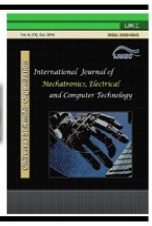
Vol. 6(20), Apr. 2016, PP. 2878-2889

Contents list available at IJMEC

International Journal of Mechatronics, Electrical and Computer Technology (IJMEC)

Journal Homepage: www.aeuso.org

PISSN: 2411-6173 - EISSN: 2305-0543



CRF-based Initialization for Nonnegative Matrix Factorization

Masoumeh Rezaei[1*], Reza Boostani[2] and Zohreh Azimifar[3]

[1] *Faculty of Computer Engineering, University of Sistan and Baluchestan, Zahedan, Iran*

[2,3]*Faculty of Computer Engineering, University of Shiraz, Shiraz, Iran*

Phone Number: +98-54-31132839

***Corresponding Author's E-mail:** mrezaei@ece.usb.ac.ir*

Abstract

NMF algorithms have been recently employed in several applications; however, the performance of NMF is highly dependent on three factors including: 1) choosing a suitable cost function, 2) using an effective initialization method and 3) determining the rank of factorized matrices. This paper is aimed at enhancing the NMF performance using CRF as an efficient initialization method for estimating initial components of NMF in addition to find the proper rank of decomposed matrices. The modified NMF is applied to JAFFE facial expression dataset and experimental results demonstrate the superiority of the proposed approach to NMF with different initialization schemes, in terms of relative error, robustness, sparsity and orthogonality.

Keywords: *Nonnegative Matrix Factorization, Conditional Random Fields, Initialization*

1. Introduction

Facial expressions play an important role in face-to-face human communications. Researches have empirically shown that during human communication, information transmitted by language, tone, and expressions account for 7%, 38%, and 55%, respectively. This statistic exhibits the importance of facial expressions in mutual understanding [34]. Facial expression recognition is widely used in actor-training-software to train inexperienced actors by giving them a feedback about how nice they can mimic different facial states or gesture such as fear, happiness, astonishment, neutral, sadness, etc. The mentioned face expressions are produced by certain voluntary contractions of the face muscles. Therefore, each face expression is considered as an information class that conveys a certain non-verbal message to the audience. To design a facial expression recognition system, a standard dataset should be provided, and discriminative component-based features should be elicited from the images for quantitatively describing the state of the face components such as the lips, eyes and eyebrows, and also the flexion or extension of the face muscles. Many papers have addressed the problem of facial expression recognition. This includes methods that use Local Binary Patterns (LBP) [27], static topographic modelling [30], texture and shape information fusion [16], Principal Component Analysis (PCA) [5], appearance modelling [1] and integration of facial expression with facial appearance models [28]. Recently, Non-negative Matrix Factorization (NMF) has been repeatedly utilized to analyse and recognize the facial expression state [21, 33, 34]. This algorithm not only preserves the part-based representation of the original image, but also guarantees nonnegative results of low-dimensional basis (W) and its corresponding weights (H). As, a few face elements are effected through each feeling state, part-based decomposition is naturally matched to analyse the facial expression. This decomposition provides discriminant features to enhance the recognition performance. In general, NMF algorithms are divided into three general classes: multiplicative algorithms, gradient descent algorithms, and alternating least square algorithms [3]. These updating algorithms iteratively find more suitable values

Article History:
Received Date: Aug. 11, 2015
Accepted Date: Feb. 26, 2016
Available Online: Apr. 23, 2016

for the *W* and *H* matrices and are terminated when the approximate equality of $A \approx WH$ with an acceptable error is satisfied. If the threshold error is chosen very low, the updating process should be terminated after passing a certain number of iteration (e.g. 3000 epochs). To improve the NMF performance, a few attempts have been made to design a problem-dependent cost function [8, 10]. The main challenge of these approaches is that the suggested cost functions are not necessarily convex; therefore, obtaining the optimal solution is impossible in a closed mathematical form. From another angle, designing a convex objective function that satisfies all constraints is a complicated task. Another approach to enhance the NMF performance is to find an initialization method in order to enable the NMF components starting from a nearly optimal point in the search space and converge to an acceptable solution in low number of iteration [4, 31, 32, 34].

In this paper, we focus on modifying the NMF performance in terms of achieving fast convergence and low relative error, increasing sparsity and orthogonality by exploiting an efficient initialization method. Due to the non-convex cost function of the standard NMF, there is no guarantee that NMF factors (*W* and *H* matrices) are optimally determined. On the other hand, NMF can provide different results corresponding to different initial values. Therefore, if the initialization values are chosen properly, there is a hope that the results would be nearly optimum in low number of epochs. Achieving fast convergence along with an acceptable error is quite suitable for real time applications.

