Journal of Applied Sciences

ISSN 1812-5654

science alert

ANSInet
an open access publisher
http://ansinet.com

Journal of Applied Sciences 11 (2): 354-359, 2011
ISSN 1812-5654 / DOI: 10.3923/jas.2011.354.359
© 2011 Asian Network for Scientific Information

An Efficient Initialization Method for Nonnegative Matrix Factorization

[1]M. Rezaei, [1]R. Boostani and [2]M. Rezaei
[1]Department of CSE and IT, Faculty of Electrical and Computer Engineering, Shiraz University, Shiraz, Iran
[2]Faculty of Computer Engineering, Shahid Bahonar University, Kerman, Iran

Abstract: Although Non-negative Matrix Factorization (NMF) has been employed in many real applications but it still suffers from three shortcomings in terms of finding a suitable initialization method, choosing an effective cost function in addition with determining the suitable reduced dimension (Factors Rank). The aim of this study is to enhance NMF performance using Fuzzy C-Means Clustering (FCM) as an efficient initialization method for estimating initial factors of NMF. In this paper, we proposed an initialization in which both W and H matrices are identified simultaneously. The proposed method was applied to JAFFE facial expression dataset and the results exhibited superiority of this method compare to other the state-of-art initialization methods on the employed dataset.

Key words: Nonnegative matrix factorization, fuzzy c-means clustering, initialization, factorization, rank reduction, facial expression

INTRODUCTION

Facial expressions play an important role in human communications such that research findings demonstrated information passed by language, tune and expressions accounts for 7, 38 and 55%, respectively during human communication. This statistic exhibits clearly the importance of facial expressions in mutual understanding among the faces (Bezdek, 1981). In order to understand the state of a face, the first step is to extract discriminative features from the face that can describe details of the face components such as lip form, eyes and eyebrow states, flexion or extension of the face muscles. In this way, a number of methods have addressed facial expression recognition such as Local Binary Patterns (LBP) (Shan *et al*., 2009), static topographic modeling (Wang and Yin, 2007), texture and shape information fusion (Kotsia *et al*., 2008), Principal Component Analysis (PCA) (Calder *et al*., 2001), appearance modeling (Abboud *et al*., 2004) and integration of facial expression with facial appearance models (Tsai *et al*., 2009).

Recently, Non-negative Matrix Factorization (NMF) has been proposed and widely used for facial expression recognition (Lee and Seung, 1999; Zhao *et al*., 2008; Ying and Zhang, 2009), because this method preserves a part-based representation of the original image and guarantees that both the resulting low-dimensional basis (W) and its accompanying weights (H) are non-negative matrices. The idea of part-based decomposition of faces simplifies the recognition of a face state due to the fact that each face state affects the aforementioned face components. In general, NMF algorithms are divided into three classes in terms of multiplicative update algorithms, gradient descent algorithms and alternating least squares algorithm (Berry *et al*., 2006). Several efforts have been made to improve the NMF performance including improvement in the NMF cost function (Cichocki *et al*., 2006; Finesso and Spreij, 2006) and development of a nearly optimum initialization method for NMF components (Xue *et al*., 2008; Wild *et al*., 2004; Zheng *et al*., 2007; Zhao *et al*., 2008).

In this study, we focus on modifying the NMF performance by exploiting an efficient initialization method. Due to the non-convex optimization formula of NMF algorithm, there is no guarantee that both components of NMF (W and H matrices) are optimally determined. On the other hand, NMF can provide different results corresponding to different initial values for W and H elements. Thus, if the initialization values are chosen properly, the results would be nearly optimum compared to the situation that initial values are determined far from the global solution.

Recently, different methods have been suggested to find suitable values for initialization of W and H (Xue *et al*., 2008; Wild *et al*., 2004; Zheng *et al*., 2007; Zhao *et al*., 2008). In this regard, Zhao *et al*. (2008) employed PCA to perform eigendecomposition of a matrix; then, they removed those eigenvectors corresponding to small eigenvalues and assigned the reduced matrices to W and H components as their initial values. Wild *et al*.

