EXEMPLAR-EMBED COMPLEX MATRIX FACTORIZATION FOR FACIAL EXPRESSION RECOGNITION

Viet-Hang Duong[1], Yuan-Shan Lee[1], Jian-Jiun Ding[2], Bach-Tung Pham[1], Manh-Quan Bui[1], Pham The Bao[3], and Jia-Ching Wang[1]

[1] Dept. of Computer Science and Information Engineering, National Central University, Jhongli, Taiwan
[2] Graduate Institute of Communication Engineering, National Taiwan University, Taipei, Taiwan
[3] Fac. of Mathematics and Computer Sciences, University of Science, Ho Chi Minh City, Viet Nam

ABSTRACT

This paper presents an image representation approach which is based on matrix factorization in the complex domain and called exemplar-embed complex matrix factorization (EE-CMF). The proposed EE-CMF approach can very effectively improve the performance of facial expression recognition. Moreover, Wirtinger's calculus was employed to determine derivatives. The gradient descent method was utilized to solve the complex optimization problem. Experiments on facial expression recognition verified the effectiveness of the proposed EE-CMF. It provides consistently better recognition results than standard NMFs.

Index Terms— Complex matrix factorization, facial expression, optimization, nonnegative matrix factorization

1. INTRODUCTION

Facial expressions contain a lot of information, such as feeling and cognitive motion. Facial expression recognition (FER) plays an important role in human communication. Age, ethnicity, gender, facial hair, the makeup style, gesture, occlusion, and environment lighting affect the performance of FER [1]. How to design an effective and robust system is a challenging topic in FER.

Feature extraction is a critical step of the FER system. Recently, many works have been done on subspace projection techniques for appearance-based feature extraction. In the subspace learning scheme, the new feature matrix is built which maps data points to a subspace. The popular subspace projection techniques, such as principal component analysis (PCA) [2], linear discriminant analysis (LDA) [3, 4, 12], and nonnegative matrix factorization (NMF) [5, 6], represent a facial image as a linear combination of low rank basis images. Lee and Seung [5, 6] found that NMF has the superior ability on parts-based representation. NMF is an unsupervised data-driven approach in which all elements of the decomposed matrix and the obtained matrix factors are forced to be nonnegative. The sparsity constraint can also be imposed on the cost function. For instance, Hoyer [7] proposed a sparse function and incorporated the sparseness into factorizing a nonnegative matrix to improve the obtained decompositions. Yuan and Oja [8] introduced a projective NMF (PNMF) which learns localized features.

Many modified NMFs have been developed to perform the FER task. Nikitidis *et al*. [9, 10] developed two extensions of the NMF that applied discriminant criteria as constraints, including the clustering based discriminant analysis (CDA) [11] and linear discriminant analysis (LDA) [12]. Lee and Chellappa suggested incorporating sparsity constraints to generate localized dictionaries from dense motion flow image sequences [13]. Apparently, most NMF frameworks require the addition of regularizes to improve FER performance. Moreover, nonnegative entries are usually compulsory for the data matrix in NMFs, which restrict the applications of NMF. Semi NMF and convex NMF algorithms have been proposed to deal with this limitation [14]. In particular, convex NMF algorithm (Con-NMF) further require that the basis vectors in NMF are convex or linear combinations of the mixed-sign data points. Besides, the interesting work of Liwicki *et al*. [15] showed the equivalence between the square Frobenius matrix norm in the complex field and the robust dissimilarity measure in the real field. These studies motivate us to propose a new model, named exemplar-embed complex matrix factorization (EE-CMF). In the proposed model, the real data is transformed to the complex domain and the complex data matrix is factorized under imitating Con-NMF frame work. The object function is minimized throughout an unconstraint complex optimization problem. In the real domain, a Con-NMF loss function is bounded optimization. Since the Cauchy-Riemann equation no longer holds, the standard complex derivative is unable to operate as usual. In complex domain, the Wirtinger calculus [16] provides an efficient tool to compute derivations and brings more advantages for complex optimization problem.

The main contributions of this work are summarized as follows.

1) An image analysis method on the complex domain, which is called EE-CMF, is proposed.

surprise sadness anger fear disgust happiness neautral

Fig. 1. Cropped face images of six facial expressions from the CK+ dataset [25].

surprise sadness anger fear disgust happiness neautral

Fig. 2. Cropped face images of six facial expressions from the JAFFE dataset [26].

