American Journal of Engineering Research (AJER)
e-ISSN: 2320-0847 p-ISSN : 2320-0936
Volume-6, Issue-12, pp-419-426
www.ajer.org
Research Paper Open Access

Pattern Recognition using Two-Dimensional Kernel Non-negative Matrix Factorization Algorithm

Wael M. Khedr[1,2]

[1](Computer Science and Information Dept., College of Science, Majmaah University, KSA.)
[2](Mathematics Dept. Faculty of Science, Zagazig University, Egypt.)
Corresponding Author: Wael m. Khedr

ABSTRACT:*Pattern recognition and classification tasks use machine learning that require a suitable representation of the pattern. Non-negative matrix factorization (NMF) has to be a useful decomposition and representation of multivariate data pattern. Two different multiplicative algorithms for NMF are analyzed. In this paper, we extend the original NMF to 2-Dimensional kernel non-negative matrix factorization (2DKNMF) to improve its performance. 2DKNMF algorithm is derived based on nonlinear kernel function to map patterns to feature space and stepwise method is applied on each individual pattern in feature space to avoid complex computations. The advantages of 2DKNMF algorithm are based on extract more useful feature hidden in the original pattern and it can process data with negative values by using some specific distribution kernel functions. Experimental results on several face databases show that 2DFNMF has better image reconstruction quality than NMF. Also the running time of 2DFNMF is less, and the recognition accuracy higher than that of NMF.*

Keywords: *Non-negative matrix factorization (NMF), Kernel NMF, 2DNMF and 2DKNMF.*

Date of Submission: 22-12-2017 Date of acceptance: 03-01-2017

I. INTRODUCTION

Learning Parts of pattern is interested in machine learning, computer vision and pattern recognition [3]. Many parts-based image representation approaches can be ascribed to a general subspace method, which has been successfully used in many high dimensional data analysis applications. Given a class of image patterns, there are many approaches to construct the subspace. One such method is principal component analysis (PCA) [4]. Recently a new subspace method called non-negative matrix factorization (NMF) [1,5] is proposed to learn the parts of objects and images. Linear and unsupervised dimensionality reduction via matrix factorization with non-negativity constraints NMF is studied when applied for feature extraction, followed by pattern recognition classification [2].

Lee and Seung [1] proposed a simple iterative algorithm for NMF and proved its convergence. NMF provides simple learning rule guaranteeing monotonically convergence to a local maximum without the need for setting any adjustable parameters. Disadvantage of NMF, the 2D image matrices must be previously transformed into 1D image vectors that may cause the loss of some structure information hiding in original 2D images. 2-Dimensional non-negative matrix factorization (2DNMF) for representing 2D images with a set of 2D bases. The key difference between 2DNMF and NMF is that the former adopt a novel representation for original images.Kernel NMF [6] can extract more useful features hide in the original data using some kernel-induced nonlinear mapping; it deal with relational data where only the relationships between objects are known; it process data with negative values by using some specific kernel functions (e.g. Gaussian).

Pattern Recognition/classification tasks use machine learning that require a suitable representation of the pattern. Typically, a useful representation can make the latent structure in the data pattern more explicit, and often reduces the dimensionality of the data pattern so that Non-negative matrix factorization (NMF) method can be applied. NMF method represents and stores pattern into one dimension, imposes the non-negativity constraints in its bases and coefficients, and it cannot disclose nonlinear structures hidden in the data pattern.

We propose the two-dimensional kernel NMF (2DKNMF), which can overcome the above limitations of NMF. First, 2DKNMF has better pattern reconstruction quality in two dimension than representation in one dimension by traditional NMF method. Second, through using kernel-induced nonlinear mapping, KNMF could extract more useful features hidden in the original data and can process data with negative values. Thus, 2DKNMF is more general than NMF. The performance 2DKNMF method is issued in the following aspect: the computational costs, enhancing the image reconstruction quality and improving the recognition accuracy with or without occlusions and noises. Evaluate the effect of the number of bases used in NMF, 2DNMF and 2DKNMF for pattern recognition without occlusion and noise.