Corresponding Author: Masoumeh Rezaei, Department of CSE and IT, Faculty of Electrical and Computer Engineering, Shiraz University, Shiraz, Iran

(2004) used spherical k-means to initialize W matrix and employed Non-negative Least Square (NNLS) algorithm (Lawson and Hanson, 1974) to calculate H matrix. This method was followed by Xue *et al.* (2008) who used divergence-based *k*-means clustering and after achieving cluster centers, to determine matrix H, in each column of this matrix, the cluster number of each sample is one and others are zero. In another attempt, Zheng *et al.* (2007) used fuzzy clustering for matrix W and then for computing matrix H, largest membership degree of each sample is set to one and others assign to zeros. In the mentioned initialization methods (except PCA), first matrix W is determined and then matrix H estimated based on W. This is the main cause that the reviewed NMF initializing algorithms tend to find a weak local optimum and then converge to it; consequently, after number of iterations, the progression of these algorithms is not significant such that in large number of iterations, results of random initialization empirically showed a better performance than these initialization methods. In this paper, we propose a method based on fuzzy clustering for initialization of both matrices. Experimental results achieved by the proposed method revealed that the NMF performance outperformed other mentioned methods and with this method, problem of local optimum for initialization methods is solved.

BACKGROUND REVIEW

Nonnegative Matrix Factorization (NMF): Nonnegative Matrix Factorization (NMF) is a low-rank approximation technique for unsupervised multivariate data decomposition, such as PCA and Independent Component Analysis (ICA) (Hyvarinen *et al.*, 2001). NMF for the first time was presented by Lee and Seung (1999), followed by a large body of research published to address the analysis, extension and applications of NMF algorithms in image processing (Lee and Seung, 1999; Xue *et al.*, 2008; Wild *et al.*, 2004; Zheng *et al.*, 2007; Zhao *et al.*, 2008; Albright *et al.*, 2006; Feng *et al.*, 2002; Finesso and Spreij, 2006; Ying and Zhang, 2009; Jeon *et al.*, 2005; Yang and He, 2010), signal processing (Chen and Cichocki, 2005; Cichocki *et al.*, 2006) and data mining (Li *et al.*, 2008; Berry *et al.*, 2006) during the last decade.

The NMF method attempts to find a solution in order to decompose a given non-negative matrix $A \epsilon R^{m \times n}$ into multiplication of two non-negative matrices $w \epsilon R^{m \times k}$ and $H \epsilon R^{k \times n}$ such that these matrices minimize the following criterion:

$$f(W,H) = \frac{1}{2}\|A - WH\|_F^2 \quad (1)$$

where, k<<min(m,n) is a positive integer that determines rank of NMF and F is the abbreviation of Frobenius norm. In some research, another divergence functions are used as cost function that the most famous one is Kullback-Leibler that identifies W and H such that these matrices minimize the following criterion:

$$f_D(W,H) = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij} \log \frac{A_{ij}}{[WH]_{ij}} - A_{ij} + [WH]_{ij} \quad (2)$$

Matrix W behaves similar to eigenvectors in PCA but with the difference that columns of matrix W are not supposed to be essentially diagonal while a non negativity constraint is preserved. In facial expression recognition problem, n is number of images and each column of matrix A is an image with size m = w×h. In other words, the original number of features is m while due to NMF decomposition the feature dimension is reduced to k. Although both PCA and SVD have smaller error in Frobenius norm than NMF (Wild *et al.*, 2004), but NMF has the following benefits in contrast to other methods that make it suitable for rank reduction:

