Combining Perceptual Features with Diffusion Distance for Face Recognition

Zhou, H., & Sadka, A. H. (2011). Combining Perceptual Features with Diffusion Distance for Face Recognition. IEEE Transactions on System, Man and Cybernetics, Part C, 41(5)(5), 577-588. [5497209]. DOI: 10.1109/TSMCC.2010.2051328

Published in:
IEEE Transactions on System, Man and Cybernetics, Part C

Combining Perceptual Features With Diffusion Distance for Face Recognition

Huiyu Zhou and Abdul H. Sadka, *Senior Member, IEEE*

***Abstract*—Face recognition and identification is a very active research area nowadays due to its importance in both human computer and social interaction. Psychological studies suggest that face recognition by human beings can be featural, configurational, and holistic. In this paper, by incorporating spatially structured features into a histogram-based face-recognition framework, we intend to pursue consistent performance of face recognition. In our proposed approach, while diffusion distance is computed over a pair of human face images, the shape descriptions of these images are built using Gabor filters that consist of a number of scales and levels. It demonstrates that the use of perceptual features by Gabor filtering in combination with diffusion distance enables the system performance to be significantly improved, compared to several classical algorithms. The oriented Gabor filters lead to discriminative image representations that are then used to classify human faces in the database.**

***Index Terms*—Configuration, diffusion distance, face recognition, holistic, perceptual features.**

I. Introduction

FACE recognition (or identification) is a very active research area nowadays due to its importance in both human computer and social interaction [16], [64]. Many applications will benefit from the success of face recognition, e.g., video surveillance, behavioral analysis, access control, and teleconferencing. Recent progress in this field has witnessed some successful systems that incorporate advanced algorithms, such as principal component analysis (PCA) (Eigenface) [57], [61], linear discriminate analysis [75], elastic bunch graph matching (Fisherface) [2], and neural networks [52].

Psychological studies suggest that face recognition by human beings can be feature and configuration based [55]. Featural information refers to as isolated facial components, such as hair, brow, eyes, nose, mouth, and cheeks. Configurational information denotes the spatial relations between the features, their interaction, and to various proportions, for example, nose length to brow length [18], [47], [49]. It is also widely accepted that holistic recognition plays an important role in face perception (e.g., [1] and [59]). The applications of these psychological studies have been commonly found in the community of computer vision, e.g., [26], [27], [71], and [76].

A. Feature Based

In recent years, the concept of face space has been used to interpret the coding of facial information [33], [63], [67]. It is discovered that face identity is represented as a locus in a multidimensional space in which the dimensions are independent perceptual attributes of faces. This approach suggests that encoding along each facial dimension is relative to properties of the average face, and facial distinctiveness is encoded as distance from the average [4], [12]. Tversky and Krantz [62] stated that faces could be distinguished by three components: eyes, mouth, and face shape. Penry [44], the inventor of the photo-fit, argued that a given face was a particular combination of individual features, and the alteration of a feature would change the whole facial appearance.

B. Configuration Based

Garner [22] proposed that the spatial relations of parts of the whole face as configurational properties (e.g., symmetry and repetition) could be used to discriminate faces. Maurer *et al.* [39] provided evidence for the separability of three types of configurational processing: detection of the first-order relations that define faces (i.e., two eyes above nose and mouth), holistic processing (glueing the features together into a Gestalt), and processing second-order relations (i.e., the spacing among features). Goffaux *et al.* [23] presented experimental results to support that low spatial frequencies play in the configural processing of faces, whereas featural processing was largely dependent on high spatial frequencies. Young *et al.* [72] demonstrated that, for adults, the encoding of relations among facial parts is sensitive to orientation.

C. Holistic Face Recognition

Yovel and Duchaine [73] explicitly supported the holistic hypothesis where both parts and spacing among them were integral aspects of face representation. Recent work has explored how one might learn to utilize holistic information [50] and the contributions of holistic processing to the analysis of facial expressions [7]. Rossion and Boremanse [51] suggested that a substantial part of the face-inversion effect could be explained by the inability to apply an experience-derived holistic representation to a face image that was rotated horizontally or beyond the orientation.

Manuscript received January 18, 2010; revised April 18, 2010; accepted May 7, 2010. Date of publication June 28, 2010; date of current version August 19, 2011. This work was supported by the European Commission under Grant FP6-045189-Specific Targeted Research Projects (STREP) retrieval of multimedia semantic units for enhanced reusability (RUSHES). This paper was recommended by Associate Editor J. Tang.

H. Zhou is with the Institute of Electronics, Communications, and Information Technology, School of Electronics, Electrical Engineering, and Computer Science, Queen's University Belfast, Belfast, BT3 9DT, U.K. (e-mail: h.zhou@ecit.qub.ac.uk).

A. H. Sadka is with the School of Engineering and Design, Brunel University, Uxbridge, UB8 3PH, U.K. (e-mail: abdul.sadka@brunel.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMCC.2010.2051328

Histogram-based features (e.g., colors, curvatures, and orientation) have been commonly used in object recognition and categorization, for example [9], [13], [31], [54], [58], and [69]. As well-recognized, histogram-based descriptors have shown extent robustness with respect to object transformation [56]. Evidence shows that combination of color, shape, and texture histograms can be utilized to improve the recognition rates against single modality approaches [40]. These recognition results also suggest that spatially structured features are desirable in a well-performed recognition system, e.g. [56].

In this paper, we incorporate spatially structured features into a histogram-based face-recognition framework, which calculates diffusion distance based on the intensity histograms [34]. In our approach, while diffusion distance is computed over a pair of human face images, the shape descriptions of these images are built using Gabor filters that have been widely studied in the literature, e.g. [20], [43], [60], and [74]. A distinctive advantage of the Gabor features is their optimality in the space-spatial frequency in two dimensions [20], while these features respond strongly to image edges [68]. Our proposed scheme demonstrates that the use of perceptual features by Gabor filtering, in combination with diffusion distance, enables the system performance to be significantly improved, particularly in the presence of illumination changes or occlusions due to pose variations. The oriented Gabor filters produce discriminative image representations that are then used to classify human faces in the database.

This paper is organized as follows. In Section II, related work is summarized, particularly with the recent developments in computer vision. In Section III, the principles of Gabor filtering and diffusion distance will be introduced. This is followed by the combination of these two features, which is reported in Section IV. Then, evaluation of the proposed strategy is performed in Section V. Finally, conclusions and future work will be given in Section VI.

II. Related Work

In this section, classical similarity measures between two images will be categorized into two groups: histogram and Gabor based. This is because these two groups are closely related to our study. Here, a histogram refers to a distribution of image pixels against specific definitions, and Gabor features describe the outcomes of applying a Gabor-filtering function to an image.

A. Histogram Based

Chapelle *et al.* [9] reported that support vector machines (SVMs) were capable of generalizing well on difficult image-classification problems, where the features referred to high-dimensional histograms. Heavy-tailed radial basis function (RBF) kernels were evaluated on the classification of images extracted from the Corel stock photo collection and shown to far outperform traditional polynomial or Gaussian RBF kernels. Schiele and Crowley [54] presented a technique where appearances of objects are represented by the joint statistics of such local neighborhood operators. Based on joint statistics, techniques have been developed for the identification of multiple objects at arbitrary positions and orientations in a cluttered scene.

Common visual codebook-generation methods used in a bag of visual words model, e.g., *k*-means or Gaussian mixture model, use the Euclidean distance to cluster features into visual code words. However, most popular visual descriptors are histograms of image measurements, such as bags of features [32], [42]. It has been shown that the histogram intersection kernel (HIK) is more effective than the Euclidean distance in supervised learning tasks with histogram features. In [69], it was demonstrated that HIK could also be used in an unsupervised manner to significantly improve the generation of visual codebooks. A histogram kernel *k*-means algorithm was proposed, which was easy to implement and worked almost as fast as *k*-means.

Rubner *et al.* [53] used the earth mover's distance (EMD) to define the difference between two distributions with sparse structures (e.g., color histograms). The only drawback of this method is the computational cost of the EMD, which is more than $O(N^3)$, where N is the number of histogram bins in the image. A fast algorithm "EMD-L_1" was then proposed for computing the EMD between a pair of histograms [35]. To perform the EMD-L_1 computation, an efficient tree-based algorithm "Tree-EMD" was presented. Tree-EMD exploited the fact that a basic feasible solution of the simplex algorithm-based solver formed a spanning tree when interpreting EMD-L_1 as a network flow-optimization problem.

