Head Yaw Estimation From Asymmetry of Facial Appearance

Bingpeng Ma, Shiguang Shan, *Member, IEEE*, Xilin Chen, *Member, IEEE*, and Wen Gao, *Senior Member, IEEE*

***Abstract*—This paper proposes a novel method to estimate the head yaw rotations based on the asymmetry of 2-D facial appearance. In traditional appearance-based pose estimation methods, features are typically extracted holistically by subspace analysis such as principal component analysis, linear discriminant analysis (LDA), etc., which are not designed to directly model the pose variations. In this paper, we argue and reveal that the asymmetry in the intensities of each row of the face image is closely relevant to the yaw rotation of the head and, at the same time, evidently insensitive to the identity of the input face. Specifically, to extract the asymmetry information, 1-D Gabor filters and Fourier transform are exploited. LDA is further applied to the asymmetry features to enhance the discrimination ability. By using the simple nearest centroid classifier, experimental results on two multipose databases show that the proposed features outperform other features. In particular, the generalization of the proposed asymmetry features is verified by the impressive performance when the training and the testing data sets are heterogeneous.**

***Index Terms*—Fourier transform, Gabor filters, head yaw estimation, linear discriminant analysis (LDA), nearest centroid (NC) classifier.**

I. Introduction

STATISTICS indicate that approximately 75% of the faces in photographs are nonfrontal [1]. However, the best known face perception systems can only deal with near-frontal faces reliably, and the performances of these systems degrade dramatically on nonfrontal faces. Therefore, pose-invariant face perception has been an active research topic for several years. To achieve the expected robustness to pose variation, one may expect to process face images differently according to their pose parameters. In this case, the pose of the input faces must be estimated as a prerequisite for sequent processes.

Pose estimation essentially means the computation of three types of rotation of a head: yaw (looking left or right), pitch (looking up or down), and roll (tilting left or right). Among them, the roll rotation can be computed easily by the relative position of the feature points, but the other two rotations are rather difficult to estimate. Because the estimation of the yaw rotation has many important applications, it attracts more attention than pitch estimation [10]. Therefore, most previous works mainly focus on the estimation of the yaw (sometimes also pitch) rotation. These methods can be categorized into two main groups [5]: model-based methods [2]–[9] and appearance-based methods [10]–[15].

The model-based methods make use of the 3-D structure of human head. Typically, they build 3-D models for human faces and attempt to match the facial features such as the face contour and the facial components of the 3-D face model with their 2-D projections. Nikolaidis and Pitas [4] propose a head pose estimation method from the distortion of the isosceles triangle formed by the two eyes and the mouth. Ji and Hu [5] propose that the shape of a 3-D face can be approximated by an ellipse and that the head pose is computed from the detected ellipse of the face. Similarly, Xiao *et al.* [9] utilize the cylindrical head model and present a robust method to recover the full motion of the head under perspective projection. Since these methods generally run very fast, they can be used in video tracking and multicamera surveillance. However, they also share some common disadvantages. First, they are sensitive to the misalignment of the facial feature points, while the accurate and robust localization of facial landmarks remains an open problem. Second, it is difficult to precisely build the head model for different person. Third, these methods generally require high resolution and image quality, which cannot be satisfied in many applications such as video surveillance.

Compared with the model-based methods, the appearance-based methods typically assume that there exists a certain relationship between the 3-D face pose and some properties of the 2-D facial image, and they use a large number of training images to infer the relationship by using statistical learning techniques. Darrell *et al.* [11] propose the use of eigenspace for head pose estimation. A separate eigenspace is computed for each face under each possible pose. The pose is determined by projecting the input image onto each eigenspace and selecting the one with the lowest residual error. In some sense, the method can be formulated as a maximum *a posteriori* estimation problem. Gong *et al.* [12], [23] study the trajectories of multiview faces in linear principal component analysis (PCA) feature

Manuscript received October 19, 2007; revised February 9, 2008 and May 21, 2008. First published October 3, 2008; current version published November 20, 2008. This work was supported in part by the National Natural Science Foundation of China under Contract 60533030 and Contract 60772071, by the Hi-Tech Research and Development Program of China under Contract 2006AA01Z122 and Contract 2007AA01Z163, and by the 100 Talents Program of the Chinese Academy of Sciences, Program of New Century Excellent Talents in University (NCET-04-0320), Natural Science Foundation of Beijing Municipal under Contract 4061001. This paper was recommended by Associate Editor X. Jiang.

B. Ma is with the ICT-ISVISION Joint Research and Development Laboratory for Face Recognition and the Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100090, China, and also with the Graduate School of Chinese Academy of Sciences, Beijing 100049, China (e-mail: bpma@jdl.ac.cn).

S. Shan and X. Chen are with the ICT-ISVISION Joint Research and Development Laboratory for Face Recognition and the Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China (e-mail: sgshan@jdl.ac.cn; xlchen@jdl.ac.cn).

W. Gao is with the School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China (e-mail: wgao@jdl.ac.cn).

Digital Object Identifier 10.1109/TSMCB.2008.928231

space and use kernel support vector machines (SVM) for pose estimation. Li *et al.* [24] exploit independent component analysis (ICA) and its variants, independent subspace analysis and topographic ICA for pose estimation. ICA takes into account higher order statistics required to characterize the view of objects and suitable for the learning of view subspaces. Chen *et al.* [10] propose the kernel-based method to deal with the nonlinearity of head pose estimation. They choose the face images of two specific head pose angles and utilize classification-based nonlinear interpolation to estimate the head poses between the two angles. Wei *et al.* [15] propose that the optimal orientation of the Gabor filters can be selected for each pose to enhance pose information and eliminate other distractive information like variable facial appearance or changing environmental illumination. In their method, a distribution-based pose model is used to model each pose cluster in Gabor eigenspace. Since the set of all facial images with various poses is intrinsically a 3-D manifold in image space, manifold learning [16]–[18] for head pose estimation is getting popular recently [19]–[22]. For instance, in [19], by thinking globally and fitting locally, Fu and Huang propose the use of the graph embedded analysis method for head pose estimation. They first construct the neighborhood weighted graph in the sense of supervised locally linear embedding [16]. The unified projection is calculated in a closed-form solution based on the graph embedding linearization and then projects new data into the embedded low-dimensional subspace with the identical projection. The head pose is finally estimated by the K-nearest neighbor classification. Intuitively, these appearance-based methods can naturally avoid the aforementioned drawbacks of the model-based methods. Therefore, they have attracted more and more attention.

However, the features used by all these methods are extracted from the entire face, which is generally vectorized as 1-D vector which lose the face structure in some sense; therefore, these features contain not only pose information but also information about identity, lighting, expression, etc. Understandably, given a representation of the same dimension, its discriminative ability will be inevitably lower if more nonpose information is preserved. In the extreme case, when the nonpose information surpasses the pose information, the performance of pose estimation will be much worse.

The key of pose estimation is to seek suitable features closely relevant to pose variations and reliably insensitive to facial variations irrelevant to pose, such as identity, lighting, expression, and the possible bias of the training set. Keeping this in mind, in this paper, we investigate the asymmetry of the facial appearance and reveal its capacity as an excellent pose-oriented face representation. Intuitively, the proposed method is based on a commonly accepted assumption that human head is bilaterally symmetrical. As an important geometric attribute of heads, the symmetry has been used to estimate the head pose in the model-based methods [25], [26]. Recently, the asymmetry of facial appearance has also been used in face recognition [27], [28] and expression recognition [29]. However, the question of how to use the asymmetry in the appearance-based pose estimation remains unanswered. With this in mind, in this paper, we propose a novel method using the asymmetry to estimate the head yaw. We show that the asymmetry is indeed closely related to the pose variations and, at the same time, independent of other facial variations, particularly the identity.

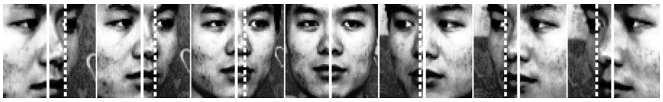


Fig. 1. Relationship between the symmetry plane of the head and the center lines of the images. The solid line is the center line of the image, and the dash line is the symmetry line.