- Elements of matrix W are nonnegative; thus basis column can be visualized (Wild *et al.*, 2004)
- Nonnegative elements of matrix H make a non-subtractive combination of basis components in contrast to other methods that allow additive and subtractive combinations of basis components. Therefore, NMF is a part-based representation but other methods results in a whole-based representation. This is a good characteristic in applicable fields such as image processing. For example, if a part of an image is damaged, defected part affects a small number of features; however, in other methods, it affects almost all of features. In addition, in many applications, we process only a part of image. For instance, in facial expression recognition, face expression is performed by some part of face
- NMF algorithm decomposes a matrix into two sparse matrices W and H which causes reduction in storage (Albright *et al.*, 2006)

Although features that achieved by NMF are part-based, the extracted components are not necessarily localized on a certain part of a face. In this study, we employed an efficient version of NMF that is termed as Local Nonnegative Matrix Factorization (LNMF) introduced by Li *et al.* (2001). LNMF identifies W and H matrices such that they minimize the following criterion:

$$f_{LNMF}(W,H)=\sum_{i=1}^{m}\sum_{j=1}^{n}(A_{ij}\log\frac{A_{ij}}{[WH]_{ij}}-A_{ij}+[WH]_{ij}+\alpha U_{ij})-\beta\sum_{i}V_{ii} \quad (3)$$

where, α, $\beta>0$ are constants and $U = W^TW$ and $V = HH^T$. With this cost function, features of LNMF are localized (Feng *et al*., 2002). Although convergence rate for LNMF is slower than NMF but the achieved features by LNMF are more localized and at the best situation the extracted components cover parts of a face.

Fuzzy C-Means Clustering (FCM): Fuzzy C-Means (FCM) is a clustering method that allows a pattern being belonged to two or more clusters with different degree of membership. FCM proposed by Bezdek (1981) is widely used in pattern recognition. It is based on minimization of the following objective function:

$$J_m=\sum_{i=1}^{N}\sum_{k=1}^{C}u_{ik}^{m}\|x_i-c_k\|^2 \quad (4)$$

where, m is a factor that is any real number higher than 1, u_{ik} is the degree of membership of input x_i in the cluster k, x_i is the i'th of d-dimensional measured data, c_k is the k'th cluster center and $\|*\|$ is any norm expressing the similarity between any measured data and the center. Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership u_{ik} and the cluster centers c_k by:

$$u_{ik}=(\sum_{j=1}^{C}\frac{\|x_i-c_k\|^{\frac{2}{m-1}}}{\|x_i-c_j\|})^{-1} \quad (5)$$

$$c_k=\frac{\sum_{i=1}^{N}u_{ik}^{m}x_i}{\sum_{i=1}^{N}u_{ik}^{m}} \quad (6)$$

This iteration will stop when:

$$\max\{|u_{ij}^{iter+1}-u_{ij}^{iter}|\}<\varepsilon$$

where, ε is a termination criterion ranged between 0 and 1 and iter is the iteration number.

Methodology: Although NMF is an efficient method for feature extraction but it does not lead to a convex solution to determine W and H matrices optimally. Hence, a proper initialization can play a very important role to achieve a better result. The suggested initialization methods mostly try to provide an accurate estimation for W matrix and to determine the H component, a different manner is considered. These updating algorithms in low number of iterations achieve an acceptable result but when the NMF initialization algorithms are allowed to iterate high number of epochs, random initialization is behaved better than all others. As far as NMF is also used as a clustering algorithm for image segmentation and document clustering (Jeon *et al*., 2005; Shahnaz *et al*., 2006); consequently, an initialization based clustering method can enhance its initialization performance (Xue *et al*., 2008; Wild *et al*., 2004; Zheng *et al*., 2007).

