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

Introduction

This work concerns the face recognition task and, in particular, the distortion of face images by partial occlusions and various expressions. The subspace-based technique is the one of most popular for finding low-dimensional representation subspaces that are embedded in a high-dimensional face space. Motivated by the work on Euler representation, we propose a **locality-preserving complex-valued Gaussian process latent variable model (LP-CGPLVM)** to learn a **complex-valued representation** of face image.

Main Contributions

1. The learned complex-valued representation supports facial recognition that is robust against partial occlusion and various expression.
2. A locality-preserving based complex prior distribution over complex-valued low-dimensional representations is developed. The MAP estimation of representation preserves not only global structure but also locality structure of face data.

Conclusion

- The potential of using complex-valued representation for occluded face images was studied.
- The results for visualizations of face images revealed that the introduced complex prior distribution makes the complex-valued representations more discriminative.
- Experimental results showed that the proposed method is more robust than baselines for facial images with simulated occlusions and practical occlusions.

Contact Information

Name Sih-Huei Chen
Lab http://mediasystem.csie.ncu.edu.tw/
Email new150019@gmail.com
Phone +886 910842195

Proposed Method

I. Robust Transformation

The cosine-based dissimilarity measure yields a shorter distance between face image and its associated occluded image than does the Euclidean norm. It can be equivalently computed using the Euler formula which maps pixel value $\mathbf{y}_n$ of image into complex data $\mathbf{z}_n$,

$$\mathbf{z}_n = \frac{1}{\sqrt{2}} e^{i\alpha\pi\mathbf{y}_n} = \frac{1}{\sqrt{2}} \left[e^{i\alpha\pi y_{n1}}, \cdots, e^{i\alpha\pi y_{nD}} \right]^{\mathrm{T}} \tag{1}$$

where α is a constant.

II. Complex-valued Facial Representation

With the robust transformation, the CGPLVM [1] is utilized to learn a complex-valued low-dimensional representation $\mathbf{W}$ of image in complex domain. The objective of CGPLVM is

$$\ln p(\mathbf{Z}|\,\mathbf{W}) = -DN \ln \pi - D \ln |\mathbf{T}_c| - \mathrm{tr}(\mathbf{T}_c^{-1}\mathbf{Z}\mathbf{Z}^{\mathrm{H}}) \tag{2}$$

where $\mathbf{T}_c$ is a kernel matrix.

III. Locality-preserving Training

To incorporate the locality-preserving term into the CGPLVM, the complex prior distribution over low-dimensional representation $\mathbf{W}$ is introduced.

$$p(\mathbf{W}) = \frac{1}{Z_d} \exp\left(-\frac{1}{\sigma_d^2} \mathrm{tr}\left(\mathbf{W}\mathbf{L}\mathbf{W}^{\mathrm{H}} \right) \right) \tag{3}$$

where $\mathbf{L} = \mathbf{E} - \mathbf{S}$ is a Laplacian matrix and $\mathbf{E}_{nn} = \sum_m \mathbf{S}_{nm}$ with $\mathbf{S}$ is computed as

$$\mathbf{S}_{nm} = \begin{cases} \exp(-\|\mathbf{y}_n - \mathbf{y}_m\|_2^2/\rho) & ; e(\mathbf{y}_n, \mathbf{y}_m) = 1 \\ 0 & ; e(\mathbf{y}_n, \mathbf{y}_m) = 0 \end{cases} \tag{4}$$

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 a constant.

IV. Prediction

For a new test image $\mathbf{z}'$, the low-dimensional representation $\mathbf{w}'$ can be estimated by optimizing the objective $\mathcal{L}$ with an uninformative prior of $\mathbf{w}'$,

$$\mathcal{L}(\mathbf{w}') = -\ln |\sigma^2(\mathbf{w}')\mathbf{I}_D| - \frac{(\mathbf{z}'-\mu(\mathbf{w}'))^{\mathrm{H}}(\mathbf{z}'-\mu(\mathbf{w}'))}{\sigma^2(\mathbf{w}')} - \frac{1}{2}\mathbf{w}'^{\mathrm{H}}\mathbf{w}' \tag{5}$$

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

Visualization

- **Database:** MHMC [2] and YaleFace database
- **Size of random block:** 60×60 to 85×85 pixels
- **Baselines:** PCA, NMF, GSNMF, GPLVM, and CGPLVM
- In YaleFace database, 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).