II. NON-NEGATIVE MATRIX FACTORIZATION (NMF) ALGORITHM

Non-negative matrix factorization is a linear non-negative approximation of multivariate data representation. Given a set of m patterns $\{A_1, A_2, \dots, A_m\}$ each A_i has two dimension $p \times q$ and is first transformed into a one dimension vector. Non-negative matrix $V(n \times m)$ contains all patterns as columns, each column of which contains $n = p \times q$ non-negative values. The matrix $V(n \times m)$ is then approximately factorizedinto two matrices W and H, such that $V \approx W \cdot H$, that is each column v_i of V can be written as linear combination of this set, i.e. $v_i \approx W \cdot h_i$. Where matrix $W(n \times r)$ contains basis vectors w_i and matrix H(r$\times m$) contains the coefficient vector h_i corresponding to vector v_i. Usually r is chosen to be smaller than $norm$ such that $r(n+m) < mn$. To find non-negative matrix factors W and H, we first need to define cost function that can be constructed using some measure of distance between two non-negative matrices V and WH. The conventional approach to find W and H is by minimizing the difference between V and WH [13] by using the Euclidean distance:

$$\min_{W,H} Z(W,H) = \frac{1}{2}\|V - WH\|^2 = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}(V_{ij} - (WH)_{ij})^2 \text{ , w.r.t } W \geq 0, H \geq 0. \quad (1)$$

Another cost function is divergence measure, the divergence $D(A\|B)$ is measured in Eq(2).Like the Euclidean distance this is lower bounded by zero, and vanishes one and only if A=B. But it cannot be called a “distance”, because it is not symmetric in A and B, soit refers to as the “divergence” of A from B. It reduces to the Kullback-Leibler divergence [1].In order to obtain W and H, it should be solve the optimization problem Minimize D (V||WH) w. r. t. W, H ≥ 0.

$$D(A\|B) = \sum_{ij}\left(A_{ij}\log\frac{A_{ij}}{B_{ij}} - A_{ij} + B_{ij}\right) \quad (2)$$

$$\min_{W,H} \sum_{i=1}^{n}\sum_{j=1}^{m}\left(V_{ij}\log\frac{V_{ij}}{(WH)_{ij}} - V_{ij} + (WH)_{ij}\right) \text{, Subject to } W \geq 0, H \geq 0. \quad (3)$$

The multiplicative update rule is given in [1] as follows:

$$W_{ij} = W_{ij}\sum_{k=1}^{m}\frac{V_{ik}}{(WH)_{ik}}H_{jk} \quad (4)$$

$$W_{ij} = \frac{W_{ij}}{\sum_{k=1}^{n} W_{kj}} \quad (5)$$

$$H_{kj} = H_{kj}\sum_{i=1}^{n} W_{ik}\frac{V_{ij}}{(WH)_{ij}} \quad (6)$$

Algorithm (1) is an iterative procedure for computing the bases W and coefficients H as the following.

Algorithm (1):Original NMF algorithm

Input: Data matrix ***V****(n×m)*, each column of which denotes pattern vector, and r rank $w.\ r.\ t.$ $r(n+m) < mn$.
Output: Construct Matrix ***W****(n×r)*and matrix ***H****(r×m)* $s.\ t.\ V \approx W \cdot H$

1. Set initial random values for ***W***and ***H.***
2. Set iteration t←1
3. Whilenot convergent based Equation (1)or(2).
4. Update the bases ***W***using Equations (4 and 5).
5. Update the coefficients ***H***using Equation (6).
6. t ← t +1.
7. End While

III. KERNEL NON-NEGATIVE MATRIX FACTORIZATION (KNMF)

The kernel NMF can extract more useful features hidden in the original data, it can solve the problem of the data where only relationships between patterns are known, and being able to process the datawith negative values by using some specific kernel functions. Before deriving kernel based NMF methods, two important definitions of kernel and kernel matrix are reviewed[6].

A kernel is a nonlinear function inner product that maps origin data from the input space to the feature space. More specifically, a kernel is defined as follows:

$$k(x,y) = \langle \emptyset(x), \emptyset(y) \rangle = (\emptyset(x))^t \cdot \emptyset(y) (7)$$

Where $\emptyset$ is a mappingfrom input **x** to a feature space F, $\emptyset : x \to \emptyset(x) \in F$. Here, $\langle \cdot,\cdot \rangle$denotes the inner product.The *kernel matrix* can measure the pair wise relationship between them and gives the similarity measures. Given a matrix $\mathbf{V}(n \times m) = \{v_1, v_2, \ldots, v_m\}$ and a kernel function **K**. The elements $\boldsymbol{K_{ij}}$ of kernel matrix of **V** is computed based on Eq.(7) so that $\boldsymbol{K}_{ij} = k(v_i, v_j) = \langle \emptyset(v), \emptyset(v) \rangle = (\emptyset(\boldsymbol{V}))^T \cdot \emptyset(\boldsymbol{V})$ and $\boldsymbol{K} \triangleq \{\boldsymbol{K_{ij}}\}^{\boldsymbol{m}}{}_{\boldsymbol{i,j=1}}$. Zhang *et al.*[7]extended the original NMF to kernel NMF (KNMF) in 2006.The nonlinear kernel mapping from the original input space to a high dimensional feature space and enables the data to representtheir correlation in the high dimensional space. Assume that the input observations are represented as matrix $\mathbf{X}(n \times m) = [x_1, x_2, \ldots, x_m]$.Let $\emptyset$ bean implicit nonlinear mapping from the original input space to ahigh dimensional feature space,$\emptyset(\mathbf{X}) = [\emptyset(x_1), \emptyset(x_2), \ldots, \emptyset(x_m)]$. So $\emptyset(\boldsymbol{V})$ is decomposed into twononnegative factors $\emptyset(\mathbf{W})$ and **H** , such that:

