathbf{L}^\mathrm{E}\mathbf{H}_\mathrm{E}{}^\mathrm{T}\right) \\ \min_{\mathbf{H}_\mathrm{I}} \sum_{i\neq j} \left\|\mathbf{h}_\mathrm{I}^i - \mathbf{h}_\mathrm{I}^j\right\|^2 \mathbf{S}_{ij}^\mathrm{I} = \min_{\mathbf{H}_\mathrm{I}} \mathrm{Tr}\left(\mathbf{H}_\mathrm{I}\mathbf{L}^\mathrm{I}\mathbf{H}_\mathrm{I}{}^\mathrm{T}\right) \end{cases} \tag{5}$$

where

$$\mathbf{S}_{ij}^\mathrm{E} = \begin{cases} \exp\left(\frac{-\|\mathbf{x}_i - \mathbf{x}_j\|_2^2}{\sigma^2}\right), \text{if } \mathrm{E}(\mathbf{x}_i) = \mathrm{E}(\mathbf{x}_j) \\ 0, \text{otherwise} \end{cases}, \text{ and}$$

$$\mathbf{S}_{ij}^\mathrm{I} = \begin{cases} \exp\left(\frac{-\|\mathbf{x}_i - \mathbf{x}_j\|_2^2}{\sigma^2}\right), \text{if } \mathrm{P}(\mathbf{x}_i) = \mathrm{P}(\mathbf{x}_j) \\ 0, \text{otherwise} \end{cases}. \tag{6}$$

In Eq.(6), $\mathrm{P}(\mathbf{x}_i)$ and $\mathrm{E}(\mathbf{x}_i)$ denote the identity label and expression label of the sample $\mathbf{x}_i$ respectively, and σ is the empirical parameter for the bandwidth of weight kernel. $\mathbf{L}$ denotes the Laplacian matrix and is derived by $\mathbf{L} = \mathbf{D} - \mathbf{S}$, where $\mathbf{D}_{ii} = \sum_{i\neq j} \mathbf{S}_{ij}$.

We next combine the constraints in Eq.(5) into Eq.(4) and define the objective function as:

$$\min_{\mathbf{W},\mathbf{H}} \|\mathbf{X} - \mathbf{WH}\|_\mathrm{F}^2 + \lambda\mathrm{Tr}\left(\mathbf{H}_\mathrm{E}\mathbf{L}^\mathrm{E}\mathbf{H}_\mathrm{E}{}^\mathrm{T}\right) + \lambda\mathrm{Tr}\left(\mathbf{H}_\mathrm{I}\mathbf{L}^\mathrm{I}\mathbf{H}_\mathrm{I}{}^\mathrm{T}\right),\ \text{s.t.}\ \mathbf{W}, \mathbf{H} \geq 0\,. \tag{7}$$

where $\mathbf{W} = [\mathbf{W}_\mathrm{I}\ \mathbf{W}_\mathrm{E}]$, $\mathrm{H} = \begin{bmatrix}\mathbf{H}_\mathrm{I} \\ \mathbf{H}_\mathrm{E}\end{bmatrix}$ and λ is an empirical parameter to control the significant of the constraints. Finally, to avoid infinitely many solutions in Eq.(7), we normalize the basis matrix $\mathbf{W}$, (i.e. $\|\mathbf{w}_i\| = 1, i = 1,2, \dots, d$) and formulate the proposed DSNMF by

$$\min_{\mathbf{W},\mathbf{H}} \|\mathbf{X} - \mathbf{WH}\|_\mathrm{F}^2 + \lambda\mathrm{Tr}\left(\widehat{\mathbf{Q}}\mathbf{H}_\mathrm{E}\mathbf{L}^\mathrm{E}\mathbf{H}_\mathrm{E}{}^\mathrm{T}\widehat{\mathbf{Q}}^\mathrm{T}\right) + \lambda\mathrm{Tr}\left(\widetilde{\mathbf{Q}}\mathbf{H}_\mathrm{I}\mathbf{L}^\mathrm{I}\mathbf{H}_\mathrm{I}{}^\mathrm{T}\widetilde{\mathbf{Q}}^\mathrm{T}\right),\ \text{s.t.}\ \mathbf{W}, \mathbf{H} \geq 0\,, \tag{8}$$