Specifically, in our study, Fourier analysis is used to represent the asymmetry, i.e., to represent the pose. By taking the intensities of each row of the face image as a 1-D signal, Fourier transform of these signals is taken as the feature extractor to extract the asymmetry features of the head. However, the signal analysis directly in the pixel domain generally suffers from noise. Moreover, local feature analysis can represent object more robust and effectively [31]. Therefore, in our method, 1-D Gabor filters are first convolved with the row signals to reduce noise and extract the local information before using the Fourier analysis. Unlike 2-D Gabor filters normally used in face recognition, 1-D Gabor filters are used to keep the asymmetry of facial appearance and reduce the computational complexity. Furthermore, linear discriminant analysis (LDA) [32] is applied after feature extraction to enhance the discriminative power and reduce the dimension. For classification, the nearest centroid (NC) classifier is exploited to validate the effectiveness of the proposed method.

The remaining part of this paper is organized as follows. In Section II, we show that the asymmetry does exist in the facial appearance from the real data. In Section III, we describe the proposed method in detail and analyze its characteristics. Section IV presents the method combining LDA with the asymmetry features. Experiments are given in Section V. Conclusion is drawn in Section VI with some discussions on the future work.

II. Asymmetry Properties

This section presents in detail the relationship between the asymmetry of the face image and yaw variations, as well as some analysis of the asymmetry of Fourier transform.

A. Asymmetry Analysis From Facial Appearance

To show the asymmetry of the face images, we illustrate the relationship between the center line of the images and the symmetry plane of heads in Fig. 1. In Fig. 1, the solid line is the center line of the 2-D images, and the dash line is the sagittal line, which we call symmetry line. The symmetry line is the intersection of the sagittal plane of head and the plane, which parallels to the image plane and passes the noise tip. Since the locations of the two eyes are scaled to a constant in the image when we crop the face, the center vertical line of the image is the center vertical line of two eyes in the 2-D image. From Fig. 1, we can find the close relationship of the two lines with pose variations. In the front-view image, the two lines are overlapping. With the pose varying from the front to the half-profile, the deviation between the two lines increases gradually. The deviation, which is related to pose variations and the asymmetry of the images at the same time, is caused by the projection from three to two dimensions. Therefore, the symmetry of the 3-D face also exists in the 2-D intensity image in some sense. From the viewpoint of the signal processing, if one takes intensities in the row of the image as the signal, it is approximately symmetric in the front view, but asymmetric when pose varies to the half-profile. To sum up, we can conclude that the asymmetry previously defined is closely related to the pose variations.

B. Asymmetry Properties of Fourier Transform

Asymmetry properties of Fourier transform are very useful in many areas. We briefly review the Fourier transform in order to analyze its asymmetry. Formally, the discrete Fourier transform of a real vector $\mathbf{y}$ with n elements is a complex vector $\mathbf{Y}$ with n elements

$$\mathbf{Y}_k = \sum_{j=0}^{n-1} e^{\frac{-2k\pi ij}{n}} \mathbf{y}_j = a_k + i \times b_k, \qquad k = 0, 1, \ldots, n-1 \quad (1)$$

where i is the imaginary unit. According to Oppenheim and Schafer [33], any sequences can be expressed as a sum of an even part (the symmetry part) and an odd part (the asymmetry part). When Fourier transform is performed on the real sequence, the even part transforms to the real part, and the odd part transforms to the imaginary part. Clearly, the asymmetry measure of the real sequence in the frequency domain should be a function of the imaginary part. A simple measure can be defined as the energy e_b of the imaginary part of the entire sequence, denoted by $e_b = (\sum_{k=0}^{n-1} b_k^2)^{1/2}$ [28]. The lower the value of e_b, the less the amount of asymmetry (and, hence, more symmetry) and vice versa. On the contrary, $e_a = (\sum_{k=0}^{n-1} a_k^2)^{1/2}$ can be used as the measure of the symmetry. To show the aforementioned property of Fourier transform, we compute e_a and e_b of the function $\sin(t)$ and $\cos(t)$ with $t \in [-\pi, \pi]$ as examples. As for the even function $\cos(t)$, when the sampling interval $\triangle = 2\pi/629$, $e_b = 1.5$, much less than $e_a = 444.5$. On the contrary, for the odd function $\sin(t)$, $e_b = 445.0$, much larger than $e_a = 1.5$. These two examples clearly illustrate that the aforementioned representation of the frequency domain is effective to measure the asymmetry.

C. Asymmetry Measure of Head Images

As discussed in Section II-A, the symmetry of facial images decreases with the increase of pose deviation from frontal view. Obviously, this observation holds for the intensity sequence of each row in the face images. Thus, it is a natural extension to define the asymmetry measure of a face image as the average of the asymmetry of all the rows. Formally, the symmetry measure of a face image is defined as follows:

$$E_a = \frac{1}{h} \sum_{j=1}^{h} e_{aj} \quad (2)$$

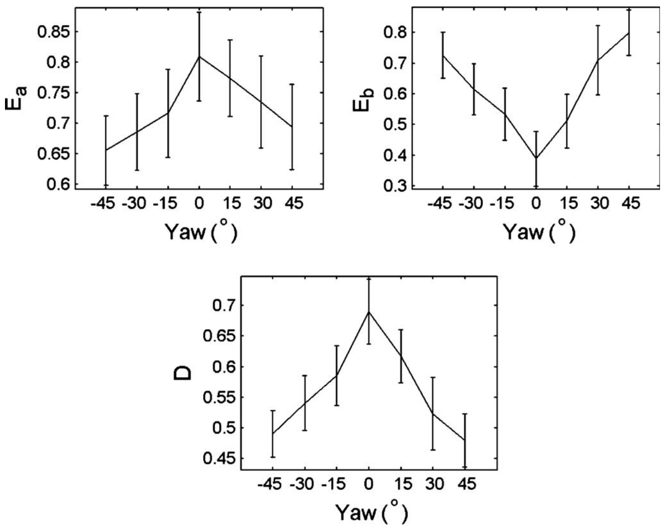


Fig. 2. Symmetry measures of the different poses on the CAS-PEAL database. The horizontal axes represent the poses, and the vertical axes represent the measures.

where e_{aj} is the symmetry measure of the intensity sequence of the jth row and h is the number of rows of the image. Similarly, the asymmetry measure E_b can be defined as follows:

$$E_b = \frac{1}{h} \sum_{j=1}^{h} e_{bj}. \quad (3)$$

To validate the effectiveness of E_a and E_b, experiments are conducted on the pose subset of the CAS-PEAL face database. The means and the standard deviations of the measure E_a and E_b of each pose are shown in Fig. 2. The horizontal axis represents the poses, whereas the vertical axis shows the measure. Since E_a and E_b reflect the two sides of the symmetry, in Fig. 2, we also show the mean and the standard deviation of a unified measure D, which is defined by combining the symmetry and asymmetry as follows:

$$D = \frac{1}{h} \sum_{j=1}^{h} \frac{e_{aj}}{e_{aj} + e_{bj}}. \quad (4)$$

Obviously, the lower the value of D, the greater the amount of asymmetry and vice versa. From Fig. 2, one can clearly see that the symmetry decreases and the asymmetry increases when the poses vary from the front to the half-profile. However, it should be noted that although the asymmetry exists in images and is related to the pose variations, we still cannot use directly the asymmetry measure to estimate the head pose. There are two reasons: 1) It is difficult to distinguish the right and the left poses only from the measure, and 2) the standard deviations are too large to estimate yaw accurately.

Fortunately, another significant hint obtained from the aforementioned analysis is that the spatial asymmetry part of the face corresponds to the imaginary part of the Fourier transform and that the symmetry part corresponds to the real part. It is just this observation that inspires us to propose the idea of extracting asymmetry features in the frequency domain. Specifically, in our method, we use the real and the imaginary parts together as a whole feature to estimate the poses, and hereinafter, they are called the asymmetry features. These features are expected to contain sufficient information that can distinguish various poses. Furthermore, they can be used in the Gabor–Fourier (GaFour) method, which is introduced in detail in the next section.

III. GaFour Method

This section presents in detail the GaFour feature and its good property to represent pose.

