A Novel Face Recognition Algorithm with Support Vector Machine Classifier

Latha Parthiban

Abstract— A novel face recognition algorithm based on Gabor texture information is proposed in this paper. Two strategies to capture it are Gabor Magnitude-based Texture Representation (GMTR) which is characterized by using the Gamma density to model the Gabor magnitude distribution and Gabor Phase-based Texture Representation (GPTR), characterized by using the Generalized Gaussian Density (GGD) to model the Gabor phase distribution. The estimated model parameters serve as texture representation. Experiments are performed on Yale, ORL and FERET databases to validate the feasibility of the method. The results show that GMTR based and GPTR-based SVM classifier both significantly outperform the widely used Gabor energy based systems and other existing subspace methods.

Index Terms— Gabor wavelet, Generalized Gaussian density, model parameters, Texture, SVM classifier.

I. INTRODUCTION

Facial recognition system is a computer application for automatically identifying or verifying a person from a digital image or a video frame from a video source. One of the ways to do this is by comparing selected facial features [22-24] from the image and a facial database. It is typically used in security systems and can be compared to other biometrics such as fingerprint or eye iris recognition systems [19,20,21]. Since the techniques used in the best face recognition systems may depend on the application of the system, one can identify at least two broad categories of face recognition systems: [13].

- Finding a person within large database of faces (e.g. in a police database. Often only one image is available per person. It is usually not necessary for recognition to be done in real time.).
- Identifying particular people in real time (e.g. location tracking system. Multiple images per person are often available for training and real time recognition is required.).

Machine recognition of faces is emerging as an active research area spanning several disciplines such as image processing, pattern recognition, computer vision and neural networks. Face recognition technology has numerous commercial and law enforcement applications [15,16]. These applications range from static matching of controlled format photographs such as passports, credit cards, photo ID’s, driver’s licenses, and mug shots to real time matching of surveillance video images [9]. The paper is organized as follows: Section 2 deals with literature survey, section 3 with Gabor texture representation, section 4 with implementation, section 5 with results and section 6 with conclusions.

II. LITERATURE SURVEY

A formal method of classifying faces was first proposed by Francis Galton in 1888 [10,11]. During the 1980’s work on face recognition remained largely dormant. Since the 1990’s, the research interest in face recognition has grown. Modern face recognition has reached an identification rate greater than 90% with well-controlled pose and illumination conditions.

Physiological evidence indicates that the brain possesses specialized ‘face recognition hardware’ in the form of face detector cells in the infero-temporal cortex and regions in the frontal right hemisphere; impairment in these areas leads to a syndrome known as prosapagnosia [14]. Interestingly, prosapagnosics, although unable to recognize familiar faces, retain their ability to visually recognize non-face objects. As a result of many studies scientists come up with the decision that face recognition is not like other object recognition.

An important finding is that human face recognition system [26-30] is disrupted by changes in lighting direction and also changes of viewpoint. We can summarize the finding of studies on human face recognition system as follows:

- The human capacity for face recognition is a dedicated process, not merely an application of the general object recognition process. Thus artificial face recognition systems should also be face specific [4].
- Distinctive faces are more easily recognized than typical ones.
- Both global and local features are used for representing and recognizing faces [5].
- Humans recognize people from their own race better than people from another race and may encode an ‘average’ face.
- Certain image transformations, such as intensity negation, strange viewpoint changes, and changes in lighting direction can severely disrupt human face recognition.[17,18]

Using the present technology it is impossible to completely model human recognition system and reach its performance. However, the human brain has its shortcomings in the total number of persons that it can accurately ‘remember’. The benefit of a computer system would be its capacity to handle large datasets of face images. Due to the simplicity and the efficiency for the feature extraction and representation, many subspace analysis methods, such as principal component analysis.
analysis (PCA)[6], linear discriminant analysis (LDA) [33], independent component analysis (ICA), and locality preserving projections (LPP), are widely used for face recognition[36]. However, the face recognition task is usually confronted with image variations in illumination condition, pose, facial expression or aging. In recent years, many methods based on Gabor filters have been introduced [1,3]. Lades et al. pioneered the application of Gabor features for face recognition by proposing the dynamic link architecture (DLA). Wiskott et al. [3] extended it to the elastic bunch graph matching (EBGM) method. EBGM represents a face as a labeled graph, in which each vertex corresponds to a manually selected facial landmark (eyes, mouth or nose, etc.) and the edge represents the connection between them. After the construction of the graph, identification can be achieved by the elastic matching between the reference graph and the probe one.