2) In complex domain, the updating rule for EE-CMF is derived based on gradient descent method.

3) A thorough experimental study on facial expression recognition is conducted. The results show the proposed EE-CMF yields better performance compared to basic and extension of the NMF.

2. PRELIMINARIES

2.1. Nonnegative Matrix Factorization

Given an $N \times M$ input data matrix $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_M)$, where M is the number of facial images and each column $\mathbf{x}_m$ corresponds to an image with size a by b ($N=a\times b$). The NMF problem is to find $\mathbf{W} \in \mathbb{R}_+^{N\times K}$ and $\mathbf{V} \in \mathbb{R}_+^{K\times M}$ that satisfies the following objective function:

$$\min_{\mathbf{W}\geq 0, \mathbf{V}\geq 0} \mathrm{O}_{NMF}(\mathbf{W},\mathbf{V}) = \frac{1}{2}\left\|\mathbf{X} - \mathbf{WV}\right\|_F^2 \tag{1}$$

The basic matrix $\mathbf{W}$ contains K vectors which are linearly combined by the coefficients in $\mathbf{V}$ to represent the data. To solve (1), Lee and Seung [5, 6] provided the iteratively updating algorithms as follows:

$$\mathbf{V}_{ij} \leftarrow \mathbf{V}_{ij}\frac{(\mathbf{W}^T\mathbf{X})_{ij}}{(\mathbf{W}^T\mathbf{W}\mathbf{V})_{ij}};\quad \mathbf{W}_{ij} \leftarrow \mathbf{W}_{ij}\frac{(\mathbf{X}\mathbf{V}^T)_{ij}}{(\mathbf{W}\mathbf{V}\mathbf{V}^T)_{ij}} \tag{2}$$

To relax the constraint of nonnegative data, Ding *et al.* [14] proposed convex nonnegative matrix factorization (Con-NMF) where mixed-sign data matrices are applied. Con-NMF imposes a constraint that the column vectors of must lie within the column space of $\mathbf{X}$ i.e. $\mathbf{W}=\mathbf{XA}$ where $\mathbf{A}$ is an auxiliary adaptive weight matrix and obtain the objective function [17]:

$$\min O_{conNMF}(\mathbf{A},\mathbf{V}) = \frac{1}{2}\left\|\mathbf{X} - \mathbf{XAV}^T\right\|_F^2 \;\; \text{s.t } \mathbf{A}\in\mathbb{R}_+^{M\times K}, \mathbf{V}\in\mathbb{R}_+^{M\times K} \tag{3}$$

The factors $\mathbf{V}$ and $\mathbf{A}$ are updated as follows [14]:

$$\begin{aligned}\mathbf{V}_{ij} &\leftarrow \mathbf{V}_{ij}\sqrt{\frac{[(\mathbf{X}^T\mathbf{X})^+\mathbf{A}]_{ij} + [\mathbf{V}\mathbf{A}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{A}]_{ij}}{[(\mathbf{X}^T\mathbf{X})^-\mathbf{A}]_{ij} + [\mathbf{V}\mathbf{A}^T(\mathbf{X}^T\mathbf{X})^+\mathbf{A}]_{ij}}} \\ \mathbf{A}_{ij} &\leftarrow \mathbf{A}_{ij}\sqrt{\frac{[(\mathbf{X}^T\mathbf{X})^+\mathbf{V}]_{ij} + [(\mathbf{X}^T\mathbf{X})^-\mathbf{A}\mathbf{V}^T\mathbf{V}]_{ij}}{[(\mathbf{X}^T\mathbf{X})^-\mathbf{V}]_{ij} + [(\mathbf{X}^T\mathbf{X})^+\mathbf{A}\mathbf{V}^T\mathbf{V}]_{ij}}}\end{aligned} \tag{4}$$

where $\mathbf{X}^T\mathbf{X} = (\mathbf{X}^T\mathbf{X})^+ - (\mathbf{X}^T\mathbf{X})^-$, $(\mathrm{a}_{ij})^+ = \max\{0, \mathrm{a}_{ij}\}$, and $(\mathrm{a}_{ij})^- = \max\{0, -\mathrm{a}_{ij}\}$.