A. One-Dimensional Gabor Filters

Although the real and imaginary parts of Fourier transform can reflect the asymmetric information of the face images with pose variations, it is still difficult to use them in yaw estimation directly. Since the intensity feature can be affected by many factors, such as lighting variations and noise, the asymmetry cannot be accurately extracted from the intensity feature directly. At the same time, due to the holistic property of Fourier transform, face representation in frequency domains loses the spatial position information in the sequence, and the noise of one pixel can influence the full frequency domain.

Considering the limitation of Fourier transform, 1-D Gabor filters are used before extracting the asymmetry from the raw image, and the GaFour method, which is the combination of Gabor filters and Fourier transform, is proposed for yaw estimation. Since Gabor filters can retain the asymmetry of image and spatial information, they can weaken the drawback of Fourier transform and make the asymmetry features more accurate and relevant to the pose.

Gabor filters are chosen for their biological relevance and technical properties. The multiscale Gabor filters have similar shapes as the receptive fields of simple cells in the primary visual cortex [34]. One-dimensional Gabor filters can be defined as

$$g_\mu(r) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-r^2}{2\sigma^2}} e^{i(2\pi\mu r)} \tag{5}$$

where μ is the modulation frequency and σ is the scale parameter which determines the width of the Gaussian envelope. The Gabor representation of a signal is the convolution of the signal with a family of Gabor filters. The convolution result $O_\mu(r)$ corresponding to the Gabor filter at frequency μ can be defined as follows:

$$O_\mu(r) = s(r) * g_\mu(r) \tag{6}$$

where $*$ denotes the convolution operator and $s(r)$ is the gray row signal of an image. Fig. 3 shows the Gabor kernels with four frequencies. In Fig. 4, we show the Gabor representations of some images.

B. GaFour Method

Based on the aforementioned analysis, we propose the GaFour method to extract the features to estimate the head pose. The flow chart of the proposed method is shown in Fig. 5.

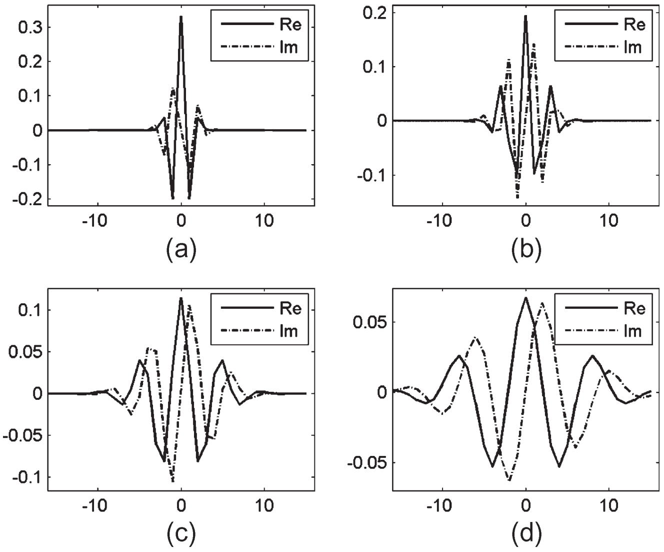


Fig. 3. Gabor kernels with four frequencies.

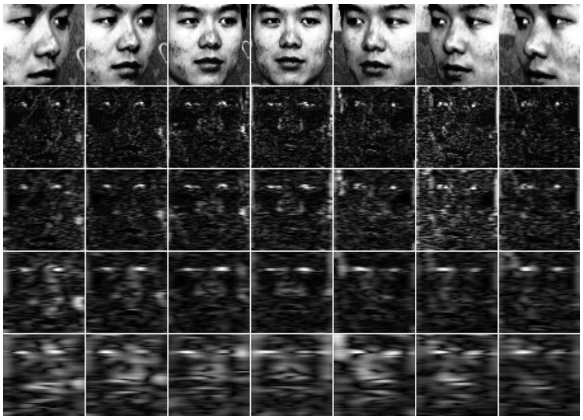


Fig. 4. Gabor representation of some sample images.

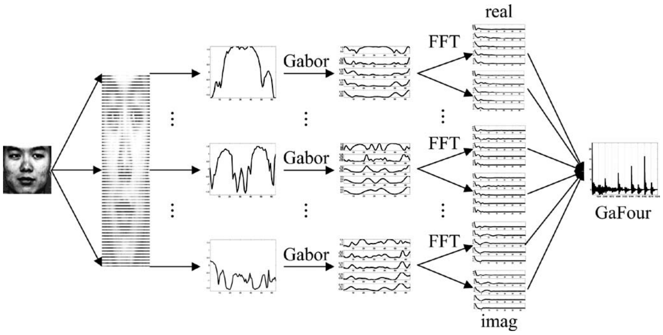


Fig. 5. Flow chart of the GaFour method.

The main procedure of the GaFour method is described briefly as follows: 1) Each row slice of the image is treated as one signal vector; thus, the image is taken as the combination of many signals. 2) For each row signal, 1-D Gabor filters with various frequencies are operated. In our case, five frequencies are exploited; thus, five magnitude signals (vectors) are generated for each row of the input image. 3) For each magnitude signal, Fourier transform is conducted, and all the vectors of the real and imaginary parts are combined together as the final asymmetry features. The asymmetry features in GaFour also

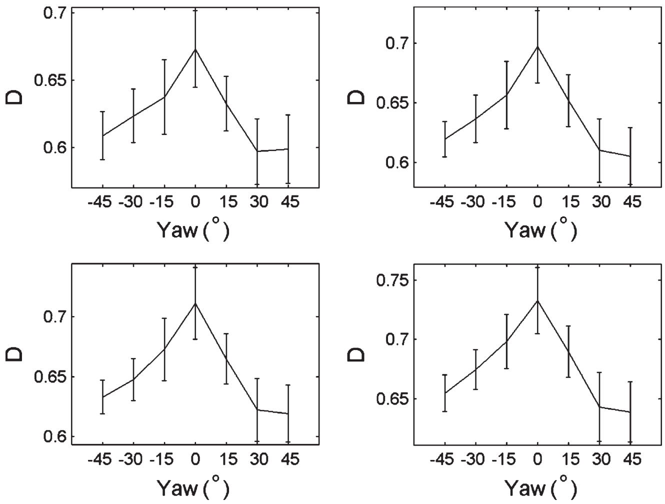


Fig. 6. Symmetry measures D of the different poses. The features are extracted by the GaFour method with four different frequencies.

include the part of the symmetry because the symmetry can be seen as the complement of the asymmetry.

C. Advantages of Using One-Dimensional Gabor Filters

Compared with Fourier transform, 1-D Gabor filters with Gaussian envelope work on the signal in a local window. The influence of the noise of one pixel is limited in the local window (rather than the full frequency domain as for the Fourier transform) and decreases with the increase of the distance between the center of the window and the noise pixel. Hence, it can achieve the optimal localization in both the spatial and frequency domains.

Compared with 1-D Gabor filters, 2-D Gabor filters have been used in face recognition and other related problems. However, there are three reasons for us to select 1-D Gabor filters, but not 2-D Gabor filters, as our feature extractors in the proposed method: 1) the asymmetry exists in the intensities of each row of the images, which can be regarded as 1-D vector; therefore, it is natural to use 1-D Gabor filters; 2) the computational cost and the dimension of the resulting features of 1-D Gabor filters are greatly lower than those of 2-D Gabor filters; and 3) the fast speed makes 1-D Gabor filters possible for a real-time pose estimation system.

D. Asymmetry Analysis of GaFour

One of the key points of the GaFour method is that the asymmetry of the face images should be kept after 1-D Gabor filters. To show that the asymmetry of the images still exists after the feature extraction by the GaFour method, we do statistics of the measure D in (4) on the same subset of CAS-PEAL database as in Fig. 2. The results are shown in Fig. 6, which illustrates the asymmetry measure D for four 1-D Gabor filters. From Fig. 6, we can see that the symmetry based on the GaFour feature decreases gradually as before when the pose varies from the front to the half-profile for all the four Gabor filters. This implies that the extracted GaFour feature can well preserve the asymmetry information of the face images. On the other hand, the Gabor transformation is a window-Fourier transformation. Therefore, when the window is relatively smaller than the scale of the object, the symmetry will still be kept.