In the last decade, several strategies have been suggested to develop an efficient initialization method [4, 31, 32, 34]. In this regard, PCA is deployed to decompose an image matrix into its eigenvectors and eigenvalues and removed those vectors corresponding to the smaller eigenvalues [34]. Finally, they assigned the remained positive eigenvectors (negative values were set to zero) corresponding to the selected eigenvalues, onto the matrices *W* and *H*, respectively. Boutsidis and Gallopoulos [4] suggested Singular Value Decomposition (SVD) to initialize NMF components under circumstances of facing with the Small-Sample-Size (SSS) problem. In another attempt two different criteria are considered to update the NMF component such that spherical *k*-means is used for initialization of matrix *W,* while Non-Negative Least Square (NNLS) algorithm is utilized to calculate the *H* matrix [31]. In a similar study, divergence-based *k*-means clustering is utilized for initialization of *W* component. For determining matrix *H*, at each column, the corresponding cluster number of each sample (image) is considered one and the other elements are set to zero [32].

As all of the mentioned approaches have their own bottlenecks, initialization of NMF components is still an open problem for research [4, 31, 32, 34]. Consequently, various attempts are still made to propose a robust method for initialization of NMF components. In order to address this problem, here, unsupervised Conditional Random Fields (CRFs) as one of the state-of-art statistical models, which has a good clustering capability, is exploited to to improve the NMF performance. Deploying this method not only leads to a faster convergence of *W* and *H* matrices but also results in a higher performance in terms of relative error, sparsity, and orthogonality. As far as CRF-based technique are successfully employed in image segmentation [11, 23], the main contribution of this research is to initialize the NMF components (*W* and *H* matrices) using unsupervised version of CRF method. Another advantage of unsupervised CRF is the relaxation of independence assumptions and providing better clustering result [22].

The rest of this paper is organized as follows. Section 2, describes the Methodology. Section 3, describes the employed dataset, pre-processing stage and evaluating criteria. Section 4, illustrates the experimental results and discusses the advantages and disadvantages of the proposed method compared to the traditional initialization methods. Finally, the paper is concluded in Section 5.

2. Methodology

In this section, the conventional methods along with the proposed method which have been implemented in this research are illustrated. In the next part, NMF method is briefly discussed, then,

the local version of NMF (LNMF) method is expressed. Next, CRF method is explained and finally the proposed CRF-NMF method is introduced.

2.1. Nonnegative Matrix Factorization(NMF)

Non-negative Matrix Factorization (NMF) is a low-rank approximation technique for unsupervised multivariate data decomposition, such as PCA and Independent Component Analysis (ICA). These methods have different constraints and interpretations [12]. NMF was firstly presented by Lee and Seung [21]. A large body of research has been published to analyse the extensions and applications of the NMF algorithm in image processing [4, 9, 10, 13, 19, 21, 31, 32, 33, 34], signal processing [6, 8, 14, 17], and data mining [3, 25, 26] during the last decade.

NMF attempts to decompose a given non-negative data matrix (e.g. Image, document) $A \in R^{m\times n}$ into a multiplication of two non-negative matrices $W \in R^{m\times k}$ and $H \in R^{k\times n}$ such that these matrices minimize the following criterion:

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

where *k*<<min(*m*,*n*) is a positive integer that determines rank of NMF and *F* is the Frobenius norm. In some studies, different similarity metrics are used as the cost function [8, 10]. To some extent, matrix *W* plays the role of eigenvectors in PCA algorithm with difference that columns of matrix *W* contain non-negative values and not supposed to be essentially diagonal. For the case of facial expression recognition, *n* is the number of images and each column of matrix $A_{m\times n}$ (all images) shows an image. In other words, the original feature dimension is *m* while by applying NMF decomposition to the matrix *A*, the number of features is reduced to *k*.

Although both PCA and SVD represent input data with smaller error in Frobenius norm than NMF [7, 31], NMF has the following benefits in comparison with the other factorization methods such as PCA, SVD, ICA, QR [2] and LU [2]. These benefits makes it suitable for rank reduction in specific applications such as data mining and image processing.

- Elements of matrix *W* are nonnegative; consequently, basis columns can be visualized [31].
- Nonnegative elements of matrix *H* make a non-subtractive combination of basis components compared to the other methods that allow additive and subtractive combinations of basis components. Therefore, NMF is a part-based representation while other methods are whole-based representation. It provides a good property in applicable fields such as image and document processing. For example, if a part of image is damaged, the defected part affects just small number of features; however, in other methods it affects almost all features. In addition, in most applications only a part of image is processed. For instance, in facial expression recognition, face expression is shown by a few parts of a face.

- NMF causes a sparse representation of input data in terms of *W* and *H* matrices leading to a drastic reduction in data storage [19].

