Computing and Informatics, Vol. 31, 2012, 779–803

EFFICIENT 3D FACE RECOGNITION WITH GABOR PATCHED SPECTRAL REGRESSION

Yue Ming, Qiuqi Ruan, Xueqiao Wang

Institute of Information Science
Beijing JiaoTong University
Beijing 100044, P. R. China
e-mail: {06120433, qqruan}@bjtu.edu.cn, silvia.wxq@gmail.com

Communicated by Iveta Zolotová

Abstract. In this paper, we utilize a novel framework for 3D face recognition, called 3D Gabor Patched Spectral Regression (3D GPSR), which can overcome some of the continuing challenges encountered with 2D or 3D facial images. In this active field, some obstacles, like expression variations, pose correction and data noise deteriorate the performance significantly. Our proposed system addresses these problems by first extracting the main facial area to remove irrelevant information corresponding to shoulders and necks. Pose correction is used to minimize the influence of large pose variations and then the normalized depth and gray images can be obtained. Due to better time-frequency characteristics and a distinctive biological background, the Gabor feature is extracted on depth images, known as 3D Gabor faces. Data noise is mainly caused by distorted meshes, varieties of subordinates and misalignment. To solve these problems, we introduce a Patched Spectral Regression strategy, which can make good use of the robustness and efficiency of accurate patched discriminant low-dimension features and minimize the effect of noise term. Computational analysis shows that spectral regression is much faster than the traditional approaches. Our experiments are based on the CASIA and FRGC 3D face databases which contain a huge number of challenging data. Experimental results show that our framework consistently outperforms the other existing methods with the distinctive characteristics of efficiency, robustness and generality.

Keywords: 3D face recognition, Gabor features, patched spectral regression, expression variations

1 INTRODUCTION

Information and communication technologies are gradually entering all aspects of our life and all industry sectors. They open a more convenient lifestyle where people interact with electronic devices embedded in environments. A prime example is the potential of using a person's face instead of intrusive biometric identifiers to not just regulate access to a controlled environment but to adapt the provided services to the preferences and needs of a recognized user.

Automatic recognition of human faces is an active research topic, which has been employed in numerous practical applications such as surveillance, automated screening, authentication and human-computer interaction compared with other biometric features. The face is an easily collectible, universal and non-intrusive biometric signal, which makes it ideal for academic studies in pattern recognition and computer vision. Over the past several decades, most research has focused on two-dimensional gray/intensity images [1]. However, from the application perspective, 2D face recognition still has some unsolved obstacles that seriously affect its wide adoption in uncontrolled environments. Currently, 2D face recognition systems are sensitive to lighting conditions, expressions, viewing angles and various of other factors, such as hair and glasses. A face is a 3D non-rigid object in nature with expression variations, and will change as age increases. A face will be occluded under the influence of hair and glasses. Illumination, image angle and distance also affect facial images. Of these problems, large pose and illumination variations commonly influence the accuracy of 2D face recognition.

With the rapid development and increasing affordability of 3D digital acquisition systems and sensors, 3D facial data provides a promising way to efficiently capture the shape and geometry information in 3D space and has the potential to improve the performance of recognition systems. The possibility of removing the influence of changes in pose, as well as invariability under diverse lighting conditions makes 3D facial recognition a more effective feature within biometrics identification. However, the introduction of new capture modalities brings new challenges for a recognition system. In this paper, we will discuss some of these challenges [14] and show how our framework addresses them.

Accuracy: Accuracy is an important indicator to show which method or framework can provide better performance. Accuracy gain in 3D systems can demonstrate whether the 3D images are more effective than 2D ones.

Efficiency: 3D facial images captured by sensors contain a huge number of data points which occupy large storage spaces at a great computational cost. The compressing techniques for high-dimensional data processing must be considered.

Regulation: Different samples collected by a scanner are represented by different numbers of 3D scattered point clouds. However, there also exist some missing points in the acquired captures because of the loss of the laser signal in different areas, mainly caused by hair, occlusions or some noises. Additionally, some

perturbations or spikes also appear in the borders or special facial areas such as nose and lower chin. All of the above issues make us intend to construct a regular and dense depth images with a fixed number of points. Also, the depth values and facial orientations need to be normalized to an average model on the different facial areas resulting in good accuracy performance.

Large pose and expression variations: The recognition performance of neutral faces has been improved substantially with 3D systems, while accuracy could degrade significantly with large pose and expression variations.

Testing Database: The evaluation platform need to show the university to the different variations. As research in face recognition will gradually focus on recognition under non-ideal conditions, such as uncooperative subjects combined with expression and pose variations, large scale testing databases will become essential.

1.1 Related Work

In view of the shortcomings of the 2D approaches, a large number of 3D and 3D + 2D multi-modal approaches have recently been proposed. Currently, literature on 3D face recognition based on the different representations of data can be classified into 4 types: point cloud representations, depth images, facial surface features or spherical representations [7]. Further overviews of the commonly used algorithms in 3D face recognition can be found in [2, 3].

The first representation, using point clouds, is based on the original shape and spatial information. ICP algorithms, as a matching method, performed an identification of the facial point clouds with 92.1 % rank-one identification on a small subset of FRGC v2 [4]. Alternatively, 3D eigenfaces extracted from the original 3D point clouds extended the measurements of similarity between different individuals [5]. Another choice is to utilize the eigenvalue and singular values of local facial area covariance to maximize the discriminated information between different subjects [6]. One barrier to recognition methods based on 3D point clouds is their high computational complexity and large storage requirements are that are driven by the large amount of data.