$$(\emptyset(V))^T \emptyset(V) = (\emptyset(V))^T W_\emptyset \cdot H \quad (9)$$

Since $K(V,V) = (\emptyset(V))^T \emptyset(V)$ and $K(V,W) = (\emptyset(V))^T W_\emptyset$ then Eq.(9) become K(V,V) = K(V,W) H. Upon equations (7, 8) and let $Y = (\emptyset(V))^T W_\emptyset$, Eq. (9) can be rewrite as $K = Y \cdot H$, where$\boldsymbol{W}_\emptyset$ is the learned bases of $\emptyset(\boldsymbol{V})$ in feature space F and $\boldsymbol{H}$ is its combining coefficients, each column of which denotes now the dimension-reduced representation for the corresponding object.

a.The Polynomial Kernel

Polynomial function can extract more useful feature hidden in the original pattern and it can process data with negative values. Polynomial function reduces the statistics and the redundancy of the representation of patterns. Buciu et al.[8] applied the polynomial functions that frequently used kernel functions in feature space. The polynomial function of input data x of d- dimension is defined in Eq. (10).

$$\boldsymbol{K}_{ij} = k(x_i, x_j) = (x_i \cdot x_j)^d \quad (10)$$

Let input data $\boldsymbol{X} \in V \subseteq \mathbb{R}^{n \times m}$ are transformed to the higher dimensional space $\mathcal{F} \subseteq \mathbb{R}^{\ell \times m}$, $\ell \gg n$. The transformed input data $F = [\emptyset(x_1), \emptyset(x_2), \ldots, \emptyset(x_m)]$where $\emptyset(X_i) = [\emptyset(X)_1, \emptyset(X)_2, \ldots, \emptyset(X)_\ell]^T \in \mathcal{F}$ and ℓ is dimensional of input vector.The cost function is shown as in Eq.(11).

$$\min_{W,H} Z(W,H) = \frac{1}{2} \|\emptyset(V) - \emptyset(W)H\|^2 \; s.t. W \geq 0 \; , H \geq 0 \quad (11)$$

Where $\emptyset(W) = [\emptyset(w_1), \emptyset(w_2), \ldots, \emptyset(w_r)]$ and vectors w_i arecalled the pre-images of the basis. Multiplicative update rules of the KNMF are obtained by a gradient descent optimizationprocedure, the kernel's gradient is given byEq.(12).

$$\frac{\partial Z}{\partial H} = (\emptyset(V) - \emptyset(W)H)\emptyset(W)^T = (\emptyset(V)\emptyset(W)^T) - (\emptyset(W)\emptyset(W)^T H) = K_{wv} - HK_{ww} \quad (12)$$

$$\therefore \Delta H \triangleq \frac{K_{wv}}{K_{ww}} \quad (13)$$

$$H(t+1) = H(t) \frac{K_{wv}}{K_{ww} H(t)} \quad (14)$$

$$\frac{\partial Z}{\partial W} = (\emptyset(V) - \emptyset(W)H)\,(VK'_{Vw} - K'_{ww} H) \quad (15)$$

$$W(t+1) = W(t) \frac{VK'_{vw}}{W(t)\Omega K'_{ww}} \quad (16)$$

$$W(t+1) = \frac{W(t+1)}{\sum_{k=1}^{n} W_{kj}} \quad (17)$$

Where,$K_{wv} = \langle \phi(w_i), \phi(v_j) \rangle$, $K_{vw} = \langle \phi(v_j), \phi(w_i) \rangle$, $K_{ww} = \langle \phi(w_i), \phi(w_j) \rangle$and thekernel matrix$\boldsymbol{K} = \langle \phi(v_i), \phi(v_j) \rangle = (v_i \cdot v_j)^d$, the derivative of $\boldsymbol{K}$ **is** $\boldsymbol{K}' = d(v_i \cdot v_j)^{d-1} = d \cdot \boldsymbol{K}^{d-1}$ for polynomial kernel. $\boldsymbol{\Omega}$is a diagonal matrix of $\boldsymbol{H}$whose elements are $\sum_{j=1}^{m} H_{ij}$, i=1,…,r. Equation(14) normalizes the basis matrix such as $w \in [0,1]$.

b.The Gaussian kernel

The Gaussian kernel function is define by$\boldsymbol{K}_{ij} = k(x_i, x_j) = exp\left(-\|x_i - x_j\|^2 / (2\sigma^2)\right)$ that empower to measure the similarity of xi and xj with the gradient$\nabla K_{ij} = -\frac{1}{2\sigma^2} K_{ij} (x_i - x_j)$. The update rules of matrices $H$and$W$ can be easily derived as in polynomial kernel [6].