Lowe proposed the scale-invariant feature transform (SIFT) for object recognition [38]. SIFT has a keypoint descriptor that used a set of 16 histograms, aligned in a 4 by 4 grid, each with eight orientation bins. The object recognition depends on the agreement of SIFT feature vectors (known as keys) that were used in a nearest neighbors approach to correspond the objects. Dalal and Triggs [13] proposed histogram of oriented gradients (HOGs) descriptors for object detection. These descriptors were extracted by dividing the image into small connected regions, called cells, and for each cell compiling a histogram of gradient directions or edge orientations for the pixels within the cell. The combination of these histograms then forms a descriptor. Similarly, shape context [3] was introduced to form a distribution of the distance between a point and the others on the same contour.

B. Gabor Filtering

Potzsch *et al.* [45] described a biologically motivated object-recognition system with Gabor wavelets as basic feature type. The region surrounding a given pixel in the image was described by the responses of a set of Gabor filters of different frequencies and orientations, all centered at the pixel position. This set of responses is called a jet. Objects are described by graphs, whose vertices were labeled by jets and whose links describe topograph.

Working toward the problem of recognizing people from their averaged gait images, Tao *et al.* [60] presented three different Gabor-function-based image representations: 1) GaborD is the sum of Gabor filter responses over directions; 2) GaborS is the sum of Gabor filter responses over scales; and 3) GaborSD is

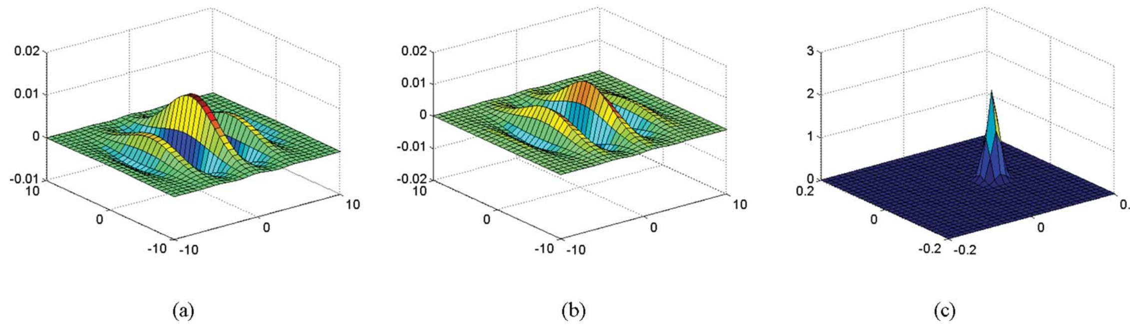


Fig. 1. Real, imaginary, and frequency components of a complex Gabor function in the spatial domain. (a) Real, (b) imaginary, and (c) frequency component. $S = 0.2$ cycles/pixel and $Q = 0°$.

the sum of Gabor filter responses over scales and directions. Ilonen *et al.* [29] presented an improved algorithm for image-feature localization. This method was based on complex-valued multiresolution Gabor features and their ranking using multiple hypothesis testing. On the other hand, Gabor-based region covariance matrices can be used as local descriptors to discriminate human faces [43].

Liu and Wechsler introduced a Gabor–Fisher classifier (GFC) for face recognition [37]. The GFC method was robust to changes of illumination and facial expression. It applied the enhanced Fisher linear discriminant model (EFM) to an augmented Gabor feature vector derived from the Gabor wavelet representation of face images. An augmented Gabor feature vector had a dimensionality further reduced using the EFM by considering both data compression and recognition (generalization) performance. The development of a GFC for multiclass problems was also proposed. Later on, a Gabor-based kernel PCA method was proposed, integrating the Gabor wavelet representation of face images and the kernel PCA method for face recognition [36]. Similar work on Gabor-based kernel PCA and its simplification was reported in [10] and [70].

III. Features for Face Recognition

In this section, we will briefly introduce the features and related techniques that will be used in our approach. In the proposed algorithm, the Gabor features can be obtained after we apply a Gabor-filtering function to the image. These features are extracted from a set of face images, which are then used to compute diffusion distance over pair-wise face images. Diffusion distance is a dissimilarity measure between histogram-based descriptors, which is defined in a temperature field. Histogram difference is measured by heat diffusion and is treated as the initial condition of a heat diffusion process. As a result, diffusion distance is derived as the sum of dissimilarities over scales. It has been justified that using diffusion distance is capable of handling deformations as well as quantization effects [34].

A. Gabor Filtering

Gabor filters include a filter bank with various scales and rotations. The filters are convolved with the image, resulting in a Gabor space. This entire process simulates the response of the 2-D receptive field profiles of the mammalian simple cortical cell [14].

In the spatial domain, a Gabor filter can be treated as a complex exponential modulated by a Gaussian function. Each Gabor consists of two functions in quadrature (out of phase by 90°), located in the real and imaginary parts of a complex function as follows (see Fig. 1 for a simulation):

$$\begin{aligned} g(x,y) = K \exp(-\pi(a^2(x-x_0)_q^2 + b^2(y-y_0)_q^2)) \\ \exp(j(2\pi(\alpha_0 x + \beta_0 y) + Q)) \end{aligned} \tag{1}$$

or in polar coordinates

$$\begin{aligned} g(x,y) = K \exp(-\pi(a^2(x-x_0)_r^2 + b^2(y-y_0)_r^2)) \\ \exp(j(2\pi S_0(x\cos\omega_0 + y\sin\omega_0) + Q)) \end{aligned} \tag{2}$$

where K is the magnitude of the Gaussian envelope, (a, b) are the two axis of the Gaussian envelope, (x, y) are the pixel position in the spatial domain, (x_0, y_0) are the location of the peak of the Gaussian envelope, (α_0, β_0) are the spatial frequencies of the sinusoid carrier in Cartesian coordinates (expressed as (S_0, ω_0) in polar coordinates), and Q is the phase of the sinusoid carrier [41].

Given a gray-level image $I(x, y)$, the convolution of $I(x, y)$ and $g(x, y)$ is given as follows:

$$G(x,y) = I(x,y) * g(x,y) \tag{3}$$

where $*$ denotes the convolution operator. The convolution can be computed efficiently by applying the fast Fourier transform (FFT), followed by point-by-point multiplications, and finally, the inverse FFT (IFFT). Therefore, we obtain the 2-D Fourier transform of (1) as follows:

$$\begin{aligned} \hat{g}(u,v) = \frac{K}{ab}\exp(j(-2\pi(x_0(u-\alpha_0) + y_0(v-\beta_0)) + Q)) \\ \exp\left(-\pi\left(\frac{(u-\alpha_0)_r^2}{a^2} + \frac{(v-\beta_0)_r^2}{b^2}\right)\right) \end{aligned} \tag{4}$$

where (u, v) denote the coordinates in the frequency domain. Or in polar coordinates

$$\begin{aligned} \parallel \hat{g}(u,v) \parallel = \frac{K}{ab}\exp\left(-\pi\left(\frac{(u-\alpha_0)_r^2}{a^2} + \frac{(v-\beta_0)_r^2}{b^2}\right)\right) \\ \text{Phase}(\hat{g}(u,v)) = -2\pi(x_0(u-\alpha_0) + y_0(v-\beta_0)) + Q. \end{aligned} \tag{5}$$

(a) (b) (c) (d) (e)

Fig. 2. Examples of applying four directions' Gabor filtering to a profile face image. (a) Original, (b) 0, (c) $\pi/4$, (d) $\pi/2$, and (e) $3\pi/4$. $S = 0.2$ cycles/pixel and $Q = 0°$.