where

$$\widetilde{\mathbf{Q}} = \mathrm{diag}(\|\mathbf{w}_1\|, \|\mathbf{w}_2\|, \dots, \|\mathbf{w}_{d-q}\|)\,, \text{ and}$$
$$\widehat{\mathbf{Q}} = \mathrm{diag}(\|\mathbf{w}_{d-q+1}\|, \|\mathbf{w}_{d-q+2}\|, \dots, \|\mathbf{w}_d\|)\,. \tag{9}$$

To solve the matrices $\mathbf{W}$ and $\mathbf{H}$ in Eq.(8), we derive the multiplicative update rule:

$$\mathbf{W}_{ij} \leftarrow \mathbf{W}_{ij} \cdot \frac{(\mathbf{XH}^\mathrm{T} + \mathbf{WY}_{\mathrm{w}-})_{ij}}{(\mathbf{WHH}^\mathrm{T} + \mathbf{WY}_{\mathrm{w}+})_{ij}}\,, \text{ and} \tag{10}$$

$$\mathbf{H}_{ij} \leftarrow \mathbf{H}_{ij} \cdot \frac{(\mathbf{W}^\mathrm{T}\mathbf{X} + \lambda\begin{bmatrix}\mathbf{H}_\mathrm{I}\mathbf{S}^\mathrm{I} \\ \mathbf{H}_\mathrm{E}\mathbf{S}^\mathrm{E}\end{bmatrix})_{ij}}{(\mathbf{W}^\mathrm{T}\mathbf{WH} + \lambda\begin{bmatrix}\mathbf{H}_\mathrm{I}\mathbf{D}^\mathrm{I} \\ \mathbf{H}_\mathrm{E}\mathbf{D}^\mathrm{E}\end{bmatrix})_{ij}}\,, \tag{11}$$

where

$$\mathbf{Y}_{\mathrm{w}+} = \begin{bmatrix}\mathbf{H}_\mathrm{I}(\lambda\mathbf{D}^\mathrm{I})\mathbf{H}_\mathrm{I}{}^\mathrm{T} & 0 \\ 0 & \mathbf{H}_\mathrm{E}(\lambda\mathbf{D}^\mathrm{E})\mathbf{H}_\mathrm{E}{}^\mathrm{T}\end{bmatrix} \text{ and}$$

$$\mathbf{Y}_{w-} = \begin{bmatrix} \mathbf{H}_I(\lambda \mathbf{S}^I)\mathbf{H}_I{}^T & 0 \\ 0 & \mathbf{H}_E(\lambda \mathbf{S}^E)\mathbf{H}_E{}^T \end{bmatrix}. \quad (12)$$

3. Experimental result

3.1. Datasets and settings

We use the Extended Cohn-Kanade (CK+) database [11] and the JAFFE database [12] to evaluate the performance of facial expression recognition. Since existing reports were usually conducted under different experimental conditions, in order to have a fair comparison, we implement all the relevant approaches under the same experimental conditions. We use 309 sequences with six basic emotion labels (angry, disgust, fear, happy, sad, and surprise) in CK+ database and 183 images in JAFFE database as our training database, respectively. For CK+, only the last frame in each sequence is selected as the expression image. All the images are aligned by eye coordinates, and then cropped into 48×44 resolution.

In all the experiments, we use the leave-one-person-out strategy to evaluate the recognition performance. The nearest neighbor classifier is used for classification. In our DSNMF, the number of basis images d is set as 60 and 35 for CK+ and JAFFE databases, respectively; the number of expression basis image q is set as $0.6d$, and the parameter λ is fixed to 1. Only the expression parts are used for expression recognition.