IV. TOW DIMENSION KERNEL NMF (2DKNMF) METHOD

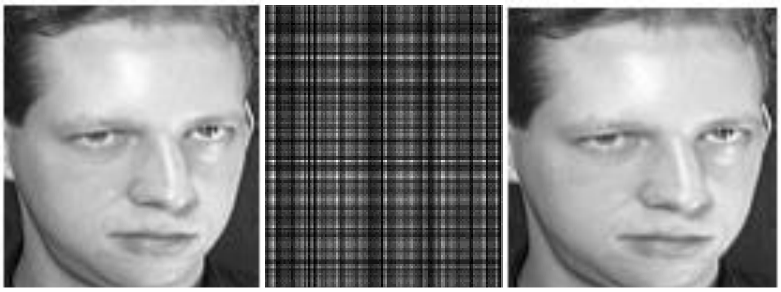
In origin NMF, a two dimensional pattern A(p×q) is first transformed into a one dimensional vector and then all the patterns are represented into a matrix V(n×m), each column of which contains n=pq nonnegative values of one of the m patterns [11]. The procedure of 2DKNMF method consists of four successive steps as followed.

First Step: Wealign the m training patterns into a matrix **V**(p×qm) = $[A_1, A_2, \dots, A_m]$, where each $A_k$denotes one of the m patterns. The procedure 2DKNMF finds nonnegative matrix W(n×r)and non-negative matrix H(r×qm). We performa kernel nonlinear function**K** on all columns patterns to construct$\emptyset(A_k)$.

Second Step:Apply KNMF algorithm based on a stepwise method individual pattern $\emptyset(A_k)$ with decomposition dimension **r** such that $\emptyset(V) \approx \emptyset(W) \cdot H$.

Third Step: Apply KNMF algorithm again on all transposed pattern$\emptyset(V)^T$with dimension s to decomposeinto matrix R(qm×s)row bases and coefficients matrix C(s×qm) , s. t.$\emptyset(V)^T \approx \emptyset(R)^T \cdot C$.

Fourth Step: Construct feature space $\{D_k\}_{k=1}^{m}$by project each pattern $A_k$onto bases matrices W and R such that: $D_k = [\emptyset(W)]^T A_k \emptyset(R_k)$. So, each feature matrix D_k will be of size ($r \times s$). Figure 4.1 shows the reconstruction of origin image from its feature extraction matrix D. The quality of the reconstructed image is measured using the peak signal-to-noise ratio (PSNR).



A B C

Figure (4.1): (a) Originimage(92 ×112) ,(b) feature extraction by 2DNMF D(70 ×70) and (c) image reconstructed from feature extraction of pattern image with PSNR=-43.0418.

A. MULTIPLICATIVE RULE BASED ON STEPWISE METHOD

In order to avoid complex computations Stepwise method divides Hinto $md\times q$sub-matrices as $H=[H_1, H_2, \dots, H_m]$, where H_k denotes the coefficients of the patterns $\emptyset(A_k)$. Since each column of $\emptyset(V)$corresponds to a column of original patterns, we also call as $\emptyset(W)$*column bases*. Thus the *k*-th pattern patterns$\emptyset(A_k)$canbe written as a weighted sum of the column bases Has follows:

$$\emptyset(A_k) \approx \emptyset(W)H_k \quad , k=1, 2, \dots, m \tag{18}$$

$$H_k(t+1) = H_k(t)\frac{K_{wv}}{K_{ww}\, H_k(t)} \tag{19}$$

$$W_k(t+1) = W_k(t)\frac{\emptyset(A_k)K'_{vw}}{W_k(t)\Omega_k K'_{ww}} \tag{20}$$

$$W(t+1) = \frac{W(t+1)}{\sum_{k=1}^{n} W_{kj}} \tag{21}$$

Where Ω_k *is a diagonal matrix whose diagonal element is* $\Omega_{ii} = \sum_i H_{ki}$, the kernel matrices $K_{wv} = \emptyset(W)^T \times \emptyset(V)$, $K_{vw} = K_{wv}^T$, $K_{ww} = \emptyset(W)^T \cdot \emptyset(W)$ and the notation K'_{vw} denote the derivative of kernel matrix K_{vw} with respected to the kernel function type.