2.2. Wirtinger's Calculus and Complex Optimization

[**Definition 1**] Let $f: \mathbb{R}\times\mathbb{R}\to\mathbb{C}$ be a function of real variables x and y such that $g(z, z^*)=f(x, y)$ where $z=x+iy$ and g is analytic with respect to z and z^*. The two "partial derivative" operators $\partial g/\partial z$ and $\partial g/\partial z^*$ are defined by:

$$\frac{\partial g}{\partial z} = \frac{1}{2}\left(\frac{\partial f}{\partial x} - i\frac{\partial f}{\partial y}\right),\; \frac{\partial g}{\partial z^*} = \frac{1}{2}\left(\frac{\partial f}{\partial x} + i\frac{\partial f}{\partial y}\right) \tag{5}$$

It is often referred to as Wirtinger's derivative [18].

[**Definition 2**] If f is a real function of a complex matrix $\mathbf{Z}$ then the complex gradient matrix is given by [18]:

$$\nabla f(\mathbf{Z}) = 2\frac{\partial f(\mathbf{Z})}{\partial \mathbf{Z}^*} = \frac{\partial f(\mathbf{Z})}{\partial \operatorname{Re}(\mathbf{Z})} + i\frac{\partial f(\mathbf{Z})}{\partial \operatorname{Im}(\mathbf{Z})} \tag{6}$$

How to find the direction of minimum rate of change will be stated in the following theorem.

[**Theorem 1**] Let $g: \mathbb{C}^{N\times M}\times\mathbb{C}^{N\times M}\to\mathbb{R}$ be a real-valued function that maps complex matrices into the real domain. $-\nabla_{\mathbf{Z}^*} g(\mathbf{Z},\mathbf{Z}^*)$ gives the direction where the function g has the minimum rate to change with respect to $\mathbf{Z}$ [19].

In fact, the first order Taylor series expansion for the real-differentiable function $g(\mathbf{Z}, \mathbf{Z}^*)$ has the form as

$$\Delta g(\mathbf{Z},\mathbf{Z}^*) \approx \left\langle \nabla_{\mathbf{Z}} g, \Delta\mathbf{Z}^*\right\rangle + \left\langle \nabla_{\mathbf{Z}^*} g, \Delta\mathbf{Z}\right\rangle = 2\operatorname{Re}\left\{\left\langle \nabla_{\mathbf{Z}^*} g, \Delta\mathbf{Z}\right\rangle\right\} \tag{7}$$

According to the Cauchy-Bunyakovsky-Schwarz inequality [20], $\left|\Delta\mathbf{Z}^H\nabla_{\mathbf{Z}^*} g\right| \leq \left\|\Delta\mathbf{Z}\right\|\left\|\nabla_{\mathbf{Z}^*} g\right\|$. The equality holds when $\Delta\mathbf{Z}$ and $\nabla_{\mathbf{Z}^*} g$ is collinear, i.e., the gradient $\nabla_{\mathbf{Z}^*} g$ defines the direction of the maximum rate of change in $g(.,.)$ with respect to $\mathbf{Z}$.

3. PROPOSED METHOD

3.1. Exemplar-Embed Complex Matrix Factorization (EE-CMF)

To get the complex data matrix $\mathbf{Z}\in\mathbb{C}^{N\times M}$, the original real data matrix $\mathbf{X}$ is first normalized and then transformed into the complex field using Euler's formula [21] by a mapping Ω from $\mathbb{R}^N$ to $\mathbb{C}^N$.

$$\Omega(\mathbf{x}_t) = \mathbf{z}_t = \frac{1}{\sqrt{2}} e^{i\alpha\pi\mathbf{x}_t} = \frac{1}{\sqrt{2}} \begin{bmatrix} e^{i\alpha\pi\mathbf{x}_t(1)} \\ \vdots \\ e^{i\alpha\pi\mathbf{x}_t(N)} \end{bmatrix} \quad (8)$$

where $\mathbf{x}_i$ denotes an N-dimensional vector, which corresponds to an input image and is sorted in the lexicographic order, $x_t(\mathrm{c}) \in [0, 1]$ and $\alpha \in \mathbb{R}^+$.