Using the basis of 2D face recognition, z-component values on the original 3D facial point clouds can be projected on X-Y plane to generate depth images in two-dimension space representing the distance between the sensor and individuals. Then, 3D face recognition can be converted to the 2D problem and a huge number of 2D methods can be introduced to the 3D systems [7]. Concentrated dimensional reduction was introduced by Faltemier et al. [8]. This method independently matched each segment within a group and then obtained the fusion results. On the other hand, geodesic distances can be calculated for isometric transformation between the selected key points [9]. However, the selection of fiducial points or interesting areas require elaborate manual processes, preventing their wide spread applications.

Facial surface features, such as curvature descriptors [10], have also been proposed for 3D face recognition. However, curvature, as a surface feature, can efficiently describe the global facial surface structure, which unfortunately cannot preserve the local discriminant information, resulting in a low recognition performance. Alternatively, spherical representations, as a novel descriptor, have been used to overcome the influence of illumination [11, 12] and pose variations in the applications of 2D or 3D face recognition [13, 7]. In addition, Kakadiaris et al. [14] obtained the deformation facial images by fitting the changes of face surface based on an annotated face model (AFM). A multistage alignment algorithm and advanced wavelet analysis resulted in robust performance with a 0.1 % False Acceptance Rate (FAR). However, they used a complex multistage approach that required calculating an AFM with a huge computational and spatial cost.

Face recognition combining 3D shape and 2D intensity information is another developing area of research. Wang et al. [15] described facial feature points by using Gabor filter responses in a 2D domain and point signatures in a 3D domain. Chang et al. [16] performed the evaluation of the recognition platform with different combinations of 2D and 3D information and showed that the combination of 2D and 3D information was more effective in characterizing an individual. Mian et al. [17] handled the facial expression problem using a fusion scheme based on spherical face representation (SFR) in 3D space and scale-invariant feature transform (SIFT) in 2D space; and then a modified ICP was used to match the different individuals and combined to achieve the final result. These results showed the potential of appearance-based methods for solving the expression problem in 3D face recognition. Due to the high dimension of the Gabor feature in depth and intensity images, dimensionality reduction with LDA was used to estimate the intrinsic feature vectors and a new hierarchical selection scheme based on Adaboost learning was proposed to construct the effective classifier [18].

1.2 Overview of Approach

In this paper, we address the major challenges of 3D face recognition systems. We have developed a 3D face recognition system based on 3D Gabor Patched Spectral Regression (3D GPSR) as shown in Figure 1. We first extract the main facial region from the irrelevant information and then align all the faces with a frontal reference image by an accurate two-step algorithm. After registration, one input 3D image is translated and rotated to the reference one. The orientation and position of all the input images are normalized, thus denoted as *regulation*. We sample a 140×120 region centered on the nose tip from the aligned 3D data and project to an X-Y plane to obtain the 3D face depth image.

D depth images are built to describe the shape and geometric features of individual faces. Each depth image can be considered as a 2D image where the gray value of each pixel is the depth value of the corresponding points. To overcome the large expressions, Gabor features are extracted from each face. The resulting 3D Gabor faces are stored into a 3$^{\text{rd}}$-order tensor, which is assembled based on

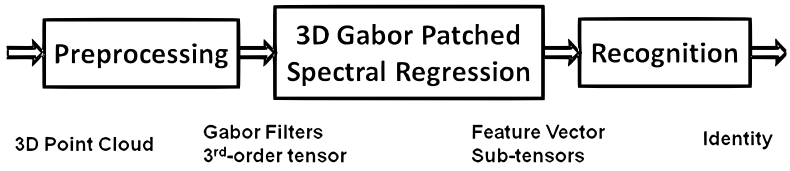


Fig. 1. The framework of the 3D face recognition system

some of the databases' elements, such as identity, illumination, pose and pixels. Then, each tensor can be divided into some local sub-tensors from the vertical and horizonal directions, called patches, to describe the geometric deformation on the different facial areas. Spectral Regression (SR) [20] is used to reduce the dimension of each patch. An intrinsic discriminated feature can be obtained and reduce the influence of noise to some degree. It will make a 3D curved surface, in higher dimension space, unfold into a plane passing through the origin. 3D GPSR can achieve a large *accuracy* gain and meets an *efficiency goal*. Any new images can be represented so that the linear combination of each patch is discriminant features. Thus, one depth image can be projected into the lower dimensional intrinsic space with a low computational and spatial cost. The processing further improves discriminant feature extraction and recognition accuracy, avoiding the calculation of a higher dimensional matrix. It also not only efficiently removes the relationship of row and column, but also reveals the non-linear characteristics of a face. The flow chart for training is shown in Figure 2. In the recognition section, we first obtain the feature vector of all patches in an input 3D image, and then we concatenate all feature vectors into one vector. Finally, recognition is achieved using a Nearest Neighbor (NN) classifier. The flow chart for face recognition is shown in Figure 3. The FRGC [19] and CASIA databases [18], containing a lot of challenging variations, are used to evaluate the performance of our 3D face recognition system.

The main contributions of this paper are as follows:

1. A preprocessing process efficiently preserves facial depth and geometric information and normalizes large pose variations.
2. Gabor features are extracted from 3D facial depth images, which approximately describe simple cell characteristics and handle large expression variations in time and frequency domain. The Gabor transformation not only enhances local properties and details the texture information of facial images, but also extracts local details effectively improving recognition, in addition to being insensitive to expression, pose and illumination variations.
3. A dimensionality reduction process based on SR subspace embedding is used to extract the discriminated feature vectors while tolerating the data noise to some degree, which can be further enhanced for improved recognition performance.

