Head Yaw Estimation via Symmetry of Regions

Bingpeng Ma[1], Annan Li[2], Xiujuan Chai[3] and Shiguang Shan[3]
[1] University of Chinese Academy of Sciences, Beijing, China mabingpeng@163.com
[2] Department of Electrical and Computer Engineering, National University of Singapore. elela@nus.edu.sg
[3] Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS),
Institute of Computing Technology, CAS, Beijing, China {chaixiujuan, sgshan}@ict.ac.cn

Abstract— **This paper proposes a novel method to estimate the head yaw rotations using the symmetry of regions. We argue and reveal that the symmetry between the two regions in the same horizontal row are closely relevant to the yaw rotation of head, while at the same time insensitive to the identity of the face. The proposed method relies on the effective combination of Gabor features and covariance descriptors. Specifically, we first extract the Gabor features of a face image, then the covariance descriptors are used to compute the symmetry of Gabor features. Since the covariance matrix can eliminate the influence which is caused by rotations and illuminations, the proposed method is robust to these variations. In addition, the proposed method can be further improved by combining it with supervised learning. Experiments on two challenging databases are conducted, on which the proposed method improves the current state-of-the-art.**

I. Introduction

During the last decade, the research of face recognition and the related problems has received more and more attentions. However, the current face recognition systems can only reliably deal with near-frontal faces, and the performances of these systems degrade dramatically on non-frontal faces. To achieve the robustness to pose variation, one may expect to process face images differently according to their pose parameters, which requires the accurate head pose(especially the yaw pose). Thus, in this paper, we focus our attention on the challenging problem of estimating the head yaw pose from face images.

The methods for head pose estimation are summarized in [1] [2], which can be categorized into two main groups: model-based methods e.g. [3] [4] and appearance-based methods e.g. [5] [6]. The model-based methods use the 3D structure of human head. Typically, they build 3D models for human faces and attempt to match the facial features such as the face contour and the facial components of the 3D face model with their 2D projections. Since these methods generally run very fast, they can be used in video tracking and multi-camera surveillance. However, the model-based methods often sensitive to the misalignment of the facial feature points and it is difficult to precisely build the head model for different person. Inevitably, the application of model-based method is limited. The appearance-based methods typically assume that there exists a certain relationship between the 3D face pose and some properties of the 2D facial image, and use a large number of training samples to infer this relationship by using statistical learning techniques.

Recently, Ma et al. proposed the GaFour method which estimates the head yaw pose by using the asymmetry of the facial appearance [7] and gained the great improvement in head pose estimation. They argue that the asymmetry of the intensities in each row of the face image is closely relevant to the yaw rotation of head. Their motivation comes from the scene that with the pose varying from the front to the half-profile, the symmetry of the face decreases gradually. Since this decrease of the symmetry is insensitive to the identity of the input face, it can be applied to estimate the head pose. Specifically, in GaFour, taking the intensities of each row of the face image as a 1D signal, 1D Gabor filters are first convolved with the row signals to reduce noise and extract the local information, and then Fourier analysis is used to represent the asymmetry features of the head, i.e. to represent the pose.

Though the asymmetry in the row signal is related to the head pose, the asymmetry based on the 1D signal is easily effected by other factors, for example, the misalignment. When there is a rotation in the plane for a frontal face image, the pixels in the same row are not symmetrically. In fact, the problem of misalignment happened inevitably since the input image in the head pose estimation problem is the output of the automatic face detectors. Generally speaking, the automatic face detectors just care about whether there is a face in the image and don't care about the misalignment. In this scene, it is easy to understand that the performance of GaFour will be decreased naturally when the misalignment happens.

In this paper, we propose a novel method to improve the accuracy of head pose estimation. In the proposed method, considering the influence of noise and illumination, 2D Gabor filters are first convolved and extracted the local information. Then, the Gabor features are divided into many regions with the same size. Covariance descriptor is taken as feature extractor to extract the symmetry of the symmetrical regions of the Gabor features. To enhance the discriminative ability and reduce the dimension of the proposed representation, furthermore, supervised learning method, specially Linear Discriminant Analysis (LDA) [8], is applied after feature extraction. To gain the final pose of the representations, the nearest centroid (NC) classifier is exploited to validate the effectiveness of the proposed method.