2.2. LNMF

Although features achieved by NMF are part based, they are not fully localized. To solve this problem, local version of NMF (LNMF) is used that is introduced by Li *et al.* [24]. LNMF identifies *W* and *H* components such that these matrices 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 f_{LNMF} is LNMF cost function and $\alpha,\beta>0$ are constants and $U=W^TW$ and $V=HH^T$. Using this cost function, LNMF features are more localized (proof see [9]). Although rate of convergence for LNMF is slower than NMF, features achieved by LNMF are highly localized. In the best situation, these features are real parts of the face. Main components of LNMF algorithm, *W* and *H*, are updated according to the following formula:

$$h_{kl}=\sqrt{h_{kl}\sum_{i}a_{il}\frac{w_{ik}}{\sum_{k}w_{ik}h_{kl}}} \quad (4)$$

$$w_{kl}=\frac{w_{kl}\sum_{j}a_{kj}\frac{h_{lj}}{\sum_{k}w_{kl}h_{lj}}}{\sum_{j}h_{lj}}$$

$$w_{kl}=\frac{w_{kl}}{\sum_{k}w_{kl}}$$

2.3. Conditional Random Fields (CRFs)

In general, probabilistic models are designed by the two following approaches: generative and discriminative models. Generative models utilize the joint probability between their inputs and outputs (labels) such as Naive Bayese (NB) and Hidden Markov Model (HMM). In both mentioned models, estimation of the joint probability distribution is needed a high computational burden [15]. To reduce the computational complexity, the second approach is presented based on conditional probability which is optimized by some criteria like Maximum Entropy (ME) and Conditional Random Fields (CRFs). Similar to Naïve Bayes, output variable of ME-based models is a single variable while CRF assigns a label vector to an input pattern by selecting the label sequence that maximizes the conditional probability; consequently, CRF can be understood as a sequential extension to the ME model known as discriminative models [15]. Fig, 1, illustrates the relation among the mentioned models.

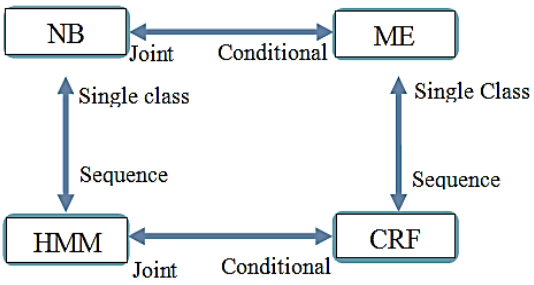


Figure 1: Relationship between Naïve bays (NB), Maximum Entropy (ME), Conditional Random Fields (CRF) and Hidden Markov Model (HMM).

Considering advantages of discriminative models, in this paper, unsupervised CRF [22] is employed. In this case, $X = \{X_1, X_2, ..., X_n\}$ represent input data, $Y = \{Y_1, Y_2, ..., Y_n\}$ is the set of labels, and *G= (V, E).* A CRF model can be expressed as a conditional probability function of *Y* given *X*:

$$p(Y \mid X) = p(X \mid Y)p(X) \propto \exp(-U(X, Y, \wedge)) \quad (5)$$

where *U* is a cost function and $\wedge$ is set of parameters that composed of m_C, *s* and k_C that describe later.

In general, CRF has a low convergence rate while impose a high computational burden if the entire dataset is used [18]. To overcome this drawback, unsupervised CRF [22] considers the neighbours of each input sample as a voting pool which contain k_C samples. In the first iteration, neighbours are generated randomly. In the next iteration, the most current *S* similar samples along with the most different sample are chosen from the previous iteration to the current generation; therefore, to set an equal population size at each iteration, k_C -*S*-1 samples are randomly selected from the remained samples. Again, the *S* most similar samples with the most dissimilar one are fed to the next generation. This process continued till the number of iterations exceeds than a user-default threshold or in two successive iterations, the resulted labels do not change.

It should be noted that k_C and *S* are problem dependent that are determined empirically. Using voting pool described in Eq. (5), the following relation can be resulted [22]:

$$p(y_i \mid x_i, x_{Ni}, y_{Ni}) \propto \exp(-U_i(y_i, y_{Ni}, x_i, x_{Ni})) \quad (6)$$

$$U_i(y_i, y_{Ni}, x_i, x_{Ni}) = \sum_{j \in N_i} W_{i,j}(y_i, y_j, x_i, x_j)$$

$$W_{i,j}(y_i, y_j, x_i, x_j) = \begin{cases} -(D - d_{i,j}) & if \quad y_i = y_j \\ (D - d_{i,j}) & if \quad y_i \neq y_j \end{cases}$$

where N_i is the voting pool (contains k_C samples), $d_{i,j}$ is the Euclidean distance between samples *i* and *j*, *D* is the estimated threshold dividing the set of euclidean distances into intra-class and inter-class distances. To estimate *D*, for each sample, its distances are calculated to m_C other randomly samples and find the minimum and maximum distances. Then, two values *dp* and *do* are calculated by taking the average of all the minimum and maximum distances, respectively. Finally, *D* is determined as:

$$D = \frac{1}{2}(dp + do) \quad (7)$$

In the first iteration, each sample is assigned to a label randomly picked from the integer range of 1 to *n*. Hence, the algorithm starts with a set of *n* singleton clusters without any prior knowledge about the number of clusters. Finally, the optimum label y_i is determined according to the following equation:

$$\hat{y}_i = \arg\min_{y_i} U_i(y_i, y_{Ni}, x_i, x_{Ni}) \quad (8)$$

2.4. Conditional Random Fields (CRFs)

Although NMF is an efficient feature extraction method, the traditional updating algorithms do not have a convex form. Therefore, NMF suffers from achieving to global optimum. To improve the NMF performance, especially in real time applications that needs a short training phase, an efficient initialization method should be employed. To address this problem, CRF is suggested to initialize the NMF components.

The main contribution of this study is to employ CRF for NMF initialization. In this version of CRF (unsupervised), clustering is performed without any prior knowledge about the number of clusters and the initial centroids [22]. This is beneficial for NMF algorithm that highly suffers from lack of prior knowledge about the suitable reduction rank (*k*). In our application, intensity of an image is considered as a random variable X_i and is utilized for clustering. For this case, each vectorized image i is considered as a random variable X_i , $X=\{X_1, X_2,..., X_n\}$ and $Y=\{Y_1, Y_2,..., Y_n\}$ is their labels. In this phase, CRF determines the rank (*k*) of NMF by its number of clusters. After the cluster centers are determined by CRF, these are arranged into columns of matrix *W*. Then for computing matrix *H*, Nonnegative Least Square (NNLS) algorithm [20] is used. Afterwards, LNMF algorithm is executed using the initialized components. What follows is the pseudo-code of CRF-NMF:

Algorithm: CRF-NMF

Inputs: A matrix that is achieved by pre-processing phase.
Outputs: W and H matrices.
Set columns of matrix A to vectors X_i *(i=1... n)*
Each sample is assigned to a label randomly picked from the integer values ranged from 1 to n
For each pattern, voting pool consists of k_C *samples that are selected randomly.Then the s most similar samples and the most different sample, are selected for next iteration.*
Determine the label of each pattern using Eq. (8)
Set i=1
While i< maximum iteration for CRF or in two successive iterations, the sample labels are not changed
    The voting pool is updated
    Update labels of each pattern with Eq. (8)
    i=i+1
End
Uses Y for label of samples then cluster centers are computed.
Set cluster centers to vectors of W matrix
H matrix is computed using NNLS algorithm.
For i=1: maximum iteration for NMF
    Update H and W matrices trough Eq. (4)
End

3. Database, preprocess and evaluation methods Database, preprocess and evaluation methods

In our experiments, Facial expression images of Japanese Female Facial Expression (JAFFE) dataset[1]are used. This dataset consists of of 213 images of 7 facial expressions containing 6 basic facial expressions and a neutral expression that posed by 10 Japanese female. At first, in the preprocessing stage, all faces from the images are cut and aligned into a fixed size (33×33). For each image, histogram equalization is applied and pixel's intensity of each matrix is normalized through [0, 1].

[1] *http://www.kasrl.org/jaffe_download.html*

Relative error, basis vectors, sparsity, and orthogonality are the well known criteria which are invoked to evaluate different methodologies. Each of them reveals the capability of a method from a specific aspect. In this paper, for sparsity, the number of elements in *W* which satisfies $W_{ij} < \|W_j\|/256$ inequality is considered. Orthogonality of W matrix is assessed through the $\sum_{i \neq j} w_i^T w_j$ equation [31]. It is obvious whatever the sparsity index inclines, the complexity of recall phase declines. Orthogonality of *W* component can reveal the independence of columns together. Relative error measures the rate of error change over the error value in successive iterations. Finally, progression of the factorized matrix basis vectors (columns of *W* matrix) can measure the convergence of NMF after initialization.

4. Experimental Results and Discussion

After the pre-processing stage, each image has 33×33 size and for all of the experiments, parameters m_C, *S* and, k_C are empirically chosen 5, 2 and 14, respectively. Maximum iteration of CRF algorithm is considered 100; however, the algorithm almost becomes stable before 100 iterations. By applying CRF to the dataset, number of clusters is determined 17 which is the suitable rank for the NMF components.