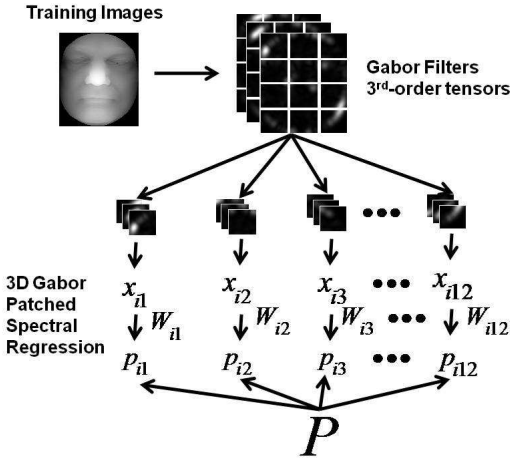


Fig. 2. The flow chart of training scheme

4. Computational analysis shows that 3D GPSR has only linear-time complexity on the dimensional reduction procedure which is a huge improvement compared to the ordinary approaches. Above all, our framework not only preserves local geometric information for 3D facial images but also adds some distinctive biological clues to the local patches. Our method is tolerant of facial expression variations and has higher recognition rates when compared with other state-of-the-art 3D face recognition methods.

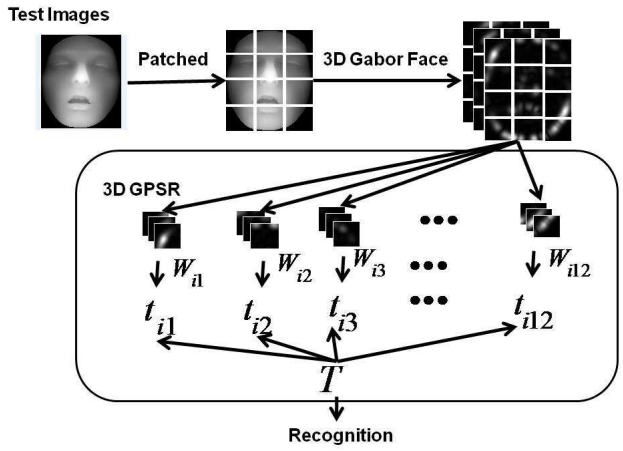


Fig. 3. The flow chart of testing scheme face recognition

The rest of this paper is organized as follows. First, we describe the face preprocessing process that permits extraction and alignment of the facial area of 3D point clouds in Section 2. In Section 3, the Gabor filters are introduced. Then, we present the 3D GPSR algorithm for 3D face recognition based on depth images and analyze the computational complexity compared to the existing methods in Section 4. Experimental results are provided in Section 5. Section 6 concludes the paper.

2 PREPROCESSING OF 3D FACE POINT CLOUDS

In this paper, one face is represented by one 3D scattered point cloud collected by a 3D laser scanner as illustrated in Figure 4 a). In the CASIA 3D face database [18], each point on a 3D facial image is described by 3D spatial coordinates and corresponding RGB color components. In the FRGC database [19], a separated 2D image is used to describe the corresponding color information. The preprocessing scheme is divided into three steps, i.e., the extraction of 3D facial surface, the alignment of the 3D face, and the acquisition of the normalized depth and intensity images. We present these tasks in more detail in the rest of the section. The process can smooth noises and interpolate missing entries in the input data which are immune to rigid transformations and overcome the influence of different resolutions.

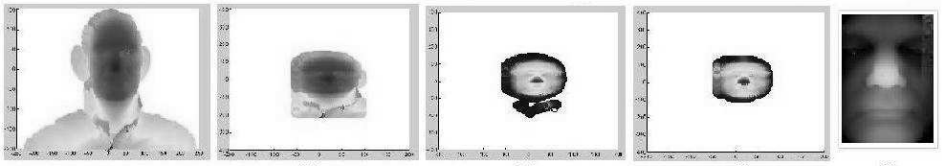
2.1 3D Facial Region Extraction

One output facial scan is typically composed of a 3D point cloud with X, Y, and Z spatial coordinates, where X and Y components form a uniform Euclidean grid and Z provides the corresponding depth value between the sensor and the individuals. Thus, there exist a lot of irrelevant information from the original 3D point cloud data, such as data corresponding to shoulders or hair occlusions, and spikes caused by a laser scanner. The main purpose of face extraction is to remove the non-facial area and retain the facial shape information. We first calculate the sum of each column in the valid point's matrix [7, 19]. By defining two lateral thresholds on the column sums, we obtain the left and right segmentation lines to remove the subject's shoulders as shown in Figure 4 b). Then, to the selected points, we further build the histogram of the z-component values [7]. We set the second-highest frequency bin of the histogram as the threshold and remove the points lower than the threshold as shown in Figure 4 c). The selected nose tip [21] can be considered as the highest point of the point cloud on the reference coordinate system. Then, a sphere centered on the nose tip with 100mm radius is used to extract the main facial area and remove the outlier points as shown in Figure 4 d).

2.2 3D Face Registration

After segmenting the main facial surface region from the original 3D scan, pose correction and alignment are performed. First, the subject's pose is corrected using

Principle Component Analysis (PCA) [28]. We compute the orthogonal eigenvectors, v_1, v_2, v_3, as the three main axes of the point cloud. Due to the properties of the facial surface, v_3 represents the normal direction of the facial surface fitted plane and v_1 corresponds to the vertical dimensional of the facial surface. Taking the nose tip as the origin, v_1 as the y-axis, and v_3 as the z-axis, the original facial points are transformed to the new coordinate system and pose correction is completed. After pose correction, alignment is constructed by an average face model (AFM), by computing the average value of each point across all training faces. The AFM is used as a reference face model, and all 3D face images are further registered by ICP [14]. Then, the registration converges monotonically to a local minimum to avoid the unwanted influences of rotation and translation. Finally, there is a refinement step [22] to minimize the z-component distance. The preprocessing procedure effectively re-samples the data and removes all irrelevant information from occlusions and spikes.



a) b) c) d) e)

