Chapter

Matrix Factorization on Complex Domain for Face Recognition

Viet-Hang Duong, Manh-Quan Bui and Jia-Ching Wang

Abstract

Matrix factorization on complex domain is a natural extension of nonnegative matrix factorization, but it is still a very new trend in face recognition. In this chapter, we present two complex matrix factorization-based models for face recognition, in which the objective functions are the real-valued functions of complex variables. Our first model aims to build a learned base, which is embedded within original space. The second model finds the base whose volume is maximized. Experimental results on datasets with and without outliers show that our proposed algorithms are more effective than competitive algorithms.

Keywords: complex matrix factorization, face recognition, nonnegative matrix factorization, projected gradient descent

1. Introduction

Face recognition is a central issue in computer vision and pattern recognition. The variations in lighting conditions, pose and viewpoint changes, facial expressions, makeup, aging, and occlusion are challenges that significantly affect recognition accuracy. Generally, the challenges in face recognition can be classified into four main categories as follows:

Illumination variations: The face of a person can appear dramatically different when illumination changes. This occurs because of spectra or source distribution and intensity changes. In practice, many two-dimensional (2D) methods show that recognition performance is notably decreased when illumination strongly occurs [1, 2]. Therefore, the problem of lighting variation is considered as one of the key challenges for face recognition system designer. Several methods have been proposed to handle variable illuminations such as extraction of illumination invariant features [3–7]; images with variable illuminations transformed to a canonical representation [8, 9]; modeling the illumination variations [10–11]; facial shapes and albedos are based on 3D face models [12].

Pose/viewpoint changes: Deformed face and self-occluded face usually occur by pose or viewpoint changes which affect the recognition process [13]. Generally, viewpoint face recognition approaches are divided into two categories: viewpoint-transformed and cross-pose based [14]. Viewpoint transformed recognition methods aim to transform the probe image to match the gallery image in the pose, whereas cross-pose-based approaches attempt to estimate the light field of the face [15, 16]. Besides, other approaches integrated 2D and 3D information [17, 18] in order to cope with pose and illumination variations.

Facial expression: Face recognition tasks are more challenging when dealing with emotional states of a person in an image. In addition, hairstyle or facial hair such as beard and mustache can change facial appearance. To handle with difficulties of expression, facial expression recognition (FER) systems, including static image FER [19–21], and dynamic sequence FER [22–24] are designed. In static-based methods, the spatial information from the current single image is extracted to obtain the feature representation. In contrary, the dynamic-based methods consider the temporal relation among adjacent frames in the sequence of input facial expression.

Occlusion: Faces may be partially occluded by other objects such as sunglasses, scarf [62], etc. Other situations of occlusion are some faces may be occluded by other faces of a group of people [25]. It is very difficult to be observed and recognized because the available part of the face is very small. Therefore, occlusion problems become harder and need to be solved in face recognition.

In face recognition, image representation (IR) techniques play an important role in improving the accuracy performance. Commonly, an IR system is to transform the input signal into a new representation which reduces its dimensionality and explicates its latent structures. Over the past decades, the subspace methods, such as principal component analysis (PCA) [26], linear discriminant analysis (LAD) [27, 28], and nonnegative matrix factorization (NMF) [29, 30] have been successfully used in feature extraction. In particularly, PCA is known as a powerful technique for dimensionality reduction and multivariate analysis. PCA seeks a linear combination of variables such that the maximum variance is extracted from the variables by projecting data onto an orthogonal base which is represented in the directions of largest variance. In image representation, eigenfaces (PCA) result in dense representations for facial images, which mainly applied the global structure of the whole facial image. Likewise, LAD finds a linear transformation that maximizes discrimination between classes.

NMF is known as an unsupervised data-driven approach in which all elements of the decomposed matrix and the obtained matrix factors are forced to be nonnegative. Furthermore, NMF is able to represent an object as various parts, for instance, a human face can be decomposed into eyes, lips, and other elements. In order to make NMF algorithms more efficient, one has proposed some constraints into the cost function such as sparsity [31, 32], orthogonally [33], discrimination [34], graph regularization [35, 36], and pixel dispersion penalty [37]. Additionally, proposing an appropriate distance metric for an NMF model plays an important role in enhancing the efficacy of the estimated linear subspace of the given data. NMF techniques commonly apply the squared Frobenius norm (Fr) or the generalized Kullback–Leibler (KL) divergence for the independent and identically distributed noise data. But in many cases, they produce an arbitrarily biased subspace when data is corrupted by outliers [38]. To overcome this drawback, L_2 and L_1 norms were proposed by Kong et al. [39] to obtain a robust NMF, in which the noise was assumed to follow the Laplacian distribution. Similarly, the earth mover's distance (EMD) and the Manhattan distance were also suggested in the work of Sandler et al. [40] and Guan et al. [41], respectively. A family of cost functions parameterized by a single shape parameter beta, called the beta-divergence [42], is commonly used on NMF approaches. Although NMFs are able to learn part-based representations and capture the Euclidean structure of high-dimensional data space, they are still limited to comprise the nonlinear sub-manifold structure behind the data.

Recently, matrix factorization techniques have been extended to complex matrix factorizations (CMFs) where the input data are complex matrices. These models have been obtaining promising results in facial expression recognition and data representation tasks [43–45]. The main idea of complex methods for face and facial expression recognition is that the original signal is projected on to the complex

field by a mapping such that the distances of two data points in the original space and projection space are equivalent. Particularly, by transforming the real values of pixel intensive to complex domain, it is shown that the squared Frobenius norm of corresponding complex vectors and the cosine dissimilarity of real-valued vectors are equivalent. As a result, the real optimization problem with cosine divergence is replaced by optimizing a complex function with the Frobenius norm. Most of the mentioned CMF models were applied to facial expression and object recognition.