After obtaining the complex training data **Z,** we employ our proposed EE-CMF to factorize **Z** into complex matrix factors. The basis matrix in EE-CMF is constructed by the linear combination of the complex training examples. Given the complex data matrix $\mathbf{Z} \in \mathbb{C}^{N\times M}$, EE-CMF factorizes **Z** into the encoding matrix $\mathbf{V} \in \mathbb{C}^{K\times M}$ and the exemplar-embed basis matrix $\mathbf{E} = \mathbf{ZW}$ where $\mathbf{W} \in \mathbb{C}^{M\times K}$. Therefore, the objective function of EE-CMF problem can be formulated as follows:

$$\min_{\mathbf{W},\mathbf{V}} \mathrm{O}_{\text{EE-}CMF}(\mathbf{W},\mathbf{V}) = \min_{\mathbf{W},\mathbf{V}} \frac{1}{2}\left\|\mathbf{Z} - \mathbf{ZWV}\right\|_F^2 \quad (9)$$

where

$$\left\|\mathbf{Z} - \mathbf{ZWV}\right\|_F^2 = Tr(\mathbf{Z} - \mathbf{ZWV})^H(\mathbf{Z} - \mathbf{ZWV})$$

$$= Tr(\mathbf{Z}^H\mathbf{Z} - \mathbf{V}^H\mathbf{W}^H\mathbf{Z}^H\mathbf{Z} - \mathbf{Z}^H\mathbf{ZWV} + \mathbf{V}^H\mathbf{W}^H\mathbf{Z}^H\mathbf{ZWV}) \quad (10)$$

3.2. Optimal Solution

In order to solve the minimization problem, we use the complex gradient descent algorithm by exploiting Wirtinger's calculus. It can be seen that (9) is a nonconvex minimization problem with respect to both variables **W** and **V**. Therefore, it is impractical to obtain the optimal solution by the conventional method. Instead, the following scheme can be used to solve the problem in (9).

- First, fix **W** and the objective function (9) is modified as a function of one variable **V** as follows:

$$\min_{\mathbf{V}} O(\mathbf{V}) = \min_{\mathbf{V}} \frac{1}{2}\left\|\mathbf{Z} - \mathbf{ZWV}\right\|_F^2 \quad (11)$$

- Then, **W** is updated based on the Moore–Penrose pseudoinverse [22], which is dented by $\dagger$, and $\mathbf{W}=(\mathbf{Z}^\dagger\mathbf{Z})\mathbf{V}^\dagger$ with fixed **V**.

To solve the subprolem (11), the function $O(\mathbf{V})$ is treated as $O(\mathbf{V},\mathbf{V}^*)$ where

$$O(\mathbf{V},\mathbf{V}^*) = \frac{1}{2} Tr[\mathbf{Z}^H\mathbf{Z} - (\mathbf{V}^*)^T\mathbf{W}^H\mathbf{Z}^H\mathbf{Z} - \mathbf{Z}^H\mathbf{ZWV} + (\mathbf{V}^*)^T\mathbf{W}^H\mathbf{Z}^H\mathbf{ZWV}] \quad (12)$$

According to Theorem 1, at a given iteration round t, the following update rule is employed:

$$\mathbf{V}^{(t+1)} = \mathbf{V}^{(t)} - 2\beta_t \nabla_{\mathbf{V}^*} O(\mathbf{V}^{(t)},\mathbf{V}^{*(t)}) \quad (13)$$

where β_t is the learning step parameter for the t^{th} iteration estimated by the Armijo rule [23]. From the Armijo rule, $\beta_t = \mu^{s_t}$, $0 < \mu < 1$, and s_t is the first non-negative integer such that the following inequality is satisfied:

$$O(\mathbf{V}^{(t+1)},\mathbf{V}^{*(t+1)}) - O(\mathbf{V}^{(t)},\mathbf{V}^{*(t)}) \leq 2\sigma \operatorname{Re}\left\{\left\langle \nabla_{\mathbf{V}^*} O(\mathbf{V}^{(t)},\mathbf{V}^{*(t)}), \mathbf{V}^{(t+1)} - \mathbf{V}^{(t)} \right\rangle\right\} \quad (14)$$

The first order partial derivative with respect to $\mathbf{V}^*$ are evaluated as follow:

$$\nabla_{\mathbf{V}^*} O(\mathbf{V},\mathbf{V}^*) = -\mathbf{W}^H\mathbf{Z}^H\mathbf{Z} + \mathbf{W}^H\mathbf{Z}^H\mathbf{ZWV} \quad (15)$$