IV. GaFour Fisher Feature for Yaw Estimation

In this section, we present our final feature extraction method and introduce its combination with the NC classifier for yaw estimation.

A. GaFour Fisher Feature

In face recognition, Gabor Fisher Classifier (GFC) [35] method has achieved very good performances on a lot of databases. The GFC method, which is robust to illumination and facial expression variability, applies LDA to the Gabor feature vector derived from the Gabor wavelet representation of face images. To encompass all the features produced by the different Gabor kernels, one concatenates the resulting Gabor wavelet features to derive an augmented Gabor feature vector. The dimensionality of the Gabor vector space is then reduced under the eigenvalue selectivity constraint of LDA to derive a low-dimensional feature representation while enhancing the discriminant ability.

Inspired by GFC, we propose the GaFour Fisher feature (GF^3) method. In the extraction of GF^3, the GaFour features are first computed from the input head image, and then, LDA is applied to the GaFour features to improve the discriminative ability. LDA has been recognized as one of the most successful methods in face recognition [36]. In LDA, for a c-class problem, the within-class scatter matrices $\mathbf{S}_{\mathrm{w}}$ and the between-class scatter matrices $\mathbf{S}_{\mathrm{b}}$ are computed as follows:

$$\mathbf{S}_{\mathrm{w}} = \frac{1}{N}\sum_{i=1}^{c}\sum_{\mathbf{x}\in\mathcal{D}_i}(\mathbf{x}-\mathbf{m}_i)(\mathbf{x}-\mathbf{m}_i)^{\mathrm{T}} \tag{7}$$

$$\mathbf{S}_{\mathrm{b}} = \frac{1}{c}\sum_{i=1}^{c} n_i(\mathbf{m}_i-\mathbf{m})(\mathbf{m}_i-\mathbf{m})^{\mathrm{T}} \tag{8}$$

where $\mathbf{m}_i$, n_i, and $\mathcal{D}_i$ are the mean vector, the total number of samples, and the sample set of the class i, respectively, $\mathbf{m}$ is the overall mean vector, and N is the total number of all the samples. $\mathbf{S}_{\mathrm{w}}$ represents the average scatter of the sample vectors $\mathbf{x}$ of different classes around their respective means $\mathbf{m}_i$. Similarly, $\mathbf{S}_{\mathrm{b}}$ represents the scatter of the conditional mean vectors $\mathbf{m}_i$ around the overall mean vector $\mathbf{m}$. Through a linear transformation, the original feature representation is projected into a new LDA subspace where $\mathbf{S}_{\mathrm{b}}$ is maximized while $\mathbf{S}_{\mathrm{w}}$ is minimized by maximizing the Fisher separation criterion $J(\mathbf{W}) = (\|\mathbf{W}^{\mathrm{T}}\mathbf{S}_{\mathrm{b}}\mathbf{W}\|/\|\mathbf{W}^{\mathrm{T}}\mathbf{S}_{\mathrm{w}}\mathbf{W}\|)$. The optimal projection matrix $\mathbf{W}$ can be obtained by solving a generalized eigenvalue problem: $\mathbf{S}_{\mathrm{b}}\mathbf{W} = \boldsymbol{\lambda}\mathbf{S}_{\mathrm{w}}\mathbf{W}$.

B. Dimension Analysis of GF^3

The feature dimension is very important in computation. A lower dimension not only can improve the computational speed but also can avoid the small sample size (3S) problem. Therefore, it is necessary for us to address the issue of GF^3 dimension.

From the flow chart of GaFour in Fig. 5, one can know that the dimension of the GaFour features is $2 \times s \times w \times h$, where s is the number of 1-D Gabor filters, and w and h are the width and the height of the image, respectively. Although the dimension of the original GaFour features is as high as $10\,240 (= 2 \times 5 \times 32 \times 32)$ in our case, it can be reduced greatly benefiting from the properties of Fourier transform.

Since the real part is the even-symmetric sequence and the imaginary part is the odd-symmetric sequence, the complex vector $\mathbf{Y}$ with n elements in Section II-B can be represented by $a_k (k = 0, \ldots, n/2)$ and $b_k (k = 1, \ldots, n/2 - 1)$ (b_0 and $b_{n/2}$ are discarded for zeros). Thus, for the 32 × 32 image, the dimension after Fourier transform is $1024 (= (32 \times 17) + (32 \times 15))$, which is the same as the original dimension of the image. Clearly, the dimension of the GaFour features can be reduced from 10 240 to $5120 (= 5 \times 1024)$ while no information is lost.

For the complex vector $\mathbf{Y}$ after Fourier transform, the low frequencies contain general and gross information of the signal, whereas the high frequencies are mostly from the details of the signal and noise. For face images, the details are mainly related to the identity of the face rather than the pose. Therefore, the high frequencies can be safely cut off. In this paper, 97% of the total energy (keeping about nine lowest frequencies) is kept, and the final dimension of the GaFour features can be reduced from 10 240 to about $2880 (= 2 \times 5 \times 9 \times 32)$.

Compared with the dimension of the GaFour features, the dimension of GF^3 is much lower, since LDA can commonly reduce the dimension to $c - 1$. For yaw estimation, c is generally very low. For instance, even if 1° yaw interval is taken as one class, there are only 181 classes from the right to the left profile. Therefore, the final dimension of GF^3 is not more than 180, which is much less than 2880, the dimension of the GaFour features.

C. GF^3 for Yaw Estimation

Since the extraction of GF^3 can be regarded as the preprocessing step for yaw estimation, it should be combined with some classifier to get the yaw of the input image. In this paper, we take the yaw estimation as a yaw classification problem, which is a possible way when the poses in the database are not continuous. In this paper, mainly NC classifier is used as the classifier to evaluate the performance of the proposed features for its simplicity. In fact, the NC classifier can be replaced by other methods, such as SVM [30], support vector regression, or relevance vector machine [39].

V. Experiments

In this section, the proposed GaFour and GF^3 are evaluated on two face database by comparing with other feature extraction methods.

We compare the performance of GaFour with the following unsupervised methods: PCA, ICA, and 2-D Gabor filters (Gabor). As one of the baseline methods in face recognition, PCA [37] is also the baseline method in appearance-based pose estimation. Since ICA has achieved better performances in pose estimation recently [24], we also show its performance by using the code of ICAFaces [38]. For the two architectures to perform ICA on images in ICAFaces, we use ICA1 and ICA2 to represent them, respectively. Since 1-D Gabor filters are applied in GaFour and GF^3, 2-D Gabor filters with five scales and eight orientations are also selected as the comparison method. For all the methods, PCA is used after feature extraction to reduce the dimension of features, and 95% of the total energy of eigenvalues is kept.

We compare the performance of GF^3 with the following supervised methods: LDA and GFC. The LDA-based baseline algorithm, similar to the Fisherfaces method [36], applies first the PCA for dimensionality reduction and then the LDA for discriminant analysis.

For all the images, the face detection method [41] 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.

A. Experiment 1: On Homogeneous Training and Testing Sets

By the term "homogeneous," we mean that the training and testing sets are sampled from the same database, i.e., the imaging conditions are similar for the training and testing sets. In all the experiments, threefold cross-validation is used to avoid overtraining. 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 overfitting 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.

1) Experiments on the CAS-PEAL Database: First, we evaluate the performances of different methods on the public CAS-PEAL database [40], which contains 21 poses combining seven yaw angles (−45°, −30°, −15°, 0°, 15°, 30°, and 45°) and three pitch angles (30°, 0°, and −30°). We use a subset containing totally 4200 images of 200 subjects whose IDs range from 401 to 600.

It is worth pointing out that, since we are mainly concerned with the yaw estimation, all the images in the data set are grouped according to the yaw regardless of the pitch. Therefore, in this experiment, we have seven yaw classes in total. Thus, for LDA, GFC, and GF^3, the reduced dimension is six, which is really low.

Table I shows the classification accuracies of different methods using the NC classifier with the Euclidean distance and the cosine similarity. From the table, three observations can be seen. First, the results of GaFour are the best in the unsupervised methods except ICA2. Second, the GF^3 method yields the best result among all the methods. Third, the results are similar for the cosine and the Euclidean distances. Therefore, only the Euclidean distance is used in the following experiments for simplicity.

