LOCALITY-PRESERVING COMPLEX-VALUED GAUSSIAN PROCESS LATENT VARIABLE MODEL FOR ROBUST FACE RECOGNITION

Sih-Huei Chen[1], *Yuan-Shan Lee*[1], *Yu-Sheng Hsu*[2], *Chung-Hsien Wu*[3], *and Jia-Ching Wang*[1]

[1]Dept. of Computer Science and Information Engineering, National Central University, Taiwan
[2]Dept. of Mathematics, National Central University, Taiwan
[3]Dept. of Computer Science and Information Engineering, National Cheng Kung University, Taiwan

ABSTRACT

Learning a low-dimensional image representation yields effective and efficient face recognition. The use of such a representation helps to weaken the curse of dimensionality. However, the traditional facial representation method is not robust against partial occlusions or variations of expression. To solve this problem, this paper proposes a more reliable, complex-valued representation of facial image. The robust representation is based on the proposed locality-preserving complex-valued Gaussian process latent variable model (LP-CGPLVM). In the LP-CGPLVM, the Euler formula is utilized to transform original facial images into the complex domain. A proper complex GP is employed to model the mapping between the complex-valued high-dimensional data and the corresponding low-dimensional representation. Moreover, the locality-preserving constraint is taken into consideration to preserve the neighborhood data structure. The experimental results indicate that our proposed method is robust against partial occlusions and various facial expressions.

Index Terms— Robust face recognition, occlusion, Gaussian process latent variable model, complex-valued representation

1. INTRODUCTION

This work concerns the face recognition problem and, in particular, the distortion of facial images by partial occlusions and various expressions. Various methods [1, 2, 3, 4] for handling partial occlusions, such as sunglasses, scarves or hands, and different expressions have been developed over the past decade. Among them, the subspace-based technique is the one of most popular for finding low-dimensional representation subspaces that are embedded in a high-dimensional face image space. An appropriate low-dimensional representation not only retains the information structures of images in high-dimensional space but also increases computational efficiency and avoids the curse of dimensionality.

One commonly used subspace-based method is nonnegative matrix factorization (NMF), which provides a part-based image representation. Zhi *et al.* [5] proposed a graph-preserving NMF (GSNMF) to represent facial images, which considered the neighborhood data structure. Similar to NMF, principal component analysis (PCA) [6] is a linear subspace method that assumes that each facial image is a linear combination of a set of orthogonal basis images. Liwicki *et al.* [7] developed the Euler PCA and applied it to face reconstruction. In Euler PCA, real-valued data are converted to complex-valued data using the Euler formula. The Frobenius norm on the complex domain has been proven to be equivalent to the cosine dissimilarity in real domain, leading to a robust measure of dissimilarity between the facial image and the associated occluded image.

Motivated by the work of Liwicki *et al.* [7], this paper proposes a locality-preserving complex-valued Gaussian process latent variable model (LP-CGPLVM) for the recognition of faces with various expression and partial occlusions. To cope with occluded face images, the real-valued pixels of a face image is firstly transformed to complex-valued data using the Euler formula. The proposed method assumes the existence of a nonlinear mapping from the complex-valued low-dimensional latent space to the complex-valued data space, which is characterized by a proper complex Gaussian process. The complex-valued Gaussian process latent variable model (CGPLVM) [8] is then employed to obtain the robust complex-valued representation. Furthermore, complex prior distribution that is based on locality-preserving projections (LPP) [9] is proposed to preserve the local data structure. The major contributions of this work are summarized as follows. (1) the learned complex-valued representation supports facial recognition that is robust against partial occlusion and various expression, and (2) the design of a complex prior distribution for complex-valued low-dimensional representations.