In this paper, similar to [70], we only utilize the magnitude of Gabor representations, which provide a measure of the local properties of an image and is less sensitive to the illumination changes. The Gabor representations, denoted as $\Phi_{i,j}$, form a feature vector for face recognition as follows:

$$\Phi = [\Phi_{1,1}, \Phi_{1,2}, \ldots, \Phi_{m,n}]^T \tag{6}$$

where m and n are numbers of scales and orientations used in the Gabor filter. Fig. 2 illustrates examples of applying four directions' Gabor filtering to a profile face image. Fig. 2(a)–(e) reveals that different image representations correspond to the oriented filters. Note that only one phase is utilized here in this example for simplicity.

B. Diffusion Distance

Histogram-based local descriptors (HBLDs) have been widely used in computer vision for shape matching, image retrieval, and categorization, e.g., [28] and [46]. Currently, most of the available approaches can only deal with the case where the corresponding histograms in different images have been aligned. As a consequence, these methods are very sensitive to distortions and quantization effects in the extracted local descriptors [34].

Consider two l-dimensional histograms $h_1(\mathbf{s})$ and $h_2(\mathbf{s})$, where $\mathbf{s} \in \mathcal{R}^l$ is a vector. The distance between them is defined as $\hat{D}(h_1,h_2)$. Using a temperature field, we treat the distance to be the evolution of a temperature field $T(\mathbf{s}, t)$ at time $t = 0$. According to the heat diffusion equation

$$\frac{\partial T}{\partial t} = \bigtriangledown^2 T \tag{7}$$

which has a unique solution as

$$T(\mathbf{s}, t) = T_0(\mathbf{s}) * f(\mathbf{s}, t) \tag{8}$$

with an initial condition

$$T_0(\mathbf{s}) = \hat{D} \tag{9}$$

where

$$f(\mathbf{s}, t) = \frac{1}{(2\pi)^{l/2} t} \exp\Big(-\frac{\mathbf{s}^T\mathbf{s}}{2t^2}\Big). \tag{10}$$

To efficiently compute the histogram distance, a distance function based on the Gaussian pyramid is adopted. This is due to the fact that the Gaussian pyramid is an efficient discretization of the diffusion process $T(\mathbf{s}, t)$. Therefore, the diffusion distance to be computed is as follows:

$$D(h_1, h_2) = \sum_{m=0}^{M} \hat{D}(|d_m(\mathbf{s})|) \tag{11}$$

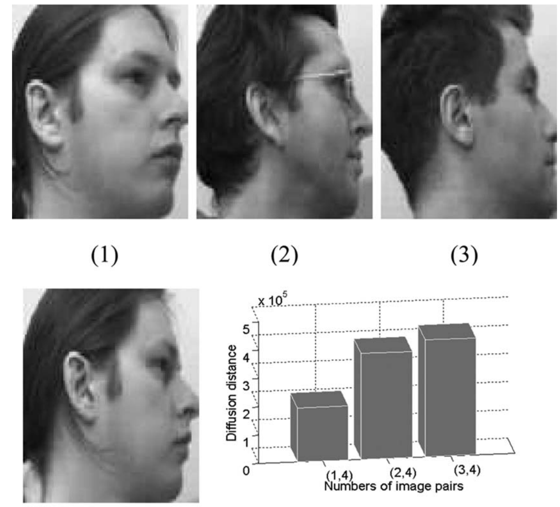
where

$$d_m(\mathbf{s}) = [d_{m-1}(\mathbf{s}) * f(\mathbf{s}, \sigma)] \downarrow_2 \tag{12}$$

where $m = 1, \ldots, M$. Equation (12), different layers of the pyramid, has an initial condition as $d_0(\mathbf{s}) = h_1(\mathbf{s}) - h_2(\mathbf{s})$. The notation $\downarrow_2$ is half-size down sampling. M is the number of the pyramid levels and σ is the standard deviation of the Gaussian filter f.

The computational complexity of $D(h_1, h_2)$ is $O(L)$, where L is the number of histogram bins. Equation (11) can be simplified as follows:

$$D(h_1, h_2) = \sum_{m=0}^{M} |d_m(\mathbf{s})|. \tag{13}$$



(1) (2) (3)

(4) (5)

Fig. 3. Image examples and computational results of diffusion distance. (1) Image 1. (2) Image 2. (3) Image 3. (4) Image 4. (5) Diffusion distance over two images. For example, (1,4) indicates image pairs 1 and 4.

Fig. 3 denotes example of the computational results of diffusion distance over three pairs of images. It is observed that the diffusion distance is antiproportional to the similarity between the images.

IV. Gabor Diffusion Distance for Face Recognition

Like pyramid-matching kernel (PMK) [25], the computation of diffusion distance involves the sum of the histogram difference over different scales of the original images. Diffusion-distance-based algorithms have not considered the spatial structure information that has been proved to be valuable in face recognition (e.g., [11]). For this reason, some approaches have been developed to extract the spatial structure information of face images. For example, Brunelli and Poggio [6] introduced a geometric-feature-based template-matching algorithm. Gao *et al.* [21] applied spatial structure circular descriptor (SSCD) for content-based 3-D model analysis. In the meantime, Gabor features can be effectively extracted from the images, being shape contexts for the recognition purpose. In this section, we investigate a new approach by incorporating the Gabor filtering into the scheme of diffusion-distance-based similarity search. First, we introduce the use of Gabor features in the scope of diffusion-distance calculation.

A. Perceptual Features for Image Representation

A filter bank of Gabor functions are normally used for descriptions of cell receptive fields in V1 of the primate visual cortex [17]. This is obtained by convolving the image I with the filter bank to produce a vector of filter response

$$\hat{I}(x,y) = I(x,y) * \tilde{f}(x,y) \tag{14}$$

which characterizes the image patch centered at (x,y) ($\tilde{f}$ is a filter function). The filter bank consist of a number of scales and orientations.

To select appropriate discrete frequencies that define the scales, it is common to use the following exponential sampling [15]:

$$F_{k_1} = c^{-k_1} F_0 \tag{15}$$

where $k_1 = \{0, \ldots, m-1\}$. F_{k_1} is the k_1th frequency and c is the frequency scaling factor (>1). In this paper, we utilize three scales for spatial frequency selectivity that has been reported to be effective in the literature [48].

The selection of discrete rotation angles θ_k follows the proposal of Kyrki *et al.* [30], where the orientations must be spaced uniformly, i.e.,

$$\theta_{k_2} = \frac{k_2 2\pi}{n} \tag{16}$$

where $k_2 = \{0, \ldots, n-1\}$. Here, n is the number of orientations to be used. As suggested, the computation can be reduced to half, since responses on angles $[\pi, 2\pi]$ are 90° shifted from responses on $[0, \pi]$ in real values. For example, we obtain the results of Gabor filtering, as illustrated in Fig. 4.

Using (4), we shall have a response matrix $\mathbf{G}$ as a result of the Gabor filtering [30]

$$\mathbf{G}(x,y) = \begin{pmatrix} r_{x,y,s_0,\theta_0} & r_{x,y,s_0,\theta_1} & \cdots & r_{x,y,s_0,\theta_{n-1}} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ r_{x,y,s_{m-1},\theta_0} & r_{x,y,s_{m-1},\theta_1} & \cdots & r_{x,y,s_{m-1},\theta_{n-1}} \end{pmatrix} \tag{17}$$

where r is the response element in the matrix, s_m are the interest frequencies, and θ_n are the orientations. To minimize the effects of illumination changes, we can normalize the response matrix as follows:

$$\parallel \mathbf{G} \parallel = \frac{\mathbf{G}}{\sqrt{\sum_{i,j} |\mathbf{G}_{i,j}|^2}}. \tag{18}$$

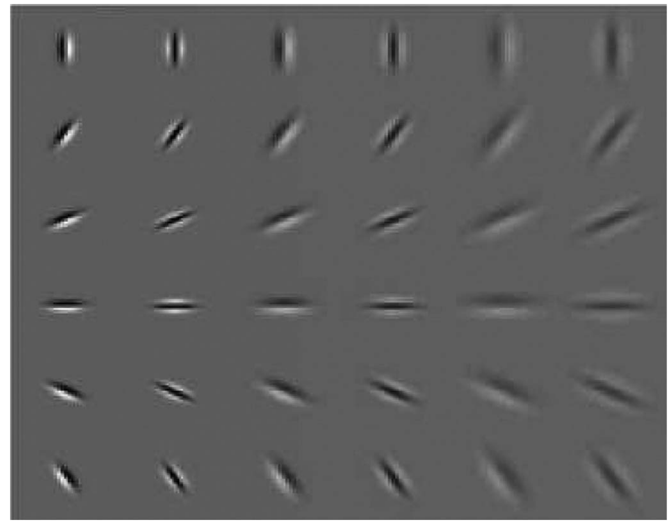


Fig. 4. Filter bank of 36 filters consisting of two phases (even and odd), three scales (spaced by half-octaves), and size orientation (uniformly spaced from 0 to π).