Compared with the symmetry of the 1D signal in GaFour, the symmetry of the 2D regions in the proposed repre-

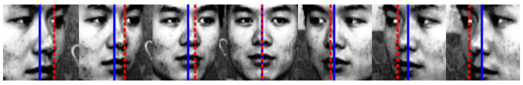

Fig. 1. The relationship between the symmetry plane of the head and the center lines of the images.

sentations is more relevant to the pose variations. On one hand, when the rotation in the plane happens, though a point and its symmetric point in the face are not in the same row of the image, their regions are still symmetrical because region is robust to the small rotation. By using the symmetry of regions, the influence of misalignment can be reduced greatly. On the other hand, the features of GaFour are extracted by Fourier transform, which means the features are not the symmetry metric directly, but just include the information of symmetry. From the viewpoint of the relationship between the pose variations and the symmetry, we argue the direct metric of the symmetry is much better than the features which includes the symmetry information. Based on this point, in this paper, the similarity of regions under covariance descriptor is taken as the symmetry metric of region.

The remaining part of this paper is organized as follows: in Section II, we show the advantage of the symmetry of 2D regions. In Section III, we describe the proposed method and introduce its improvement by combining with supervised methods. Experiments are given in Section IV. Conclusions are drawn in Section V with some discussions on the future work.

II. Relationship between Pose and Symmetry of Regions

The relationship between the center line of the images and the symmetry plane of heads are shown in Fig. 1 [7]. In Fig. 1, the dash line denotes the center line of the 2D images while the solid line denotes the symmetry line of 3D faces. With the pose varying from the front to the half-profile, the deviation between the two lines increases gradually. From this sense, in [7], the authors conclude that the symmetry is closely relevant to pose variations. More details can be found in [7].

Though the symmetry is relevant to the pose variations, the symmetry defined in [7] is based on the 1D signals. In this paper, we argue that the symmetry based on the 2D region is more relevant to the pose variations. To demonstrate this, we conduct the following experiment using the images from CAS-PEAL face database [15].

First, we define the symmetry measure E_{1D} of 1D signals for a face image as follows:

$$E_{1D} = \frac{1}{wh} \sum_{i=1}^{h} \sum_{j=1}^{w} (I_{ij} - I_{ij}^T)^2 \qquad (1)$$

where I_{ij} is the image intensity at position (i, j), w and h are the width and the height of the input image, respectively. The image I is flipped horizontally and we can gain a new

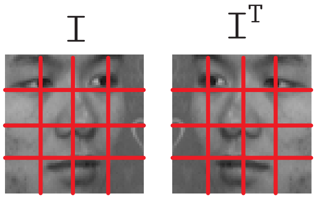

Fig. 2. A face image and its' symmetric image.

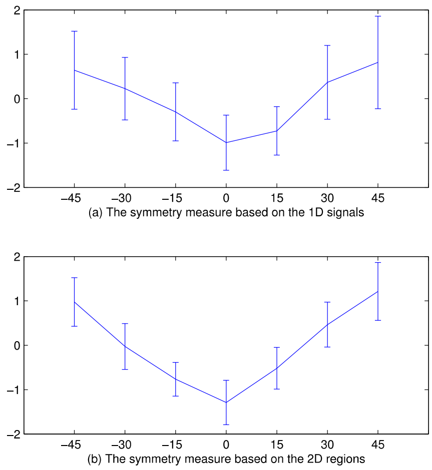

Fig. 3. The symmetry measures of the different poses on the CAS-PEAL database. The horizontal axes represent the poses and the vertical axes represent the measures. (a)The symmetry of the 1D signals. (b)The symmetry of 2D regions. This figure show that the symmetry of 2D regions is more related to the pose varies than that of 1D signals.

face image I^T. In Fig. 2, we show the image I and its correlative I^T. In fact, image I has the same information with image I^T because the pixel (x, y) of I is the same with the pixel $(w + 1 - x, y)$ of I^T. By using I^T, the regions with the same position of two image is much easier to be accepted by human brain than the different position of one image. Obviously, the lower the value of E_{1D}, the greater the amount of symmetry and vice versa. The means and the standard deviations of the symmetry measure E_{1D} of different poses are shown in Fig. 3 (a). The horizontal axis represents the poses while the vertical axis shows the measure of symmetry. From the figure, we can know with pose varying from the front the half-profile, E_{1D} is increased, which means the symmetry is decreased.