2. PRELIMINARIES

2.1. Gaussian Process Latent Variable Model (GPLVM)

In this section, we briefly review the GPLVM [10]. GPLVM is a nonlinear method for finding a low-dimensional manifold in high-dimensional data using a GP prior. Given high-dimensional data $\mathbf{y}_n \in \mathbb{R}^D, n = 1, ..., N$, we de-

978-1-5386-4658-8/18/$31.00 ©2018 IEEE

note the corresponding low-dimensional representation as $\mathbf{x}_n \in \mathbb{R}^Q$, by $Q \ll D$. The logarithm marginal likelihood $\mathcal{L}(\mathbf{X})$ of the image data $\mathbf{Y} = [\mathbf{y}_1, ..., \mathbf{y}_N]^{\mathrm{T}}$ conditioned on $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_N]^{\mathrm{T}}$ is written as follows.

$$\mathcal{L}(\mathbf{X}) = -\frac{DN}{2}\ln 2\pi - \frac{D}{2}\ln|\mathbf{T}| - \frac{1}{2}\mathrm{tr}(\mathbf{T}^{-1}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}) \quad (1)$$

where $\mathbf{T} = \mathbf{K} + \beta^{-1}\mathbf{I}_N$, β is the parameter of Gaussian noise and $\mathbf{K}$ is a kernel matrix with $\mathbf{K}_{nm} = k(\mathbf{x}_n, \mathbf{x}_m)$ that expresses the relationship among data. For example, a radial basis function (RBF) kernel is defined as $k(\mathbf{x}_n, \mathbf{x}_m) = \theta_1 \exp(-\theta_2 \|\mathbf{x}_n - \mathbf{x}_m\|^2)$, where $\theta = \{\theta_1, \theta_2\}$ are hyperparameters in the model. The low-dimensional representations $\mathbf{X}$ and the hyperparameters θ can be learned by maximizing Eq. (1).

Unlike PCA and NMF methods, GPLVM applies a probabilistic framework that can compensate for variation in estimation of low-dimensional representations. In the field of pattern recognition and image processing, the GPLVM has been used for visualization [10]. Besides, it can be used with uncertain input data [11], making it more robust than non-probabilistic methods when data are missing.

2.2. Locality-preserving Projections (LPP)

LPP [9] is a linear dimensionality reduction algorithm that preserves the locality of high-dimensional data. It has been successfully used to represent facial images [5]. The locality-preserving objective is,

$$\sum_{nm}(\mathbf{x}_n - \mathbf{x}_m)\mathbf{S}_{nm} \quad (2)$$

where $\mathbf{x}_n$ is the low-dimensional representation of input data $\mathbf{y}_n$ and $\mathbf{S}$ is a similarity matrix, which can be constructed using a Gaussian kernel:

$$\mathbf{S}_{nm} = \begin{cases} \exp(-\|\mathbf{y}_n - \mathbf{y}_m\|_2^2/\rho) & ; e(\mathbf{y}_n, \mathbf{y}_m) = 1 \\ 0 & ; e(\mathbf{y}_n, \mathbf{y}_m) = 0 \end{cases} \quad (3)$$

where $e(\mathbf{y}_n, \mathbf{y}_m) = 1$ represents that $\mathbf{y}_n$ and $\mathbf{y}_m$ belong to the same subject and ρ is an empirical parameter.

Based on LPP, Zhong *et al.* [12] proposed a Gaussian process latent random field (GPLRF), which developed a prior distribution to impose the locality constraint on the representation of data in latent space. Eleftheriadis *et al.* [13] presented a discriminative shared GPLVM (DS-GPLVM) that is based on the work of Zhong *et al.* [12] and applied it to recognize facial expressions.

3. LOCALITY-PRESERVING COMPLEX-VALUED GPLVM (LP-CGPLVM)

The goal of this paper is to preserve the advantages of the GPLVM and to extend them to robust facial recognition with partial occlusions and various expressions. Based on our previous work [8], we propose a locality-preserving CGPLVM (LP-CGPLVM). Unlike the work of Eleftheriadis *et al.* [13], we introduce a locality-preserving prior distribution in the complex domain. Two perspectives of robustness are considered herein . First, the robust dissimilarity measure based on the Euler formula. Second, the robustness of dealing with uncertain input data. In this section, the proposed method is divided into robust transformation, complex-valued facial representation and locality-preserving training.