Fig. 4. Main steps in facial region preprocessing. a) Original 3D point cloud; b) after lateral thresholding; c) after depth histogram thresholding; d) after facial region extracting; e) normalized depth image

2.3 Normalized Depth and Intensity Image Acquisition

We have cropped the original 3D facial images centered at the detected nose tip as discussed above. Then, the Z component values of the cropped region are projected on the X-Y plane to obtain a depth image with 140×120 resolution and the corresponding intensity image is normalized to [0, 255] with 140×120 pixels as shown in Figure 4 e). For the rough surfaces on the depth images, the mean filter is computed for the 5×5 sub-window around it for smoothing. For some special cases, if the difference between the pixel and the mean is larger than a preset threshold, the pixel is treated as an outlier and is removed. We fill the small holes by linear interpolation of the neighboring pixels [18]. The intensity images are processed using a similar method and histogram equalization is used to reduce the influence of illumination variation. Intensity images can also be obtained from 2D color images in the FRGC database.

3 GABOR WAVELET

Due to the consistent forms of kernel Gaussian functions in time and frequency domains, Gabor Wavelet transformation has the ability to analyze signal local properties in different areas and shows excellent locality in both time and frequency domains. Gaussian functions can simultaneously capture spatial frequency scalar and local structural characteristics of image local areas in multiple directions. Gabor functions can approximately describe simple cell characteristics of the visual cortex of vertebrate brains. Image representation in a visual cortex exists as time and spatial-frequency components and one image can be decomposed into local symmetric and anti-symmetric basis functions, which completely anastomoses Gabor functions. Due to better time-frequency characteristics and a distinctive biological background, Gabor features of facial images are insensitive to geometric deformation and noise. Gabor functions bring many unique advantages when dealing with pose and expression variations. Spatial locality and orientation selectivity are the desirable characteristics of Gabor wavelets, which are optimally localized in the local information in the frequency domains. The Gabor filters can be defined as follows [32]:

$$\psi_{\mu,v}(z) = \frac{\|k_{\mu,v}\|^2}{\sigma^2} e^{\left(-\|k_{\mu,v}\|^2 \|z\|^2 / 2\sigma^2\right)} \left[e^{ik_{\mu,v}z} - e^{-\sigma^2/2}\right] \tag{1}$$

where μ and v denote the parameters of orientation and scale of the Gabor filters, $z = (x, y)$, $\| \bullet \|$ is the norm operator, and the wave vector $k_{\mu,v}$ is defined as follows:

$$k_{\mu,v} = k_v e^{i\phi_\mu} \tag{2}$$

where $k_v = k_{\max}/f^v$ and $\phi_\mu = \pi\mu/8$. $k_{\max}$ is the maximum frequency, and f is the spacing factor between different scales of the filters in the frequency domain. In most face recognition cases, these parameters for a Gabor wavelet $\sigma = 2\pi$, $k_{\max} = \pi/2$, $f = \sqrt{2}$ are used by researchers [32, 33].

The Gabor wavelets can be generated from the scaling and rotation of the mother wavelet via the wave vector $k_{\mu,v}$ as demonstrated in Equation (1). Thus, they are all self-similar in shape. Zhong et al. [35] show the Fourier spectrum in the frequency domain of a 3D facial image. They point out that most energy is distributed in the low frequency area corresponding to the global information of a face's image. However, middle and high frequency areas can reflect more detailed discriminative information and benefit recognition. Therefore, we need more middle and high bands to obtain efficient discriminant features.

The Gabor wavelet representation of a 3D facial depth image is a convolution of the image with a family of Gabor filters as defined by (1), called a 3D Gabor face. Let $D(x, y)$ be the depth level distribution of a 3D face, the convolution of image $D(x, y)$ and a Gabor filter $\psi_{\mu,v}$ can be denoted as follows:

$$G_{\mu,v}(x, y) = D(x, y) * \psi_{\mu,v}(x, y) \tag{3}$$

where $*$ denotes the convolution operator, and $G_{\mu,v}$ is the convolution result corresponding to the Gabor filter at orientation μ and scale v. In most applications [32, 33], the researchers are willing to use Gabor wavelets with five different scales, $v \in \{0, \cdots, 4\}$ and eight orientations, $\mu \in \{0, \cdots, 7\}$. It is clear that 3D Gabor faces contain more detailed shape and geometry information than the original 3D depth image.

The Gabor filters can efficiently decompose the original images into different scales and orientations to exhibit desirable local details.

Gabor wavelet transforms are invariant to rotation, scale and translations in feature extraction [34]. At the same time, Gabor functions also enhance image low-level features like edges, peaks, valleys, and ridges, which are equivalent to enhancing key facial element information such as the nose, eyes, and mouth plus local characteristics like dimples, melanotic nevus and scars. They not only preserve global facial information but also enhance local characteristics. When the pose, expression and position of a face change, local changes are less than global changes. This results in a very robust face representation. However, 40 ($= 5 \times 8$) images generated from the Gabor filters require a huge number of storage space and computational cost. Since the different local scales and orientation features reflect the different desirable characteristics of the images, in our case, we only choose those optimal discriminate features for recognition. Therefore, $G_{3,1}$ with three orientations and one scale comprises the 3D Gabor face of the original 3D depth image in our framework, which can capture more discriminative information and enhance low-level image features compared with the other methods. Each 3D depth face corresponds to three 3D Gabor faces, which can be stored as a 3$^{\text{rd}}$-order tensor. We also use it in experiments in Section 5.