We also define the symmetry measure E_{2D} based on the similarity of the regions. First, we divide a face image into many regions. We name a region using the position of the left-upper pixel. Then the symmetry measure E_{2D} of the input face image is defined by:

$$E_{2D} = \frac{1}{n} \sum_{ij} sym(R_{ij}, R_{ij}^T) \qquad (2)$$

where $sym(\cdot)$ means the similarity of two regions. R_{ij} is the region in the original input image I and R_{ij}^T is the region in image I^T. In the experiment, we take the covariance descriptor as the function of $sym(\cdot)$, which is used in the proposed method and we introduce it in the following section. Obviously, the lower the value of E_{2D}, the greater the amount of symmetry and vice versa. We repeat the experiment and show the means and the standard deviations of the symmetry measure E_{2D} of different pose in Fig. 3 (b). From Fig. 3 (b), we can also clearly see that the symmetry decreases when the pose variations from the front to the half-profile, which means that the symmetry of 2D regions is also related to the pose varies and can be applied in head pose estimation. Compared with Fig. 3 (a), the means of E_{2D} are more closed to a straight line while the standard deviations of E_{2D} are much smaller than those of E_{1D} when pose varies from the front to the half-profile. From the comparison, we can conclude that the symmetry based on the 2D regions is more relevant to the pose variations than based on the 1D row signals. So, in this paper, we try to use the symmetry of 2D regions to improve the accuracy of head pose estimation.

III. Covariance Descriptor of Gabor filters

In this section, we introduce the proposed method named Covariance Descriptor of Gabor filters (CovGa). Fig. 4 shows the flowchart of the proposed CovGa. CovGa is a two stages representation: Gabor features are first extracted and then the symmetrical regions of Gabor filters are encoded by covariance descriptors. In the following, we first introduce the two stages. Then, we introduce how to improve the discriminative ability of CovGa.

A. Gabor filters in CovGa

In the proposed method, we do not compute the symmetry on the intensity feature directly. But considering the advantages of Gabor filters in face recognition and the related areas, we exploit the multi-resolution and multi-orientation Gabor filters to de-composite the input face images for sequential feature extraction. Then, we compute the symmetry of the regions on each Gabor representation.

The Gabor filters are inspired by the human visual system and their kernels are very similar to the 2-D receptive field profiles of the mammalian cortical simple cells. For an image $I(x, y)$, we compute its convolution with Gabor filters accordingly to the following equation [16]:

$$G(\mu, \nu) = I(x, y) * \psi_{\mu,\nu}(z) \tag{3}$$

where:

$$\psi_{\mu,\nu}(z) = \frac{\|k_{\mu,\nu}\|^2}{\sigma^2} \mathrm{e}^{\left(\frac{-\|k_{\mu,\nu}\|^2 \|z\|^2}{2\sigma^2}\right)} \left[e^{ik_{\mu,\nu} z} - \mathrm{e}^{\frac{-\sigma^2}{2}} \right] \tag{4}$$

$$k_{\mu,\nu} = k_\nu e^{i\phi_\mu}, k_\nu = 2^{-\frac{\nu+2}{2}} \pi, \phi_\mu = \mu \frac{\pi}{8} \tag{5}$$

where μ and ν are the scale and orientation parameters, respectively. In our work, μ is quantized into 16 scales while ν is quantized into 8 orientations.

In our method, the number of scales is fixed to 16 and two neighborhood scales (within the same orientation) are grouped into one band (we therefore have 8 different bands). We then apply MAX pooling over two consecutive scales:

$$G_{ij} = \max(G(2i - 1, j), G(2i, j)) \tag{6}$$

The MAX pooling operation increases the tolerance to small scale changes which often occurs in head images since head images are misaligned or only roughly aligned.