Generally, DLA is superior to other face recognition techniques in terms of rotation invariant, but the matching process is computationally expensive. Not constructing a graph, Kalocsai et al. [38] employed dozens of Gabor filters to obtain face representation by convoluting the facial landmarks with these filters. However, setting the landmarks needs manual operation. On the contrary, a quite different approach involves convolving the whole face image with 40 Gabor filters, resulting in an overwhelming high-dimensional feature space. In order to reduce the dimensionality of the feature vector, a popular strategy is to down-sample the filtered images with a factor. The down-sampled Gabor features are concatenated to form an augmented feature vector, which is then projected into a low-dimensional linear subspace by LDA. Shen and Bai [40] extended the Gabor feature-based linear space to the kernel space and achieved better classification performance. However, one disadvantage of down-sampling is that a great number of discriminative Gabor features are discarded. Another technique used for dimensionality reduction is to select the optimal Gabor features. Considering the memory requirement and computational cost, only a small number of training samples can be used in practice. Recently, the AdaBoost algorithm has been employed to select the most discriminative Gabor features. However, for all the methods, it is very time-consuming to select the most useful ones from so many Gabor features.

All the above mentioned methods take directly the Gabor magnitude as face feature, which can be denoted as Gabor-based energy feature (GEF) [31-38]. Another disadvantage shared by the methods using GEF is that face matching requires alignment of corresponding pairs of pixels ideally. The histogram method to represent the features detected is not adequate in many applications as it loses structure information of the object. Fortunately, multi-resolution such as Gabor wavelet transformation of images can avoid this problem. There are other disadvantages if the histogram serves as image feature. Generally, the histogram method requires hundreds of histogram bins to capture accurately the image information. Thus it leads to impractical complexity in both storage of image features and face matching. Additionally, choosing an appropriate number of histogram bins is difficult, in most cases it has to resort to extensive experiments.

### III. GABOR TEXTURE REPRESENTATION

#### A. Gabor Wavelets

Gabor wavelets are widely and successfully used in computer vision, such as face recognition, due to their biological relevance and computational properties. The Gabor kernels are similar to the receptive field profiles of the mammalian cortical simple cells. The Gabor wavelet representation captures salient visual properties including spatial localization, orientation selectivity, and spatial frequency characteristic. In the spatial domain, a Gabor wavelet is a complex exponential modulated by a Gaussian function, which has been used to model the receptive field of the orientation-selective simple cells. The Gabor wavelets can be defined as follows [1,2]

$$\Psi_{u,v}(z) = \frac{1}{\sigma^2} e^{-\frac{(z-\mu)^2}{2\sigma^2}} e^{i\mu z}$$

where $u$ and $v$ define the orientation and scale of the Gabor kernels, $z=(x,y)$, $\mu$ denotes the norm operator, and the wave vector $k_{u,v}$ is defined as follows:

$$k_{u,v} = K f^v e^{i\phi_v}$$

where $K = \frac{K_{max}}{f^v}$, $\phi_v = \frac{\pi}{8}$, $K_{max}$ the maximum frequency and $f$ is the spacing factor between kernels in the frequency domain.

The Gabor kernels are all self-similar since they can be generated from one filter, the mother wavelet, by scaling and rotation via the wave vector $k_{u,v}$. Each kernel is a product of a Gaussian envelope and a complex plane wave, while the first term in the square bracket determines the oscillatory part of the kernel and the second term compensates for the DC value. The effect of the DC term becomes negligible when the parameter $r$, which determines the ratio of the Gaussian window width to wavelength, has sufficiently large value.

#### B. Gabor magnitude and phase representation

The Gabor representation of a face image can be obtained by convolving the image with Gabor kernels. Let $I(x,y)$ be the gray level distribution of an image, the convolution of the image and a Gabor kernel $\Psi_{u,v}$ is defined as follows

$$O_{u,v}(x,y) = I(x,y) * \Psi_{u,v}(x,y)$$

where $*$ denotes the convolution operator, and $O_{u,v}(x,y)$ is the convolution result corresponding to the Gabor kernel at scale $v$ and orientation $u$. Generally, a set of Gabor kernels with different frequencies and orientations are required ($v=0,1,\ldots,V-1; u=0,1,\ldots,U-1$). Therefore, the set $S= \{O_u(x,y): u\in\{0,1,\ldots,U-1\}; v\in\{0,1,\ldots,V-1\}\}$ forms the Gabor wavelet representation of image $I(x,y)$. The output $O_{u,v}(x,y)$ is a complex, from which the magnitude $M_{u,v}(x,y)$ and the phase $P_{u,v}(x,y)$ can be computed as follows:
\[ M_{u,v}(x,y) = \sqrt{\left(\text{Re}(O_{u,v}(x,y))\right)^2 + \left(\text{Im}(O_{u,v}(x,y))\right)^2} \]

\[ P_{u,v}(x,y) = \tan \left(\frac{\text{Im}(O_{u,v}(x,y))}{\text{Re}(O_{u,v}(x,y))}\right), \quad -\Pi \leq P_{u,v}(x,y) \leq \Pi \]