4 3D GABOR PATCHED SPECTRAL REGRESSION

Each point cloud has been transformed into a regular 3D depth image. Pose correction is used to normalize the large pose variations of different individuals and Gabor features can represent the expression variations of the different facial regions. Beside these two aspects, data noise is another obstacle to accuracy improvement. Zhong et al. [35] summarize the sources of the data noises into three categories: from the 3D acquisition system, from the occlusions and from mis-registration. Obviously, all 3D faces have a similar geometric shape and most of the registration noises and occlusions can be reduced by the preprocessing. However, there are still some small noises widely distributed in a huge facial space. In our framework, local sub-tensor division is adopted to describe the different facial regions and Spectral Regression is used to minimize the effect of noises. We divide 3D Gabor faces into many local sub-tensors as illustrated in Figure 2. The strategy can efficiently reduce the level of noises based on the local areas of the 3D face and extract the low-dimensional discriminative features, with the lower computational complexity simultaneously. We denote this method as 3D Gabor Patched Spectral Regression (3D GPSR). This

approach preserves the local neighbor structure of a facial manifold and increases global discriminant information. Tensors, as constraint conditions, nicely retain the ability of local preservation and at the same time increase separability, which overcomes expression and pose variations to some extent. Next, we will give a detailed discussion about 3D Spectral Regression and its computational analysis.

4.1 3D Spectral Regression

Here, we consider a 3D depth image in comparison with one 2D gray image. Each point of the 3D depth image corresponds to a pixel of a gray image. Z-component values of the points can be treated as the pixel value in 2D images [5]. As a result, we can process the depth image as we would a 2D image. Each depth image can be unfolded into one dimension vector.

Spectral Regression (SR) was first applied to discriminant feature extraction by Cai et al. [20]. In our case, given m face depth images as explanatory variables $X \subset R^{n\times m}$, together with their corresponding response values $Y \subset R^{d\times m}$, the goal of regression is to estimate the regression matrix A. To better reflect the relationship of explanatory and response variables among different samples, it is instructive to construct a graph embedding framework to generate the response matrix based on Laplacianface [23]. The graph embedding method is utilized to provide the dataset mapping as the graph vertices for feature extraction. The process can preserve local embedded structure and obtain the best subspace that reflects the basic manifold structure of the images, thus decreasing the influence of expression and pose variations. It is more suitable for analyzing and discriminating facial data than other algorithms.

First, we focus on the generation of the response variables Y. Here, suppose we have c classes of data and the k-th class has m_k samples, where $m_1 + \cdots + m_c = m$. We denote G is an undirected weighted graph, including m vertices. The i-th node corresponds to the sample label c_i. In our experiment, we utilize the weighted graph [20] matrix $W = [W_{ij}]_{m\times m}$ as follows:

$$W_{ij} = \begin{cases} 0, & \text{if there is no edge between } i \text{ and } j \\ 1/l_k, & \text{if } x_i \text{ and } x_j \text{ both belong to the } k^{\text{th}} \text{ class} \\ \delta \cdot s(i,j), & \text{otherwise} \end{cases} \tag{4}$$

where $0 < \delta \leq 1$ is a parameter to balance the weights between supervised information and unsupervised neighbor information. For the same class, we only use label information as the weights and for the neighboring but different classes, we obtain the weights by Heat function $s(i,j)$. In the supervised training mode, we calculate the eigenvector of the Laplacian operator, search for optimal local embedding and avoid removing the relationship of row and column, keeping us from extracting identification features and increasing the computational complexity.

$$s(i,j) = \mathrm{e}^{-\frac{\left\|x_i - x_j\right\|^2}{2\sigma^2}}, \sigma \in R \tag{5}$$

Then, we denote the response variables $Y = [y_1, y_2, \ldots, y_m]^T$ as the map from the graph to the real line. Regression analysis aims at seeking a regression matrix A to minimize a "distance", which indicates that when samples i and j are close, then y_i and y_j are close as well [20]. Based on graph embedding, the optimal Y can be obtained by minimizing the objective function as follows

$$\min_Y tr(YLY^T), \text{ s.t.} YTY^T = I, \tag{6}$$

where $L = T - W$ is the graph Laplacian [20, 23] and T is a diagonal matrix with diagonal elements whose entries are column sum of W, $t_{ii} = \sum_j W_{ij}$.

In order to remove an arbitrary scaling factor, we introduce the constraint $YTY^T = I$ in the graph embedding. It is easy to see that the above optimization problem has the following equivalent variation:

$$\max_Y tr(YWY^T), \text{ s.t.} YTY^T = I. \tag{7}$$

The optimal d-column matrix Y corresponds to the generalized eigenvectors associated with the d largest eigenvalues in $WY = \lambda TY$.

The simplest linear model for regression can be denoted as a linear combination of the explanatory variables to their corresponding response variables, i.e., $Y = AX$. However, in our practical applications, we cannot obtain sufficient clean data. In order to overcome the influence of the noise to 3D face recognition, the regression equation has been extended to a small noise term E so that

$$Y = f(X) = AX + E, \tag{8}$$

where E can be treated as a zero mean Gaussian random variable. In order to control over-fitting, a ridge regularizer is incorporated into the framework [20]. Then, the objective function can be demonstrated as follows:

$$\min_Y \|Y - AX\|_F^2 + \alpha \|A\|_F^2 . \tag{9}$$

With some simple algebraic formulations [24, 25], the regularized least square problem can be solved as follows,

$$A = (XX^T + \alpha I)^{-1} XY, \tag{10}$$

where I is a $n \times n$ identity matrix.

Finally, the discriminative feature vectors can be extracted by SR embedding into a d low-dimensional subspace by

$$X \to Z = AX. \tag{11}$$

3D face samples were mapped to a lower feature space from a higher observation space via non-linear mapping bringing out the intrinsic lower dimensional structure

hidden in the higher observational data. This has many positive benefits, including compressing data, thereby reducing storage requirements, removing unnecessary noise, extracting effective features for recognition and realizing visualization of the higher dimensional data.