B. Covariance Descriptor in CovGa

The core of the proposed CovGa is how to measure the symmetry of the regions from the 2D images. In this paper, we employ the covariance descriptor as the metric of the symmetry of the regions. Covariance descriptor was firstly proposed by Tuzel et al. for object detection [9], and then widely used in other fields such as pedestrian detection [10] and object tracking. Covariance descriptor is able to capture shape, location and color information. It is shown that the performance of the covariance features is superior to other methods as rotations and illuminations changes are absorbed by the covariance matrix.

In the second step of CovGa, first, for each pixel of the Gabor features G_{ij}, a 7-dimensional feature vector $f_{ij}(x, y)$ is computed to capture the spatial, intensity, texture and shape statistics:

$$f_{ij}(x, y) = [x, y, G_{ij}, G_{ij_x}, G_{ij_y}, G_{ij_{xx}}, G_{ij_{yy}}] \tag{7}$$

where x and y are the pixel coordinate, G_{ij} is the intensity information at position (x, y), G_{ij_x} and G_{ij_y} are the gradient of image G_{ij} in direction x and y, respectively. $G_{ij_{xx}}$ and $G_{ij_{yy}}$ are the second-order gradient of image G_{ij} in direction x and y, respectively, which can be gain by the convolution between $[-1; 2; -1]$ and the image G_{ij}.

Then, Gabor features are divided into small overlapping rectangular regions. The covariance descriptor of the region is computed as:

$$C_{ij,r} = \frac{1}{n-1} \sum_{(x,y)\in r} (f_{ij}(x, y) - \bar{f_{ij}})(f_{ij}(x, y) - \bar{f_{ij}})^T \tag{8}$$

where $\bar{f_{ij}}$ is the mean of all the f_{ij} in the region r and n is the number of the pixels in the region r.

For region r and its' symmetrical region r^T in the row, we compute their similarity $d_{ij,r}$ under covariance descriptors and take it as the symmetry of regions:

$$d_{ij,r} = d(C_{ij,r}, C_{ij,r^T}) = \sqrt{\sum_{p=1}^{P} \ln^2 \lambda_p(C_{ij,r}, C_{ij,r^T})} \tag{9}$$

where $\lambda_p(C_{ij,r}, C_{ij,r^T})$ is the p-th generalized eigenvalues of $C_{ij,r}$ and C_{ij,r^T}. The symmetrical region r^T is the r-th region of G_{ij}^T and G_{ij}^T is the flipped image of G_{ij}.

Finally, the symmetry metrics are concatenated to form the image representation:

$$D = (d_{1,1,1}, \cdots, d_{1,1,R}, \cdots, d_{1,K,R}, \cdots, d_{M,K,R}) \tag{10}$$

where R is the number of regions, M and K are the number of Gabor bands and orientations, respectively.

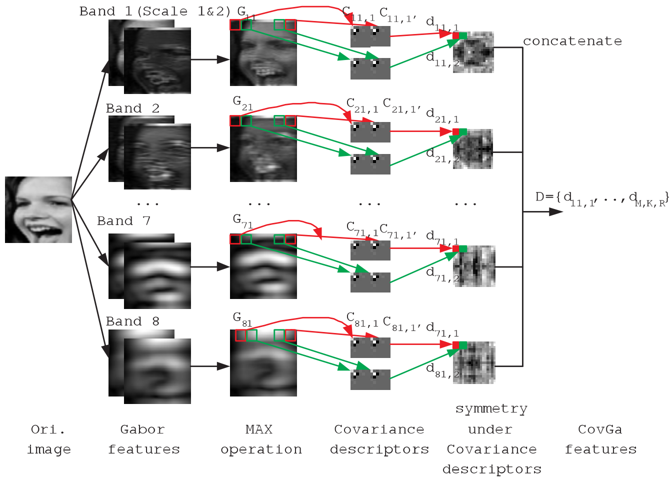

Fig. 4. The flowchart of the proposed CovGa.

In CovGa, the distance between two face images I_i and I_j is obtained by computing the Euclidian distance between their representations D_i and D_j:

$$d(I_i, I_j) = ||D_i - D_j|| \tag{11}$$

Since Gabor filters and covariance descriptors are both known to be tolerant to illumination variations, the CovGa representation is also robust to illumination variations.