As to the NC classifier, since there are totally about $400 (= 600/3 \times 2)$ samples for each angle in the training set, only one centroid cannot reflect sufficiently the distribution of

TABLE I
ACCURACY (IN PERCENT) OF POSE CLASSIFICATION ON THE CAS-PEAL DATABASE

	Methods and Yaw Classification Correct Rates (%)							
Distance	PCA	ICA1	ICA2	LDA	Gabor	GFC	GaFour	GF^3
Euclidean	65.07	67.63	83.35	86.38	64.97	88.66	79.65	91.42
Cosine	64.51	67.34	83.66	86.35	65.26	88.66	79.43	91.01

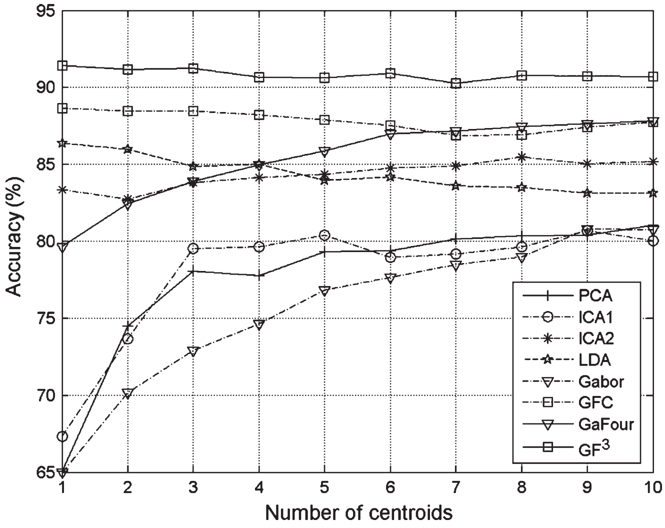


Fig. 7. Accuracy on the CAS-PEAL database. The x-axis is the centroid number of each pose, and the y-axis is the accuracy.

the samples. Therefore, for each angle, k-mean method is used to find k centroids from the training samples, and then, the NC classifier is used to predict the yaw class of the testing samples. The accuracies with various k's ranging from one to ten are shown in Fig. 7. From this figure, it can be seen that the accuracies of GaFour are among the best in the unsupervised methods and that the accuracies of GF^3 are always the best for all k's. 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 LDA, GFC, and GF^3, 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.

2) Experiments on the Multipose Database: Another database that we used is a private multipose database created by ourselves. Unlike the CAS-PEAL database, the poses (particularly in yaw) are almost continuous in the multipose database. The database consists of 7731 images of 102 subjects taken under normal indoor lighting conditions and fixed background with a Sony EVI-D31 camera. The yaw and the pitch angles range within $[-90°, +90°]$ and $[-50°, +50°]$ with intervals of 1°, respectively. The number of images for each people is different ranging from 22 to 142. To slightly reduce the complexity of the experiment, a subset of the database is used, in which the images are selected with the following rules: 1) the yaw angles are limited in the range from $-50°$ to $50°$, and there are no limitations for the pitch angles; 2) the sample number is 60 for each class (i.e., yaw angle); and 3) all the 102 subjects are included in the subset, and the number of the images for each subject is almost the same. Some images of one subject with the results of face detection are shown in Fig. 8. In the experiment, each angle in yaw is taken as one separate class, which means that there are totally 101 classes. Therefore, the dimension of LDA-based features will be not more than $100(= 101 - 1)$. Please also note that the threefold cross-validation evaluation method similar to that on CAS-PEAL is also used here. Therefore, the persons in the training and testing sets are also completely different.

Since the yaw angles are nearly continuous in this data set, unlike the accuracy used in CAS-PEAL experiment, we use the error mean as the performance measurement. The error mean m is computed based on the following equation:

$$m = \frac{1}{N} \sum_{i=1}^{N} |p'_i - p_i| \tag{9}$$

where N is the total number of the testing samples, and p_i and p'_i are the ground-truth and the predicted angles of the ith sample, respectively. Thus, the unit of this measurement is actually rotation degree in yaw.

As to the number of centroids for NC, considering that the sample number is about $40(= 60/3 \times 2)$ for each class/angle in the training set, the maximal centroid number for each angle is limited to seven, which is different from that in the experiment on the CAS-PEAL database.

The error mean m against the number of centroids for NC of different features is shown in Fig. 9. From the figure, it can be seen that the results of GaFour are always better than the other unsupervised methods and that the results of GF^3 are always the best among all the methods for all k's.

From the experiments on the two databases, we can draw the conclusion that the asymmetry of GaFour features are indeed effective to reflect the pose information and that its discriminant ability can be improved further by combining with the LDA to form the GF^3.

3) Robustness for the Image Size: To test the robustness of the methods to the image resolution, we repeat the experiments with three different scales: 16 × 16, 32 × 32, and 64 × 64, which are down sampling from the CAS-PEAL database. The best accuracies for different features are shown in Table II with k's ranging from one to ten.

From Table II, we can draw three conclusions. First, for all the image sizes, the results of GaFour are still the best among the unsupervised methods. Second, for Gabor and GFC, the accuracies are improved greatly when the image resolution varies from low to high, whereas the accuracies of the other method are not very sensitive to the change of the image sizes. Thus, we guess that the suitable scales of 2-D Gabor filter are very critical. Third, the accuracies of GaFour and GF^3 are more robust to the image sizes. For GaFour and GF^3, the

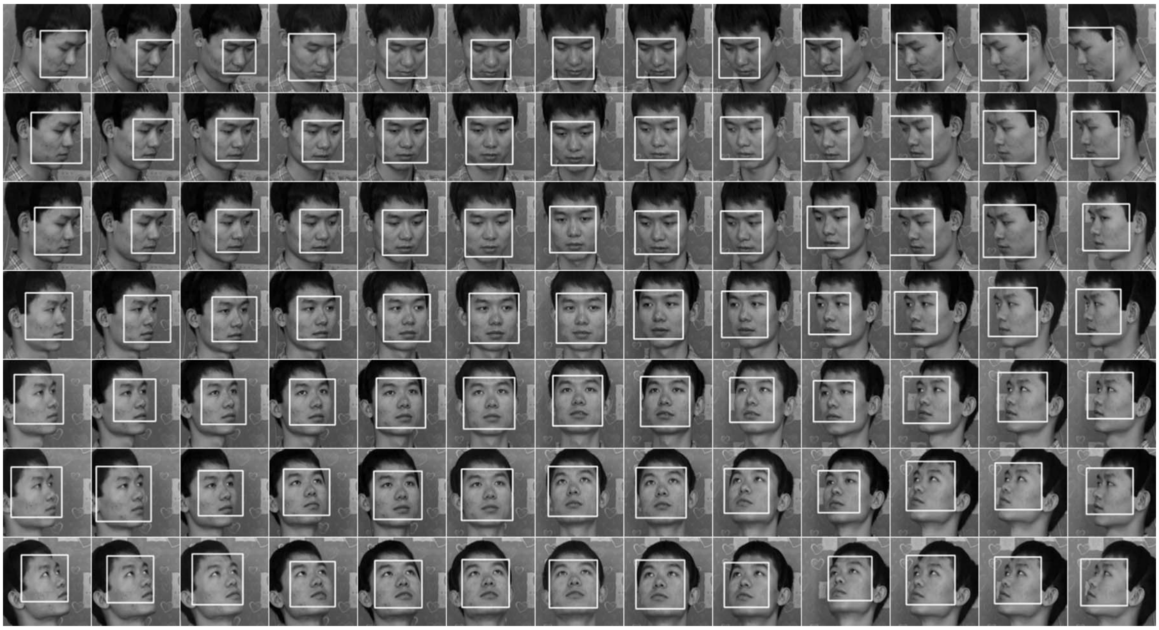


Fig. 8. Face images of one subject in the multipose database.