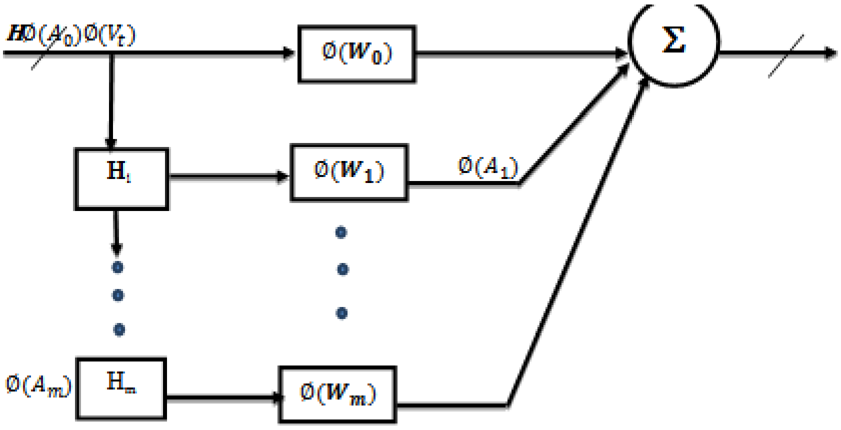


Figure 4.2: Illustration of stepwise method of 2DKNMF

B. Face Recognition using 2DKNMF Method

The2DKNMF method is used to construct feature space $\{D_k\}_{k=1}^{m}$ of dataset$\{A_k\}_{k=1}^{m}$by project each pattern A_k into two bases matrices. The feature space of training images are then used as prototypes. Construct C-centroids of the clustering based on feature space. A test face image $\boldsymbol{A}$ to be classified is represented by its projection onto the feature space as $D = [\emptyset(W)]^T A\emptyset(R_k)$ and then calculate the distance based on norm distance between tested image and each prototype as follows: $dist(D, D_k) = \|D - \bar{D}_k\|$. The tested image is classified to the class to which the closest prototype belongs.

Algorithm (2):2DKNMF Algorithm
Training Procedure**:** **Input:**$p\times q$matrices$\{A_k\}_{k=1}^{m}$, and rank r $w.r.t.$ $d(n+m) < mn$. **Output:**Construct $p\times r$ *column bases* W and m $r\times q$sub-matrices as $\{H_k\}_{k=1}^{m}$ $s.t.\ V \approx W \cdot H$ 1. Align the m training images into a $p\times q$matrix V=$[A_1, A_2, \dots, A_m]$, 2. Initiate all matrices *column bases* W andcoefficients matrices H= $[H_1, H_2, \dots, H_m]$with rank r. 3. Perform polynomial kernel function as $\emptyset(V) = \emptyset\{A_1, A_2, \dots, A_m\}$to constructmatrices$\{\emptyset(A_k)\}_{k=1}^{m}$. 4. Set Iteration t ←1 5. While not convergent 6. For k = 1 to m 7. Perform Stepwise method for each matrix of $\{\emptyset(A_k)\}_{k=1}^{m}$ 8. UpdateH_k using $H_k(t+1) = H_k(t)\frac{K_{wv}}{K_{ww}H_k(t)}$ 9. UpdateW_k using $W_k(t+1) = W_k(t)\frac{\emptyset(A_k)K'_{vw}}{W_k(t)\Omega_k K'_{ww}}$ 10. End For 11. Accumulate matrix H *from* $\{H_k\}_{k=1}^{m}$. 12. Compute $E_t = \{W \times H\}_t$ 13. Compute the$Divergence(V\|E_t)$ 14. t ← t +1. 15. End While
Testing Procedure: **For Pattern Recognition**

Assume matrix H contains C-class features which the columns of H are divided into C- blocks, namely $H = \{H_k\}_{k=1}^{C}$

Step (1) :Compute the pseudo inverse of matrixK_{ww} and the mean vector$\bar{h}_k$ of H_k where $k = 1, 2, \ldots, C$.

Step 2 : For the testing pattern Y :

- Compute kernel matrix$K_{wy} = \phi(W)^T \phi(Y)$.
- Extract features of testing patternY by$h = K_{ww} \cdot K_{wy}$.
- Compute norm distances between**h** and each$\bar{h}_k$ by $d_k = \|h - \bar{h}_k\| \forall k = 1,2,\ldots,C$.