In this chapter, we present two complex matrix factorization-based models for face recognition. In the following sections, we denote M-dimensional column vector $\mathbf{y} = (y_1, ..., y_M)^T \in \mathbb{R}_+^M$ to be an observed sample. Let $\mathbf{Y}$ be a dataset comprising of N-observations; $\mathbf{Y}$ is expressed in the matrix form as $\mathbf{Y} = (\mathbf{y}_1, ..., \mathbf{y}_N) \in \mathbb{R}_+^{M \times N}$, where $\mathbb{R}_+$ denotes the set of nonnegative real numbers. In the proposed models, the real data set $\mathbf{Y}$ is transformed to the complex domain, and the complex data matrix $\mathbf{Z}$ is factorized under imitating NMF frameworks. The contributions of this chapter are summarized as follows:

1. The image analysis methods on the complex domain, which are called structured complex matrix factorization (StCMF) and constrained complex matrix factorization (CoCMF), are proposed.

2. In complex domain, the updating rule for StCMF and CoCMF is derived based on gradient descent method.

3. A thorough experimental study on face recognition is conducted, the results show that the proposed StCMF and CoCMF yield better performance compared to extensions of the real NMFs.

2. Background

2.1 Nonnegative matrix factorization

Assume that we are given an initial data matrix $\mathbf{Y} \in \mathbb{R}_+^{M \times N}$ and a positive integer $K \ll \min\{\mathbf{M}, \mathbf{N}\}$. NMF methods aim to find a basis matrix $\mathbf{U} \in \mathbb{R}_+^{M \times K}$ and a coding variable matrix $\mathbf{V} \in \mathbb{R}_+^{K \times N}$, such that $\mathbf{Y} \approx \mathbf{UV}$. The standard NMF is usually formulated as an optimization:

$$\min_{\mathbf{U},\,\mathbf{V}} D(\mathbf{Y} \| \mathbf{UV}) \text{ s.t.} \mathbf{U} \geq 0, \mathbf{V} \geq 0 \tag{1}$$

where $D(\mathbf{Y} \| \mathbf{UV})$ is a divergence function to measure the distance between $\mathbf{Y}$ and $\mathbf{UV}$.

Most NMF techniques estimate the linear subspace of the given data by the Frobenius norm (F) or the generalized Kullback–Leibler (KL) divergence which have the following forms:

$$D_F(\mathbf{A} \| \mathbf{B}) = \|\mathbf{A} - \mathbf{B}\|_F^2 = \sum_{i,j} \left(\mathbf{A}_{ij} - \mathbf{B}_{ij}\right)^2 \tag{2}$$

$$D_{KL}(\mathbf{A} \| \mathbf{B}) = \lim_{\beta \to 0} D_\beta(\mathbf{A} \| \mathbf{B}) = \sum_{i,j} \mathbf{A}_{ij} \log \frac{\mathbf{A}_{ij}}{\mathbf{B}_{ij}} - \mathbf{A}_{ij} + \mathbf{B}_{ij} \tag{3}$$

The problem (1) is non-convex; thus, it may result in several local minimal solutions. To find an optimization solution, the iterative methods are commonly used.

Generally, there are three classes of algorithms for solving this problem including multiplicative update, gradient descent, and alternating nonnegative least squares algorithms. The most popular approach to solve (1) is the multiplicative update rules proposed by Lee and Seung [30]. For example, the iteratively updating rules of a Frobenius NMF cost function are given by

$$\mathbf{V}_{ij}^{(t)} \leftarrow \mathbf{V}_{ij}^{(t-1)} \frac{\left(\mathbf{U}^{(t-1)T}\mathbf{Y}\right)_{ij}}{\mathbf{U}_{ij}^{(t-1)^T}\left(\mathbf{U}^{(t-1)}\mathbf{V}\right)_{ij}}; \tag{4}$$

$$\mathbf{U}_{ij}^{(t)} \leftarrow \mathbf{U}_{ij}^{(t-1)} \frac{\left(\mathbf{Y}\mathbf{V}^{(t-1)T}\right)_{ij}}{\left(\mathbf{U}^{(t-1)}\mathbf{V}\right)_{ij}\mathbf{V}_{ij}^{(t-1)^T}}; \tag{5}$$

2.2 The cosine divergence

Given the representations of two images, $\mathbf{I}_t$ and $\mathbf{I}_s$ are M-dimensional vectors $\mathbf{y}_t$, $\mathbf{y}_s$ in the lexicographic order, respectively. First, $\mathbf{y}_t, \mathbf{y}_s \in \mathbb{R}^M$ is normalized to get the values $\mathbf{y}_t(c), \mathbf{y}_s(c) \in [0,1]$, where c is the element vector index or the vector spatial location. The correlation between images $\mathbf{I}_t$ and $\mathbf{I}_s$ through the cosine dissimilarity between $\mathbf{y}_t$ and $\mathbf{y}_s$, is introduced by

$$D_C(\mathbf{y}_t, \mathbf{y}_s) = \sum_{c=1}^{M} \left\{1\text{- }\cos\left(\alpha\pi\,\mathbf{y}_t(\mathrm{c})\text{-}\alpha\pi\,\mathbf{y}_s(\mathrm{c})\right)\right\} \tag{6}$$

One of interesting properties of the cosine distance measurement is suppression outlier which is proved in [46]. The comparison between Frobenius norm and cosine divergence is showed in **Figure 1**. Liwiki et al. [46] show that the Frobenius distance between the original and the same subject is smaller; in contrary, a large distance between the original image and the image of a different person or occlusion image results from the cosine-based measure.