B. Bayesian Model

We now look at the available features out of the Gabor filters and their corresponding model distribution $p(\beta)$, where β is a set of model parameters of the distribution. In face recognition, a face is associated with the feature appearance $\mathcal{A}$ and spatial structures χ. Let $\mathcal{A}_d$ and χ_d be the appearance of the detected feature and spatial structures, respectively. The similarity between two histograms (or images) can be determined by a Bayesian decision $\mathcal{B}$ ($\mathbf{S}$ indicate that two faces refer to the same person and $\bar{\mathbf{S}}$ is the opposite)

$$\begin{aligned} \mathcal{B} &= \frac{p(\mathbf{S}|\mathcal{A}, \mathcal{A}_d, \chi, \chi_d)}{p(\bar{\mathbf{S}}|\mathcal{A}, \mathcal{A}_d, \chi, \chi_d)} \\ &= \frac{p(\mathcal{A}, \chi|\mathcal{A}_d, \chi_d, \mathbf{S})p(\mathbf{S})}{p(\mathcal{A}, \chi|\mathcal{A}_d, \chi_d, \bar{\mathbf{S}})p(\bar{\mathbf{S}})} \\ &\approx \frac{\int p(\mathcal{A}, \chi|\beta, \mathbf{S})p(\beta|\mathcal{A}_d, \chi_d, \mathbf{S})d\beta}{\int p(\mathcal{A}, \chi|\beta, \bar{\mathbf{S}})p(\beta|\mathcal{A}_d, \chi_d, \bar{\mathbf{S}})d\beta}. \end{aligned} \tag{19}$$

The larger the $\mathcal{B}$ becomes, the more similarity between the two faces holds. In addition, we can have an approximation $p(\beta|\mathcal{A}_d, \chi_d, \bar{\mathbf{S}}) \approx p(\beta|\mathcal{A}_d, \chi_d)$.

The model parameters β is subject to a posterior distribution $p(\beta|\mathcal{A}_d, \chi_d)$ that consist of parameters $\mathcal{A}_d$ and χ_d. The ratio between $\mathbf{S}$ and $\bar{\mathbf{S}}$ can be considered to be 1 in practice. We consider a mixture model of constellation models [65] with J components. Let c_J be a mixing coefficient, $\mu_{\mathcal{A}_d}$ and μ_{χ_d} be means of appearance and structures, respectively, $\Omega_{\mathcal{A}_d}$ and Ω_{χ_d} be precision matrices of appearance and structures. Using the definition of [19] while assuming the independence of image patches, we have

$$\begin{aligned} & p(\beta|\mathcal{A}_d, \chi_d) \\ &\quad = p(c_J) \prod p(\Omega_{\mathcal{A}_d}) p(\mu_{\mathcal{A}_d}|\Omega_{\mathcal{A}_d}) p(\Omega_{\chi_d}) p(\mu_{\chi_d}|\Omega_{\chi_d}) \end{aligned} \tag{20}$$

where $p(c_J)$ is a symmetric Dirichlet. We have

$$\begin{cases} p(\Omega_{\mathcal{A}_d}) = \mathcal{W}(\Omega_{\mathcal{A}_d}|\alpha_{\mathcal{A}_d}, \Gamma_{\mathcal{A}_d}) \\ p(\Omega_{\chi_d}) = \mathcal{W}(\Omega_{\chi_d}|\alpha_{\chi_d}, \Gamma_{\chi_d}) \end{cases} \tag{21}$$

which are two Wishart distributions that act as the weights, depending on the Gabor responses, and

$$p(\mu_{\chi_d}|\Omega_{\chi_d}) = \mathcal{N}(\chi_d|\mathbf{m}_{\chi_d}, \lambda_{\chi_d}\Omega_{\chi_d}) \tag{22}$$

where $\mathcal{N}$ is a Gaussian with parameters $(\mathbf{m}_{\chi_d}, \lambda_{\chi_d})$. Unlike [19], which can adapt itself according to the residual tails, we here define $p(\mathcal{A}, \chi|\mathcal{A}_d, \chi_d, \mathbf{S})$ to be a multivariate Gaussian rather than Student's T distribution

$$\begin{aligned} p(\mathcal{A}, \chi|\mathcal{A}_d, \chi_d, \mathbf{S}) \\ = \sum_i \sum_j c_0 \mathcal{N}(\mathcal{A}_j|\alpha_{\mathcal{A}_i}, \mathbf{m}_{\mathcal{A}_i}, \Gamma_{\mathcal{A}_i}) \mathcal{N}(\chi_j|\alpha_{\chi_i}, \mathbf{m}_{\chi_i}, \Gamma_{\chi_i}) \end{aligned} \tag{23}$$

where c_0 is a constant. Using Gaussian instead of Student's distributions is driven by the fact that, first, the computational efficiency can be improved; second, less parameterization is required (the latter needs to define confidence intervals that can vary in different circumstances). The case with $\bar{\mathbf{S}}$ is very similar and omitted here.

C. Similarity Measure

Similarity measure is necessary for face recognition. For histogram similarity, it is often to compute "distances" between two faces, and then, use a defined distance metric for similarity measure. On the other hand, kernel-based schemes can be used for face recognition, where an inner product is performed. Our proposed approach, alike the former, is to describe a set of Gabor features in a histogram.

Perceptual features can be extracted using the Gabor filtering (see the previous section). The logarithm of $\mathcal{B}$ directly links (19) back to the computation of diffusion distance between two faces. In the previous section, recognition for two face images have been discussed. For face recognition of multiclass face images, we first conduct pair-wise similarity measures, which is followed by a classification procedure (k-means) that categorizes the overall faces into corresponding classes. We could use other classifiers that may or may not end up with better recognition rates. Nevertheless, our main attention in this study is to justify the superiority of the proposed feature-extraction scheme to the state of the art, rather than the classifier's performance. Therefore, the proposed algorithm for multiclass face recognition is shown in the following tabulation (see Algorithm 1).

Algorithm 1 The proposed Gabor-diffusion distance based face recognition algorithm.

1. Initialisation of Gabor phases, scales and orientation parameters
2. For scales n_2 = 1:3
3. For orientations n_3 = 0:$\frac{\pi}{6}$:π

repeat

  4. Gabor filtering using Eqs. (4) and (17);
  5. Computation of oriented diffusion distance by Eq. (11):

  repeat

    6. Pyramid formation;
    7. Histogram differencing;
    8. Summation of the differences;

  until

  9. Similarity measure by Eq. (19).

until

10. EndFor
11. EndFor
12. Apply k-means classification to the available similarity matrix.

V. Experimental Work

In this section, we describe the experiments on two databases for face recognition. The first experiment will be performed on the Sheffield face images, which is a publicly accessible database [24]. This database consists of 564 images of 20 individual persons (mixed race/gender/appearance). Each person is photographed in a range of poses from profile to frontal views—each in a separate directory labeled 1a, 1b, ...,1t and the face images are numbered consecutively as they were acquired. Each image is of approximately 220-by-220 pixels with 256-bit gray scale. Example images of this database are illustrated in Fig. 5.

Fig. 5. Image examples of the Sheffield face database.

Fig. 6. Image examples of the MIT-CBCL Face database.

The second experiment will use the MIT-CBCL face-recognition database that contains face images of ten subjects. We will only use the test set in this database, which has 200 images per subject. The illumination, pose (up to about 30° of rotation in depth that causes partial occlusions), and the background changes do apply [66]. Example images of this database can be found in Fig. 6, where occlusions due to the pose changes often appear.