3.2. Evaluation of facial representation

Figure 2 shows the basis images obtained by our proposed DSNMF. In Figure 2(a), the expression basis images ($\mathbf{W}_E$) tend to describe local-parts on faces; in Figure 2(b), most of the identity basis images ($\mathbf{W}_I$) characterize the global facial structure. The results in Figure 2 verify that facial expression mainly involves local appearance changes, while facial identity information usually contains more global structure in expressive facial images.

In Figure 3, we use only the expression basis ($\mathbf{W}_E$) to reconstruct the expressive images. The reconstructed images of the same expression, though for different persons, look very similar and contain almost no identity information. On the other hand, in Figure 4, while we use only the identity basis ($\mathbf{W}_I$) to reconstruct expressive images, the reconstructed images for the same person with different expressions all resemble their own neutral faces. These two simulations show that our proposed method effectively decomposes expression images into their corresponding identity and expression parts.

3.3. Facial expression recognition

We compare the proposed DSNMF with the following facial representations: NMF, GSNMF [13], PNGE [14], and sparse representation using difference image (SR-Diff) [7]. GSNMF and PNGE are the state-of-the-art extension of NMF, while the SR-Diff is the approach using neutral face image.

Table 1 shows the recognition rate of different methods. Because our experiment setting is more challenging than that in [13], GSNMF only achieves similar performance to NMF; in other words, the additional label information is insufficient to extract expression-related feature. PNGE [14] performs worse than other methods because the projective constraint of NMF is too strong. In addition, their performance may be greatly influenced by human identity information. In contrast, by decomposing identity and expression information, our proposed method successfully extract person-independent expression features. Therefore, as shown in table 1, the proposed method significantly outperforms the other approaches on both two databases. Furthermore, our method outperforms [7] by 10.1% without using neutral images. These experiments showed that our method indeed reduces the influence from personal variations and is able to extract person-independent expression features.

Table 1. Accuracy(%) of facial expression recognition for the CK+ and JAFFE databases.

	CK+	JAFFE
NMF	72.82	41.53
GSNMF [13]	71.84	43.17
PNGE [14]	51.78	42.62
SR-Diff [7]	80.91	45.90
Our DSNMF	**90.92**	**53.01**

Figure 1. Images of different persons from the CK+ database. (a) Expressive images, (b) Neutral face images, and (c) Difference images.

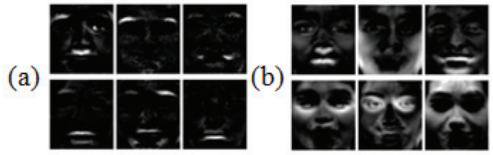

Figure 2. Basis of our DSNMF. (a) Expression basis ($\mathbf{W}_E$), (b) Identity basis ($\mathbf{W}_I$).

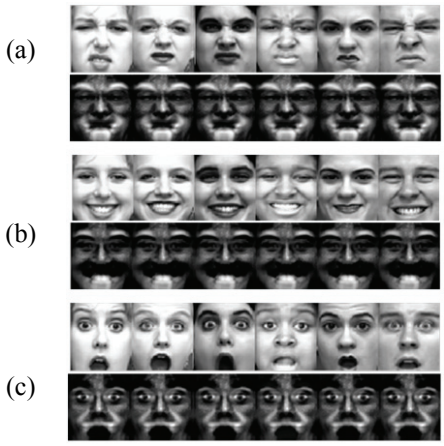

Figure 3. Reconstructed images of our DSNMF using only the expression basis ($\mathbf{W}_E$). Each image column is the same person with different expressions. (a) Disgust, (b) Happy, (c) Surprise.