Wild *et al.* (2004) used spherical k-means to initialize W matrix and then NNLS algorithm is used for identifying H matrix, this method followed by Xue *et al.* (2008) used divergence-based k-means clustering and after achieving to cluster centers for computing H matrix used from following criterion:

$$H_{ij}=\begin{cases}1 \text{ if } i=\arg\min_i d(A_i-W_i)\\ 0 \text{ else}\end{cases} \text{for } j=1,\ldots,n \quad (7)$$

Zheng *et al*. (2007) also used from fuzzy clustering for achieving W matrix and then for achieving to H matrix used from following criterion:

$$H_{ij}=\begin{cases}1 \text{ if } i=\arg\max_i (U_{ij})\\ 0 \text{ else}\end{cases} \text{for } j=1,\ldots,n \quad (8)$$

As far as these methods emphasis first in computing W matrix and then H matrix; the possibility of falling in local optimum is high for both NMF components. To overcome this drawback, in this study, we stress in achieving W and H matrix simultaneously and we use U as H matrix that causes a better initial point that lead to a higher performance in low iterations but also in high iterations produces a better result than random initialization method.

For the facial expression recognition task, we must first extract discriminative information. To do this, first, pure face are detected and scaled in a fixed geometry size. Then for each image, we apply histogram equalization then each image (matrix A) is decomposed and is normalized in the interval of 0 to 1. Then, FCM is employed and set $W_0 = C$ and $H_0 = U$. Next, we use LNMF algorithm in which H and W are updated according to the following Eq:

$$h_{kl}=\sqrt{h_{kl}\sum_{i}a_{il}\frac{w_{ik}}{\sum_{k}w_{ik}h_{kl}}} \quad (9)$$

$$w_{kl} = \frac{w_{kl} \sum_j a_{kj} \frac{h_{lj}}{\sum_k w_{kl} h_{lj}}}{\sum_j h_{lj}} \quad (10)$$

$$w_{kl} = \frac{w_{kl}}{\sum_k w_{kl}} \quad (11)$$

Experimental results in the next section show that initialization of NMF is improved using the proposed method.

RESULTS AND DISCUSSION

In all experiments JAFFE dataset (http://www.kasrl.org/jaffe download.html) is used containing 213 images include 7 facial expressions consisting 6 basic facial expressions and neutral expression that posed by 10 Japanese female models. After the preprocessing phase (described above), each image has 33×33 size and for all experiments, fuzzifier parameters and rank of NMF are empirically chosen 17 and 1.53, respectively. Moreover, number of iteration for all clustering methods is set to 100.

Comparison of initialization methods in term of relative error: As mentioned in the introduction, different methods have been suggested to find suitable values for initialization of W and H (Xue *et al*., 2008; Wild *et al*., 2004; Zheng *et al*., 2007; Zhao *et al*., 2008). In Table 1, we compared our method to the mentioned methods and the obtained results revealed the supremacy of the proposed method in better initializing than the former method. For comparison, we used from relative error that is the ratio of $\|A\text{-}WH\|^2_f/\|A\|^2_f$ (Wild *et al*., 2004).

Comparison of initialization methods in terms of sparsity and orthogonality: In this experiment, we investigate orthogonality and sparsity of all initialization methods (Fig. 1a, b). The results exhibited that our method outperformed its rival initialization methods and show that in other methods after some iteration orthogonality and sparsity progression are converged to a constant but our method even in large iterations behaved better than random initialization.

As shown in Fig. 1, PCA does not converge to local optimum but results of the proposed method outperformed those achieved by PCA. In this experiment, for sparsity we use from number of elements in W that $W_{ij} < \|W_j\|/256$ and orthogonality of this matrix is $\sum_{i \neq j} w_i^T w_j$ (Wild *et al*., 2004).