The condition in (14) guarantees the decrease of the function value in each iteration. Finally, one can choose a pre-defined threshold ε and set the stopping condition as follows:

$$\left\|\nabla_{\mathbf{V}^*} O(\mathbf{V},\mathbf{V}^*)\right\|_F \leq \varepsilon \quad (16)$$

4. EXPERIMENTS

In this section, we evaluated the performance of the proposed EE-CMF framework for FER. The classification capability of the derived encoding coefficient vector was compared with various NMF-based methods. We obtained the basic matrix $\mathbf{W}_{tr}$ from $\mathbf{W}_{tr} = \mathbf{Z}_{tr}(\mathbf{V}_{tr}\mathbf{Z}_{tr})^\dagger$ in the training phase. The test sample $\mathbf{z}_{te}$ was encoded by $\mathbf{v}_{te} = (\mathbf{ZW}_{tr})^\dagger\mathbf{z}_{te}$. Classification was performed by the nearest neighbor classifier after projection.

4.1. Data Description, Baselines, and Experiment Settings

The proposed model was evaluated on two publicly available databases: the Cohn-Kanade (CK) [25] and the JAFFE [26] datasets. There are seven facial expressions in these datasets, including one neutral state and six basic expressions that contains happiness, sadness, surprise, anger, disgust, and fear. Each facial image in two databases was cropped and resized to have fixed size of 32 × 32 pixels. Figures 1 and 2 show some of the images in the two datasets.

The proposed algorithm was compared to the following popular NMF algorithms: (1) basic NMF [6]; (2) semi-NMF [14]; (3) convex NMF [14] (Con-NMF); (4) weighted NMF (We-NMF) [27], which assigns binary weights to the data matrix; (5) Ne-NMF [28], which is an efficient solver that applies Nesterov's optimal gradient method in the optimization process.

To satisfy (12), the rate of reducing the step size μ was set with the sufficient decrease condition at 0.01. The stopping criterion was as in (15) where the relative tolerance ε was 10^{-4} or at most 10000 iterations.

TABLE I
FACIAL EXPRESSION RECOGNITION RATE (%) USING THE CK DATABASE WITH DIFFERENT SUBSPACE DIMENSIONALITIES
(Case 1: No. Training =1)

No. Base	EE-CMF	NMF	Semi-NMF	Con-NMF	We-NMF	Ne-NMF
20	**95.43**	85.41	75.12	55.25	85.06	94.3
30	**92.25**	90.99	86.88	65.89	91.07	89.3
40	91.24	93.88	91.90	73.55	**94.17**	90.27
50	**95.06**	94.50	93.82	80.27	94.75	91.80
60	**96.14**	95.06	95.00	87.40	94.92	92.34
70	**96.59**	95.18	95.58	90.42	95.62	93.31
80	**96.74**	95.93	96.34	91.57	95.58	93.27
90	**96.63**	95.95	96.59	92.42	95.87	93.26
100	**96.78**	96.03	96.57	92.03	95.62	94.17
Ave.	**95.21**	93.66	91.98	80.98	93.63	92.45

TABLE II
FACIAL EXPRESSION RECOGNITION RATE (%) USING THE CK DATABASE WITH DIFFERENT SUBSPACE DIMENSIONALITIES
(Case 1: No. Training =2)

No. Base	EE-CMF	NMF	Semi-NMF	Con-NMF	We-NMF	Ne-NMF
20	93.14	84.44	66.75	69.31	84.49	**94.75**
30	94.02	91.38	79.86	69.94	91.32	**94.08**
40	**96.12**	94.55	89.95	71.74	94.60	94.41
50	**97.58**	96.03	93.66	70.47	95.15	94.52
60	96.89	**97.02**	95.70	76.78	96.47	94.38
70	**97.27**	96.70	96.12	82.26	97.11	93.72
80	97.77	**97.91**	97.14	86.28	97.66	94.21
90	**98.21**	97.88	97.96	89.84	97.49	94.16
100	**98.35**	97.85	98.02	92.78	97.47	93.28
Ave.	**96.59**	94.86	90.57	78.82	94.64	94.17

TABLE III
FACIAL EXPRESSION RECOGNITION RATE (%) USING THE JAFFE DATABASE WITH DIFFERENT SUBSPACE DIMENSIONALITIES
(Case 1: No. Training =1)