4.2 Computational Analysis

In this subsection, we will compare the computational complexity among the traditional methods and SR. For our case of 3D face recognition based on depth images, we consider the number of the features (n) is much larger than the number of samples (m).

For PCA [20, 28], the algebraic steps can be divided into two steps, including calculating the eigenvectors of covariance matrix XX^T and estimating the projective functions. First, the calculation of XX^T requires $\frac{1}{2}m^2n$ complexity and computing its eigenvectors needs $\frac{9}{2}m^3$ complexity. SVD decomposition is used to estimate the projective function, requiring m^2n complexity [36]. Thus, the total time complexity of PCA is $\frac{3}{2}m^2n + \frac{9}{2}m^3$.

Compared to PCA, LPP [20, 23] needs two extra steps. The first one is the construction of the p-nearest neighbor graph with $\frac{1}{2}m^2n + 2mn + m^2 \log m$ complexity. The second is sorting the pairwise distances, requiring $\frac{1}{2}m^2n + 2mn + m^2 \log m$ complexity. Plus the eigenvector computation and SVD decomposition, the total time complexity of LPP is $\frac{5}{2}m^2n + 4mn + 2m^2 \log m + \frac{9}{2}m^3$. Orthogonal LPP (OLPP) [29] has a similar analysis.

Spectral regression computation involves two steps: response generation and regularized least squares. The complexity of the first step is mainly based on the Gram-Schmidt method with $mc^2 - \frac{1}{3}c^3$ complexity [20]. For the second step, the problem can be efficiently solved by the iterative LSQR [25]. In our case, the number of class (c) is much less than the number of samples (m). Usually, we set $s = 5$. With s iterations, the complexity cost is $s(2mn + 3m + 5n)$ for each class. Overall, the total computational complexity of SR is $cs(2mn + 3m + 5n)$.

We summarize the complexity analysis of SR in Table 1, together with PCA and LPP [20]. The computational advantage of SR over traditional PCA and LPP, especially for large scale high dimensional data (with large m and n) is clear. SR has only linear-time complexity compared to the cubic-time complexity in other methods.

PCA	$\frac{3}{2}m^2n + \frac{9}{2}m^3$
LPP	$\frac{5}{2}m^2n + 4mn + 2m^2 \log m + \frac{9}{2}m^3$
SR	$cs(2mn + 3m + 5n)$

Table 1. Computational complexity analysis

4.3 Local Patches Strategy

A 3D facial image contains many basic shape elements and geometric information. If we use 3D SR on the whole images, we cannot accurately describe the main facial areas which contain more discriminative information, such as eyes, nose and mouth, also losing some spatial information. The images are essentially in the form of tensors and the row and column information cannot completely substitute for 1D vectors. Computational complexity and storage cost lead to dimension disaster and over-fitting. Simple vectorization usually destroys some relevant inherent data, hides redundant information, and preserves higher-order dependencies existing in the data. As a result, we cannot obtain more compact representations of the original data. To solve these problems, 3D Gabor faces, stored as the 3^{rd}-order tensor, are divided into some small local sub-tensors, called patches. For each patch, SR is applied to dimensional reduction to obtain the intrinsic feature vectors and then concatenate the features of different patches into one vector to represent the whole image, called 3D GPSR. In this way, we not only reflect the scale and orientation information in a depth image, but also emphasize the characteristics of the different facial parts. To better improve the recognition accuracy, we also review some literature on how the size of local patches will influence the recognition results [35]. Overall, in regards to reducing the data noises and preserving spatial information, the best strategy was found to be dividing the vertical direction into four parts and the horizontal one into three parts.

In our framework, the process mainly consists of two sections. In the learning section, Gabor transformation is used on each facial depth images to obtain a 3^{rd}-tensor face as discussed in Section 3. Then, twelve local patches can be obtained by dividing these 3^{rd}-tensors. Given m training images, for $k^{\text{th}}(1 \leq k \leq 12)$ patch, we unfold the local 3^{rd}-tensors into the vectors and store them as a matrix $X_k = [x_{k1}, x_{k2} \dots x_{km}]$. Then, 3D GPSR embedding regression matrix estimation is applied to this matrix. Finally, the low-dimensional feature vectors $P_k = [p_{k1}, p_{k2} \dots p_{k_m}](p_{ki} \in R^{s \times m}, s < n)$ can be calculated as follows:

$$P = \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_k \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1m} \\ z_{21} & z_{22} & \cdots & z_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ z_{k1} & z_{k1} & \cdots & z_{km} \end{bmatrix} = [Z_1, Z_2 \cdots Z_m]. \tag{12}$$

In the recognition section, we also divide the testing images into many local patches as discussed above. Then, we map them into the 3D GPSR subspace via the regression matrix obtained by the learning process. After 3D GPSR embedding is completed, the feature vectors of test images F can be obtained.

For the purpose of face recognition, we need to seek a minimizing distance between training and testing images. The label of training samples with the minimized

distance is as the recognition result. In our experiments, the Euclidean Distance is utilized to calculate the distance as follows:

$$D(P_i, T) = \|P_i - F\| . \tag{13}$$

While the simplest Nearest Neighbor (NN) classifier is used in our framework for 3D face recognition, the result is the output with the shortest distance between the feature vector of the trained image and the feature vector of the tested image in the embedding subspace.

The GPSR framework has many advantages. First of all, since 3D GPSR unfolds an image into a high-dimensional vector in multiple scales and orientation space, the matrix may encounter some difficulties, for example singular or over-fitting. By detaching the image into many patches, we efficiently avoid the problems. Furthermore, by Gabor filters, we can add the spatial and frequency information to each patch, which can extract the shape features of different facial parts. Finally, based on SR, we can preserve the local information of the corresponding patches and obtain the compact feature vectors. All of these benefits result in a better recognition rate than other existing methods. The detailed analysis will be presented in Section 5.