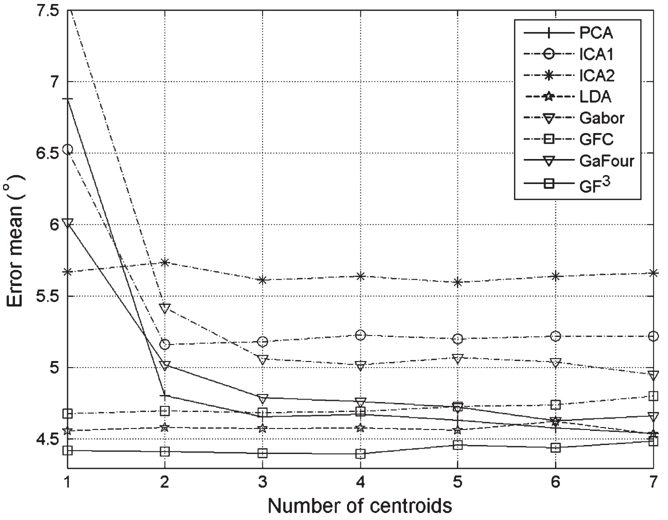


Fig. 9. Error mean on the multipose database. The x-axis is the centroid number of each pose, and the y-axis is the error mean.

asymmetry features are computed from the row and can be seen as the global information on the row while the details of the images are only the supplement to the asymmetry. Based on the characteristic of the robustness to the image sizes, we can use the low-resolution image in pose estimation to reduce the time and space requirement.

4) Distribution Analysis of Different Features: In pattern classification, the most important thing for features is its separability for various classes (yaw in our case). To intuitively show the effectiveness of GaFour and GF^3 for this purpose, the following experiment is designed.

Images with 0° in pitch and with four angles in yaw, namely, 0°, -15°, -30°, and -45°, are considered. All the images come from the CAS-PEAL database. Different features are processed using PCA. For each PCA subspace, the leading three dimensions are kept for visualization. Fig. 10 shows the coordinates of the 800 testing images in the 3-D embedding for different methods. In the figure, the filled circles, the filled triangles, the filled diamonds, and the filled blocks are the projections of the samples with yaw of 0°, -15°, -30°, and -45°, respectively, in their own subspace.

From Fig. 10, it can be seen that the samples of four yaw angles are separated in order in the embedding of GF^3, while points of different poses are heavily overlapped in the embedding of other methods. This experiment shows that the separability of GF^3 is impressively better than the other methods.

5) Comparison With SVM: Since LDA is applied to extract the discriminant features and reduce the data dimensionality, we also compare it with the widely used SVM with/without LDA dimensionality reduction. In SVM, radial basis function kernel is used. Generally, SVM is used for two-class problems. We use "one against one" approach to solve the c-class problem. The results are shown in Table III.

From Table III, it can be known that on the CAS-PEAL database, the 92.86% accuracy of GaFour+SVM is slightly better than the 91.42% accuracy of GF^3 + NC. On the multipose database, the 5.83° error mean of GaFour+SVM is worse than the 4.40° error mean of GF^3 + NC. The following reasons explain our emphasis on GF^3 + NC.

First, the dimension of GF^3 is much lower than that of the original GaFour, which can lead to much less storage requirement and higher speed for memory access. The details of the dimension analysis have been introduced in Section IV-B.

Second, compared with the SVM, the computational cost of the simple classifier combining LDA and NC is much less. This advantage will become more evident when considered together with the much lower dimension.

Third, to say the least, the main contribution of this paper is proposing the asymmetry feature and its GaFour-based description for yaw estimation. Therefore, many other classifiers can be applied for the sequential classification.

TABLE II
MAXIMUM ACCURACY (IN PERCENT) OF THE DIFFERENT k'S WHEN THE IMAGE SIZE VARIES

	Accuracy of different methods (%)							
Size	PCA	ICA1	ICA2	LDA	Gabor	GFC	GaFour	GF^3
16 × 16	79.94	81.91	84.36	82.36	66.07	73.52	84.09	89.43
32 × 32	81.04	80.69	85.49	86.38	80.78	88.66	87.82	91.42
64 × 64	81.48	80.79	84.36	86.30	85.44	91.47	88.15	89.96

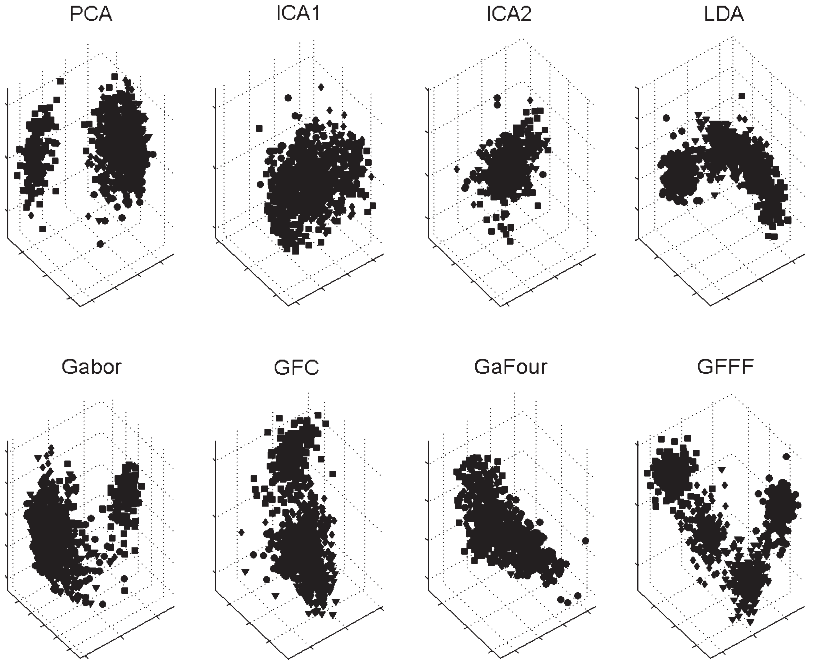


Fig. 10. Three-dimensional embedding of different features given by PCA. The filled circles, the filled triangles, the filled diamonds, and the filled blocks are the projection of the samples with pose 0°, −15°, −30°, and −45°, respectively.

TABLE III
PERFORMANCE COMPARISON OF GF^3 + NC AND GaFour+SVM

Database	Measurement	GF^3+NC	GaFour+SVM
CAS-PEAL	Accuracy	91.42%	92.86%
Multi-Pose	Error Mean	4.40°	5.83°

B. Experiment 2: On Heterogeneous Training and Testing Sets

As we have mentioned, the experiments in Section V-A are conducted on homogeneous training and testing sets. Easy to understand, this kind of classification task is relatively easy; therefore, as we have seen in the last section, the performances of most methods are satisfactory. However, in practical applications, the situation might not be so perfect, i.e., the imaging conditions of the testing images might not be homogeneous with those of the training ones. We call this kind of situation heterogeneous testing. Heterogeneous testing data are very common in the real world, because system developers may hardly know what kind of data will be presented to the system in practical applications. Therefore, it is very significant to evaluate a system by using some testing data heterogeneous with the training data. Essentially, heterogeneous testing is highly related to the generalizability problem in pattern recognition. Aiming at this goal, the following experiments are designed.

In this experiment, the multipose database is used as the training data set, whereas the CAS-PEAL database is used as the testing data set. The images in the multipose database are captured by Sony EVI-D31 camera, whereas the images in the CAS-PEAL database are captured by a simple USB camera. Inevitably, the two databases are constructed with different population under different imaging conditions, such as different lighting conditions and backgrounds. Therefore, the two databases are quite heterogeneous overall in terms of camera parameters, population, lighting conditions, expressions, and background, which are very appropriate for the aforementioned purpose.

The error mean m of different methods with different k's is shown in Fig. 11. First, from the figure, one can surprisingly find that the performances of various methods have become quite different, which apparently forms contrast with the experimental results in the aforementioned experiments. The error mean of all the methods is larger than 7.0°, whereas the error mean is not more than 5.6° in Section V-A2. The comparison shows that the heterogeneous data affect the performances of yaw estimation greatly. Second, we must point out that the performance of LDA is much worse although it can get better results when the training and testing data sets are homogeneous. It seems that the discriminant ability of LDA is confused by the nonpose information of the heterogeneous databases.