It is still an open problem to select the Gabor wavelet parameters, \(u,v,\sigma,k_{\text{max}}\) and \(f\), and up to now only experimental consideration rules the choice. In most cases, \(u = 8; \quad v = 5; \quad \sigma = 2\pi; \quad k_{\text{max}} = \pi/2\), and \(f = \sqrt{2} \pi\) are used. However, one can choose the parameters that are more suitable for a given application and an appropriate choice can be very helpful. To reduce the effects caused by illumination variances, the filtered images are normalized so that the Gabor responses for one point in all frequencies and orientations have unit variance. Once the magnitude and phase values are obtained using the above formulae it is then modeled into a histogram using gamma distribution for representing magnitude and Gaussian distribution to represent phase. From the histogram used to model the phase and the magnitude the values of \(\alpha\) and \(\beta\) are estimated using the maximum likelihood estimation (MLE) process where \(\alpha\) represents the width of the curve and \(\beta\) represents the shape of the curve that defines the histogram. Each individual image is modeled for the \(\alpha\) and \(\beta\) values as a whole and also split into two or more non overlapping sub images and each of these sub images are also modeled. This is to ensure higher efficiency. The values of the sub images are referred to as the Local Gabor Magnitude Texture Representation (LGMTR) and Local Gabor Phase Texture Representation (LGPTTR) and the overall image’s values are referred to as the Global Gabor Magnitude Texture Representation (GGMTR) and Global Gabor Phase Texture Representation (GGPTR).

IV. IMPLEMENTATION

The whole methodology can be implemented in four phases as shown in Fig.1. The various phases are explained in subsections A,B, C and D.

A. Preprocessing of the images (Phase – I)

In this phase, the images from the database are classified into training and testing set. For example, the ORL database that contains 10 images of each person is split into 7 images for training and the other 3 are used for testing. Thus the values are entered into the reference matrix using the training images and this is then used as a reference to compare the values generated from the testing images. An input image function is obtained from the database and this function is applied to the Gabor function to get the Gabor output. The same image is convolved with the Gabor output to get the Gabor magnitude and phase. The database is split into training and testing sets. The flowchart of phase-1 is shown in Fig.2. 

B. Representing the Gabor magnitude and Gabor phase(Phase - 2)

In this phase, the Gabor Magnitude is represented using the Gamma distribution and the Gabor phase is represented using the Gaussian distribution. The maximum likelihood estimation or the MLE is used to determine the parameters required to model the distributions. [3-5]. The image of the person used for extracting Gabor features and Gabor output is shown in Fig. 3. The Gabor magnitude features (Fig. 4) are represented using histograms and gamma distribution is used to model it.
C. Getting local and global Gabor features (Phase 3)

Once the magnitude and the phase are modeled using gamma and Gaussian distributions respectively, the image is subdivided into a series of smaller images and the local Gabor magnitude (Fig. 6) and phase representations (Fig. 7) are obtained. These are combined together to obtain the Gabor magnitude phase texture representation.

D. Using the Support Vector Machine to classify the image (Phase 4)

The SVM [7] classifier is designed and it classifies the image based on the extracted parameters and recognizes the image content (face). The SVM maps the features to higher dimensional space and then uses an optimal hyperplane in the mapped space. This implies that though the original features carry adequate information for good classification, mapping to a higher dimensional feature space could potentially provide better discriminatory clues that are not present in the original feature space. The selection of suitable kernel function appears to be a trial-and-error process. One would not know the suitability of a kernel function and performance of the SVM until one has tried and tested with representative data.

V. EXPERIMENTAL RESULTS

In this section, we evaluate the performances of the proposed GMTR-based, GPTR-based and GMPTR-based SVM methods for face recognition task based on different face databases. The databases used in our study include the Yale database, the ORL database and the FERET database. The Yale database is used to test the proposed methods under varying facial expressions and illumination conditions, the ORL database is used as a baseline study, and the FERET database is employed to assess the performances when there are facial variations over time, facial expression and illumination condition. In all the experiments, original images are cropped and then normalized to the size of 64 * 64 by bilinear interpolation.
A. Results on Yale database

The Yale face database contains 165 images of 15 individuals, each person has 11 different images under various lighting conditions (center-light, left-light, right-light), facial expressions (happy, normal, sad, sleepy, surprised, wink) and wearing glasses (glasses, no-glasses).

We have to first determine that how many sub-images the filtered image should be partitioned into to obtain good texture representation in terms of recognition rate. A proper size of the sub-image can balance the spatial locality and compactness of the representation. A too large size may degrade the system due to the loss of much spatial information. On the other hand, a too small size would make the system sensitive to the noise and expensive computationally. Experiments are performed on the Yale face database, and then the best parameters determined are employed for other face databases. After repeated experimentation we found that 64 × 64 images produce the best results.