3.1. Robust Transformation

The cosine-based dissimilarity measure is robust for occluded facial images [14]. It yields a shorter distance between the facial image and the associated image with partial occlusion than does the Euclidean norm. Given two images $\mathbf{y}_n, \mathbf{y}_m \in \mathbb{R}^D$ with pixel values from 0 to 1, the cosine-based dissimilarity measure is defined by

$$d(\mathbf{y}_n, \mathbf{y}_m) = \sum_{d=1}^{D}\left[1 - \cos\left(\alpha\pi(y_{nd} - y_{md})\right)\right] \quad (4)$$

where $\alpha \in \mathbb{R}^+$.

Following the work of Liwicki *et al.* [7], the real-valued pixels of an image can be transformed into the complex domain to deal with the occluded facial image. This is feasible because the cosine dissimilarity measure in real domain is equivalent to the Euclidean norm between the complex-valued data [7]. The facial image $\mathbf{y}_n$ with the values from 0 to 1 can be mapped to the complex domain using the Euler representation, which is given by

$$\mathbf{z}_n = \frac{1}{\sqrt{2}}e^{i\alpha\pi\mathbf{y}_n} = \frac{1}{\sqrt{2}}\begin{bmatrix} e^{i\alpha\pi y_{n1}} \\ \vdots \\ e^{i\alpha\pi y_{nD}} \end{bmatrix} \quad (5)$$

Notably, for $\alpha < 2$, the mapping is one-to-one. In this work, α was set to 1.5. The relationship between Eq. (4) and Eq. (5) can be written as follows

$$\begin{aligned} \|\mathbf{z}_n - \mathbf{z}_m\|_2^2 &= \left\|\frac{1}{\sqrt{2}}e^{i\alpha\pi\mathbf{y}_n} - \frac{1}{\sqrt{2}}e^{i\alpha\pi\mathbf{y}_m}\right\|_2^2 \\ &= \sum_{d=1}^{D}\left[1 - \cos(\alpha\pi y_{nd} - \alpha\pi y_{md})\right] \\ &= d(\mathbf{y}_n, \mathbf{y}_m) \end{aligned} \quad (6)$$

3.2. Complex-valued Facial Representation

With the above robust property, a CGPLVM [8] is utilized to learn a low-dimensional representation of data in the complex domain. The principle behind the CGPLVM is that complex-valued data can be represented by low-dimensional latent space and a proper complex GP. Given complex-valued data $\mathbf{z}_n \in \mathbb{C}^D$, the CGPLVM is defined as

$$z_{nd} = g_d(\mathbf{w}_n) + \varepsilon_{nd} \tag{7}$$

where $\mathbf{w}_n \in \mathbb{C}^Q$, with $Q \ll D$, is the corresponding low-dimensional representation and $\varepsilon_{nd} \sim \mathcal{CN}(0, \beta^{-1}, 0)$. The functions $g_d, d = 1, ..., D$ are drawn from an independent proper complex GP. That is, $\mathbf{g}_d = [g_d(\mathbf{w}_1), ..., g_d(\mathbf{w}_N)]^{\mathrm{T}} \sim \mathcal{CN}(\mathbf{0}, k_c(\mathbf{w}_n, \mathbf{w}_m), \mathbf{0})$, k_c is a kernel function which specifies the similarity among the complex-valued latent variables $\mathbf{w}_n$. In this work, k_c is obtained as the sum of two real kernel functions, k_{rr} and k_{ii} [15]. The k_{rr} and k_{ii} are chosen as

$$k(\mathbf{w}_n, \mathbf{w}_m) = \theta_1 \exp\left(-\theta_2 (\mathbf{w}_n - \mathbf{w}_m)^{\mathrm{H}} (\mathbf{w}_n - \mathbf{w}_m)\right) + \theta_3 \tag{8}$$

where $(\cdot)^{\mathrm{H}}$ is the Hermitian transpose and $\theta = \{\theta_1, \theta_2, \theta_3\}$ are hyperparameters.

The log marginal likelihood of complex-valued data $\mathbf{Z} = [\mathbf{z}_1, ..., \mathbf{z}_N]^{\mathrm{T}} \in \mathbb{C}^{N \times D}$ given the latent variables $\mathbf{W} = [\mathbf{w}_1, ..., \mathbf{w}_N]^{\mathrm{T}} \in \mathbb{C}^{N \times Q}$ is

$$\ln p(\mathbf{Z}|\, \mathbf{W}) = -DN \ln \pi - D \ln |\mathbf{T}_c| - \mathrm{tr}(\mathbf{T}_c^{-1} \mathbf{Z}\mathbf{Z}^{\mathrm{H}}) \tag{9}$$

where $\mathbf{T}_c = \mathbf{K}_c + \beta^{-1}\mathbf{I}_N$. Learning in the CGPLVM comprises maximizing Eq. (9) with respect to the low-dimensional representations and the hyperparameters.