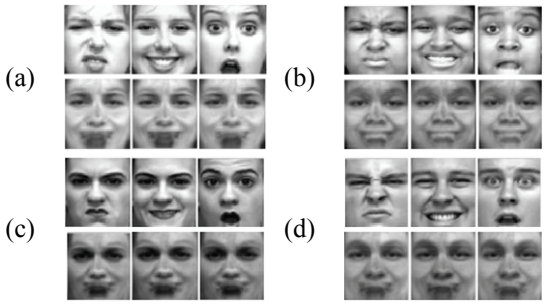

Figure 4. Reconstruction using only the identity basis ($\mathbf{W}_I$) of our DSNMF. (a)-(d) first row are the same person with different expressions and second row are the reconstructed images.

4. Conclusion

In this paper, we propose a novel approach to extract person-invariant expression features for expression recognition. The proposed DSNMF decomposes face image into two subspaces: identity and expression part. The expression features effectively characterize the expression-related facial appearance change and contain less non-expression information. Our experiments on CK+ and JAFFE databases show that the proposed DSNMF outperforms other state-of-the-art extensions of NMF. Furthermore, the DSNMF even improves the recognition rate by 10.1% without using any neutral face images. By decomposing identity and expression information, DSNMF successfully extracts person-invariant expression features and are more robust to appearance variations across different persons. In addition, the proposed DSNMF is feasible to recognize expressions across different individuals even when the testing person is not included in the training data.

References

[1] O. Rudovic, I. Patras, and M. Pantic, "Coupled Gaussian Process Regression for Pose-Invariant Facial Expression Recognition," *European Conference on Computer Vision,* vol. 6312, pp. 350-363, 2010.

[2] G. Zhao and M. Pietikainen, "Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions," *IEEE Trans. PAMI,* vol. 29, pp. 915-928, 2007.

[3] B. Fasel and J. Luttin, "Automatic Facial Expression Analysis: Survey," *Pattern Recognition,* vol. 36, no. 1, pp. 259-275, 2003.

[4] P. Martins and J. Batista, "Identity and Expression Recognition on Low Dimensional Manifolds," *IEEE Intl. Conf. Image Processing,* 2009.

[5] Kai-Tai Song and Yi-Wen Chen, "A Design for Integrated Face and Facial Expression Recognition," *IEEE Conf. Industrial Electronics Society,* 2012.

[6] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, "Classifying Facial Actions," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 21, no. 10, pp. 974-989, 1999.

[7] S. Zafeiriou, and M. Petrou, "Sparse Representations For Facial Expressions Recognition via L1 optimization," *IEEE Conf. Computer Vision and Pattern Recognition Workshops,* 2010.

[8] J.Zou, Q.Ji, and G.Nagy, "A Comparative Study of Local Matching Approach for Face Recognition," *IEEE Trans. Image Processing*, vol. 16, no. 10, pp. 2617-2628, 2007.

[9] B. Heisele, P. Ho, J. Wu, and T. Poggio, "Face Rrecognition: Component-based Versus Global Approaches," *Computer Vision and Image Understanding,* vol. 91, pp. 6–12, 2003.

[10] D. D. Lee and H. S. Seung, "Learning the Parts of Objects by Non-negative Matrix Factorization," *Nature,* vol. 401, pp.788–791, 1999.

[11] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete expression dataset for action unit and emotion-specified expression," *IEEE Conf. Computer Vision and Pattern Recognition Workshops,* 2010.

[12] M. J. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding Facial Expressions with Gabor Wavelets," *IEEE Intl. Conf. Automatic Face and Gesture Recognition,* pp.200-205, 1998.

[13] R. Zhi, M. Flierl, Q. Ruan, and W. B. Kleijn, "Graph-Preserving Sparse Nonnegative Matrix Factorization with Application to Facial Expression Recognition," *IEEE Trans. SMC-Part B: Cybernetics,* vol. 41, no. 1, pp. 38-52, 2011.

[14] X. Liu, S. Yan, and H. Jin, "Projective Nonnegative Graph Embedding," *IEEE Trans. Image Processing,* vol. 19, no. 5, pp. 1126-1137, 2010.