To evaluate the proposed perceptual features-based face-recognition algorithm, we compare the proposed approach with several competing methods. These techniques (and the proposed one) consists of the classical diffusion-distance-based

Fig. 7. Illustration of the face images superimposed by the detected SIFT keypoints, where the length and orientation of an arrow indicate the scale of a keypoint and the corresponding orientation. Better view in color.

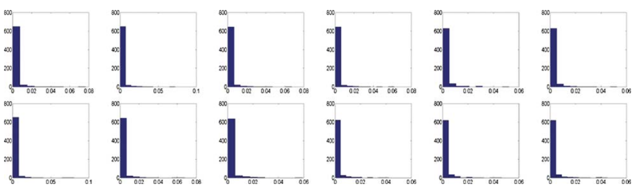


Fig. 8. Histogram illustration of PHOG features extracted from Fig. 5.

voting ("DD-V"), SIFT matching [38] plus k-means classification ("SIFT-KM"), pyramid histogram of orientation gradients (PHOGs) [5] plus multiclass SVM ("PHOG-SVM"), inner-distance shape context [35] plus k-means classification ("IDSC-KM") algorithms, and the proposed Gabor diffusion distance plus k-means classification (abbreviated as "GDD-KM"). All the algorithms involved in the evaluation have not been optimized for the efficiency purpose.

A. Image Representations

SIFT feature detection intends to find the key locations that are defined as maxima and minima of the outcomes of difference of Gaussians function applied to a smoothed and resampled image. Low-contrast candidate points and edge response points along an edge will be discarded. Dominant orientations are then assigned to localized keypoints. SIFT keypoints of Fig. 5 are denoted in Fig. 7. For the sake of feature matching, Lowe proposed to use a modification of the k-d tree algorithm that allows us to assign the nearest neighbors with high probability as the best matches.

The PHOG descriptors [5] can be used to describe object shapes. The motivation of using these descriptors is driven by the fact that human beings do consider the shape properties when they attempt to classify different faces, e.g. hair, nose, or mouth. The underlying feature of these PHOG descriptors is an edge map that can be calculated using the Canny's edge-detection method [8]. The extracted vector represents the image by local shapes and spatial layouts of the shapes. Local shapes are captured by the distribution of the edge orientations within a region and spatial layouts by tiling the image into regions

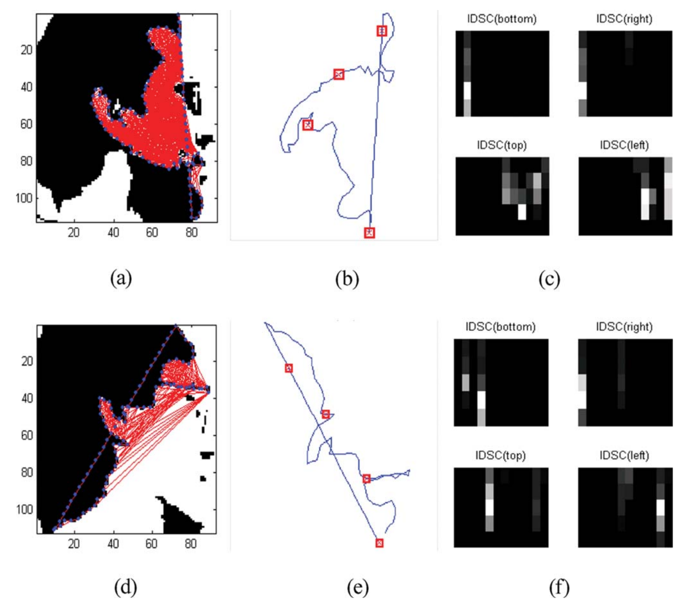


Fig. 9. Illustration of inner distance shape-context features, where (a) and (d) are shape path, (b) and (e) are longest contour, and (c) and (f) are inner distance shape context.

at different resolutions. Therefore, the descriptors are a set of histograms of orientation gradients over each image subregion at each resolution level. Fig. 8 shows histograms of PHOG features related to the subfigures of Fig. 5. It is observed that the two rows posses very similar histograms due to the similarity of the two faces in shape.

IDSC deploys the inner distance to build shape descriptors [35]. The inner distance is articulation insensitive, while being optimal for complicated shapes. The articulation invariant signatures of 2-D shapes can be generated by combining the inner distance and multidimensional scaling (MDS). Afterward, the shape context is integrated with the inner distance to formulate a new descriptor. This is followed by a dynamic-programming-based optimization method that pursues shape matching. The IDSC descriptors can be extended to capture texture information. In this paper, before calculating the inner

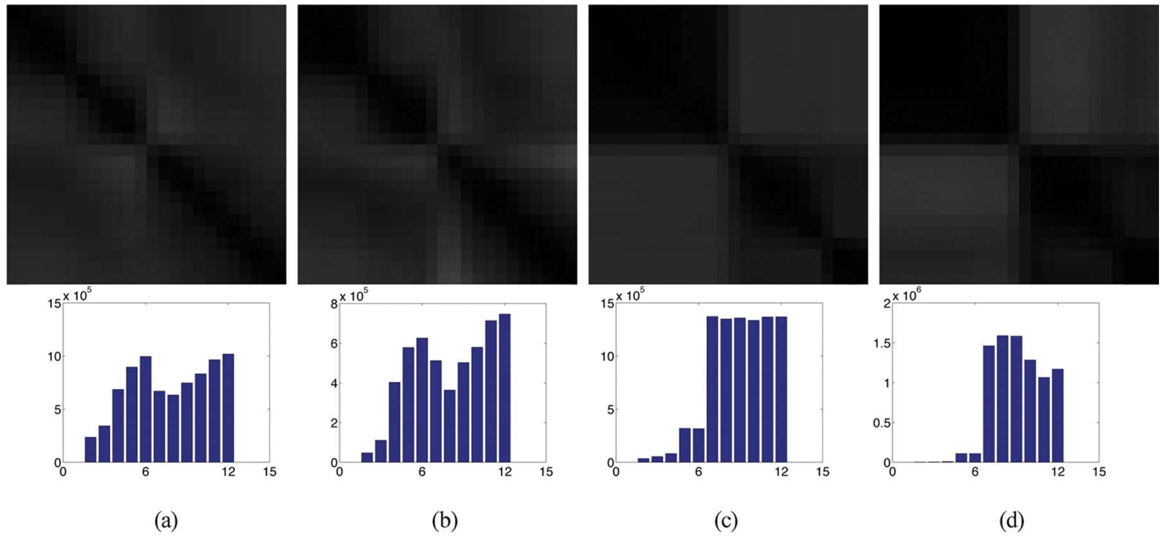


Fig. 10. Illustration of diffusion distance maps by different algorithms. (a) and (c) Classical diffusion distance. (b) and (d) Proposed Gabor diffusion distance. (a) and (b) Result from Fig. 5. (c) and (d) Results from Fig. 6. Histograms indicate the calculated diffusion distance against image numbers (1, 2, . . . , 12).

distance, we apply a k-means clustering algorithm to produce binary images from the original face images. Fig. 9 denotes the inner-distance shape-context features of the first column images of Fig. 5. Fig. 9(c)–(f) illustrates the resulting IDSCs of two different faces.

In Fig. 10, given two groups of face images similar to as shown in Fig. 5, the proposed Gabor diffusion-distance method leads to more discriminative representations [see Fig. 10(a) and (b)]. Fig. 6 shows two sets of significantly varied faces and the proposed Gabor diffusion-distance method produces large variations [see Fig. 10(c) and (d)]. In the meantime, the histograms of the computed diffusion distance against image numbers further enhance these observations, where the proposed Gabor diffusion-distance method reduces the difference within the same class while augmenting the difference between the two classes. Therefore, the proposed feature-representation algorithm is better than the classical diffusion-distance scheme in these examples in terms of the discriminative capability.

B. Classification Accuracy

1) Experiment 1—Sheffield Faces: The proposed strategy is evaluated here using the Sheffield face database. Features can be extracted, and then, categorized using the previously mentioned algorithms, i.e., "GDD-KM," "DD-V," "SIFT-KM," "PHOG-SVM," and "IDSC-KM." The overall classification operates without any supervision. The classification performance can be measured using confusion matrices and mean recognition rates over the individuals.