Step 3: Assign test pattern Y to K-class such that Minimum $\{d_k\}_{k=1}^{C}$

V. EXPERIMENTS AND RESULTS

In this section, we compare three methods traditional NMF, kernel NMF and 2DKNMF. ORL database [10] contains a set of faces taken between April 1992 and April 1994 at the Olivetti Research Laboratory in Cambridge, UK. There are 10 different images of 40 distinct subjects. For some of the subjects, the images were taken at different times, varying lighting slightly, facial expressions (open/closed eyes, smiling/non-smiling) and facial details (glasses/no-glasses). All the images are taken against a dark homogeneous background and the subjects are in up-right, frontal position (with tolerance for some side movement). The size of each image is 92×112, 8-bit grey levels.

Feature Extraction by traditional NMF method the final Residual is 2.5941e+03

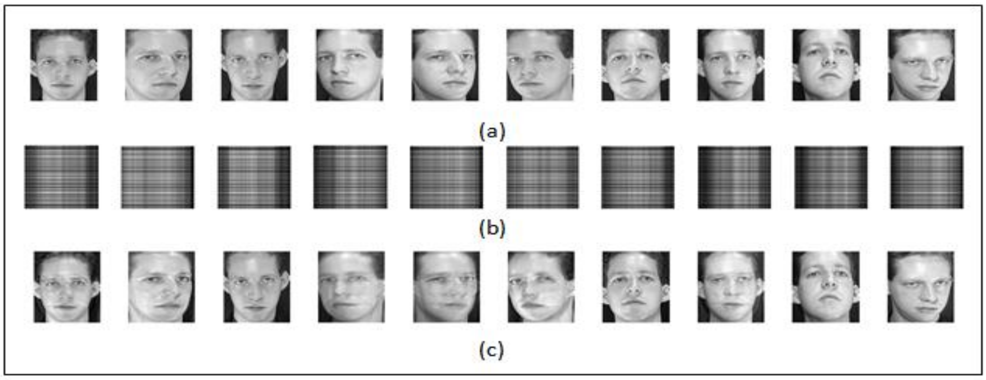


Figure (5.1): (a) Originimage(92 ×112) ,(b) feature extraction by NMF D(70 ×70) and (c) image reconstructed from feature extraction of pattern image with PSNR=-43.0418.

Feature Extraction by 2DKNMF method based on Gaussian kernel function (rbf ,σ=1). The Final Residual is0.4535941e+02.

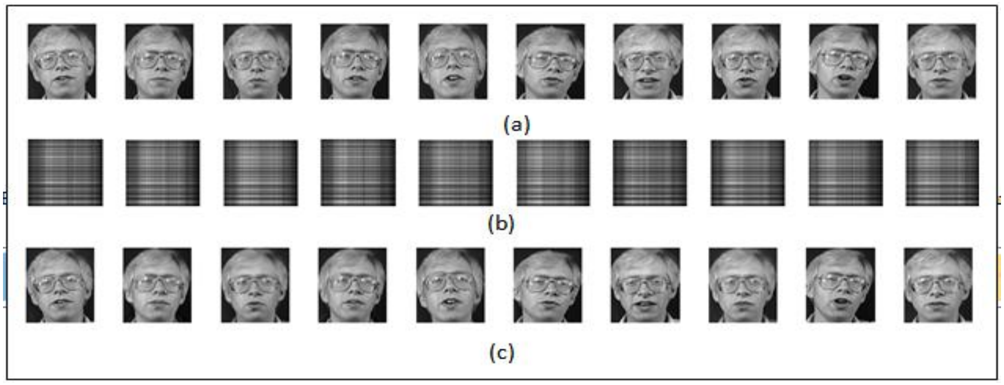


Figure (5.2): (a) Originimage(92 ×112) ,(b) feature extraction by 2DNMF D(70 ×70) and (c) image reconstructed from feature extraction of pattern image with PSNR=-23.68.

2DFNMF Method: Extract Feature space $\{D_k\}_{k=1}^{m}$ of 25training dataset images has been done by the following steps.

Step(1) : $\emptyset(V) = W \times H$

Step(2): $\emptyset(V)^T = R \times C$

Step(3): $D_k = W_k^T \times V_k^T \times R$

The results of training processing have been finished after 300iterations The final residual is 3.4641e+00. This is shown in figs.(5.3, 5.4).