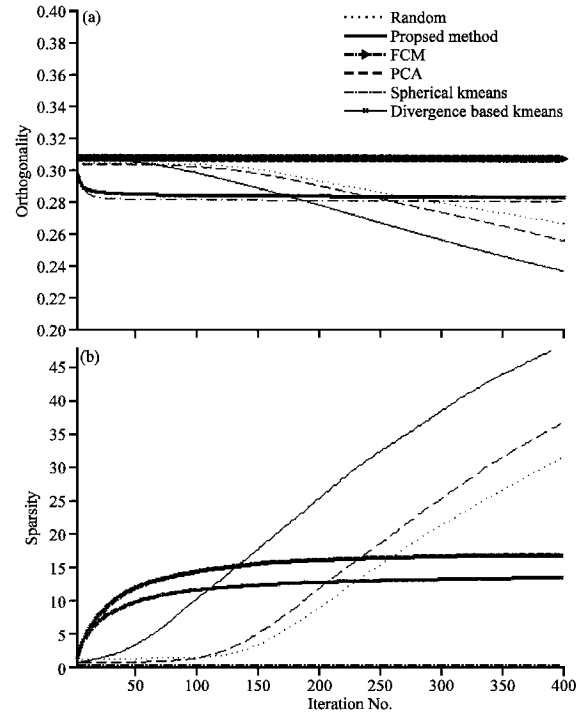

Fig. 1: Comparison of initialization methods in terms of sparsity and orthogonality. (a) Comparison of initialization methods in term of orthogonality and (c) Comparison of initialization methods in term of sparsity

Table 1: Comparison of initialization methods in term of relative error

Initialization methods	Relative error
Random	6.6867
PCA	1.9410
Spherical k-means	0.4520
Divergence-based k-means	0.4488
FCM[5]	0.1922
The proposed method	0.1916

Local optimum: In this experiment we represent the progression of the left factor W in NMF in equal number of iterations using different initialization methods. For this case, we compare our method to NMF with FCM initialization (Zheng *et al*., 2007) and random initialization.

As mentioned in the former section, FCM initialization, after proposed method, resulted in higher performance than other employed initialization methods. As stated before, in LNMF algorithm achieving to local parts of a face is the goal and this experiment showed that basis vectors achieved by FCM (Zheng *et al*., 2007) converges to a local minimum and progression of this method and other initialization methods are very poor (Fig. 2a-e). The highlighted points can be categorized as follows: with our method not only in low number of