2.3 Euler's formula and a space transformation

Let us consider two mappings:
$g : \mathbb{R}^M \rightarrow \mathbb{R}^{2M}$ such that

$$g(\mathbf{y}_t) = \frac{1}{\sqrt{N}}\left[\cos(\mathbf{y}_t)^T \ \sin(\mathbf{y}_t)^T\right]^T; \forall \mathbf{y}_t \in \mathbb{R}^N \tag{7}$$

where
$$\cos(\mathbf{y}_t) = \left[\cos(\mathbf{y}_t(1)), \cos(\mathbf{y}_t(2)), ..., \cos(\mathbf{y}_t(M))\right]^T \tag{8}$$

$$\sin(\mathbf{y}_t) = \left[\sin(\mathbf{y}_t(1)), \sin(\mathbf{y}_t(2)), ..., \sin(\mathbf{y}_t(M))\right]^T \tag{9}$$

Figure 1.
Sample images for making comparison between dissimilarity measures.

$$\left\|g\left(\mathbf{y}_t\right)\right\| = 1 \tag{10}$$

and $h : \mathbb{R}^M \to \mathbb{C}^M$ is defined by

$$\mathbf{z}_t = h\left(\mathbf{y}_t\right) = \frac{1}{\sqrt{2}} e^{\,i\alpha\pi\mathbf{y}_t} = \frac{1}{\sqrt{2}} \begin{bmatrix} e^{i\alpha\pi\mathbf{y}_t(1)} \\ \vdots \\ e^{i\alpha\pi\mathbf{y}_t(M)} \end{bmatrix} \tag{11}$$

The nonlinear function h is to transform the real-valued features to complex feature space. In other words, a complex vector space with M-dimensions can be regarded as a $2M$-dimensional real vector space.

It is proven that the cosine dissimilarity distance of a pair of data in the input real space equals to the Frobenius distance of the corresponding data in complex domain [47]. This observation is the first motivation of StCMF and CoCMF by mapping the samples into the complex space with a nonlinear mapping function h and performing matrix factorization in this complex feature space.

2.4 Wirtinger calculus

Any function of a complex variable z can be defined as $f(z)|_{z=x+iy} = F(x,y) = U(x,y) + i\,V(x,y)$, where $i^2 = -1$ and x, $\mathrm{y} \in \mathbb{R}$. Palka et al. [48] defined the complex differentiability as follows:

Definition 1. Let $\mathrm{A} \subset \mathbb{C}$ be an open set. The function $f : \mathrm{A} \to \mathbb{C}$ is said to be differentiable at $z_0 \in \mathrm{A}$ if there is a limit $\lim_{z \to z_0} \frac{f(z)-f(z_0)}{z-z_0}$ which exists independently on the manner where $z \to z_0$.

A necessary condition for f being holomorphic is that the Cauchy-Riemann equations hold, that is, $\frac{\partial U}{\partial x} = \frac{\partial V}{\partial y}$ and $\frac{\partial U}{\partial y} = -\frac{\partial V}{\partial x}$; otherwise, it is nonholomorphic. In statistical signal processing, the functions of interest are real-valued and have complex arguments z and hence are not analytic on complex plane. In this case we can use Wirtinger calculus [49], which writes the expansions in conjugate coordinate system by considering the function $f(z)$ as a bivariate function $f(z, z^*)$ and treating z and z^* as independent arguments.

Definition 2. The pair of partial derivative operators for function $f(z) = f(z, z^*)$ referred to as the Wirtinger derivative [49] is defined by

$$\frac{\partial f}{\partial z} = \frac{1}{2}\left(\frac{\partial f}{\partial x} - i\frac{\partial f}{\partial y}\right), \quad \frac{\partial f}{\partial z^*} = \frac{1}{2}\left(\frac{\partial f}{\partial x} + i\frac{\partial f}{\partial y}\right) \tag{12}$$

In case of real-valued function of complex variables, we also have one special property which is useful for optimization theory described later.

Lemma 1. The differential df of a real-valued function $f : \mathrm{A} \to \mathbb{R}$ with complex valued $z \in A \subset \mathbb{C}$ can be expressed as

$$df = 2\,\mathrm{Re}\left(\frac{\partial f(z)}{\partial z^*} dz\right) \tag{13}$$

3. Complex matrix factorization

Let the input data matrix $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_N)$ contain N data vectors as columns. As described in previous sections, the elements of real matrix $\mathbf{Y}$ are normalized and transformed into a complex number field to yield the complex data matrix $\mathbf{Z}$. Two unconstraint and constrained optimization problems in an unordered complex field is introduced in the following sections, respectively.

3.1 Structured complex matrix factorization (StCMF)

The idea of structured complex matrix factorization (StCMF) is to build a learned base which is embedded within original space. The basis matrix in StCMF is constructed by the linear combination of the complex training examples. Given the complex data matrix $\mathbf{Z} \in \mathbb{C}^{M \times N}$, StCMF factorizes $\mathbf{Z}$ into the encoding matrix $\mathbf{V} \in \mathbb{C}^{K \times N}$ and the exemplar-embed basis matrix $\mathbf{U} = \mathbf{ZW}$ where $\mathbf{W} \in \mathbb{C}^{M \times K}$. Therefore, the objective function of StCMF problem can be formulated as follows:

$$\min_{\mathbf{W},\,\mathbf{V}} f_{StCMF}(\mathbf{W}, \mathbf{V}) = \min_{\mathbf{W},\,\mathbf{V}} \frac{1}{2} \|\mathbf{Z} - \mathbf{ZWV}\|_F^2 \tag{14}$$

where $\|.\|_F$ denotes the Frobenius norm and $K \ll \min\{N, M\}$

$$\begin{aligned} \text{and } \|\mathbf{Z} - \mathbf{ZWV}\|_F^2 &= Tr(\mathbf{Z} - \mathbf{ZWV})^H (\mathbf{Z} - \mathbf{ZWV}) \\ &= Tr\left(\mathbf{Z}^H\mathbf{Z} - \mathbf{V}^H\mathbf{W}^H\mathbf{Z}^H\mathbf{Z} - \mathbf{Z}^H\mathbf{ZWV} + \mathbf{V}^H\mathbf{W}^H\mathbf{Z}^H\mathbf{ZWV}\right) \end{aligned}$$