Here, we use two examples to demonstrate how the obtained confusion matrices present the systematic performance. Table I shows the results of the confusion matrix using the "PHOG-SVM" algorithm. We observe that, in general, this algorithm leads to very low-recognition rates. For example, the diagonal elements of classes 6 and 11 possess the lowest rates, which is

Fig. 11. Examples of classified Sheffield face images. First three columns—correctly categorized face images. Remaining columns—incorrectly categorized face images.

21% due to the strong intra- and intersimilarity of face images. These two groups have been confused with classes 13 and 1, respectively. The highest rates (89%) appear on class 19, which is somehow influenced by class 20. Most of the remaining recognition rates are less than 50%. Comparably, Table II illustrates the results of "GDD-KM" with the range of (47%, 84%), where most values are around 60%–70%.

Taking an insight into this case, we focus on the comparison of the five algorithms. In Table III, it is observed that the "GDD-KM" algorithm has significant improvements against the others in terms of the classification rates. The performance is closely related to the extracted features that are discriminant for specific categories. Exemplary categorization results are shown in Fig. 11, where "correctly" and "incorrectly" categorized images demonstrate the degree of classification difficulty. It is found that the miscategorization of the face images on the lower row is due to the fact that the images shown on the two rows are of very similar shapes and textures (e.g., glasses and beards).

2) Experiment 2—MIT-CBCL Faces: All the algorithms are evaluated using the MIT-CBCL face database in this section. Similar to the aforementioned discussion, we use two examples to show the performance comparison of "PHOG-SVM" and "GDD-KM." Tables IV and V denote the confusion matrices of "PHOG-SVM" and "GDD-KM," respectively. The average categorization rate of "PHOG-SVM" is approximately $50\% \pm 7\%$, while that of "GDD-KM" is $76\% \pm 22\%$. Comparing the five

TABLE I
CONFUSION MATRIX OF SHEFFIELD FACES BY "PHOG-SVM"

Percentages	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	**26**	0	5	0	5	11	0	11	11	0	26	16	0	0	0	0	0	5	0	16
2	5	**42**	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0
3	0	0	**5**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	**32**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	11	0	0	11	**53**	0	0	5	0	0	5	0	0	0	0	0	0	0	0	5
6	5	0	0	0	0	**21**	0	0	0	0	5	0	0	0	0	0	0	0	0	0
7	5	5	0	11	0	0	**32**	0	0	0	0	0	0	0	0	0	0	0	0	0
8	5	0	0	0	11	0	0	**26**	0	0	0	5	0	0	0	0	0	0	0	0
9	0	0	0	5	5	0	26	21	**37**	0	0	5	0	0	0	0	0	0	0	0
10	0	5	5	0	0	11	0	0	0	**32**	5	0	0	0	0	0	0	0	0	0
11	0	0	0	0	5	11	5	0	0	32	**21**	0	0	0	0	0	0	0	0	0
12	11	0	0	5	5	0	5	11	16	0	0	**32**	0	0	0	0	0	0	0	5
13	0	0	5	0	0	21	0	0	5	16	0	0	**47**	0	0	0	0	0	0	5
14	11	5	11	5	11	5	0	0	21	0	5	11	0	**37**	0	0	0	0	0	0
15	5	21	5	5	0	0	0	11	0	5	16	0	0	21	**37**	0	0	0	0	0
16	16	16	16	0	0	5	0	0	0	0	5	5	11	21	16	**63**	0	0	0	0
17	0	0	21	11	0	5	0	0	5	5	0	0	0	5	26	5	**37**	0	0	0
18	0	0	16	11	0	5	0	16	5	5	0	0	0	11	21	0	26	**84**	0	0
19	0	5	11	5	0	0	11	0	0	0	11	0	21	0	0	26	37	5	**89**	0
20	0	0	0	0	5	5	21	0	0	5	0	21	21	5	0	5	0	5	11	**68**

TABLE II
CONFUSION MATRIX OF SHEFFIELD FACES BY "GDD-KM"

Percentages	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	**74**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	**58**	5	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	11	0
3	0	21	**79**	0	0	0	0	0	0	0	0	0	16	0	0	0	0	0	0	0
4	0	11	0	**84**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	**47**	0	0	21	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	11	**74**	0	0	0	11	16	0	0	0	5	0	0	0	0	0
7	0	0	0	0	0	0	**79**	16	0	0	0	0	0	0	0	0	5	0	0	0
8	0	0	0	0	0	0	0	**63**	0	0	0	0	0	0	0	0	0	21	0	0
9	0	0	0	0	0	0	0	0	**74**	11	0	0	0	11	0	5	0	0	0	11
10	0	0	0	0	0	11	0	0	0	**68**	16	0	0	0	5	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	**47**	0	0	0	16	0	0	0	0	0
12	0	11	16	0	32	0	0	0	0	0	0	**84**	0	0	0	5	16	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	**84**	0	0	0	0	0	11	11
14	0	0	0	0	0	0	0	0	0	0	0	0	0	**63**	0	0	0	0	0	11
15	0	0	0	16	0	16	0	0	0	0	0	0	0	11	**68**	11	0	0	5	0
16	0	0	0	0	0	0	0	0	0	0	11	16	0	0	0	**58**	0	0	0	0
17	0	0	0	0	0	0	16	0	0	11	11	0	0	0	5	11	**79**	0	0	0
18	26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	**79**	0	11
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	**74**	5
20	0	0	0	0	0	0	5	0	26	0	0	0	0	16	0	0	0	0	0	**53**

TABLE III
MEAN AND STANDARD DEVIATION OF RECOGNITION RATES OF SHEFFIELD FACES BY DIFFERENT ALGORITHMS

	"GDD-KM"	"DD-V"	"SIFT-KM"	"PHOG-SVM"	"IDSC-KM"
Rates(%)	69±12	65±11	39±7	41±21	61±18

TABLE IV
CONFUSION MATRIX OF MIT-CBCL FACES BY "PHOG-SVM"

Percentages	1	2	3	4	5	6	7	8	9	10
1	**48**	2	2	2	0	0	0	2	0	4
2	16	**42**	2	8	2	0	0	0	0	0
3	22	26	**40**	0	2	0	2	0	16	8
4	0	10	20	**52**	0	2	0	2	0	11
5	0	10	6	20	**54**	2	0	2	0	8
6	0	0	0	4	34	**40**	8	4	2	4
7	12	4	12	4	2	30	**54**	6	2	0
8	2	6	0	10	0	8	26	**54**	2	9
9	0	0	14	0	0	6	8	30	**56**	0
10	0	0	4	0	6	12	2	0	22	**56**

TABLE V
CONFUSION MATRIX OF MIT-CBCL FACES BY "GDD-KM"

Percentages	1	2	3	4	5	6	7	8	9	10
1	**92**	0	4	0	0	0	0	0	6	0
2	0	**50**	0	0	0	8	0	0	0	0
3	2	0	**92**	0	4	0	32	0	0	0
4	2	0	0	**32**	0	0	0	0	0	0
5	0	12	0	68	**62**	0	4	0	0	0
6	4	0	0	0	0	**92**	2	0	0	0
7	0	38	0	0	6	0	**62**	0	0	8
8	0	0	0	0	0	0	0	**90**	0	0
9	0	0	0	0	0	0	0	10	**94**	2
10	0	0	4	0	28	0	0	0	0	**90**

TABLE VI
MEAN AND STANDARD DEVIATION OF RECOGNITION RATES OF MIT-CBCL FACES BY DIFFERENT ALGORITHMS

	"GDD-KM"	"DD-V"	"SIFT-KM"	"PHOG-SVM"	"IDSC-KM"
Rates(%)	76±22	67±13	43±9	50±7	68±11

algorithms as a whole, we can obtain Table VI, which reveals that the proposed "GDD-KM" has the best classification capability. Interestingly, "IDSC-KM" has the second best performance, where most distinctive shapes in the categories have captured the attention of "IDSC-KM." Fig. 12 shows some image examples that have been correctly or incorrectly classified. The misclassification attributes to the significant loss of the facial features and the proposed algorithm cannot separate it from the other images.