5 EXPERIMENTS

In this section, we evaluate the performance of our proposed framework and compare it with commonly used methods. The experiments were based on the CASIA [18] and FRGC [19] 3D face databases. The first one was collected in Fast Mode during August and September 2004 with a non-contact 3D digitizer, Minolta VIVID 910. This database captured images of 123 subjects, with each individual having about 37 or 38 3D point clouds. The total number of range images is 4 625, including the variations in expressions, poses, occlusions and illumination, for each one also combined with other variations. This database contains a huge number of images with complex variations and the selected images can be used to evaluate any algorithms. The FRGC database is more recent. It is a common evaluation platform for 3D face recognition, organized by NIST [19]. Based on the time of acquisition, the database can be divided into Spring 2003, Fall 2003 and Spring 2004, with over 4 900 range images from 557 people. In each section, one range image and its corresponding color image were captured by a Minolta VIVID 900/910 series sensor in the fine mode every one or two weeks. The images have 640×480 resolution with a large number of challenging variations, such as expressions, occlusions and spikes. The subject pool is approximately 57 % males and 43 % females ranging in age from 18 to over 28 years [14, 19]. For each database, the data set was divided into two subsets in our experiments: the training set and the testing set. Based on these two databases, the experiments demonstrated the efficiency of our framework on the challenging facial variations and the sensor information.

In this section, we evaluate the performance of the proposed framework in both recognition and verification scenarios. Following the procedure described in Sec-

tion 2, we preprocessed all the point clouds in the CASIA 3D face database and the FRGC database to obtain the normalized depth images and intensity images. From these normalized images, we extracted different discriminative features to represent an individual.

All of our experiments have been implemented in Matlab 7.5 and run on a P4 2.1 GHz Windows XP machines with 2GB memory.

5.1 Recognition Results

First, we show the experimental results for the recognition scenario. In these experiments, we use the CASIA database. We demonstrate the outstanding performance of our proposed framework by comparing experiments involving different algorithms, different facial images and different 3D face databases.

5.1.1 Experiments with Different Algorithms

In these experiments, we made detailed comparisons with some existing methods for 3D face recognition to show the relative performance of the proposed algorithm. The algorithms considered include surface curvature (SC) based on point cloud representations [10], point signature (PS) with combined 2D and 3D information [26], COSMOS shape index (CO) with facial surface features [27], Annotated Deformable Model (ADM) [14], sparse spherical representations (SSR) [7] and our proposed framework for 3D face recognition. The different features were extracted for each sample to identify an individual. Table 2 shows the recognition accuracy in the different test sets as described below.

The test set was further divided into seven subsets to evaluate the performance of different features with pose and expression variations [18]. For each subset, we randomly selected the images as test sets and the recognition results indicate the average accuracy across the experiments with different selected images. The number of test images is in brackets.

1. IV (400 images): Illumination variations under neutral expression and front pose, including up, down, left and right lighting.
2. EV (500 images): Expression variations under office light and front pose, including smile, laugh, anger, surprise and eyes closed.
3. EVR (500 images): Expression variations under lighting from the right side and front pose, including smile, laugh, anger, surprise and closed eyes.
4. SPV (700 images): Small pose variations under office light and neutral expression, including 20–30 degrees views of right, left, up, down, tilt left and tilt right.
5. LPV (200 images): Large pose variations under office light and neutral expression, including 50–60 degrees views of right and left.

6. SPVS (700 images): Small pose variations under smile and office light, including 20–30 degrees views of right, left, up, down, tilt right and tilt left.
7. LPVS (200 images): Large pose variations under smile and office light, including 50–60 degrees views of right and left.

Test Sets	SC	PS	CO	ADM	SSR	3D GPSR
IV	59.98 %	53.67 %	48.21 %	98.33 %	96.47 %	98.37 %
EV	52.93 %	49.56 %	45.74 %	95.73 %	93.08 %	95.18 %
EVR	54.53 %	48.12 %	46.02 %	96.14 %	94.23 %	96.35 %
SPV	55.34 %	50.55 %	45.27 %	93.97 %	92.83 %	94.02 %
LPV	39.53 %	33.56 %	32.64 %	56.85 %	60.99 %	75.35 %
SPVS	51.39 %	47.05 %	43.76 %	90.38 %	82.15 %	85.37 %
LPVS	38.43 %	30.18 %	31.79 %	52.14 %	58.43 %	70.16 %

Table 2. Recognition rate using different features

As shown in Table 2, we can draw the following conclusions. The highest recognition rate was up to 98.37 %, which was obtained by our framework in the IV test set. The last three methods evaluated have significantly better performances than the first three ones. As the facial surface descriptor, curvature can efficiently describe the global facial surface structure. However, it cannot preserve the local discriminant information, resulting in a lower recognition performance. PS combined 2D and 3D information residing in the high computation complexity that can achieve good results in the small testing size. For a larger data size, it cannot give satisfactory results. CO, as facial surface feature, needs to select the fiducial points or interesting areas which are often obtained manually and prevent its wide application. Moreover, for multiple classes, its performance degrades significantly. With expression variations, our framework and ADM [14] performed more robust than other methods. Kakadiaris et al. [14] used wavelets and Pyramid transformation to describe the scale information of a human face in the presence of expression variations and reported the performance was 97.0 % verification as a 0.1 % False Acceptance Rate (FAR) in the Face Recognition Grand Challenge. Relative to Kakadiaris's algorithm, our method extracted the Gabor features to reflect not only the scale information but also the orientation changes. The results showed that our framework can significantly reduce the influence of the expressions and illumination. On the other hand, AFM [14] requires a huge computation cost, which took about 23 seconds to obtain the recognition results with our PC configurations. Our method can obtain the compact low-dimensional features based on SR and efficiently save the computational complexity.