3.2 Constrained complex matrix factorization (CoCMF)

Considering a dataset of N complex vectors $\mathbf{Z} = [\mathbf{Z}_1, \mathbf{Z}_2, ..., \mathbf{Z}_N]$, each of $\mathbf{Z}_i$ represents a data instance. The proposed CoCMF model decomposes $\mathbf{Z}$ into a product of two matrices $\mathbf{W}$ and $\mathbf{V}$ such that each instance $\mathbf{Z}_i$ is a convex combination of latent components $\mathbf{W}$. We call $\mathbf{V}$ and $\mathbf{W}$ the encoding matrix and the basis matrix, respectively. Geometrically, the data points $\mathbf{Z}_i$, i = 1, 2, ..., N all lie in or on the surface of a simplicial cone $\mathbf{S_W}$, whose vertices correspond to the columns of $\mathbf{W}$ and

$$\mathbf{S_W} = \left\{ \mathbf{z} : \mathbf{z} = \sum_{i=1}^{K} \mathbf{W}_i \mathbf{v}_i; \mathbf{v}_i \in \mathbb{R}_+ \right\} \tag{15}$$

Note that $\mathbf{S_W}$ lies in the positive orthant and the volume of $\mathbf{S_W}$ (Vol ($\mathbf{S_W}$)) is given by the following formula [48]:

$$Vol(S_{\mathbf{W}}) = \frac{|\det(\mathbf{W})|}{(K-1)!} \tag{16}$$

In [51], Zhou et al. illustrated that the small-cone constraint on the bases $\mathbf{W}$ will impose suitable sparseness on $\mathbf{V}$. Inversely, the large-cone penalty will result in sparseness on the bases of factorization and the reconstruction errors on the training data, and the test data will be simultaneously decreased [50, 52]. Therefore, all observed data can be reconstructed by linearly combining the bases of a dictionary. Combining the goals of enlarging the volume of the simplex base, the constrained complex matrix factorization (CoCMF) problem is formulated as follows:

$$\min_{\mathbf{W},\,\mathbf{V}} f_{CoCMF}(\mathbf{W},\mathbf{V}) = \min_{\mathbf{W},\,\mathbf{V}} \frac{1}{2}\|\mathbf{Z}-\mathbf{W}\mathbf{V}\|_F^2 - \frac{|\det(\mathbf{W})|}{(K-1)!} \tag{17}$$

$$\text{s.t } \mathbf{W} \in \mathbb{C}^{M\times K},\ \mathbf{V} \in \mathbb{R}_+^{K\times N} \text{ and } \sum_{i=1}^{K} \mathbf{V}_{ij} = 1\ \forall j$$

Since $0 < \det(\mathbf{W}^{\mathrm{T}}\mathbf{W}) \leq 1$ holds under the assumptions $1^{\mathrm{T}}\mathbf{W}_i = 1$. To simply the model, in this work, the log-determinant function is exploited to modify the volume penalty, and Eq. (17) can be written as the following form:

$$\min_{\mathbf{W},\,\mathbf{V}} f_{CoCMF}(\mathbf{W},\mathbf{V}) = \min_{\mathbf{W},\,\mathbf{V}} \frac{1}{2}\|\mathbf{Z}-\mathbf{W}\mathbf{V}\|_F^2 - \log\left(\det\left(\mathbf{W}^T\mathbf{W}\right)\right) \tag{18}$$

$$\text{s.t } \mathbf{W} \in \mathbb{C}^{M\times K},\ \mathbf{V} \in \mathbb{R}_+^{K\times N}\ \sum_{i=1}^{K} \mathbf{V}_{ij} = 1, \text{ and } \sum_{i=1}^{K} \left|\mathbf{W}_{ij}\right| = 1\forall j$$

3.3 Complex matrix factorization via projected gradient descent

It can be seen that (12) and (16) are non-convex minimization problems with respect to both variables $\mathbf{W}$ and $\mathbf{V}$, so they are impractical to obtain the optimal solution. These NP-hard problems can be tackled by applying the block coordinate descent (BCD) with two matrix blocks [53] to obtain a local solution. The specific problems (14) and (18) were solved by the following scheme:

Fixing $\mathbf{W}$ and solving the following one variable optimization problems

$$\min_{\mathbf{V}} f_{StCMF\_V}(\mathbf{V}) = \min_{\mathbf{V}} \frac{1}{2}\|\mathbf{Z}-\mathbf{Z}\mathbf{W}\mathbf{V}\|_F^2 \tag{19}$$

$$\min_{\mathbf{V}} f_{CoCMF\_V}(\mathbf{V}) = \min_{\mathbf{V}} \frac{1}{2}\|\mathbf{Z}-\mathbf{W}\mathbf{V}\|_F^2 \tag{20}$$

$$\text{s.t } \mathbf{V} \in \mathbb{R}_+^{K\times N},\ \sum_{i=1}^{K} \mathbf{V}_{ij} = 1\ \forall j$$

Then, $\mathbf{W}$ is updated based on the Moore-Penrose pseudoinverse [54], which is dented by † and $\mathbf{W} = (\mathbf{Z}^{\dagger}\mathbf{Z})\mathbf{V}^{\dagger}$ for Eq. (14) and $\mathbf{W} = \mathbf{Z}\mathbf{V}^{\dagger}$ for Eq. (18) with fixed $\mathbf{V}$.

Taking advanced of Wirtinger calculus, the gradient is evaluated in the forms

---

Algorithm 1: Complex projected gradient (CPG) with Armijo rule

---

Input: $\mathbf{Z}$, $\mathbf{W}$

Output: $\mathbf{v}$

1. Initialize any feasible $\mathbf{V}_0$, $0 < \beta < 1$, $0 < \sigma < 1$

2. Iterations, for k = 1, 2, ...

$$\mathbf{V}_{k+1} = P[\mathbf{V}_k - \alpha_k \nabla_{\mathbf{V}^*} f(\mathbf{W},\mathbf{V}_k)]$$

where $\alpha_k = \mu^{t_k}$, t_k is the first nonnegative integer such that

$$f(\mathbf{W},\mathbf{V}_{k+1}) - f(\mathbf{W},\mathbf{V}_k) \leq 2\sigma \mathrm{Re}\{\langle \nabla_{\mathbf{V}^*} f(\mathbf{W},\mathbf{V}_k), \mathbf{V}_{k+1} - \mathbf{V}_k\rangle\}$$