4.1. Image Results achieved using just CRF(without using NMF)

It is expected to rapidly converge to local optimum when CRF initializes the NMF matrices. In the other words, when CRF is applied to the pre-processed face images for initializing *W* and *H* components, (without further applying NMF), we expect to achieve images that are similar to their original ones. In this way, *W* and *H* matrices are initialized by CRF and after composing matrix *W*×*H*, left images in Fig. 2, are obtained while those on the right are original images. Experimental results depict that by using just CRF leads to a very good estimation of images, where the resulted images are denoised and looks clear, that demonstrates its susceptibility for initialization of NMF.

Figure 2: left images are estimated using CRF and right ones are the counterpart original images.

4.2. Comparing NMF with Random and CRF initializations in term of basis vectors

In this experiment, the progression of the left factor, *W,* in NMF is exhibited using different initialization methods. In practice, rank of NMF determines the number of basis vectors which is determined by CRF, as it mentioned before, 17 clusters are determined. In the first comparison phase, progression of the basis vectors (columns of *W*) by the proposed method is compared to standard NMF [21] which uses random initialization. The superiority of the proposed method to the random initialization appeared when the performance of both algorithms passes a certain number of iterations. The comparison results demonstrate the significance of the estimated basis vectors (main face components) achieved by CRF compared to that of standard NMF in low number of iterations, that is proper for real time applications. For better visualization, obtained images by infinite norm on the JAFFE dataset are depicted in Fig. 3.

As we expected, in high iterations, as shown in Fig. 3, results of standard NMF outperformed that of the proposed CRF-NMF. The achieved results seem logical because all of the initialization methods mostly find a local optimum; therefore, after a few epochs their progression would not be significant. Nevertheless, in real time problems, using CRF, provide acceptable face components in a low number of iterations (very fast convergence) but its results is not brilliant in high iterations while so far random initialization results mostly lead to estimate better face components than other initialization methods [4, 31, 32, 34] when updating algorithm is highly iterated (e.g. 3000).

4.3. Comparison of CRF-NMF and other initialization methods in term of relative error

As mentioned in the introduction part, several methods have been suggested to find suitable values for initialization of *W* and *H* [4, 31, 32, 30]. In this experiment, the proposed method is compared to the SVD, PCA, divergence-based *k*-means, spherical *k*-means, and standard NMF (random initialization). decompositions methods on the JAFFE database. The obtained results (brought in Table 1) illustrate that the proposed CRF-based initialization outperformed the other initialization methods in terms of relative error. It should be pointed that relative error is chosen as the ratio of $\|A-WH\|_F^2 / \|A\|_F^2$ [31].

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

Initialization method	CRF	Random	Divergence based *k*-means	Spherical *k*-means	SVD	PCA
Relative error	0.4447	6.6867	0.4488	0.4520	0.8796	1.9410

4.4. Comparison of CRF-NMF and standard NMF in term of sparsity and orthogonality

In this part, sparsity and orthogonality are assessed and depicted for the standard NMF and the proposed CRF-NMF in various iterations. To evaluate the algorithm in terms of sparsity, we consider number of elements in *W* that satisfies the $W_{ij} < \|W_j\|/256$ inequality, and orthogonality of this matrix is assessed by the $\sum_{i \neq j} w_i^T w_j$ equation [31]. As shown in Fig. 4, results of CRF-NMF outperformed of the standard NMF in terms of sparsity and orthogonality. In high iterations, as we expected, the results of standard NMF outperformed that of the proposed CRF-NMF. As it is explained before, the achieved results seem logical, because all of the suggested initialization methods mostly find a local optimum and highly suffer from lack of diversity in their optimization algorithm. while the standard NMF incorporate randomness through its search leading to improve the local minimum problem compared to the other initialization methods.

4.5. Image denoising using CRF initialization

This is an interesting experiment in which each image is first contaminated by noise and then decomposed into *W* and *H* components using the proposed CRF initialization. Afterward, the primary noisy images are reconstructed just using multiplication of *W* and *H* without employing NMF. In this way, Gaussian noise is added into 10 randomly selected images from the dataset and the reconstructed images (by multiplication of *W***H* components) are shown in Fig. 5. As we can see, the reconstructed images carry less noise compare to its primary noisy image. This experiment shows that noise effect is significantly diminished when decomposing is done by CRF and then abruptly reconstructs the images. In other word, CRF can be used as a de-noising algorithm to recognize the facial expression states. This property rises from this fact that CRF estimates a rough representation of the primary image (leading to determine a nearly local optimum solution) at the cost of losing the detailed information like noises.