VI. DISCUSSIONS AND CONCLUSION

In this paper, we have presented a new face-recognition algorithm by combining Gabor features within the scope of

Fig. 12. Examples of classified MIT-CBCL face images. First three columns— correctly categorized face images. Remaining columns— incorrectly categorized face image.

diffusion-distance calculation. This strategy starts from the Gabor filtering that consists of three scales and six orientations. It is followed by the calculation of diffusion distance based on a Bayesian model. This proposed algorithm has been compared against several state-of-the-art techniques. The experimental results show that the proposed face-recognition scheme has the best performance in accuracy.

Most classical HBLDs can only be used to deal with the aligned shapes and are sensitive to distortion and quantization of the images. To compound these problems, in this paper, Gabor features are generated to represent the local characteristics. These features are driven by the nature of human perception, being of a good capability to differentiate the face images used in this study. To enhance the performance of face recognition, a Bayesian model was used in this study to determine the similarity degree between two histograms. The rationale of using this Bayesian decision model is to ensure the maximization of likely estimation across different histograms. The available experimental results have justified the successful usage of the Gabor features and the Bayesian model. In spite of the promising results, we also observed less satisfactory outcomes of the recognition. For example, the recognition rates of the proposed algorithm in handling the occlusions due to dramatical pose changes.

In the future, work will proceed in various areas. For example, some categories of the face databases have witnessed some less optimal classification rates by the proposed algorithm, e.g., when the subjects are bearded or wear glasses. This issue may be solved if substantial efforts can be made to model the appearance, e.g., textures in these areas. On the other hand, Gabor filtering process is time consuming for the time being, but an appropriate optimization process of the proposed algorithm will most likely lead to better computational efficiency.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers and the handling editor for their invaluable comments that significantly help improve the quality of this manuscript.

REFERENCES

[1] J. Bartlett and J. Searcy, "Inversion and configural of faces," *Cogn. Psychol.*, vol. 25, pp. 281–316, 1993.
[2] P. Belhumeur, J. Hespanha, and D. 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.
[3] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 509–522, Apr. 2002.

[4] P. Benson and D. Perrett, "Perception and recognition of photographic quality facial caricatures: Implications for the recognition of natural images," *Euro. J. Cogn. Psychol.*, vol. 3, pp. 105–135, 1991.

[5] A. Bosch, A. Zisserman, and X. Munoz, "Urepreenting shape with a spatial pyramid kernel," in *Proc. CIVR*, 2007, pp. 401–408.

[6] R. Brunelli and T. Poggio, "Face recognition: Features versus templates," *IEEE Trans. Pattern. Anal. Mach. Intell.*, vol. 15, no. 10, pp. 1042–1052, Oct. 1993.

[7] A. Calder, A. Young, J. Keane, and M. Dean, "Configural information in facial expression perception," *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 26, pp. 527–551, 2000.

[8] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Vis.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[9] O. Chapelle, P. Haffner, and V. N. Vapnik, "Support vector machines for histogram-based image classification," *IEEE Trans. Neural Netw.*, vol. 10, no. 5, pp. 1055–1064, May 1999.

[10] W.-P. Choi, S.-H. Tse, K.-W. Wong, and K.-M. Lam, "Simplified gabor wavelets for human face recognition," *Pattern Recognit.*, vol. 41, no. 3, pp. 1186–1199, 2008.

[11] N. Costen, D. Parker, and I. Craw, "Spatial content and spatial quantisation effects in face recognition," *Perception*, vol. 23, pp. 129–146, 1994.

[12] S. Dakin and R. Watt, "Biological "bar codes" in human faces," *J. Vis.*, vol. 9, pp. 1–10, 2009.

[13] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2005, pp. 886–893.

[14] J. Daugman, "Two-dimensional spectral analysis of cortical receptive field profiles," *Vis. Res.*, vol. 20, no. 10, pp. 847–856, 1980.

[15] J. Daugman, "Complete discrtete 2-d gabor transformations by neural networks for image analysis and compression," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 7, pp. 1169–1179, Jul. 1988.

[16] J. R. del Solar and P. Navarrete, "Eigenspace-based face recognition: A comparative study of different approaches," *IEEE Trans. Syst., Man, Cybern.C, Appl. Rev.*, vol. 35, no. 3, pp. 315–325, Aug. 2005.

[17] R. DeValois and K. K. DeValois, *Spatial Vision*. Oxford, U.K.: Oxford Univ. Press, 1988.

[18] R. Diamond and S. Carey, "Why faces are and are not special: An effect of expertise," *J. Exp. Psych.: Gen.*, vol. 115, pp. 107–117, 1986.

[19] L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories," *Comput. Vis. Image Understanding*, vol. 106, no. 1, pp. 59–70, 2007.

[20] D. Gabor, "Theory of communication," *J. IEE*, vol. 93, pp. 429–459, 1946.

[21] Y. Gao, Q. Dai, and N.-Y. Zhang, "3d model comparison using spatial structure circular descriptor," *Pattern Recognit.*, vol. 43, no. 3, pp. 1142–1151, 2010.

[22] W. Garner, "Aspects of a stimulus: Features, dimensions, and Configurations," in *Cognition and Categorization*. E. H. Rosch and B. B. Lloyd, Eds. Hillsdale, NJ: Erlbaum, 1978, pp. 99–133.

[23] V. Goffaux, B. Hault, C. Michel, Q. Vuong, and B. Rossion, "The respective role of low and high spatial frequencies in supporting configural and featural processing of faces," *Perception*, vol. 34, pp. 77–86, 2005.

[24] D. Graham and N. Allinson, "Characterizing virtual eigensignatures for general purpose face recognition," in *Face Recognition: From Theory to Applications* (ser. NATO ASI Series F, Computer and Systems Sciences), H. Wechsler, P. J. Phillips, V. Bruce, F. Fogelman-Soulie, and T. S. Huang, Eds., Berlin, Germany: Springer-Verlag, 1998, pp. 446–456.

[25] K. Grauman and T. Darrell, "The pyramid match kernel: Discriminative classification with sets of image features," in *Proc. Int. Conf. Comput. Vis.*, 2005, pp. 1458–1465.

[26] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang, "Face recognition using laplacianfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 328–340, Mar. 2005.

[27] R.-L. Hsu and A. Jain, "Semantic face matching," presented at the IEEE Conf. Multimedia Expo., Lausanne, Switzerland, 2002, pp. 145–148.

[28] C. Huang, C. Chen, and P. Chung, "Contrast context histogram—An efficient discriminating local descriptor for object recognition and image matching," *Pattern Recognit.*, vol. 41, no. 10, pp. 3071–3077, 2008.

[29] J. Iloen, J.-K. Kamarainen, P. Paalanen, M. Hamouz, J. Kittler, and H. Kalviainen, "Image feature localization by multiple hypothesis testing of gabor features," *IEEE Trans. Image Process.*, vol. 17, no. 3, pp. 311–325, Mar. 2008.

[30] V. Kyrki, J.-K. Kamarainen, and H. Kälviäinen, "Simple gabor feature space for invariant object recognition," *Pattern Recognit. Lett.*, vol. 25, no. 3, pp. 311–318, 2004.

[31] I. Laptev, "Improvements of object detection using boosted histograms," in *Proc. Br. Mach. Vis. Conf.*, 2005, pp. 2588–2599.

[32] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bag of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2006, pp. 2169–2178.

[33] X. Li and Y. Pang, "Deterministic column-based matrix decomposition," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 1, pp. 145–149, Jan. 2010.

[34] H. Lin and K. Okada, "Diffusion distance for histogram comparison," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2006, pp. 246–253.

[35] H. Lin and K. Okada, "An efficient earth mover's distance algorithm for robust histogram comparison," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 5, pp. 840–853, May 2007.

[36] C. Liu, "Gabor-based kernel pca with fractional power polynomial models for face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 5, pp. 572–581, May 2004.

[37] 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.

[38] D. Lowe, "Object recognition from local scale-invariant features," in *Proc. IEEE Int. Conf. Comput. Vis.*, 1999, pp. 1150–1157.

[39] D. Maurer, R. Le Grand, and C. Mondloch, "The many faces of configural processing," *Trends Cogn. Sci.*, vol. 6, pp. 255–260, 2002.

[40] B. Mel, "SEEMORE: Combining color, shape, and texture histogramming in a neurally inspired approach to visual object recognition," *Neural Comput.*, vol. 9, pp. 777–804, 1997.