$$\nabla_{\mathbf{V}} f_{StCMF\_V}(\mathbf{V}) = -\mathbf{W}^H\mathbf{Z}^H\mathbf{Z} + \mathbf{W}^H\mathbf{Z}^H\mathbf{Z}\mathbf{W}\mathbf{V} \tag{21}$$

$$\nabla_{\mathbf{V}} f_{CoCMF\_V}(\overline{\mathbf{V}}) = \mathbf{W}^H\mathbf{W}\,\overline{\mathbf{V}} - \mathbf{W}^H\mathbf{Z} \tag{22}$$

$$\text{where } \overline{\mathbf{V}} = \left[\frac{\mathbf{V}_1}{\|\mathbf{V}_1\|_1}, \frac{\mathbf{V}_2}{\|\mathbf{V}_2\|_1}, \ldots, \frac{\mathbf{V}_N}{\|\mathbf{V}_N\|_1}\right]; \overline{\mathbf{V}} \geq 0 \tag{23}$$

We summarize the projected gradient method for optimizing (21) and (22) in **Algorithm 1**.

4. Experiments

To investigate the recognition performance of the proposed StCMF and CoCMF methods, we have conducted extensive experiments on the ORL dataset [55] and the Georgia Tech face dataset [56] in two scenarios for face recognitions including holistic face and key point occluded face.

First, we give brief description about the data collections and experiment setting. Second, the performance comparisons and corresponding results are shown.

4.1 Datasets and experiment setting

The ORL dataset contains 400 grayscale images corresponding to 40 people's face. The images were captured at different times, under different lighting conditions, with different facial expression (open or close eyes, smiling or non-smiling) and facial details (glasses or no glasses). All the face images are manually aligned and cropped. For the computational efficiency, each cropped image is resized to 28×23 for face recognition without occlusion and 32×32 pixels for face recognition with occlusion. **Figure 2** shows some instances of such random face on ORL dataset.

The Georgia Tech face dataset (GT) contains images of 50 people taken during 1999 and stored in JPEG format. For each individual, there are 15 color images captured at resolution of 640×480 pixels. Most of the images were taken in two different sessions to take into account the variations in illumination conditions, facial expression, and appearance. In our experiments, original images are normalized, cropped and scaled into 31×23 pixels, and finally converted into gray level images. Examples of GT dataset are shown in **Figure 3**.

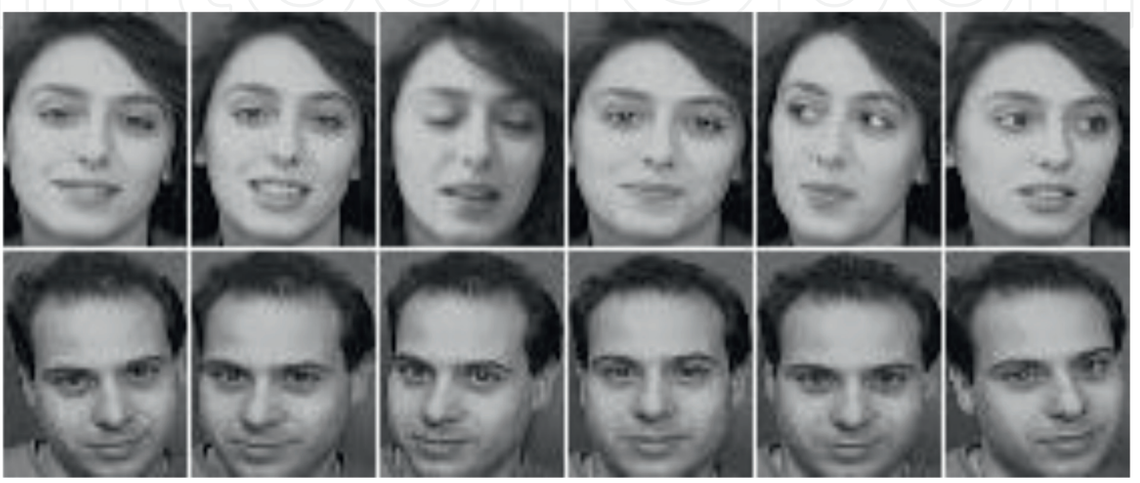

Figure 2.
Sample facial images from ORL dataset [55].

Figure 3.
Sample facial images from GT dataset [56].

We use the nearest neighbor (NN) classifier for all face recognition with/without occlusion experiments. The platform was a 3.0 GHz Pentium V with 1024 MB RAM running Windows. Code was written in MATLAB.

4.2 Performance and comparison

4.2.1 Face recognition on ORL dataset

For this case, in order to evaluate the performance of the proposed StCMF and CoCMF, we make the comparisons with seven representative algorithms, namely, NMF [29], P-NMF [57], P-NMF (Fr) [58], P-NMF (KL) [58], OPNMF (Fr) [59], OPNMF (KL) [59], NNDSVD-NMF [60], and GPNMF [60]. Different training numbers ranging from five to nine images were randomly chosen from each individual to construct the training set, and the rest images constitute the test set which was used to estimate the accuracy of face recognition [61]. The learning basic images in all selected algorithms are $K = 40$, and the mean recognition rate are described in **Table 1**.

Table 1 shows the detailed recognition accuracies of compared algorithms. As can be seen, our algorithms significantly outperform the other algorithms in all the cases. Almost algorithms achieve the best accuracy when the number of training face images per class is eight exceptionally our proposed methods and GPNMF. Besides, there is the same trend between the number of training images and accuracy rate; that is, the lower training numbers lead to a decreasing rate of

No. Trains	StCMF	CoCMF	GPNMF	NMF	PNMF	P-NMF (Fr)	P-NMF (KL)	OPNMF (Fr)	OPNMF (KL)	NNDSVD-NMF
5	**90.85**	**90.30**	86.5	84.5	82.4	83.7	85.0	80.0	79.0	43.0
6	**91.75**	**92.25**	87.5	84.4	85.81	85	84.4	83.0	82.0	39.3
7	**91.17**	**94.75**	87.5	83.3	87.33	85.6	85.9	84.4	80.0	36.8
8	**93.75**	**93.88**	88.75	88.75	88.5	88.8	88.0	84.3	83.0	40.8
9	**97.50**	**95.50**	92.5	85	90.75	87.25	87.5	84.0	83.0	42.3
Avg.	**93.00**	**93.34**	88.55	85.19	86.96	86.07	86.16	83.14	81.4	40.44

Table 1.
Face recognition accuracy on the ORL dataset with different train numbers.

recognition. StCMF achieves the best performance (97.50%) when the number of training samples is chosen largest. However, CoCMF achieves higher improvement in general.