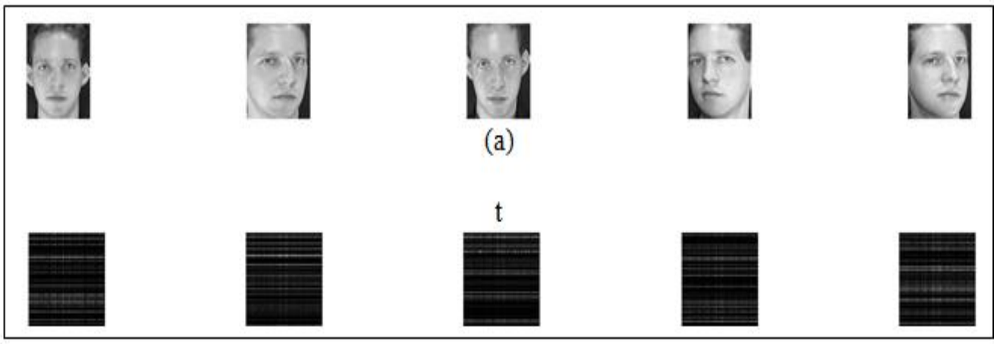


Figure (5.3): (a) Originimage (92 ×112) and (b) Feature extraction D(70 ×70)by2DFNMF method .

Extract Feature space $\{D_k\}_{k=1}^{m}$ of 25 training dataset images is shown in fig.(5.4). Where each image is transformed into matrix D with decompose r=70.

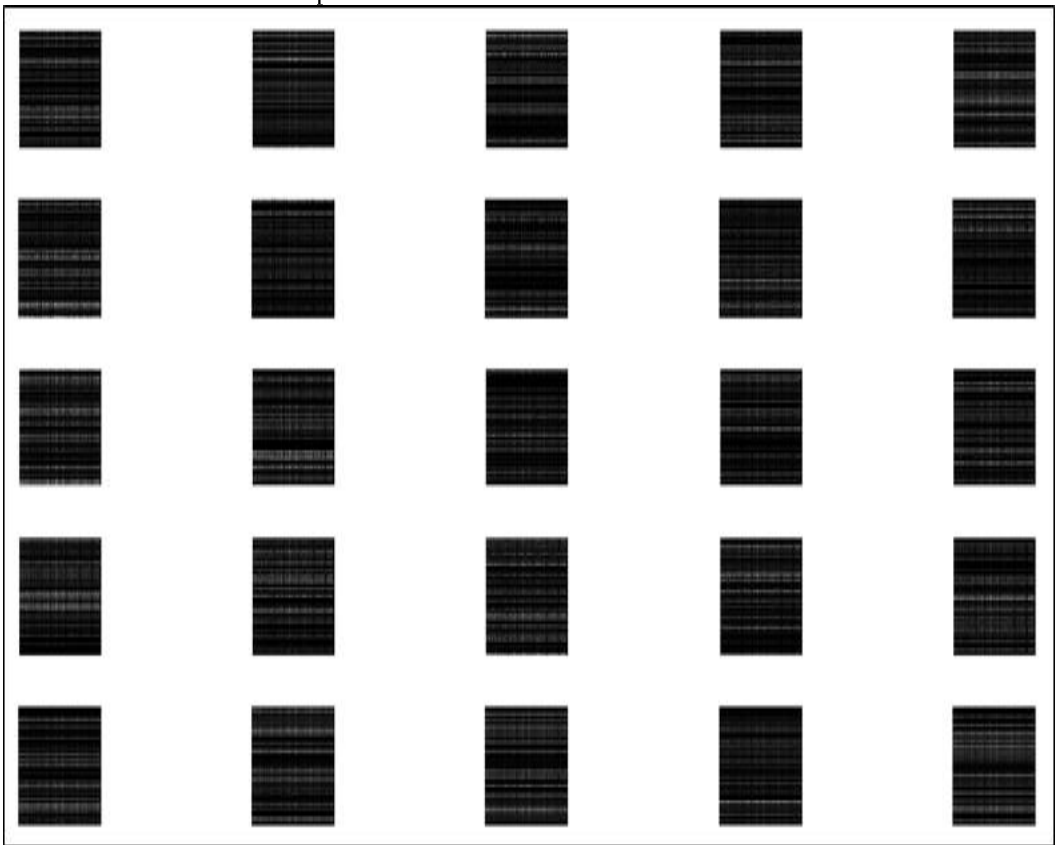


Figure (5.4): Feature extraction D(70 ×70) is computed by 2DFNMF method.

Fig.5.6 shows the comparisons of performances of NMF, 2DNMF and 2DKNMF under number of different bases dimension using PSNR.Experimental results on two face databases convince our claim that 2DKNMF improves 2DNMF.