C. Enhancement of CovGa

The dimensionality the original CovGa representation is very high because of the multi-scales and multi-orientations of Gabor filters. Reducing the dimensionality makes the method more efficient. Here we show that the simple method, such as Principal Component Analysis (PCA) [11], can work well for the proposed representation.

Besides dimension reduction, head pose estimation evidently needs discriminating features rather than pure representation. Therefore, we need combing the discriminant analysis method with the CovGa representation in order to improve the recognition performance. Generally, the performance of the supervised method is much better than that of the unsupervised method. So, we can use the supervised method, such as Linear Discriminant Analysis (LDA) [12], Marginal Fisher Analysis (MFA) [13] or Locality Sensitive Discriminant Analysis (LSDA) [14], to improve the accuracy of head pose estimation. In this paper, we only select LDA for its simplify and then propose the method Supervised Covariance Descriptor of Gabor filters (sCovGa). In fact, LDA can be replaced by other supervised methods in sCovGa.

D. CovGa for Head Pose Estimation

Since the extraction of CovGa (sCovGa) can be regarded as the preprocessing step for yaw estimation, it should be combined with the classifier to get the yaw pose of the input image. In this paper, NC is selected as the classifier to evaluate the performance of the proposed features. In NC classifier, for the training samples with the same pose, k-mean method is applied to find their k centroids. Then we

Fig. 5. The face images of one subject in the CAS-PEAL database.

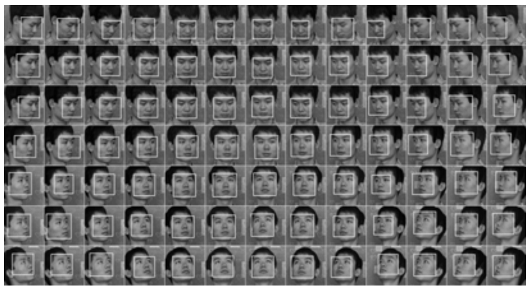

Fig. 6. The face images of one subject in the Multi-Pose database.

compute the distance between the input feature and each class centroid, and take the class with the smallest Euclidean distance as the output label. Compared with the Nearest Neighbor (NN) classifier, NC classifier can eliminate the error caused by the identify since the image difference of the same people with the near poses maybe less than the image difference of the different people with the same pose.

IV. Experiments

In this section, we design the experiments to show the effectiveness of CovGa an sCovGa.

We compare the performance of CovGa with the following unsupervised methods: PCA, GaFour and HOG. We also compare the performance of sCovGa with the supervised method of LDA, GFFF and sHOG. GFFF (sHOG) is the supervised method of GaFour (HOG) by using LDA. As one of the baseline methods in face recognition, PCA [11] is also the baseline method in appearance-based pose estimation.

In paper [7], the authors have shown that the performance of GaFour and GFFF are much better than that of other methods, such as ICA and Gabor. So the results of these methods are not shown in this paper. For all the methods, PCA is used after feature extraction to reduce the dimension of features and 95% of total energy of eigenvalues is kept. For the supervised methods, we apply PCA first for dimensionality reduction and then LDA for discriminant analysis. In the experiment, the region of CovGa is set to 8×8 with overlapping 4×4. To improve the performance of CovGa, we also compute the symmetry of regions in the same column.

For all the images, the face detection method [17] is applied to locate the face region from the input images, and then all the face regions are normalized to the same size of 32×32. Finally, histogram equalization is used to reduce the influence of lighting variations.

In all the experiments, 3-fold cross-validation is used to avoid over-training. Specifically, we rank all the images by subjects and divide them into three subsets. Two subsets are taken as the training set and the other subset is taken as the testing set. In this way, the persons for training and testing are totally different, thus avoiding the over-fitting in identity. Testing is repeated three times, by taking each subset as the testing set. The reported results are the average of all the tests.

A. *Experiments on the CAS-PEAL database*

We first evaluate the performances of different methods on the public CAS-PEAL database [15]. The CAS-PEAL database contains twenty-one poses combining seven yaw poses ($[-45°, 45°]$ with intervals of $15°$ and three pitch poses ($30°$, $0°$ and $-30°$). We use a subset containing totally $4,200$ images of 200 subjects whose IDs range from 401 through 600. Considering the cross-validation, there are totally about $400(= 600/3 \times 2)$ samples for each pose in the training set. Some images in the CAS-PEAL database are shown in Fig. 5. The accuracies when the center number varies from 1 to 20 are shown in Fig. 7. The x-axis represents the center number of each class and the y-axis represents the accuracy.