It is observed that the above-selected algorithms employ a different kind of measurements such as Frobenius (Fr) and Kullback–Leibler (KL) and add more graph to regularize as well as adjust basic NMF to projective NMF. In a reprocessing image, centered aligning image technique is applied for other methods to enhance effective recognition rate that cannot be focused on our StCMF and CoCMF models. However, the best recognition rate of all obtained by our proposed CoCMF method which has extra regularizes term.

One of the difficulties in NMF is the estimation of the number of components or K. The choice of K results in a compromise between data fitting and model complexity; that is, a greater K leads to a better data approximation, but a smaller K makes a model being easier to estimate and fewer parameters to transmit. In almost NMFs, K is typically chosen such that it is larger than the estimated number of sources and follows the constraint $(N + M)K \ll NM$. This limit of NMFs illustrated by the observation that among all results, the lowest rate belongs to NNDSVD-NMF, one NMF method utilizes SVD to get initialization which results from significant independency of NNDSVD-NMF on the number of bases K.

4.2.2 Face recognition on GT dataset

Table 2 shows the recognition rates versus feature dimension by the competing methods on GT dataset**.** GT dataset exists with many challenging samples that are harder to recognize. Thus, the performance of all methods is lower than those of ORL dataset. In this dataset, the implement was done similarly as those in the previous section in choosing algorithms to compare as well as dividing randomly into two different sets, each containing a different number of testing and training images. In our experiments, we set K = 50 and range the number training being five odd numbers as {5, 7, 9, 11, 13}. The experimental results show that as the number of training images increases, the efficiency of the recognition system also increases. We can see that CoCMF method achieves the best performance and StCMF holds the second place in overall. All the methods obtain their best results when 13 training samples are used (the largest number of training sample in our experiment). In this case, the highest recognition rate belongs to the StCMF method again.

4.2.3 Face recognition on occluded ORL images

For a more convincing experimental assessment of the power of our proposed models in occlusion processing, we test the performance on occluded images of

No. Trains	StCMF	CoCMF	GPNMF	NMF	PNMF	P-NMF (Fr)	P-NMF (KL)	OPNMF (Fr)	OPNMF (KL)	NNDSVD-NMF
5	**39.64**	**59.40**	59.14	54.70	46.84	58.90	57.97	57.89	48.08	23.80
7	**54.80**	**62.25**	60.96	59.38	52.50	60.20	60.88	60.44	48.68	23.83
9	**75.20**	**69.67**	62.5	62.40	54.93	64.03	63.35	62.48	48.84	24.30
11	**69.50**	**70.50**	65.37	65.20	57.25	63.75	63.38	63.17	49.36	27.35
13	**77.60**	**73.00**	69.00	67.40	61.60	65.60	64.05	63.50	49.50	30.20
Avg.	**63.35**	**66.96**	63.39	61.82	54.63	62.50	61.93	61.50	48.90	25.90

Table 2.
Face recognition accuracy on the GT dataset with different train numbers.

Figure 4.
Occluded face samples from ORL dataset with patch sizes of 15 × 15, 20 × 20, 25 × 25, 30 × 30, and 35 × 35.

Occluded Size	StCMF	CoCMF	GPNMF	NMF	PNMF	P-NMF (Fr)	P-NMF (KL)	OPNMF (Fr)	OPNMF (KL)	NNDSVD-NMF
15×15	**79.58**	**80.21**	75.16	74.32	72.55	69.16	71.25	74.18	45.16	54.46
20×20	**72.08**	**73.79**	64.52	65.45	62.15	67.52	71.23	65.00	41.52	25.62
25×25	**70.00**	**71.17**	65.54	55.18	52.38	65.54	62.19	55.00	35.54	19.83
30×30	**52.08**	**61.54**	54.53	45.62	43.87	48.53	55.21	45.89	28.53	13.22
35×35	**39.17**	**41.00**	43.25	33.63	31.06	43.25	38.79	33.39	23.25	16.13
Avg.	**62.58**	**65.54**	60.60	54.84	52.40	58.80	59.73	54.69	34.80	25.85

Table 3.
Face recognition accuracy on the occluded ORL image with different occlusion sizes.

ORL database. In cropped 112 × 92 dimension test image gallery, occlusion was simulated by using a sheltering patch with different size ranges in set {10 × 10, 15 × 15, 20 × 20, 25 × 25, 30 × 30} and placed at random locations before resized in 28 × 21. **Figure 4** shows examples of occluded ORL images.

In this experiment, we take randomly the training images with the ratio 4:6 for training/testing and test several times on each sort of percent of randomly occluded test image. **Table 3** shows the detailed recognition accuracy on all selected algorithms and our proposed methods. It can be seen that the recognition rate of all methods is increased when the size of occlusion batch is decreased. Obviously, StCMF and CoCMF outperform other tested approaches even if occlusion. This reveals that StCMF and CoCMF are more robust outlier than the other.