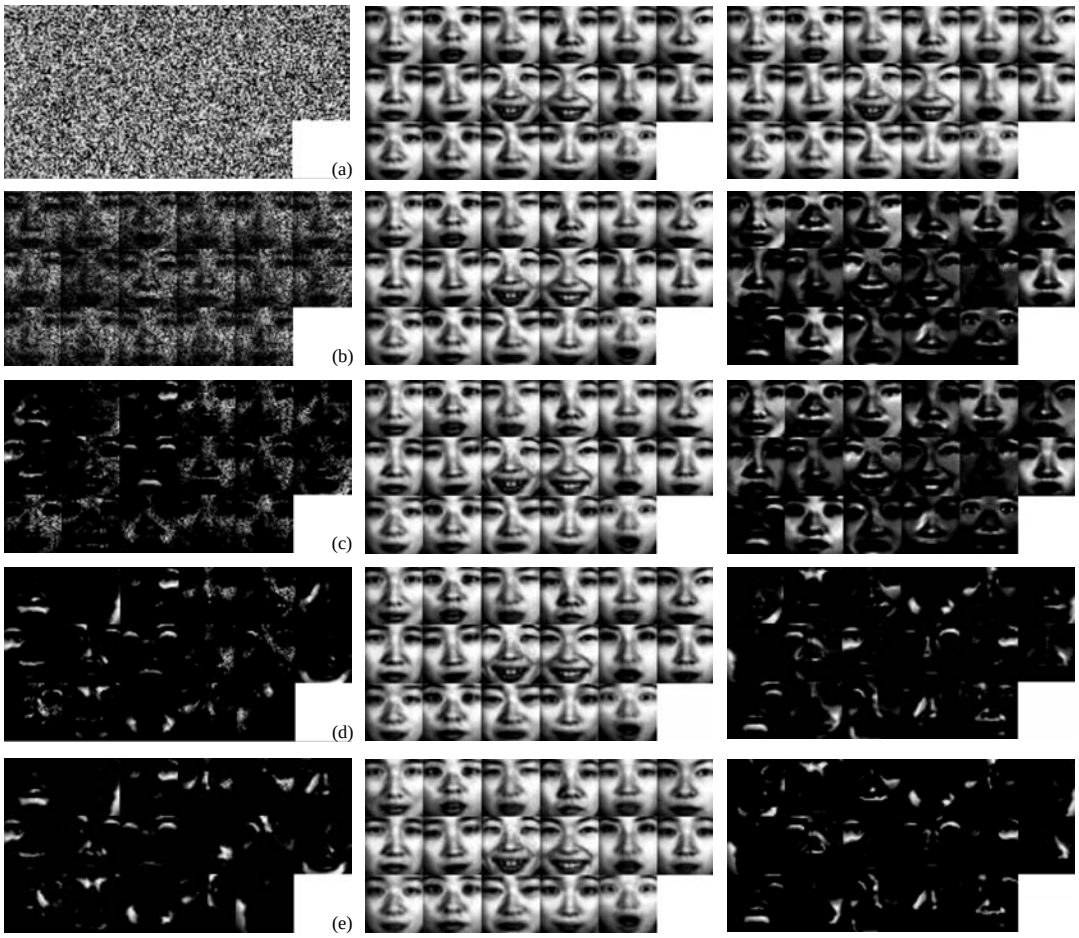

Fig. 2: The progression of basis vector (columns of W matrix), left images achieved basis vectors by proposed method a n d right images are achieved by random initialization and middle are achieved by FCM (Zheng *et al*., 2007) method. (a) Initial Basis vectors, (b) Basis vectors after 100 iterations, (c) Basis vectors after 400 iterations, (d) Basis vectors after 1000 iterations and (e) Basis vectors after 4000 iterations

iterations, face components are better separated than the random initialization, but also in high iterations (4000 iteration), our method achieves to good parts. For a better visualization we use from infinitive norm for face representation.

CONCLUSIONS

NMF is a part based representation that has been applied to many applicable such as dimension reduction, image segmentation, image compression and document clustering. In the last decade, many efforts have been made to improve the NMF efficiency (Xue *et al*., 2008; Wild *et al*., 2004; Zheng *et al*., 2007; Zhao *et al*., 2008; Feng *et al*., 2002; Cichocki *et al*., 2006; Li *et al*., 2001; Finesso and Spreij, 2006; Li *et al*., 2008). The main focus of these algorithms is enhancing initialization of NMF (Xue *et al*., 2008; Wild *et al*., 2004; Zheng *et al*., 2007; Zhao *et al*., 2008) due to its high sensitivity to initial values. As far as the suggested initialization methods converge to local optima, for this reason in this study we used from FCM. Using FCM for initialization, not only we achieve acceptable results but also in high number of iterations, the proposed revealed its superiority compare to the other initialization methods.

ACKNOWLEDGMENT

This research has been supported by Iran Telecommunication Research Center (ITRC) and the grant number is 500/4709.

REFERENCES

Abboud, B., F. Davoine and M. Dang, 2004. Facial expression recognition and synthesis based on an appearance model. Signal Process. Image Commun., 19: 723-740.

Albright, R., J. Cox, D. Duling, A. Langville and C. Meyer, 2006. Algorithms, initializations and convergence for the nonnegative matrix factorization. http://meyer. math.ncsu.edu/Meyer/Abstracts/ InitNMF.html.

Berry, M.W., M. Browne, A.N. Langville, V.P. Pauca and R.J. Plemmons, 2006. Algorithms and applications for approximate nonnegative matrix factorization. Computat. Stat. Data Anal., 52: 155-173.

Bezdek, J.C., 1981. Pattern recognition with fuzzy objective functions algorithms. 1st Edn., Plenum, New York.

Calder, A.J., A.M. Burton, P. Miller, A.W. Young and S. Akamatsu, 2001. A principal component analysis of facial expressions. Vision Res., 41: 1179-1208.