3.3. Locality-preserving Training

Notably, the CGPLVM focuses on preserving the global data structure. To incorporate the locality-preserving constraint into the CGPLVM, the complex prior distribution of the low-dimensional representation $\mathbf{W}$ is introduced. Just as in the GPLRF [12], the locality-preserving term is defined to preserve the neighborhood data structure, the introduced complex prior distribution does the same things for complex-valued data, which is written as

$$p(\mathbf{W}) = \frac{1}{Z_d} \exp\left(-\frac{1}{\sigma_d^2} \mathrm{tr}\left(\mathbf{W}\mathbf{L}\mathbf{W}^{\mathrm{H}}\right)\right) \tag{10}$$

where $\mathbf{L} = \mathbf{E} - \mathbf{S}$ is a Laplacian matrix, $\mathbf{S}$ is defined as Eq. (3) and $\mathbf{E}_{nn} = \sum_m \mathbf{S}_{nm}$. The MAP estimation of the complex-valued representation $\mathbf{W}$ can be obtained by gradient descent-based methods. For a new test image $\mathbf{z}'$, the prediction of low-dimensional representation $\mathbf{w}'$ can be estimated by optimizing the likelihood $p(\mathbf{z}', \mathbf{w}'|\, \mathbf{Z}, \mathbf{W}, \theta)$ with an uninformative prior of $\mathbf{w}'$, which is written as follows.

$$\mathcal{L}(\mathbf{w}') = -\ln\left|\sigma^2(\mathbf{w}')\mathbf{I}_D\right| - \frac{(\mathbf{z}' - \mu(\mathbf{w}'))^{\mathrm{H}}(\mathbf{z}' - \mu(\mathbf{w}'))}{\sigma^2(\mathbf{w}')} - \frac{1}{2}\mathbf{w}'^{\mathrm{H}}\mathbf{w}' \tag{11}$$

where $\mu(\mathbf{w}') = \mathbf{Z}^{\mathrm{H}}\mathbf{T}_c^{-1}\mathbf{k}$ and $\sigma^2(\mathbf{w}') = k_c(\mathbf{w}', \mathbf{w}') - \mathbf{k}^{\mathrm{H}}\mathbf{T}_c^{-1}\mathbf{k}$, with $\mathbf{k} = [k_c(\mathbf{w}_1, \mathbf{w}'), ..., k_c(\mathbf{w}_N, \mathbf{w}')]^{\mathrm{T}}$.

(a)

(b)

Fig. 1. (a) Example images from the MHMC face database. (b) Images with randomly masked occlusions with block sizes of 60×60 to 85×85 pixels.

4. EXPERIMENTAL RESULTS

4.1. Experimental Data and Settings

In this work, the performance of the proposed method when applied to the MHMC [16] and YaleFace database, was evaluated. The MHMC database includes 4 subjects, each associated with four facial expressions– angry, happy, neutral and sad. Each subject in the database spoke 30 sentences related to a topic about a mood, which were recorded. Fig. 1 (a) shows an example from the MHMC database. In the experiment herein, a total of 480 color frontal images of four subjects, with 30 images per expression, were collected from the MHMC database. The original image resolution was 240×320 pixels. Based on the coordinates of the eyes, the original images were cropped to 240×150 to remove the background. Then, the images were converted to grey scale and downsampled to 40×25. Two fifths of the non-occluded images were used as training samples. The rest were masked by a random block with a size of 60×60 to 85×85 pixels and used as test samples, as shown in Fig. 1 (b). We also conducted robust face recognition using a database with practical occlusion. The YaleFace refers to 15 subjects. Each subject consists of 11 grey scale images. The images were collected with different settings, including illumination, facial expression and occlusion (glasses). Fig. 2 displays some samples that are associated with one subject from the YaleFace database. The resolution of the original images is 243×320 pixels. Each image is manually cropped and then resized to 40×27 pixels. To make this database more challenging, M non-occluded images were randomly selected and masked using a block of size 55×85. An occluded image (glasses) and M artificially occluded images ($M = 3, 4, 5$) from each subject are used for testing. The remaining N images are used for training ($N = 5, 6, 7$).