5. Summary and discussion

In this paper, we have proposed a new approach to complex matrix factorization to face recognition. Preliminary experimental results show that StCMF and CoCMF achieve promising results for face recognition by utilizing the robustness of cosine-based dissimilarity and extend the main spirits of NMF from real number field to complex field which adds flexible constraints for the real-valued function of complex variables. We have also noted how strong is the proficiency of StCMF as well as CoCMF on face recognition task. Our proposed methods are simple frameworks which do not need more complicated regularizes like NMFs in the real domain. We believe that this capability of proposed methods will be stable in other application tasks. In future work, three aspects of the proposed system will be centered on. First, we add more regularized rules into objective function to a range of further application such as speech and sound processing. Second, we employ other classifiers such as complex neural network or complex SVM to treat well the complex-valued feature. Last, kernel methods will be exploited in both feature extraction and classification of StCMF and CoCMF constructed paradigm to develop the performance of nonlinear contexts.

Acknowledgements

This research is partially supported by the Ministry of Science and Technology under Grant Number 108-2634-F-008-004 through Pervasive Artificial Intelligence Research (PAIR) Labs, Taiwan.

Author details

Viet-Hang Duong[1], Manh-Quan Bui[2] and Jia-Ching Wang[2,3]*

1 Faculty of Information Technology, BacLieu University, VietNam

2 Department of Computer Science Information Engineering, National Central University, Taiwan

3 Pervasive Artificial Intelligence Research (PAIR) Labs, Taiwan

*Address all correspondence to: jcw@csie.ncu.edu.tw

References

[1] Batur AU, Hayes MH. Segmented linear subspaces for illumination robust face recognition. International Journal of Computer Vision. 2004;**57**(1):49-66

[2] Chen T, Yin W, Zhou X, Comaniciu D, Huang T. Total variation models for variable lighting face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2006;**28**(9): 1519-1524

[3] Gao XY, Maylor KHL. Face recognition using line edge map. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2002;**24**(6): 764-779

[4] Guo B, Lam KM, Lin KH, Siu WC. Human face recognition based on spatially weighted Hausdorff distance. Pattern Recognition Letters. 2003;**24**: 499-507

[5] Adini Y, Moses Y, Ullman S. Face recognition: The problem of compensating for changes in illumination direction. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2001;**19**(7): 721-732

[6] Lee KC, Ho J, Kriegman D. Acquiring linear subspaces for face recognition under variable lighting. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2005;**27**(5): 684-698

[7] Han H, Shan S, Chen X, Gao W. A comparative study on illumination preprocessing in face recognition. Pattern Recognition. 2013;**46**(6): 1691-1699

[8] Zhao W, Chellappa R. SFS based view synthesis for robust face recognition. In: Proc. the 4th Conference on Automatic Face and Gesture Recognition. 2000

[9] Shashua A, Tammy RR. The quotient image: Class-based rerendering and recognition with varying illuminations. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2001;**23**(2): 129-139

[10] Georghiades AS, Belhumeur PN, Kriegman DJ. From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2001;**23**(2):643-660

[11] Ishiyama R, Sakamoto S. Geodesic illumination basis: compensating for illumination variations in any pose for face recognition. In: Proc. the 16th International Conference on Pattern Recognition. Vol. 4. 2002. pp. 297-301

[12] Gao W, Shan SG, Chai XJ, Fu XW. Virtual face generation for illumination and pose intensitive face recognition. In: IEEE International Conference on Acoustics, Speech, and Signal Processing. Vol. 4. pp. 776-779

[13] Ho H, Chellappa R. Pose-invariant face recognition using Markov random fields. IEEE Transactions on Image Processing. 2013;**22**(4):1573-1584

[14] Blanz V, Grother P, Phillips PJ, Vetter T. Face recognition based on frontal views generated from non-frontal images. In: IEEE Conf. Computer Vision and Pattern Recognition. 2005. pp. 454-461

[15] Gross R, Matthews I, Baker S. Appearance-based face recognition and light-fields. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2004;**26**(4):449-465

[16] Wu G, Masia B, Jarabo A, Zhang Y, Wang L, Dai Q, et al. Light field image processing: An overview. IEEE Journal

of Selected Topics in Signal Processing. Special Issue on Light Field Image Processing. 2017

[17] Malassiotis S, Strintzis M. Robust face recognition using 2D and 3D data: Pose and illumination compensation. Pattern Recognition. 2005;**38**(12): 2537-2548

[18] Asthana A, Marks T, Jones M, Tieu K, Rohith M. Fully automatic pose-invariant face recognition via 3D pose normalization. IEEE International Conference on Computer Vision (ICCV 2011). 2011:937-944

[19] Shan C, Gong S, McOwan PW. Facial expression recognition based on local binary patterns: A comprehensive study. Image and Vision Computing. 2009;**27**(6):803-816

[20] Liu P, Han S, Meng Z, Tong Y. Facial expression recognition via a boosted deep belief network. In: Proc. the IEEE Conference on Computer Vision and Pattern Recognition. 2014. pp. 1805-1812

[21] Mollahosseini A, Chan D, Mahoor MH. Going deeper in facial expression recognition using deep neural networks. In: Proc. IEEE Winter Conference on Applications of Computer Vision. 2016. pp. 1-10

[22] Zhao G, Pietikainen M. Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2007;**29**(6):915-928

[23] Jung H, Lee S, Yim J, Park S, Kim J. Joint fine-tuning in deep neural networks for facial expression recognition. In: Proc. IEEE International Conference on Computer Vision (ICCV). 2015. pp. 2983-2991

[24] Zhao X, Liang X, Liu L, Li T, Han Y, Vasconcelos N, et al. Peak-piloted deep network for facial expression recognition. In: European Conference on Computer Vision. Springer; 2016. pp. 425-442

[25] Aleix MM. Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2002;**24**(6):748-763

[26] Fukunaga K. Statistical Pattern Recognition. Acadamic; 1990

[27] Hyvarinen A, Karhunen J, Oja E. Independent Component Analysis. Wiley Interscience; 2001

[28] Belhumeur PN, Hespanha JP, Kriegman DJ. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1997;**19**(7):711-720

[29] Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. Nature. 1999;**401**(6755): 755-791

[30] Lee DD, Seung HS. Algorithms for non-negative matrix factorization. In: Proc. NIPS. 2000. pp. 556-562