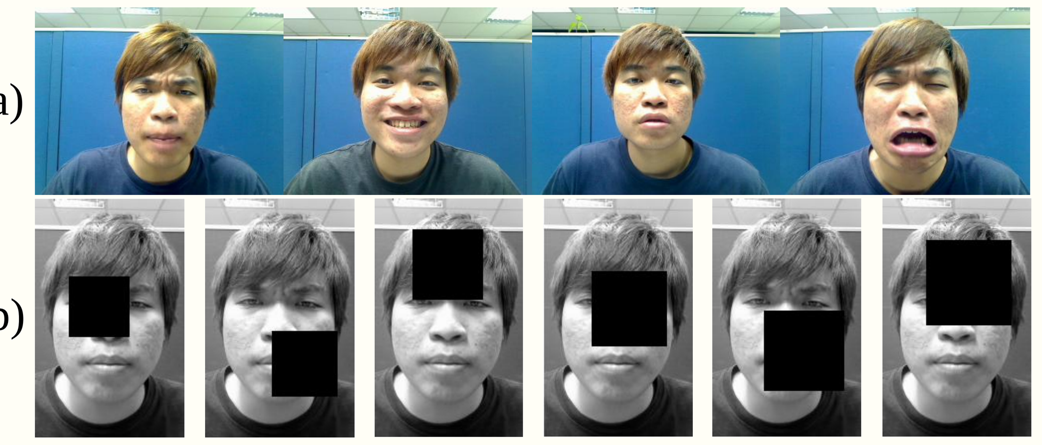

Figure 1: (a) Example images from MHMC database. (b) Images with randomly masked occlusions with block sizes of 60×60 to 85×85 pixels.

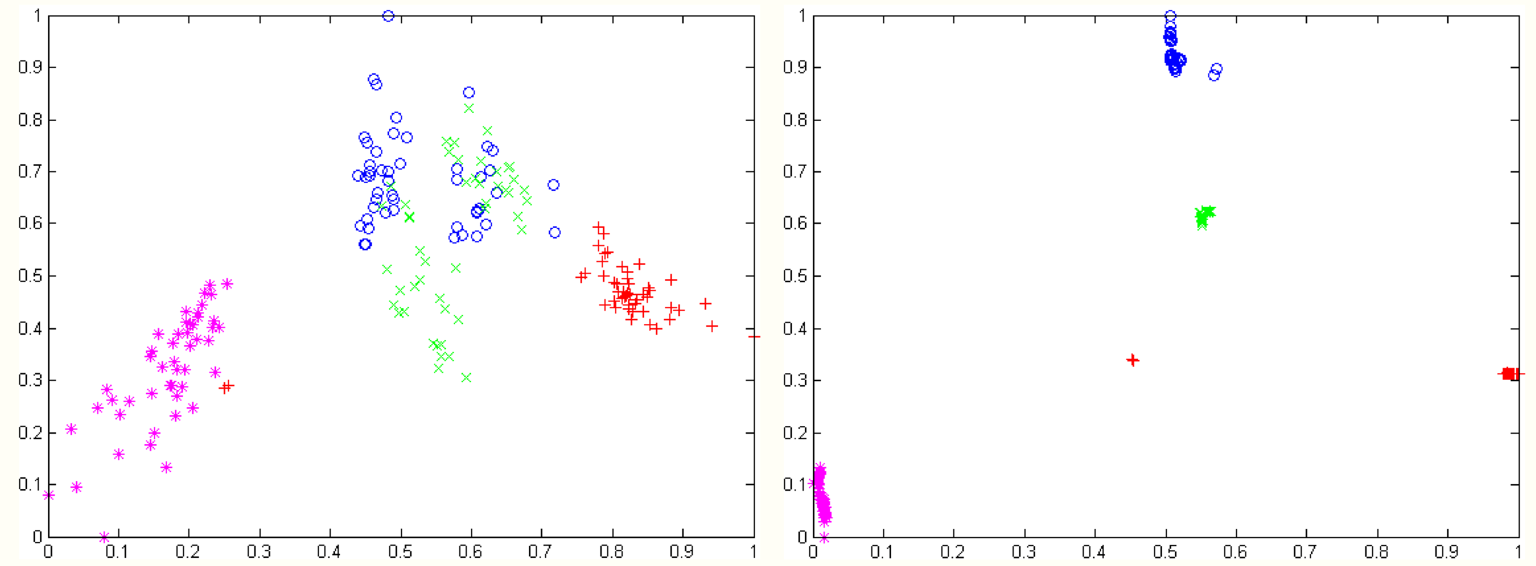

Figure 2: Visualizations of training images in 2-D latent space: (left) CGPLVM, (right) LP-CGPLVM.