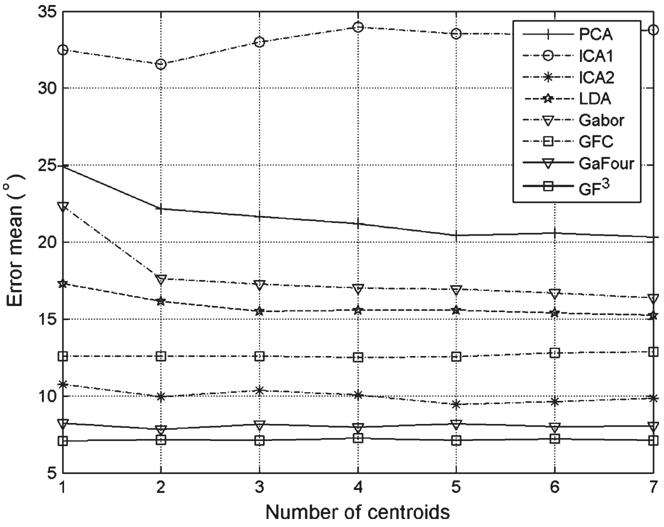


Fig. 11. Error mean of yaw estimation. The training and the testing sets are heterogeneous. The x-axis is the centroid number of each angle, and the y-axis is the error mean.

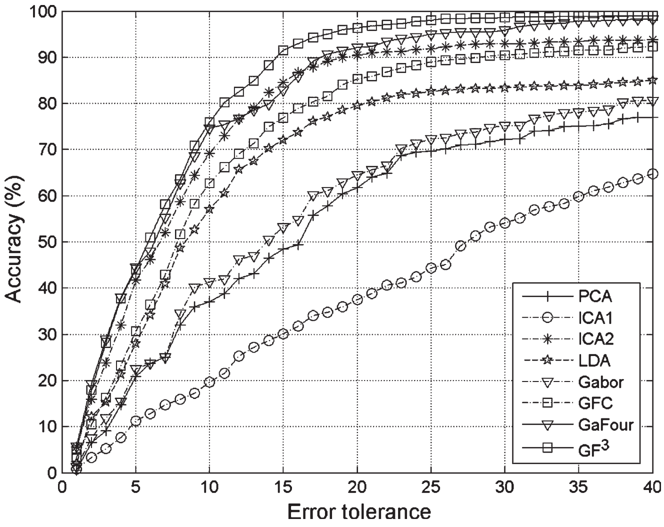


Fig. 12. Accuracy with the error tolerance. The x-axis represents the error tolerance, and the y-axis represents the accuracy.

Finally, in this kind of evaluation, the results of GaFour and GF^3 outperform all the other methods impressively although they are just comparable to other methods in the aforementioned experiments. We attribute the superiority to the extracted asymmetry features. Since the features in GaFour and GF^3 are more related to pose variation, the effects caused by the difference between the training and testing data are decreased greatly.

Besides the error mean, the accuracies with the error tolerances are also important in pose estimation, which shows the error distribution of samples. The accuracy A_δ with the error tolerance δ can be computed as follows:

$$A_\delta = \frac{1}{N}\sum_{i=1}^{N} A'_{\delta,i} \tag{10}$$

where N is the total number of testing samples, and

$$A'_{\delta,i} = \begin{cases} 1, & |p_i - p'_i| \le \delta \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

Intuitively, the accuracy is the correct rate when one accepts an error less than δ degree as the correct estimation. The A_δ of different methods is shown in Fig. 12. The figure shows that the accuracies of GF^3 are always higher than those of other methods with all the error tolerance δ. For GF^3, the accuracy is about 44%, 76%, 92%, and 96% when the error tolerance δ is 5°, 10°, 15°, and 20°, respectively, i.e., the estimation error of 44% testing samples is less than 5° and the error of 96% samples is less than 20°. For both GaFour and GF^3, the accuracies approximate to 100% when δ is 40°. However, for the other methods, the accuracies are not more than 94% when δ is 40°. The results further show that the error distribution of GaFour and GF^3 is impressively much better than that of the other methods.

To sum up, these results strongly support the following observations. The proposed method can better generalize to the unseen heterogeneous data. It is also more robust than the other benchmark methods when the data sets have more variations in lighting, population, backgrounds, and cameras. This further implies that the pattern representation based on the image asymmetry is more heavily related to pose and thus provides a good choice for yaw estimation.

Besides the effectiveness, the computation of GaFour and GF^3 is also efficient. We compute the time needed for GaFour and GF^3 using Matlab 7.0.4 on a PC (CPU Pentium IV at 3.4 GHz, 2-GB memory). The average computing time for GaFour feature extraction of the 32 × 32 images is 3 ms, and the time of classification using the NC classifier is about 0.08 ms, which indicates that our methods can be used in the real-time system.

VI. Discussion and Conclusion

Based on the assumption that features for pose estimation should be closely relevant to pose but less relevant to other properties (e.g., the identity) of the face image, in this paper, we proposed a novel face representation method based on the asymmetry of the face image for head yaw estimation. We show that the Fourier transform can provide good asymmetry measures consistent with the pose variations; therefore, the asymmetry feature is extracted as the resulting real and imaginary parts of Fourier transform of the input face image. In order to eliminate the influence of lighting and noise, multiple-scale 1-D Gabor filters are further exploited to filter the images before Fourier transform. In addition, LDA is finally applied to enhance the discriminant ability and reduce the feature dimension. Extensive experimental results illustrate the advantages of our method in robustness, effectiveness, efficiency, and generalizability. In particular, in the heterogeneous testing, the proposed methods impressively outperform the other well-known methods.

There are also several aspects to be further studied in the future. First, this paper only discusses the effectiveness of the asymmetry for estimating the yaw angles, but it is still an open problem to extend the proposed method to estimate simultaneously the yaw, pitch, and roll angles. Motivated by the effectiveness of horizontal row signal for yaw estimation, similarly, we might also expect good performance for pitch estimation based on vertical columns. We will validate this idea in the future. Second, the dimension of the final GaFour feature is still high even though it has been reduced; therefore, more work should be done to further reduce the dimensionality. Finally, it is also worth studying how to set the optimal parameters of 1-D Gabor filters for different rows in order to emphasize some regions and further reduce the dimensionality.

ACKNOWLEDGMENT

The authors would like to thank M. Bartlett for providing the ICAFaces code.

REFERENCES

[1] A. Kuchinsky, C. Pering, M. L. Creech, D. Freeze, B. Sera, and J. Gwizdka, "FotoFile: A consumer multimedia organization and retrieval system," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.—The CHI Is the Limit*, 1999, pp. 496–503.

[2] T. F. Cootes, K. Walker, and C. J. Taylor, "View-based active appearance model," in *Proc. 4th IEEE Int. Conf. Autom. Face Gesture Recog.*, 2000, pp. 227–232.

[3] B. Braathen, M. S. Bartlett, and J. R. Movellan, "3-D head pose estimation from video by stochastic particle filtering," in *Proc. 8th Annu. Joint Symp. Neural Comput.*, 2001.

[4] A. Nikolaidis and I. Pitas, "Facial feature extraction and determination of pose," in *Proc. NOBLESSE Workshop Nonlinear Model Based Image Anal.*, 1998, pp. 257–262.

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

[6] Y. Ebisawa and Y. Nurikabe, "Face pose estimation based on 3D detection of pupils and nostrils," in *Proc. IEEE Int. Conf. Virtual Environ., Human–Comput. Interfaces, Meas. Syst.*, 2005, pp. 92–97.

[7] S. Yan, Z. Zhang, Y. Fu, Y. Hu, J. Tu, and T. S. Huang, "Learning a person-independent representation for precise 3D pose estimation," in *Proc. CLEAR Eval. Workshop*, 2007, pp. 297–306.

[8] N. Krüger, M. Pötzsch, and C. von der Malsburg, "Determination of face position and pose with a learned representation based on labeled graphs," *Image Vis. Comput.*, vol. 15, no. 8, pp. 665–673, Aug. 1997.

[9] J. Xiao, T. Moriyama, T. Kanade, and J. F. Cohn, "Robust full-motion recovery of head by dynamic templates and re-registration techniques," *Int. J. Imaging Syst. Technol.*, vol. 13, pp. 85–94, Sep. 2003.

[10] L. Chen, L. Zhang, Y. Hu, M. Li, and H. Zhang, "Head pose estimation using Fisher manifold learning," in *Proc. IEEE Int. Workshop Anal. Model. Faces Gestures*, 2003, pp. 203–207.