[31] Hoyer P. Non-negative sparse coding. In: Proc. IEEE Neural Networks for Signal Processing. 2002. pp. 557-565

[32] Hoyer P. Non-negative matrix factorization with sparseness constraints. Journal of Machine Learning Research. 2004;**5**:1457-1469

[33] Li H, Adal T, Wang W, Emge D, Cichocki A. NMF with orthogonality constraints and its application to Raman spectroscopy. VLSI. 2007;**48**:83-97

[34] Guan N, Tao D, Luo Z, Yuan B. Manifold regularized discriminative nonnegative matrix factorization with fast gradient descent. IEEE Transactions on Image Processing. 2011;**20**(7): 2030-2048

[35] Cai D, He XF, Wu X, Han JW. Non-negative matrix factorization on manifold. In: Proc. IEEE Int'l Data Mining (ICDM '08). 2008. pp. 63-72

[36] Cai D, He XF, Wu X, Han JW, Huang TS. Graph regularized non-negative matrix factorization for data representation. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2011;**33**(8):1548-1560

[37] Duong VH, Lee YS, Pham BT, Mathulaprangsan S, Bao PT, Wang JC. Spatial dispersion constrained nmf for monaural source separation. In: Proc. the 10th International Symposium on Chinese Spoken Language Processing (ICSLP). 2016

[38] Cichocki A, Zdunek R, Amari S. Csisz'ar's divergences for non-negative matrix factorization: Family of new algorithms. In: Proc. Int. Conf. Independent Component Analysis and Signal Separation. 2006. pp. 32-39

[39] Kong D, Ding C, Huang H. Robust nonnegative matrix factorization using $L_{2,1}$ norm. In: Proc. ACM Int. Conf. Information and Knowledge Management. 2011. pp. 673-682

[40] Sandler R, Lindenbaum M. Nonnegative matrix factorization with earth mover's distance metric for image analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2011;**33**(8):1590-1602

[41] Guan N, Tao D, Luo Z, Shawe-Taylor J. MahNMF: Manhattan non-negative matrix factorization [Online]. 2012. Available from: http://arxiv.org/abs/1207.3438

[42] Cichocki A, Cruces S, Amari S. Generalized alpha-beta divergences and their application to robust nonnegative matrix factorization. Entropy. 2011; **13**(1):134-170

[43] Duong VH, Lee YS, Pham BT, Mathulaprangsan S, Bao PT, Wang JC. Complex matrix factorization for face recognition [Online]. 2016. Available from: https://arxiv.org/ftp/arxiv/papers/1612/1612.02513.pdf

[44] Duong VH, Lee YS, Pham Ding JJ, Pham BT, Bui MQ, Bao PT, et al. Exemplar-embed complex matrix factorization for facial expression recognition. In: Proc the 42nd International Conference on Acoustics, Speech and Signal Processing (ICASSP 2017). 2017

[45] Duong VH, Bui MQ, Ding JJ, Lee YS, Pham BT, Bao PT, et al. A new approach of matrix factorization on complex domain for data representation. IEICE Transactions on Information and Systems. 2017;**E100-D** (12):3059-3063

[46] Liwicki S, Tzimiropoulos G, Zafeiriou S, Pantic M. Euler principal component analysis. International Journal of Computer Vision. 2013;**1**: 498-518

[47] Duong VH, Lee YS, Ding JJ, Pham BT, Bui MQ, Bao PT, et al. Projective complex matrix factorization for facial expression recognition. EURASIP Journal on Advances in Signal Processing. 2018;**10**

[48] Palka BP. An Introduction to complex function theory. Springer; 1991

[49] Wirtinger. Wirtinger Zur formalin Theorie de Funktionen von mehr komplexen Ver anderlichen. Mathematische Annalen. 1927;**97**:357-375

[50] Strang G. Linear Algebra and Its Applications. 4th ed. Belmont, Ca: Thomson, Brooks/Cole; 2006

[51] Zhou G, Xie S, Yang Z, Yang JM, He Z. Minimum volume constrained nonnegative matrix factorization: enhanced ability of learning parts. IEEE Transactions on Neural Networks. 2011; **22**(10):1626-1637

[52] Liu T, Gong M, Tao D. Large-cone nonnegative matrix factorization. IEEE Transactions on Neural Networks and Learning Systems. 2016. DOI: 10.1109/TNNLS.2016.2574748

[53] Kim J, He Y, Park H. Algorithms for nonnegative matrix and tensor factorizations: A unified view based on block coordinate descent framework. Global Optimization. 2013;**58**:285-319

[54] Barata JCA, Hussein MS. The Moore-Penrose pseudoinverse: A tutorial review of the theory. Brazilian Journal of Physics. 2012;**42**:146-165

[55] The ORL Dataset of Face. Website: http://www.cl.cam.ac.uk/research/dtg/attarchive/facedataset.html

[56] Dataset by Georgia Institute of Technology. Website: http://www.anefian.com/research/facereco.html

[57] Lin CJ. Projected gradient methods for non-negative matrix factorization. Neural Computation. 2007;**19**:2756-2779

[58] Yang Z, Yuan Z, Laaksonen J. Projective non-negative matrix factorization with applications to facial image processing. International Journal of Pattern Recognition and Artificial Intelligence. 2007;**21**(8):1353-1362

[59] Yang Z, Oja E. Linear and nonlinear projective non-negative matrix factorization. IEEE Transactions on Neural Networks and Learning Systems. 2010;**21**(5):734-749

[60] Boutsidis C, Gallopoulos E. SVD based initialization: A head start for nonnegative matrix factorization. Pattern Recognition. 2008;**41**(4):1350-1362

[61] Liu Y, Jia C, Li B, Pang S, Yu Z. Graph regularized projective non-negative matrix factorization for face recognition. Journal of Computer Information Systems. 2013;**9**(5):2047-2055

[62] Sharif M, Sajjad M, Jawad JM, Younas JM, Mudassar R. Face recognition for disguised variations using gabor feature extraction. Australian Journal of Basic and Applied Sciences. 2011;**5**(6):1648-1656