No. Base	EE-CMF	NMF	Semi-NMF	Con-NMF	We-NMF	Ne-NMF
20	**66.99**	65.24	59.65	47.06	63.36	66.85
30	66.36	68.11	64.27	54.06	**68.32**	61.05
40	**72.31**	70.84	66.36	54.06	68.95	63.28
50	**72.03**	71.68	68.6	47.06	69.02	61.82
60	**72.45**	71.12	71.05	46.36	72.38	63.71
70	**72.31**	69.79	70.70	27.34	69.16	62.52
Ave.	**70.41**	69.46	66.77	45.99	68.53	63.21

TABLE IV
FACIAL EXPRESSION RECOGNITION RATE (%) USING THE JAFFE DATABASE WITH DIFFERENT SUBSPACE DIMENSIONALITIES
(Case 1: No. Training =2)

No. Base	EE-CMF	NMF	Semi-NMF	Con-NMF	We-NMF	Ne-NMF
20	**79.86**	70.55	57.80	55.08	67.39	74.79
30	**79.86**	73.70	66.57	57.36	75.48	74.93
40	**77.53**	75.61	72.33	58.49	77.26	73.70
50	77.53	78.49	74.52	60.60	**80.00**	75.34
60	**81.40**	79.31	74.93	61.60	80.55	70.68
70	**85.78**	80.68	78.22	60.59	82.19	72.05
Ave.	**80.33**	76.39	70.73	58.95	77.15	73.58

4.2. Performance and Comparison

4.2.1. Facial expression recognition on the Cohn–Kanade dataset

Regarding to facial expression images of the CK database, we conducted two experimental cases with different numbers of training data for each expression. The first case was to take one image among five frames for training and the rest images for testing. In the case 2, two images of each expression from each person were collected to form the training dataset. The average recognition rates with different subspace dimensionalities are shown in Table I.

From Table I, there is the same trend between the number of training images and accuracy rate in most algorithms. That is, a smaller number of training data (i.e., smaller dimensionality) leads to a lower recognition rate. As shown in Table II, the facial expression recognition rates obtained in case 1 is lower than those in case 2 and the recognition rate increases as the dimensionality grows. Compared to other methods, the proposed EE-CMF framework yields the best results. The average recognition rates of NMF, Semi-NMF, Con-NMF, We-NMF, Ne-NMF, and the proposed EE-CMF in Case 2 are 94. 86%, 90.57%, 78.82%, 94. 64%, 94.17%, and 96.59%, respectively.

4.2.2. Facial expression recognition on the JAFFE dataset

The JAFFE dataset was also adopted to evaluate the FER performance. In the first case, one image of each expression per person was taken at random to construct the training data and the rest images were used in the test phase. Similarly, in case 2, two images were obtained for the training set. However, the JAFFE database is more challenging and many images within it are difficult to recognize. Tables III and IV show the recognition rate of each. All of the experimental results in Tables I-IV demonstrate that the proposed EE-CMF framework has much better performance than other existing methods.

5. CONCLUSION

This work proposes the exemplar-embed complex matrix factorization (EE-CMF), a subspace learning framework, in the complex domain. After transforming training images into the complex domain, the basis matrix is constructed by the linear combination of the complex training examples. The gradient descent method with Wirtinger's calculus was used to solve the complex matrix factorization problems. The proposed EE-CMF framework was tested on two facial expression datasets and very accurate recognition results are yielded. The proposed framework is much superior to the traditional and extension of NMF algorithms. Incorporating more constraints to widen applications and further improving the performance will be the future work. The complexification for tensor factorization is also a direction for future research.

6. REFERENCES

[1] B. Fasel and J. Luttine, “Automatic facial expression analysis: a survey,” *Pattern Recognition*, vol. 36, no. 1, pp. 259-275, 2003.

[2] K. Fukunaga, *Statistical pattern recognition*, Acadamic, 1990.

[3] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent component analysis*, Wiley Interscience, 2001.

[4] S. Noushath, G. H. Kumar, and P. Shivakumara, “Diagonal fisher linear discriminant analysis for efficient face recognition,” *Neurocomputing*, vol. 69, pp. 1711-1716, 2006.

[5] D. D Lee and H. S. Seung, “Learning the parts of objects by non-negative matrix factorization,” *Nature*, vol. 401, no. 6755, pp. 755-791, 1999.