As we see in Fig. 5, in first row original images are shown, second row depicts noisy images, and the third row exhibits the refined images by CRF initialization and reconstruction as explained in 4.1

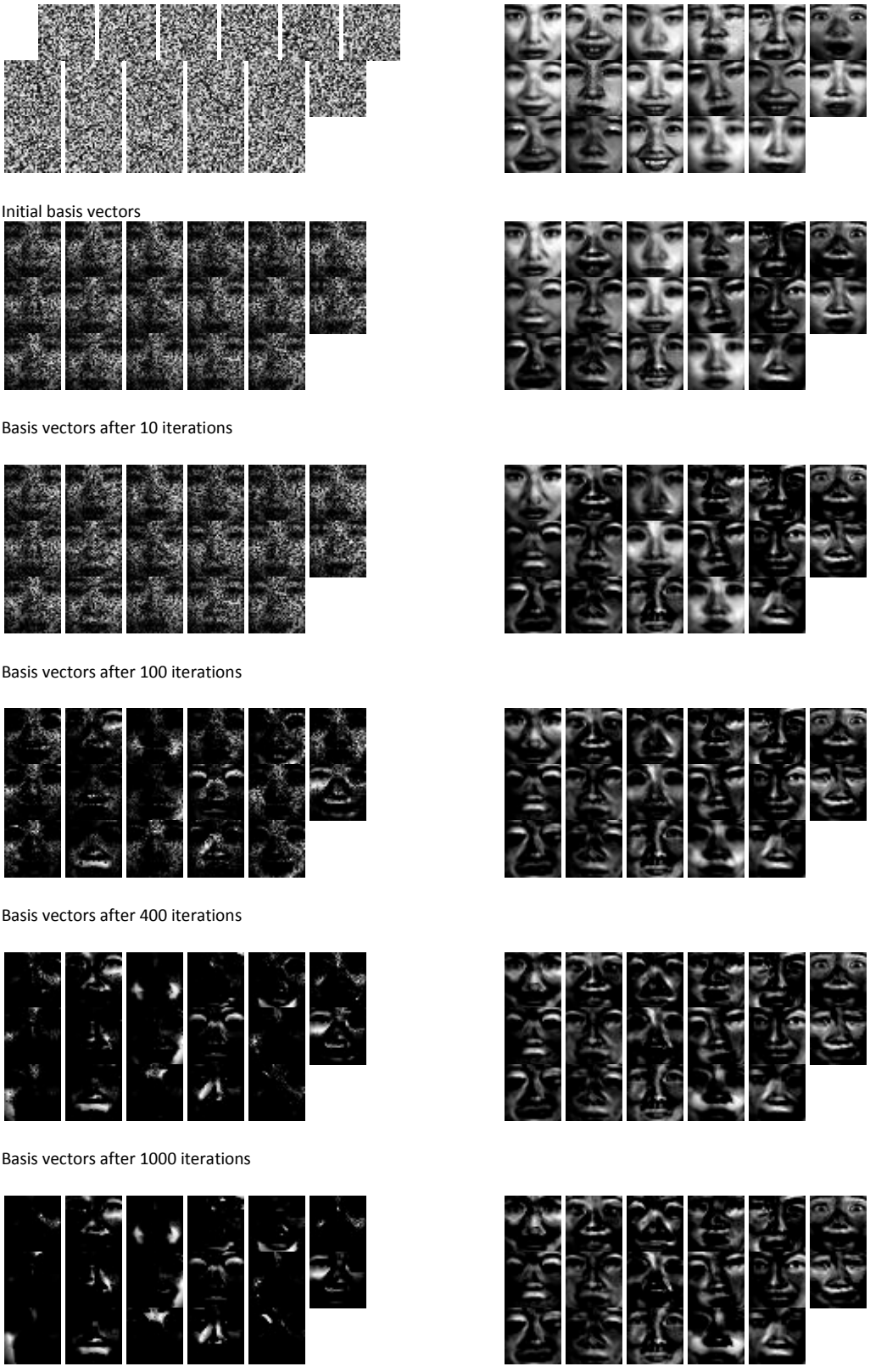


Figure 3: Above face components show the progression of basis vectors (columns of W matrix). Left and right images are basis vectors achieved by CRF and random initialization, respectively.