The proposed method was compared to other subspace-based representation methods– PCA, NMF, GSNMF [5], GPLVM [10] and CGPLVM [8]. In PCA and NMF-based methods, the training images were used to learn the basis. Then, the low-dimensional representations of the test images were obtained using the basis. In GPLVM and CGPLVM,

Fig. 2. Some samples from the YaleFace database.

the RBF based kernel [8] was used and formed using the training images. Given the pre-trained kernel matrix, the low-dimensional representations of the test images were learned based on the representations of the training images. The 1 nearest-neighbor (1-NN) was adopted as the classifier.

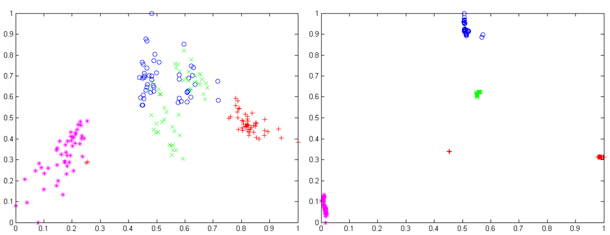

Fig. 3. Visualization of training images in 2-D latent space: (left) CGPLVM, (right) LP-CGPLVM.

4.2. Results and Discussions

In the first experiment, the effectiveness of using the locality-preserving complex prior distribution was investigated on visualization experiment. For clarity, only four subjects from the MHMC database were displayed. Fig. 3 shows relevant results obtained using both the CGPLVM and the proposed LP-CGPLVM with the non-occluded face data. In Fig. 3, the 2-D complex-valued representations were transformed to real-valued representations by applying the inverse function of Eq. (5). The value of σ_d^2 was set to 10^{-2}. The experiments indicated that the LP-CGPLVM has a greater discriminative locality ability than the CGPLVM.

In the second experiment, the robustness of the proposed method was evaluated using artificial occlusion images and practical occlusion images. Fig. 4 compares the proposed method with the baselines when applied to the MHMC database with the latent dimension Q set to 15. Table 1 presents the recognition results obtained using the YaleFace database with the latent dimension Q of 30. Three observations are made. (1) the recognition rate of the proposed robust complex-valued representation exceeds those of the other real-valued representation methods on all occlusion block sizes. (2) owing to the nonlinear nature of GP, the recognition rates of GPLVM, CGPLVM and LP-CGPLVM are better than those of the PCA and NMF, and (3) a comparison between the CGPLVM and the LP-CGPLVM confirmed the power of the locality-preserving term.

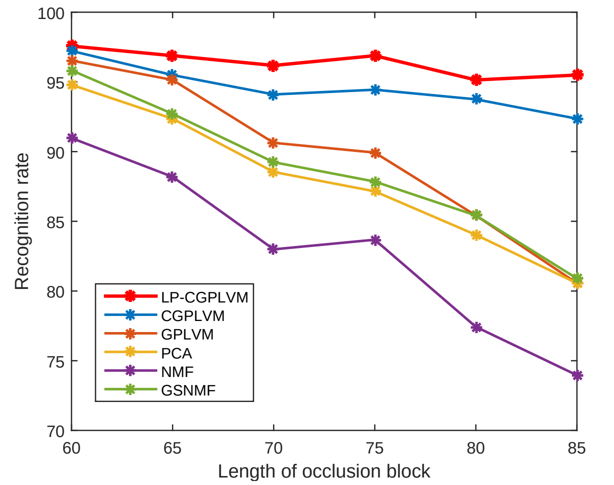

Fig. 4. Recognition results obtained using different methods with various occlusion block sizes on MHMC database.