From the above figure, we can see that on the CAS-PEAL database, the results of CovGa are much better than those of PCA and HOG, while similar to those of GaFour. But after using LDA, the results of sCovGa are the best of all methods. By combining with the supervised method, the advantage of the symmetry of 2D regions is more clearly. For the unsupervised methods, it can also be seen that the accuracy increases with the increase of k when k is very small. However, for the supervised methods, such as sCovGa, the accuracies are nearly equal for different k's, which actually implies the excellent compactness of each class in the feature space obtained by LDA. Especially, for different number of centers, the accuracies of sCovGa are const to 93.2% while the results of GFFF are 90.5%. The robustness of sCovGa to the number of centers shows the good discriminate-ability of sCovGa. In the real system, we can just use 5 or 6 centers for each pose, which can decrease the computing

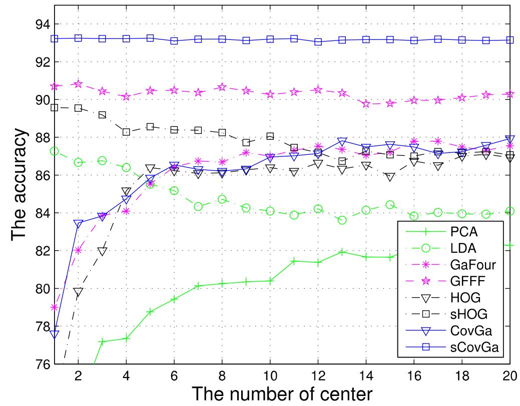

Fig. 7. The accuracies on the CAS-PEAL database. The x-axis represents the center number of each class and the y-axis represents the accuracy.

cost of matching the gallery sample and the probe sample. We attribute the improvement of CovGa and sCovGa to two aspects. First, by using the symmetry of regions, the feature is more related to the pose variations. Second, in computing the symmetry of region, we use covariance descriptors, in which rotations and illuminations changes are absorbed by the covariance matrix.

B. *Experiments on the Multi-Pose database*

The second experiment is on the private Multi-Poses database, which consists of $3,030$ images of 102 subjects taken under normal indoor lighting conditions and fixed background. The yaw poses and the pitch poses range within $[-50°, +50°]$ with intervals of $1°$. The sample number is 30 for each class (i.e. yaw pose). Since the poses of this database are near continuous, we use the error mean between the predicted label and the ground-truth, not the accuracy, as the measure of the performance. Some images of the results of face detection are shown in Fig. 6. Considering that the sample number is about to $40(= 60/3 \times 2)$ for each class (pose) in the training set, the maximal centroid number for each pose is limited to 7, which is different from that in the experiment on the CAS-PEAL database. The error means when the center number varies from 1 to 7 are shown in Fig. 8.

From the above figure, we can see on the MultiPoses database the advantage of the proposed method is more obviously. The results of CovGa are best of all the unsupervised methods and the results of sCovGa are best of all the methods. Especially, for different number of centers, the error means of sCovGa are near to $4.1°$ while those of GFFF are $4.7°$. The error means of CovGa are near to $4.5°$ which even better than those of the supervised methods. The good performance of the proposed method show that the proposed method can improve the performance of head pose estimation by using the symmetry of regions.

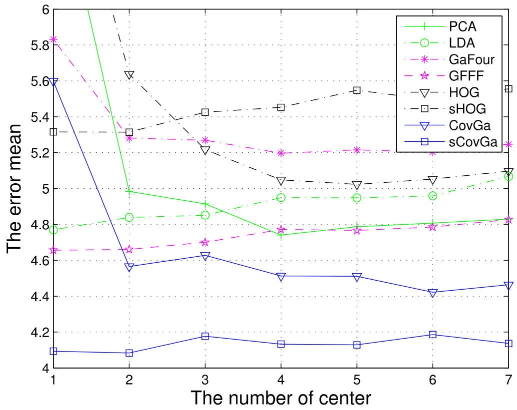

Fig. 8. The error means on the Multi-Poses database. The x-axis represents the center number of each class and the y-axis represents the error mean.