<table>
<thead>
<tr>
<th>Method</th>
<th>No of training samples per subject</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3</td>
</tr>
<tr>
<td>PCA</td>
<td>70.08</td>
</tr>
<tr>
<td>SVM</td>
<td>86.88</td>
</tr>
<tr>
<td>GMTR+SVM</td>
<td>86.23</td>
</tr>
<tr>
<td>GPTR+SVM</td>
<td>88.34</td>
</tr>
<tr>
<td>GMPTR+SVM</td>
<td>89.00</td>
</tr>
</tbody>
</table>

Table 1: Comparison of recognition rate and dimensionality in Yale database

From Table 1, we can gain the following conclusions:
- The proposed GMTR-based, GPTR-based and GMPTR based SVM methods outperform PCA and SVM methods.
- The performance of GPTR-based SVM is comparable to that of GMTR-based SVM, which indicates that the Gabor phase is not useless for face classification and may contain as much discriminative information as the Gabor magnitude contains.

The GMPTR based SVM gives the best performance consistently. This means that the magnitude and the phase parts in the Gabor filtered image have provided different information for the classification and they are mutually complementary. The graph Fig.8 shows that GMPTR-based SVM gives the best performance.

B. Results on ORL database

In ORL database, there are 10 different gray images for each of 40 distinct subjects. For some subjects, the images were taken at different time, under various lighting conditions, facial expressions (open or closed eyes, smiling or not smiling) and facial details (glasses or no-glasses).

<table>
<thead>
<tr>
<th>Method</th>
<th>No of training samples per subject</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3</td>
</tr>
<tr>
<td>PCA</td>
<td>84.57</td>
</tr>
<tr>
<td>SVM</td>
<td>90.89</td>
</tr>
<tr>
<td>GMTR+SVM</td>
<td>92.10</td>
</tr>
<tr>
<td>GPTR+SVM</td>
<td>93.15</td>
</tr>
<tr>
<td>GMPTR+SVM</td>
<td>94.32</td>
</tr>
</tbody>
</table>

Table 2: Comparison of recognition rate and dimensionality in ORL database

All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movements. In this experiment, a random subset with 3, 4, 5 images per individual is selected to form the training set, and the rest of the database is used for the testing set. For each given n, we average the results over 10 random splits. In general, the performances of all these methods vary with the number of dimensions of reduced space. The top recognition rate achieved by each method and the corresponding dimensionality of reduced space are demonstrated in Table 2. It is shown that the GPTR-based and GMPTR-based SVM achieve better performances than the
other methods. The graph Fig.9 shows that GMPTR-based SVM gives the best performance.

C. Results on FERET database

The FERET database [8, 12, 25] is a standard test bed for face recognition techniques. In this experiment, we select 720 face images from 120 subjects with six images per person, which were acquired under variations over time, facial expression and illumination condition. 3, 4 images of each person are randomly chosen for training, while the remaining images are used for testing. Table 3 shows the top recognition rate and the corresponding dimensionality of reduced space for each method. As can be seen, most algorithms can achieve a higher recognition rate when more training samples are used. However, the performance of PCA method deteriorates when the number of training samples increases from 3 to 4. The possible reason is that more variations due to the aging, lighting or facial expression are retained in the projection vectors when more training samples are used. The superiorities of GMTR-based, GPTR-based and GMPTR-based SVM to the other methods are obvious. Particularly, the performance of GPTR-based SVM is very satisfactory. The graph Fig.10 shows that GMPTR-based SVM gives the best performance.

![Graph showing recognition rate vs number of trained images in FERET database](image)

VI. CONCLUSIONS

This paper proposes a new algorithm based on Gabor texture information for face recognition, which is captured by two kinds of strategies: GMTR and GPTR. They are obtained by using Gamma Distribution and Gaussian Distribution to model the magnitude and the phase distributions in the Gabor filtered image, respectively. The main contributions of our algorithm are as follows: (1) Different from the conventional Gabor-based methods that take directly the magnitude as image feature, GMTR explores texture information from the Gabor magnitude, which results in a more compact representation for face images. (2) GPTR is obtained from the Gabor phase that is discarded traditionally. (3) To validate the complementariness of GMTR and GPTR, we also obtain GMPTR by fusing them at feature level and evaluate the performance of it. As shown in our experimental results on Yale, ORL and FERET databases, GMTR-based and GPTR-based SVM are superior to PCA and SVM and can always achieve the best performance in terms of recognition accuracy. It can be concluded that the Gabor phase is not useless, even its discriminative power is comparable to that of the Gabor magnitude, and they provide complementary information for classification.

REFERENCES