5. CONCLUSIONS AND FUTURE WORK

This paper proposes a novel and robust means of representing facial images. The potential of using the complex-valued representation for occluded facial images was studied. Unlike state-of-the-art image representation methods, the proposed LP-CGPLVM uses a robust dissimilarity measure that is based on the Euler formula and then analyzes the complex-valued data rather than real-valued pixels. The updating rule of the complex-valued representation is derived. The results of the first experiment revealed that the introduced complex prior distribution of low-dimensional latent variables makes the complex-valued representations more discriminative. The second experiment demonstrated that the proposed complex-valued representation is more robust than the real-valued representation for facial images with simulated occlusions and practical occlusions. In the future, we would like to apply the proposed method to other types of images.

Table 1. Recognition rate of proposed method and baselines with various numbers of training samples (N) on YaleFace.

N	5	6	7
PCA	86.67	89.33	93.33
NMF	87.78	92.00	90.00
GSNMF	88.89	93.33	96.67
GPLVM	86.67	90.67	93.33
CGPLVM	90.00	94.67	96.67
LP-CGPLVM	**91.11**	**96.00**	**98.33**

6. REFERENCES

[1] X. Wei and C. T. Li, "Fixation and saccade based face recognition from single image per person with various occlusions and expressions," in *Proc. CVPR*, 2013, pp. 70–75.

[2] X. X. Li, D. Q. Dai, X. F. Zhang, and C. X. Ren, "Structured sparse error coding for face recognition with occlusion," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1889–1900, May 2013.

[3] Y. Wen, W. Liu, M. Yang, and M. Li, "Efficient misalignment-robust face recognition via locality-constrained representation," in *Proc. ICIP*, 2016, pp. 3021–3025.

[4] E. J. He, J. A. Fernandez, B. V. K. V. Kumar, and M. Alkanhal, "Masked correlation filters for partially occluded face recognition," in *Proc. ICASSP*, 2016, pp. 1293–1297.

[5] R. Zhi, M. Flierl, Q. Ruan, and W. B. Kleijn, "Graph-preserving sparse nonnegative matrix factorization with application to facial expression recognition," *IEEE Trans. Syst. Man. Cybern. B, Cybern.*, vol. 41, no. 1, pp. 38–52, Feb. 2011.

[6] I. T. Jolliffe, *Principal Component Analysis*, Springer-Verlag, New York, 1986.

[7] S. Liwicki, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "Euler principal component analysis," *Int. J. Comput. Vis.*, vol. 101, no. 3, pp. 498–518, 2013.

[8] S. H. Chen, Y. S. Lee, and J. C. Wang, "Phase-incorporating speech enhancement based on complex-valued Gaussian process latent variable model," *arXiv preprint arXiv:abs/1612.09150v2*, 2016.

[9] X. He and P. Niyogi, "Locality preserving projections," in *Proc. NIPS*, 2003, pp. 153–160.

[10] N. Lawrence, "Gaussian process latent variable models for visualisation of high dimensional data," in *Proc. NIPS*, 2004, pp. 329–336.

[11] M. K. Titsias and N. Lawrence, "Bayesian Gaussian process latent variable model," in *AISTATS*, 2010.

[12] G. Zhong, W. J. Li, D. Y. Yeung, X. Hou, and C. L. Liu, "Gaussian process latent random field," in *Proc. AAAI*, 2010, pp. 679–684.

[13] S. Eleftheriadis, O. Rudovic, and M. Pantic, "Discriminative shared Gaussian processes for multiview and view-invariant facial expression recognition," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 189–204, Jan 2015.

[14] A. Fitch, A. Kadyrov, W. Christmas, and J. Kittler, "Fast robust correlation," *IEEE Trans. Image Process.*, vol. 14, no. 8, pp. 1063–1073, 2005.

[15] R. Boloix-Tortosa, F. J. Payan-Somet, E. Arias-de-Reyna, and J. José Murillo-Fuentes, "Proper complex Gaussian processes for regression," *arXiv preprint arXiv:abs/1502.04868*, 2015.

[16] J. C. Lin, C. H. Wu, and W. L. Wei, "Error weighted semi-coupled hidden Markov model for audio-visual emotion recognition," *IEEE Trans. Multimedia*, vol. 14, no. 1, pp. 142–156, Feb. 2012.