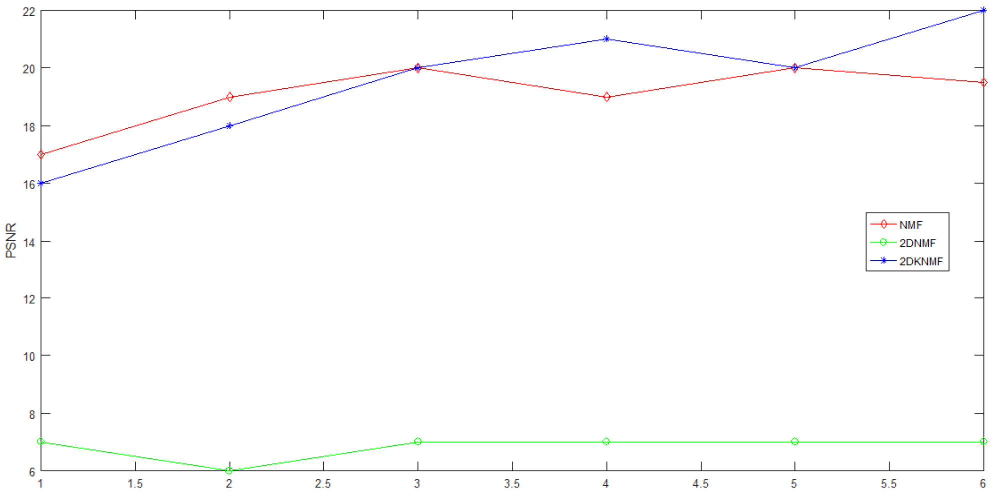


Figure (5.6): Comparisons of performances of NMF, 2DNMF and 2DKNMF under number of different bases dimension using PSNR.

VI. CONCLUSION

In this paper, we have proposed an extend NMF method, 2-D kernel non-negative matrix factorization(2DKNMF), for face representation and recognition. This work is aimed to improve theperformance of traditional NMF in the followingaspect: reducing the computational costs, enhancing the image reconstruction qualityand improving the recognition accuracy with or without occlusions and noises. Weachieved our goal through using a novel image representation method, i.e. using 2Dbases instead of traditional 1D bases.Experimental results on two face databasesconvince our claim that 2DKNMF improves 2DNMF on the above three aspects.

VII. ACKNOWLEDGEMENT

Sincere gratitude is expressed to Deanship of Scientific Research, Al Majmaah University, Saudi Arabia for financial support, fruitful cooperation and continuous help throughout this work.

REFERENCES

[1]. D.D. Lee and H.S. Seung. Learning the parts of objects by non-negative matrix factorization. Nature, 401(6755): 1999, 788–791.
[2]. I. Buciu and I. Pitas. Application of non-negative and local non-negative matrix factorization to facial expression recognition. In Proc. 17th Int. Conf. on Pattern Recognition, Cambridge, UK, 2004 ,pages 288–291.
[3]. Li, S.Z., Hou, X.W., Zhang, H.J., Cheng, Q.S.: Learning spatially localized, parts-based representation. In: CVPR, Hawaii, 2001.
[4]. Jolliffe, I.T.: Principal Component Analysis. Springer-Verlag, New York, 1986
[5]. W. Khedr and M. E. Abd El Aziz Signal Separation using Non-negative Matrix Factorization Based on R1-norm, Life Science Journal;9(4) , 2012,ppt. 703-707 .
[6]. HyekyoungLee, AndrzejCichocki ,SeungjinChoi Kernel nonnegative matrix factorization for spectral EEG feature extraction,Neurocomputing 72 (2009) 3182–3190.
[7]. D. Zhang, Z. H. Zhou, and S. Chen, "Non-negative matrix factorization on kernels," Trends in Artificial Intelligence, Aug. 2006, pp. 404-412.
[8]. Lee D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization.Nature 401(1999) 788-791
[9]. Buciu, N. Nikolaidis, and I. Pitas, "Nonnegative matrix factorization in polynomial feature space," IEEE Transactions on Neural Networks, vol. 19,Apr. 2008, pp. 1090- 1100.
[10]. F. Samaria and A. Harter "Parameterisation of a stochastic model for human face identification" 2nd IEEE Workshop on Applications of Computer Vision December 1994, Sarasota (Florida).
[11]. Chen, Sonqcan; Zhi-Hua Zhou "Two-Dimensional Non-negative Matrix Factorization for Face Representation and Recognition", Analysis and Modeling of faces and Gestures, 2005, pp350-363.

Wael M. Khedr American Journal of Engineering Research (AJER). "Pattern Recognition using Two-Dimensional Kernel Non-Negative Matrix Factorization Algorithm ." American Journal of Engineering Research (AJER), vol. 06, no. 12, 2017, pp. 419-426.