Chen, Z. and A. Cichocki, 2005. Nonnegative matrix factorization with temporal smoothness and/or spatial decorrelation constraints. Technical Report, Laboratory for Advanced Brain Signal Processing. RIKEN.Tokyo, Japan. http://citeseerx.ist.psu.edu/ viewdoc/summary?doi=10.1.1.63.7094.

Cichocki, A., R. Zdunek and S. Amari, 2006. Csiszar's divergences for non-negative matrix factorization: Family of new algorithms. Lecture Notes Comput. Sci., 3889: 32-39.

Feng, T., S.Z. Li, H.Y. Shum and H. Zhang, 2002. Local non-negative matrix factorization as a visual representation. Proceedings of the 2nd International Conference on Development and Learning (DL'02), USA., pp: 178-183.

Finesso, L. and L. Spreij, 2006. Nonnegative matrix factorization and I-divergence alternating minimization. Linear Algebra Appl., 416: 270-287.

Hyvarinen, A., J. Karhunen and E. Oja, 2001. Independent Component Analysis. Wiley Interscience, UK.

Jeon, B.K., Y.B. Jung and K.S. Hong, 2005. Image segmentation by unsupervised sparse clustering. Proceedings of the 7th IEEE Workshop on Application of Computer Vision, Jun. 5-7, Breckenridge, pp: 2-7.

Kotsia, I., S. Zafeiriou and I. Pitas, 2008. Texture and shape information fusion for facial expression and facial action unit recognition. Pattern Recognition, 41: 833-851.

Lawson, C.L. and R.J. Hanson, 1974. Solving Least Squares Problems. Prentice-Hall, Englewood Cliffs, New Jersey.

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

Li, S.Z., X.W. Hou, H.J. Zhang and Q.S. Cheng, 2001. Learning spatially localized, parts-based representation. Proc. IEEE Int. Conf. Comput. Vision Pattern Recognition, 1: 207-212.

Li, Z., H. Peng and X. Wu, 2008. A new descriptive clustering algorithm based on nonnegative matrix factorization. Proceedings of the IEEE International Conference Granular Computing, Aug. 26-28, Hangzhou, pp: 407-412.

Shahnaz, F., M.W. Berry, V.P. Pauca and R.J. Plemmons, 2006. Document clustering using nonnegative matrix factorization. Inform. Process. Manage. Int. J., 42: 373-386.

Shan, C., S. Gong and P.W. McOwan, 2009. Facial expression recognition based on local binary patterns: A comprehensive study. Image Vision Comput., 27: 803-816.

Tsai, P., L. Cao, T. Hintz and T. Jan, 2009. A bi-modal face recognition framework integrating facial expression with facial appearance. Pattern Recognition Lett., 30: 1096-1109.

Wang, J. and L. Yin, 2007. Static topographic modelling for facial expression recognition and analysis. Comput. Vision Image Understand., 108: 19-34.

Wild, S., J. Curry and A. Dougherty, 2004. Improving non-negative matrix factorizations through structured initialization. Pattern Recognition, 37: 2217-2232.

Xue, Y., C. Tong, Y. Chen and W.S. Chen, 2008. Clustering-based initialization for nonnegative matrix factorization. Applied Math. Comput., 205: 525-536.

Yang, H. and G. He, 2010. Online face recognition algorithm via nonnegative matrix factorization. Inform. Technol. J., 9: 1719-1724.

Ying, Z. and G. Zhang, 2009. Facial expression recognition based on NMF and SVM. Proceedings of the IEEE International Conference on Information Technology (IT'09), USA., pp: 458-462.

Zhao, L., G. Zhang and X. Xu, 2008. Facial expression recognition based on PCA and NMF. Proceedings of the 7th World Congress on Intelligent Control and Automation, June 25-27, Chongqing, pp: 6826-6829.

Zheng, Z., J. Yang and Y. Zhu, 2007. Initialization enhancer for non-negative matrix factorization. Eng. Appl. Artif. Intell., 20: 101-110.