[6] D. D. Lee and H. S. Seung, “Algorithms for non-negative matrix factorization,” in *Proc. NIPS*, 2000, pp. 556-562.

[7] P. Hoyer, “Non-negative matrix factorization with sparseness constraints,” *J. Mach. Learn*., vol. 5, pp. 1457-1469, 2004.

[8] Z. Yuan and E. Oja, “Projective nonnegative matrix factorization for image compression and feature extraction,” in *Proc. 14th Scandinavian Conf. Image Anal.*, 2005, pp. 333-342.

[9] S. Nikitidis, A. Tefas, and I. Pitas, “Using subclasses in discriminant non-negative subspace learning for facial expression recognition,” in *Proc. European Signal Processing Conference (EUSIPCO)*, 2011, pp. 1964-1968.

[10] S. Nikitidis, A. Tefas, and I. Pitas, “Facial expression recognition using clustering discriminant non-negative matrix factorization,” in *Proc. IEEE International Conference on Image Processing*, 2011, pp. 3001-3004.

[11] X. Chen and T. Huang, “Facial expression recognition: a clustering-based approach,” *Pattern Recognition Letters*, vol. 24, pp. 1295-1302, 2003.

[12] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “Eigenfaces vs. fisherfaces: recognition using class specific linear projection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, 1997.

[13] C. S. Lee and R. Chellappa, “Sparse localized facial motion dictionary learning for facial expression recognition,” in *Proc. IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)*, 2014, pp. 3548-3552.

[14] C. Ding, T. Li, and M. I. Jordan, “Convex and semi-nonnegative matrix factorizations,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 45-55, Jan. 2010.

[15] S. Liwicki, G. Tzimiropoulos, S Zafeiriou, and M. Pantic, “Euler principal component analysis,” *Int. J. Comput. Vis*., vol. 1, pp. 498-518, 2013.

[16] W. Wirtinger, “Zur formalen theorie der funktionen von mehrkomplexen veranderlichen,” *Mathematische Annalen*, vol. 97, no. 1, pp. 357-375, 1927.

[17] W. Hu, K. Choi, P. Wang, Y. Jiang, and S. Wang, “Convex nonnegative matrix factorization with manifold regularization,” *Neural Netw*., vol. 63, pp. 94-103, 2015.

[18] M. F. Amin, M. I. Amin, A. Al-Nuaimi, and K. Murase, “Wirtinger calculus based gradient descent and Levenberg Marquardt learning algorithms in complex-valued neural networks,” in *Proc. Int. Conf. Neural Information Processing ((ICONIP)*, 2011, pp. 550-559.

[19] A. Hjorunges and D. Gesbert, “Complex-valued matrix differentiation: techniques and key results," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2740-2746, 2007.

[20] Strang and Gilbert. *Linear algebra and its applications (4th ed.)*. Stamford, pp. 154-155, 2005.

[21] Moskowitz and Martin, *A course in complex analysis in one variable*, World Scientific Publishing Co., pp. 7, 2002.

[22] J. C. A. Barata and M. S. Hussein, “The Moore–Penrose pseudoinverse: a tutorial review of the theory,” *Brazilian Journal of Physics*, vol. 42, pp. 146-165, 2012.

[23] C. J. Lin, “Projected gradient methods for non-negative matrix factorization,” *Neural Comput*., vol. 19, pp. 2756-2779, 2007.

[24] K. B. Petersen and M. S. Pedersen, “The matrix cookbook,” 2008. Available: http://www2.imm.dtu.dk/pubdb/p.php?3274

[25] T. Kanade, J. Cohn, and Y. Tian, “Comprehensive database for facial expression analysis,” in *Proc. 4th IEEE Int. Conf. Autom. Face Gesture Recog.*, 2000, pp. 46-53.

[26] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, “Coding facial expressions with Gabor wavelets,” in *Proc. 3rd IEEE Conf. Autom. Face Gesture Recog*, 1998, pp. 200-205.

[27] D. Wang, T. Li, and C. Ding, “Weighted feature subset nonnegative matrix factorization and its applications to document understanding,” in *Proc. IEEE ICDM*, 2010, pp. 541-550.

[28] N. Guan, D. Tao, Z. Luo, and B. Yuan, “NeNMF: an optimal gradient method for non-negative matrix factorization”, *IEEE Trans. Signal Processing*, vol. 60, no. 6, pp. 2882-2898, 2012.