V. Conclusions

Based on the relationship between the symmetry of the face image and the head pose, we propose in this paper a novel face representation method for head yaw estimation. Compared with the symmetry of 1D signals, the symmetry of 2D regions is much more related to the pose variations of head while robust to misalignment. To extract the symmetry of regions which locate in the same horizontal position, we use covariance descriptor on the Gabor representations. The results on two databases show the effectiveness of the proposed method.

There are also several aspects to be studied in the future. First, considering the necessary of the real-time system, the covariance descriptor should be replaced by its fast version. Second, in this paper, we select LDA for its simpleness. In fact, we should investigate which supervised method is more suitable to combine with CovGa for improving the accuracy of the proposed method further.

Acknowledgment

This paper is partially supported by National Natural Science Foundation of China under contract No. 61003103, 61001193 and 61173065, Research Fund for the Doctoral Program of Higher Education under contract No. 20100142120029 and Hubei Provincial Natural Science Foundation under contract No. 2010CDB02304. The authors would like to thank the reviewers for their helpful comments and suggestions.

References

[1] E. Murphy-Chutorian and M. M. Trivedi, "Head pose estimation in computer vision: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 4, pp. 607–626, April 2009.

[2] Q. Ji and R. Hu, "3D face pose estimation and tracking from a monocular camera," *Image and Vision Computing*, vol. 20, no. 7, pp. 499–511, 2002.

[3] T. F. Cootes, K. Walker, and C. J. Taylor, "View-based active appearance model," *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 227–232, 2000.

[4] J. Xiao, T. Moriyama, T. Kanade, and J. F. Cohn, "Robust full-motion recovery of head by dynamic templates and re-registration techniques," *International Journal of Imaging Systems and Technology*, vol. 13, pp. 85–94, 2003.

[5] L. Chen, L. Zhang, Y. Hu, M. Li, and H. Zhang, "Head pose estimation using fisher manifold learning," *Proc. IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, pp. 203–207, 2003.

[6] Y. Wei, L. Fradet, and T. Tan, "Head pose estimation using Gabor eigenspace modeling," *Proc. IEEE International Conference on Image Processing*, pp. 281–284, 2002.

[7] B. Ma, S. Shan, X. Chen, and W. Gao, "Head yaw estimation from asymmetry of facial appearance," *IEEE Transactions on Systems, Man, and Cybernetics–Part B: Cybernetics*, vol. 38, no. 6, pp. 1501–1512, 2008.

[8] C. Liu and H. Wechsler, "Robust coding schemes for indexing and retrieval from large face databases," *IEEE Transactions on Image Processing*, vol. 9, no. 6, pp. 132–137, 2000.

[9] O. Tuzel, F. Porikli, and P. Meer, "Region covariance: a fast descriptor for detection and classification," *Proc. European Conference on Computer Vision*, pp. 589–600, 2006.

[10] O. Tuzel, F. Porikli, and P. Meer, "Pedestrian detection via classification on riemannian manifolds," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 10, pp. 1713–1727, 2008.

[11] M. A. Turk and A. P. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[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. 20, no. 7, pp. 711–720, 1997.

[13] S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, and S. Lin, "Graph embedding and extension: A general framework for dimensionality reduction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, pp. 40–51, 2007.

[14] D. Cai, J. Han, X. He, K. Zhou, and H. Bao, "Locality sensitive discriminant analysis," *Proceedings of the International Joint Conference on Artifical Intelligence*, pp. 1713–1726, 2007.

[15] W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang, and D. Zhao, "The CAS-PEAL large-scale chinese face database and baseline evaluations," *IEEE Transactions on Systems, Man and Cybernetics, Part A*, vol. 38, no. 1, pp. 149–161, 2008.

[16] L. Wiskott, J. M. Fellous, Norbert Krüger, and C. V. D. Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 775–779, 1997.

[17] S. Yan, S. Shan, X. Chen, W. Gao, and J. Chen,"Matrix-Structural Learning (MSL) of cascaded classifier from enormous training set," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2007.