[41] J. Movellan. (2010, Jan.). Tutorial on gabor filters. [Online]. Available: http://mplab.ucsd.edu/tutorials/gabor.pdf

[42] D. Nister and H. Stewenius, "Scalable recognition with a vocabulary tree," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2006, pp. 2161–2168.

[43] Y. Pang, Y. Yuan, and X. Li, "Gabor-based region covariance matrices for face recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 7, pp. 989–993, Jul. 2008.

[44] J. Penry, *Looking at Faces and Remembering Them: A Guide to Facial Identification*. London, U.K.: Blek Books, 1971.

[45] M. Potzsch, N. Kruger, and C. von der Malsburg, "Improving object recognition by transforming Gabor flter responses," *Netw.: Comput. Neural Syst.*, vol. 7, no. 2, pp. 341–347, 1996.

[46] R. Rahmani, S. Goldman, H. Zhang, S. Cholleti, and J. Fritts, "Localized content-based image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 11, pp. 1902–1912, Apr. 2008.

[47] S. Rakover, "Featural vs. configurational information in faces: A conceptual and empirical analysis," *Br. J. Psychol.*, vol. 93, pp. 1–30, 2002.

[48] L. Renninger and J. Malik, "When is scene identification just texture recognition?," *Vis. Res.*, vol. 44, no. 19, pp. 2301–2311, 2004.

[49] G. Rhodes, S. Brake, and A. Atkinson, "What's lost in inverted faces," *Cognition*, vol. 47, pp. 25–57, 1993.

[50] R. Robbins and E. McKone, "Can holistic processing be learned for inverted faces?," *Cognition*, vol. 88, pp. 79–107, 2003.

[51] B. Rossion and A. Boremanse, "Nonlinear relationship between holistic processing of individual faces and picture-pane rotation: Evidence from the face composite illusion," *J. Vis.*, vol. 8, pp. 1–13, 2008.

[52] H. Rowley, S. Baluja, and T. Kanade, "Neural network-based face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 1, pp. 23–38, Jan. 1996.

[53] Y. Rubner, C. Tomasi, and L. Guibas, "The earth mover's distance as a metric for image retrieval," *Int. J. Comput. Vis.*, vol. 40, no. 2, pp. 99–121, 2000.

[54] B. Schiele and J. Crowley, "Recognition without correspondence using multidimensional receptive field histograms," *Int. J. Comput. Vis.*, vol. 36, pp. 31–50, 2000.

[55] A. Schwaninger, S. Ryf, and F. Hofer, "Configuration information is processed differently in perception and recognition of faces," *Vis. Res.*, vol. 43, pp. 1501–1505, 2003.

[56] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio, "Robust object recognition with cortex-like mechanisms," *IEEE Trans. Pattern Recognit. Mach. Intell.*, vol. 29, no. 3, pp. 411–426, Mar. 2007.

[57] L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human faces," *J. Opt. Soc. Amer. A*, vol. 4, no. 3, pp. 519–524, 1987.

[58] L. W. T. Serre and T. Poggio, "Object recognition with features inspired by visual cortex," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2005, pp. 994–1000.
[59] J. Tanaka and M. Farah, "Parts and wholes in face recognition," *Quart. J. Exp. Psychol.*, vol. 46A, pp. 225–245, 1993.
[60] D. Tao, X. Li, X. Wu, and S. Maybank, "General tensor discriminant analysis and gabor features for gait recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1700–1715, Oct. 2007.
[61] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cogn. Neurosci.*, vol. 3, no. 1, pp. 71–86, 1991.
[62] A. Tversky and D. Krantz, "Similarity of schematic faces: A test of interdimensional additivity," *Percept. Psychophys.*, vol. 5, pp. 124–128, 1969.
[63] T. Valentine, "A unified account of the effects of distinctiveness, inversion, and race in the recognition," *Quart. J. Exp. Psychol. A: Hum. Exp. Psychol.*, vol. 43, pp. 161–204, 1991.
[64] X. Wang and X. Tang, "Hallucinating face by eigentransformation," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 35, no. 3, pp. 425–434, Aug. 2005.
[65] M. Weber, M. Welling, and P. Perona, "Unsupervised learning of models for recognition," in *Proc. Eur. Conf. Comput. Vis.*, 2000, pp. 18–32.
[66] B. Weyrauch, J. Huang, B. Heisele, and V. Blanz, "Component-based face recognition with 3d morphable models," in *Proc. First IEEE Workshop Face Process. Video*, 2004, p. 85.
[67] H. Wilson, G. Loffler, and F. Wilkinson, "Synthetic faces, face cubes, and the geometry of face space," *Vis. Res.*, vol. 42, pp. 2909–2923, 2002.
[68] P. Wolfrum, C. Wolff, J. Lucke, and C. von der Malsburg, "A recurrent dynamic model for correspondence-based face recognition," *J. Vis.*, vol. 8, no. 7, pp. 1–18, 2008.
[69] J. Wu and J. Rehg, "Beyond the Euclidean distance: Creating effective visual codebooks using the histogram intersection kernel," presented at the IEEE Int. Conf. Comput. Vis., Kyoto, Japan, 2009.
[70] X. Xie and K.-M. Lam, "Gabor-based kernel pca with doubly nonlinear mapping for face recognition with a single face image," *IEEE Trans. Image Process.*, vol. 15, no. 9, pp. 2481–2492, Sep. 2006.
[71] D. Xu, S. Yan, and J. Luo, "Face recognition using spatially constrained earth mover's distance," *IEEE Trans. Image Process.*, vol. 17, no. 11, pp. 2256–2260, Nov. 2008.
[72] A. Young, D. Hellawell, and D. Hay, "Configurational information in face perception," *Perception*, vol. 16, pp. 747–759, 1987.
[73] G. Yovel and B. Duchaine, "Specialised face perception mechanisms extract both part and spacing information: Evidence from developmental prosopagnosia," *J. Cogn. Neurosci.*, vol. 18, pp. 580–593, 2006.
[74] B. Zhang, S. Shan, X. Chen, and W. Gao, "Histogram of Gabor phase patterns (HGPP): A novel object representation approach for face recognition," *IEEE Trans. Image Process.*, vol. 16, no. 1, pp. 57–68, Jan. 2007.
[75] W. Zhao, R. Chellappa, and P. Phillips, "Subspace linear discriminant analysis for face recognition," Univ. Maryland, College Park, MD, Tech. Rep. CVL-Reports-1999/TR4009, 1999.
[76] H. Zhou, Y. Yuan, and A. Sadka, "Application of semantic features in face recognition," *Pattern Recognit.*, vol. 41, no. 10, pp. 3251–3256, 2008.

Huiyu Zhou received the B.Eng. degree in radio technology from Huangzhong University of Science and Technology, China, in 1990, the M.Sc. degree in biomedical engineering from the University of Dundee, Dundee, U.K., and the Ph.D. degree in computer vision from Heriot-Watt University, Edinburgh, U.K.

He is currently at the Institute of Electronics, Communications, and Information Technology, School of Electronics, Electrical Engineering, and Computer Science, Queen's University Belfast, Belfast, U.K. He has authored or coauthored widely in international journals and conferences. His research interests include computer vision, human motion analysis, intelligent systems, and human–computer interface.

Abdul H. Sadka (SM'08) received the Ph.D. degree in electrical and electronic engineering from Surrey University, Surrey, U.K., in 1997.

He is currently the Chair, the Head of electronic and computer engineering, and the Director of the Centre for Media Communications Research, Brunel University West London, Uxbridge, U.K. He has 15-year experience in academic leadership and excellence. He is an Internationally Renowned Expert in visual media processing and communications with an extensive track record of scientific achievements and peer-recognized research excellence. He was a Scientific Advisor and a Consultant to several key companies in the international telecommunications sector. As principal investigator, he has received nearly £2.5M worth of research grants and contracts. He has authored or coauthored widely in international journals and conferences. He is the author of a highly regarded book on "*Compressed Video Communications*" (Wiley, 2002). He holds three patents in the video transport and compression area.

Dr. Sadka is a Fellow of the Institution of Engineering and Technology.