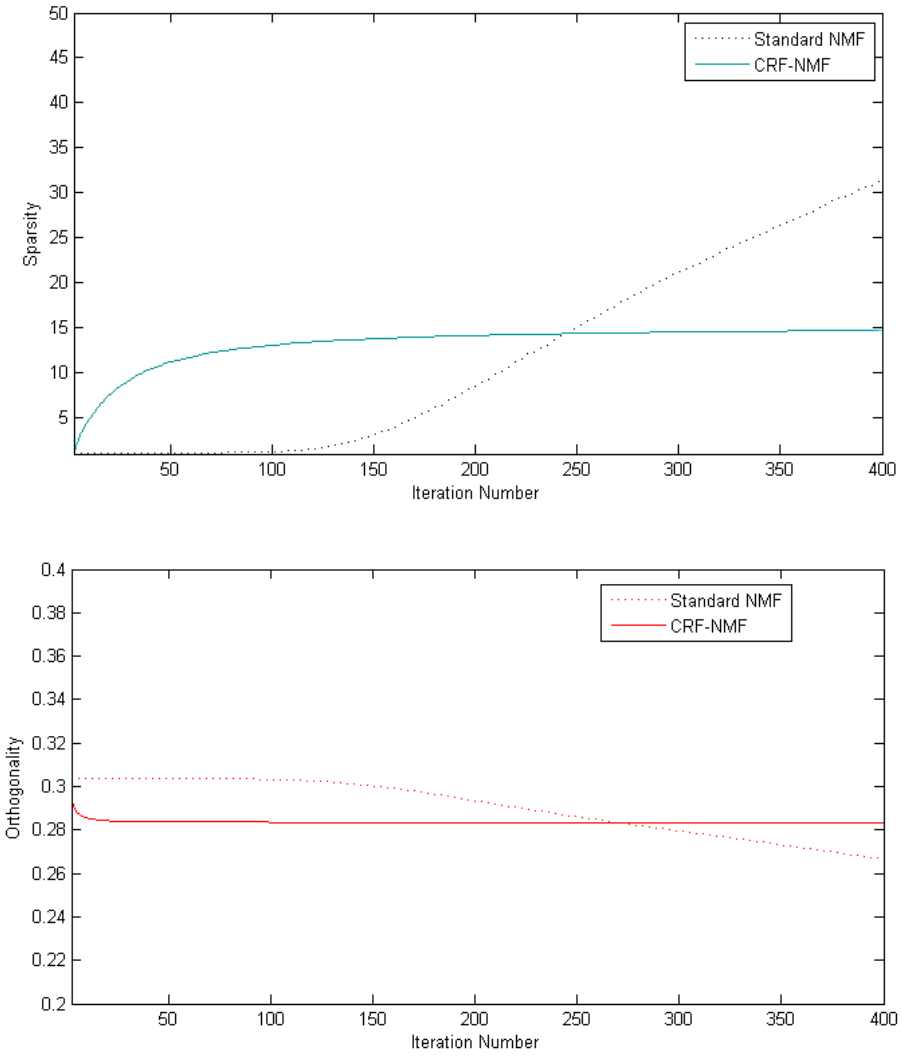


Figure 4: Comparison of standard NMF and CRF-NMF in terms of sparsity and orthogonality

Figure 5: Above face components show de-noising capability of the proposed CRF initialization method. Images in first row are original images, images in second row are noisy images and images in third row are images that are achieved by CRF initialization.

Conclusion and Summary

NMF is a part-based representation that has been successfully applied to many applications; however, it highly suffers from its improper initialization. To solve this shortcoming, CRF is employed to initialize the NMF components with the objective of faster convergence along with diminishing the relative error and also increasing the spasity and orthogonality. Due to the probabilistic nature of CRF, the achieved results empirically prove that the proposed scheme provides stable and fairly accurate

results. Moreover, another benefit of the proposed initialization method is to determine the suitable rank factor for NMF which is a challenging parameter for NMF components. In addition, initialization using CRF leads to diminish the noise level from the primary images. Therefore, the proposed method is a fast, accurate and robust method. Even in noisy environment that can be considered as an alternative method instead of random initialization (gold-standard) method especially when we are compelled to stop NMF in early stages for real time applications.

References

[1] Abboud, B., Davoine, F., Dang M., Facial expression recognition and synthesis based on an appearance model, Signal Processing, 19(8), 723-740(2004).

[2] Bau, D.T., Lloyd N., "Numerical linear algebra, Philadelphia. Society for Industrial and Applied Mathematics", ISBN 978-0-89871-361-9 (1997).

[3] Berry, M.W., Browne, M., Langville, A.N., Pauca, V., Plemmons, R., "Algorithms and Applications for Approximate Nonnegative Matrix Factorization. Computational statistics and data analysis", 52(1), 155-173, (2006).

[4] Boutsidis, C., Gallopoulos, E., "On SVD-based initialization for nonnegative matrix factorization. Pattern Recognition", 41(4), 1350-1362, (2005).

[5] Calder, A.J., Burton, A.M., Miller, P., Young, A.W., Akamatsu S., "A principal component analysis of facial expressions". Vision Research. 41(9), 1179-1208, (2001).

[6] Chen Z., Cichocki A., "Nonnegative Matrix Factorization with temporal smoothness and/or spatial decorrelation constraints". Technical Report, Laboratory for Advanced Brain Signal Processing, RIKEN, Tokyo, Japan, (2005).

[7] Cichocki, A., " Unsupervised Learning Algorithms and Latent Variable Models: PCA/SVD, CCA/PLS, ICA, NMF, etc", Academic Press Library in Signal Processing,, 1:1151–1238, (2014).