Experimental Results

Table 1: Recognition rates for 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**

- Recognition rate of the proposed robust complex-valued representation exceeds those of the other real-valued representation methods on all occlusion block sizes.
- Comparison between the CGPLVM and the LP-CGPLVM confirmed the power of the locality-preserving term.

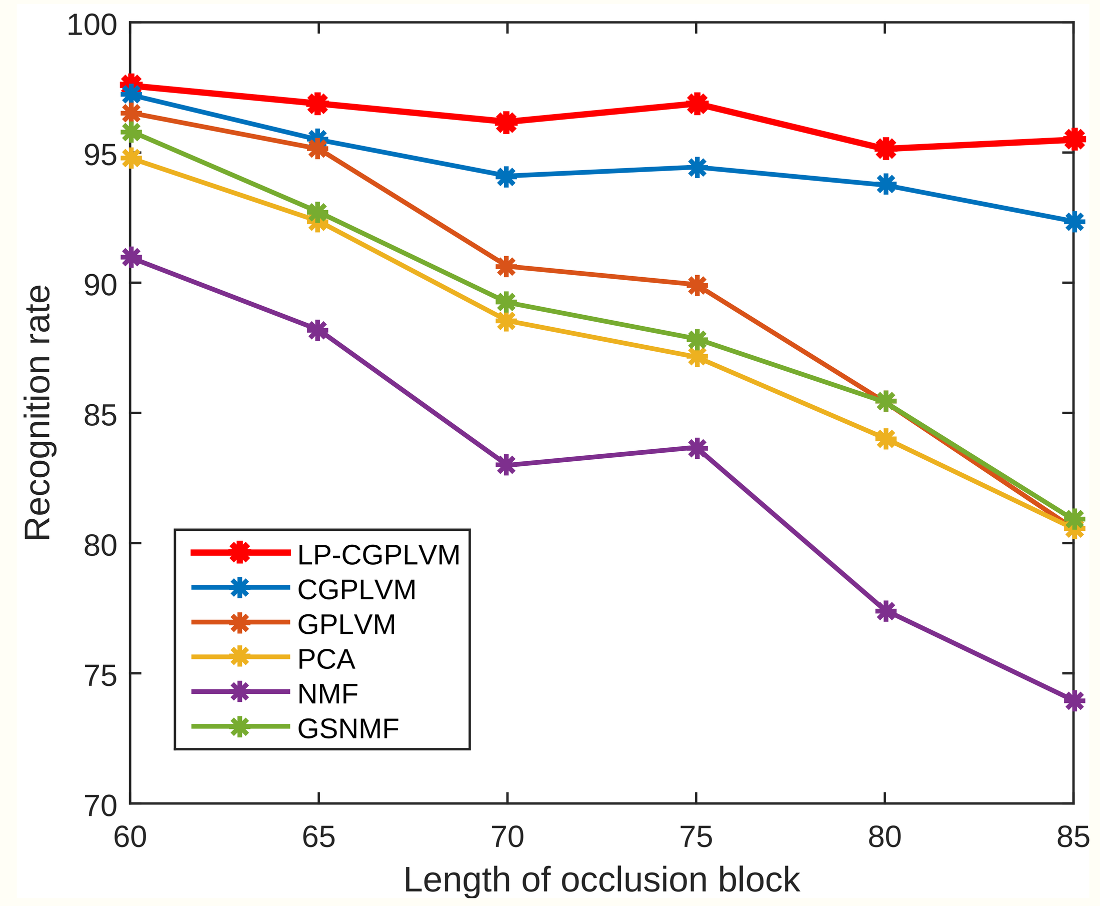

Figure 3: Recognition results obtained using different methods with various occlusion block sizes on MHMC database.

References

[1] 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.

[2] 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*, 14(1):142–156, Feb. 2012.