[11] T. Darrell, B. Moghaddam, and A. P. Pentland, "Active face tracking and pose estimation in an interactive room," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 1996, pp. 67–72.

[12] Y. Li, S. Gong, and H. Liddell, "Support vector regression and classification based multi-view face detection and recognition," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recog.*, 2000, pp. 300–305.

[13] S. Z. Li, Q. Fu, L. Gu, B. Scholkopf, Y. Cheng, and H. Zhang, "Kernel machine based learning for multi-view face detection and pose estimation," in *Proc. 8th IEEE Int. Conf. Comput. Vis.*, 2001, pp. 674–679.

[14] V. Krüger, S. Bruns, and G. Sommer, "Efficient head pose estimation with Gabor wavelet networks," in *Proc. 11th Brit. Mach. Vis. Conf.*, 2000.

[15] Y. Wei, L. Fradet, and T. Tan, "Head pose estimation using Gabor eigenspace modeling," in *Proc. IEEE Int. Conf. Image Process.*, 2002, pp. I-281–I-284.

[16] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.

[17] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.

[18] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," *Adv. Neural Inf. Process. Syst.*, vol. 15, pp. 585–591, 2001.

[19] Y. Fu and T. S. Huang, "Graph embedded analysis for head pose estimation," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recog.*, 2006, pp. 3–8.

[20] J. Tu, Y. Fu, Y. Hu, and T. S. Huang, "Evaluation of head pose estimation for studio data," in *Multimodal Technologies for Perception of Humans*, vol. 4122, R. Stiefelhagen and J. Garofolo, Eds. Berlin, Germany: Springer-Verlag, 2007, pp. 281–290.

[21] N. Hu, W. Huang, and S. Ranganath, "Head pose estimation by non-linear embedding and mapping," in *Proc. IEEE Int. Conf. Image Process.*, 2005, p. II-342-5.

[22] B. Raytchev, I. Yoda, and K. Sakaue, "Head pose estimation by nonlinear manifold learning," in *Proc. Int. Conf. Pattern Recog.*, 2004, vol. 4, pp. 462–466.

[23] J. N. S. Kwong and S. Gong, "Learning support vector machines for a multi-view face model," in *Proc. Brit. Mach. Vis. Conf.*, 1999, pp. 503–512.

[24] S. Z. Li, X. Lu, X. Hou, X. Peng, and Q. Cheng, "Learning multiview face subspaces and facial pose estimation using independent component analysis," *IEEE Trans. Image Process.*, vol. 14, no. 6, pp. 705–712, Jun. 2005.

[25] Y. Hu, L. Chen, Y. Zhou, and H. Zhang, "Estimating face pose by facial asymmetry and geometry," in *Proc. 6th Int. Conf. Autom. Face Gesture Recog.*, 2004, pp. 651–656.

[26] M. Gruending and O. Hellwich, "3D head pose estimation with symmetry based illumination model in low resolution video," in *Deutsche Arbeitsgemeinschaft für Mustererkennung*. Berlin, Germany: Springer-Verlag, 2004, pp. 45–53.

[27] Y. Liu, K. L. Schmidt, J. F. Cohn, and S. Mitra, "Facial asymmetry quantification for expression invariant human identification," *Comput. Vis. Image Understanding*, vol. 91, no. 1/2, pp. 138–159, Jul. 2003.

[28] S. Mitra and M. Savvides, "Analyzing asymmetry biometric in the frequency domain for face recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2005, pp. 953–956.

[29] S. Mitra and Y. Liu, "Local facial asymmetry for expression classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2004, pp. 889–894.

[30] C. Cortes and V. Vapnik, "Support vector network," *Mach. Learn.*, vol. 20, pp. 273–297, 1995.

[31] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005.

[32] C. Liu and H. Wechsler, "Robust coding schemes for indexing and retrieval from large face databases," *IEEE Trans. Image Process.*, vol. 9, no. 1, pp. 132–137, Jan. 2000.

[33] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.

[34] L. Wiskott, J.-M. Fellous, and C. von der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 775–779, Jul. 1997.

[35] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition," *IEEE Trans. Image Process.*, vol. 11, no. 4, pp. 467–476, Apr. 2002.

[36] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

[37] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cognit. Neurosci.*, vol. 3, no. 1, pp. 71–86, 1991.

[38] M. S. Bartlett, *Face Image Analysis by Unsupervised Learning*. Norwell, MA: Kluwer, 2001.

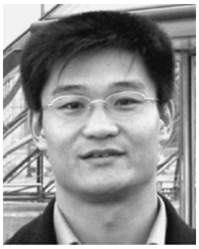
[39] M. E. Tipping, "Sparse bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, 2001.

[40] 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 Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 38, no. 1, pp. 149–161, Jan. 2008.

[41] S. Yan, S. Shan, X. Chen, W. Gao, and J. Chen, "Matrix-structural learning (MSL) of cascaded classifier from enormous training set," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2007.

Bingpeng Ma received the B.S. degree in mechanics and the M.S. degree in mathematics from the Huazhong University of Science and Technology, Wuhan, China, in 1998 and 2003, respectively. Since 2003, he has been working toward the Ph.D. degree in the ICT-ISVISION Joint Research and Development Laboratory for Face Recognition, Institute of Computing Technology (ICT), Chinese Academy of Sciences (CAS), Beijing.

He is also currently with the Key Laboratory of Intelligent Information Processing, ICT, CAS, and with the Graduate School of CAS. His research interests include the areas of mathematical programming, machine learning, and pattern recognition.



Shiguang Shan (M'04) received the M.S. degree in computer science from the Harbin Institute of Technology, Harbin, China, in 1999, and the Ph.D. degree in computer science from the Institute of Technology (ICT), Chinese Academy of Sciences (CAS), Beijing, in 2004.

He is currently a Vice Researcher and serves as the Vice Director of the Digital Media Center and the ICT-ISVISION Joint Research and Development Laboratory for Face Recognition, ICT, CAS, where he is also with the Key Laboratory of Intelligent Information Processing. His research interests include image analysis, pattern recognition, and computer vision. He is particularly focused on face-recognition-related research topics.

Xilin Chen (M'00) received the B.S., M.S., and Ph.D. degrees in computer science from the Harbin Institute of Technology, Harbin, China, in 1988, 1991, and 1994, respectively.

He was a Professor with the Harbin Institute of Technology from 1999 to 2005. He was a Visiting Scholar with Carnegie Mellon University, Pittsburgh, PA, from 2001 to 2004. Since August 2004, he has been with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, where he is also with the Key Laboratory of Intelligent Information Processing and the ICT-ISVISION Joint Research and Development Laboratory for Face Recognition. His research interests include image processing, pattern recognition, computer vision, and multimodal interfaces.

Dr. Chen has served as a program committee member for more than 20 international and national conferences. He has received several awards, including the China's State Scientific and Technological Progress Award in 2000, 2003, and 2005 for his research work.

Wen Gao (M'92–SM'05) received the M.S. degree in computer science from the Harbin Institute of Technology, Harbin, China, in 1985, and the Ph.D. degree in electronics engineering from the University of Tokyo, Tokyo, Japan, in 1991.

He was a Professor in computer science with the Harbin Institute of Technology from 1991 to 1995 and with the Institute of Computing Technology (ICT), Chinese Academy of Sciences (CAS), Beijing, China, from 1996 to 2005. He is currently a Professor with the School of Electronics Engineering and Computer Science, Peking University, Beijing. He is also an Adjoint Professor of ICT, CAS. He has been leading research efforts to develop systems and technologies for video coding, face recognition, sign language recognition and synthesis, and multimedia retrieval. He has published four books and over 500 technical articles in refereed journals and proceedings in the areas of signal processing, image and video communication, computer vision, multimodal interfaces, pattern recognition, and bioinformatics.

Dr. Gao serves the academic society as the General Cochair of IEEE ICME07 and as the Head of the Chinese delegation to the Moving Picture Expert Group of the International Standard Organization since 1997. He is also the Chairman of the working group responsible for setting a national audio video coding standard for China. He has received many awards, including five national awards for his research achievements and activities.