[8] Cichocki, A., Zdunek, R., Amari, S., "Csiszar's divergences for Non-negative Matrix Factorization: Family of new algorithms. Independent Component Analysis and Blind Signal Separation", 3889, 32 – 39, (2006).

[9] Feng, T., Li, S.Z., Shum, H.Y., Zhang H., "Local Non-Negative Matrix Factorization as a Visual Representation". Development and Learning, 178-183, (2002).

[10] Finesso, L., Spreij, L., "Nonnegative matrix factorization and I-divergence alternating minimization". Elsevier. Linear Algebra and its Applications, 416(2-3), 270-287, (2006).

[11] Geng, X., Zhao, J., "Interactive Image Segmentation with Conditional Random Fields. Nautral Computation", 2, 96-101, (2008).

[12] Hyvarinen A., Karhunen J., Oja E., "Independent Component Analysis". New York: Wiley, (2001).

[13] Jeon B. K., Jung Y. B., Hong K. S., "Image segmentation by unsupervised sparse clustering. Application of computer vision", 2-7, (2005).

[14] Kang, T. G., Kwon, K., Shin, J. W., Kim, N. S., "NMF-based Target Source Separation Using Deep Neural Network", Signal Processing Letters , 22(2), 229 – 233, (2015).

[15] Klinger R., Tomanek K., "Classical probabilistic models and conditional random fields". Technische University at Dortmund, Dortmund, Electronic Publication, (2007).

[16] Kotsia, I., Zafeiriou, S., Nikolaidis, N., Pitas I., "Texture and shape information fusion for facial expression and facial action unit recognition". Pattern Recognition. 41(30), 833-85, 2008.

[17] Kwon, K., Shin, J. W., Kim, N. S., NMF-Based Speech Enhancement Using Bases Update, Signal Processing Letters,, 22(4), 450 – 454, 2014.

[18] Lafferty, J., McCallum, A., Pereira, F., "Conditional random fields: Probabilistic models for segmenting and labelling sequence data". In Proc. ICML-01, 282.289, (2001).

[19] Langville A., Meyer C., Albright R.., Cox J., Duling, D., "Algorithms, initializations, and convergence for the nonnegative matrix factorization". Preprint, (2006).

[20] Lawson, C.L., Hanson R.J., "Solving Least Squares Problems", Prentice–Hall, Englewood CliMs, NJ, (1974).

[21] Lee, D., Seung, H., "Learning the Parts of Objects by Non-Negative Matrix Factorization". Nature 401, 788–791, (1999).

[22] Li, C.T., Yuan Y, Wilson R., "Unsupervised clustering of conditional random fields approach for clustering gene expression time series". Bioinformatics. 24(21), 2467-2473, (2008).

[23] Li, M., Bai, M., Wang, C., Xiao, B., "Conditional random field for text segmentation from images with complex background". Pattern Recognition Letters. 31(16), 2295-2308, (2010).

[24] Li, S.Z., Hou, X.W., Zhang H.J., "Learning spatially localized, parts-based representation". Computer Vision and Pattern Recognition, 207-212, (2001).

[25] Li, Z., Peng, H., Wu, X., "A New Descriptive Clustering Algorithm Based on Nonnegative Matrix Factorization". Granular Computing, 407 – 412, (2008).

[26] Shahnaz, F., Berry, M.W., Pauca, V.P., Plemmons, R.P., "Document clustering using nonnegative matrix factorization". Information Processing and Management. 373-386, (2006).

[27] Shan, C., Gong, S., McOwan, P.W., "Facial expression recognition based on Local Binary Patterns: A comprehensive study". Image and Vision Computing. 27(6), 803-816, (2009).

[28] Tsai, P., Cao, L., Hintz, T., Jan, T., "A bi-modal face recognition framework integrating facial expression with facial appearance". Pattern Recognition Letters. 30(12), 1096-1109, (2009).

[29] Wang, J. J. Y., Gao, X., Max–Min distance nonnegative matrix factorization, Neural Networks, 61,75–84, 2015.

[30] Wang, J., Yin, L., "Static topographic modelling for facial expression recognition and analysis". Computer Vision and Image Understanding. 108(1-2), 19-34, (2007).

[31] Wild, S., Curry, J., Dougherty A., "Improving non-negative matrix factorizations through structured initialization". Pattern Recognition, 37(11), 2217-2232, (2004).

[32] Xue, Y., Tong, Ch., Chen, Y., Chen, W., "Clustering-based initialization for non-negative matrix factorization". Applied Mathematics and Computation, 205(2), 525-536, (2008).

[33] Ying, Z., Zhang, G., "Facial Expression Recognition Based on NMF and SVM". International Forum on Information Technology, 458 – 462, (2009).

[34] Zhao, L., Zhang, G., Xu, X., "Facial Expression Recognition Based on PCA and NMF". Intelligent control and automation. 6826 – 6829, (2008).