Facial pose variation is another major factor affecting the recognition performance. For large pose variations, SSR [7] and our method gave the best results. SSR [7] utilized the sparse spherical representation to construct a dictionary for facial pose description and reduced the dimensions based on LDA resulting in a better performance in the large challenging databases. However, the construction is

a time-consuming process, requiring a large-scale processor with about 18 seconds for a scanner. The preprocessing of our framework can efficiently correct the pose variations and its implementation is simpler. Noise and the size of the test set degrade the performance significantly. Reducing dimensions based on 3D GPSR extracted the discriminant feature vectors and efficiently decreased the level of noise. Thus, our approach is robust with respect to occlusions, the age of the subject, image backgrounds, distances, expressions and poses, overcoming the influence of these external factors quite well.

In our system, unlike most of the existing algorithms, we use a simpler and efficient method, taking less than 5 seconds on our PC with 2.1 GHz CPU and 2 G RAM. A face is a 3D non-rigid object in nature with expression variations, and will change as age increases. A face will be occluded under the influence of hair and glasses. Illumination, image angles and distances also affect facial images. In our system, the processing further improved discriminant feature extraction and recognition accuracy, avoiding the calculation of a higher dimensional matrix. It also not only efficiently removes the relationship of row and column, but also reveals the non-linear characteristics of a face. The Gabor transformation not only enhances local properties and details the shape information of facial images, but also extracts local details which effectively improves recognition, in addition to being insensitive to expression, pose and illumination variations. 3D face samples were mapped to a low-dimensional feature space from a higher observation space via mapping, bringing out intrinsic lower dimensional structure hidden in higher observational data. This has many positive benefits, including compressing data which reduces storage requirements, removes unnecessary noise, and extracts effective features for recognition. Therefore, 3D GPSR consistently outperforms other existing algorithms.

5.1.2 Comparison with 2D Face Recognition Technologies

In this experiment, we made detailed comparisons between 3D face recognition based on 3D depth images and 2D face recognition based on 2D intensity images to demonstrate the performance of the proposed algorithm. We chose 1 845 facial images of 123 subjects in the CASIA database, where each subject had 15 scans. For each individual, the first five face images of each person were selected for training and the rest of the images were used for testing. The results shown in Table 3 allow us to draw the following conclusions:

1. Depth information is more robust than the gray images. Shape variation and facial surface structure are the most important indicators for discriminating an individual and the feature vectors reflecting the shape variations can improve the recognition rate distinctively.
2. Comparing the results generated from our framework and other traditional methods, we can see that our framework has consistently higher accuracy in both 2D and 3D facial images. This shows that our method can efficiently eliminate the impact of expression variations.

3. GPSR used the least time in all cases as shown in Table 3, and it is possible to save more computational complexity and storage space with high accuracy.

Due to the limitations of face recognition, 2D face recognition still encounters many unsolved difficulties to develop a robust face recognition system, including poses, expressions, illumination, aging and subordinates. From the experimental results, we can intuitively see that the actual 3D information has no relation to view and illumination. In conclusion, compared to 2D face recognition, 3D face recognition has higher accuracy and can overcome existing problems with 2D face recognition.

Method	3D Depth Faces			2D Gray Faces		
	Accuracy	Time(s)	Dim	Accuracy	Time(s)	Dim
PCA	87.32 %	81	82	61.3 %	77	102
LPP	91.87 %	85	115	63.9 %	81	118
OLPP	96.99 %	102	49	75.77 %	98	46
GPSR	98.78 %	51	102	80.89 %	49	108

Table 3. Recognition rate using different reducing dimension schemes and data resources

5.1.3 Experiments with Different 3D Face Databases

In these experiments, 3D face recognition algorithms were evaluated on a more challenging database, called FRGC [19]. We show the generality of our proposed framework compared with some other standard state-of-the-art solutions in both recognition and verification scenarios of 100 subjects, for each subject with ten scans. We preprocess the original 3D point clouds as discussed in Section 2 for each scan. The number of vertices after preprocessing typically varied between 30 000 and 40 000 and was uniformly sampled to the depth images with 140×120 resolution.

We defined several test configurations for our experimental evaluation. First, we compared our method with subspace methods which included 3D GPSR, 3D eigenfaces (3D PCA) [28], 3D Laplacianfaces (3D LPP) [23], and 3D Orthogonal Laplacianfaces (3D OLPP) [29]. For each subject, the first 2, 3, 4, 5 face images of each person were selected for training, and the rest was used for testing. For each given set of training samples, we chose the best result of the dimension parameters for each method. Table 4 shows the results.

We can see from the table that the new algorithm presented in this paper achieved the highest recognition rate among the four methods. Extraordinarily, when five images were used for training, the new method achieved the highest recognition accuracy rate of 97.99 %.

Furthermore, compared with other algorithms, we also find that the training section in 3D GPSR does not need many training images to achieve better performance. Increasing the training sets significantly improves the performance. This is

Method	Accurate recognition rate			
	2	3	4	5
3D PCA	84.12 %	86.28 %	86.67 %	87.18 %
3D LPP	85.38 %	87.04 %	88.17 %	89.39 %
3D OLPP	93.99 %	94.43 %	95.83 %	96.72 %
3D GPSR	93.13 %	95.43 %	96.67 %	97.99 %

Table 4. Recognition rate using FRGC 3D face database

because we applied the tensor descriptor and added the scale and orientation information into the training sets. We only use a small number of training images and computational time to obtain good recognition results, which means that 3D GPSR can efficiently describe 3D face characteristics.

Second, we made a comparison between 3D GPSR and two appearance-based methods, which contain a local binary pattern (LBP) [30] and learned visual codebook (LVC) [31]. LBP is an efficient texture descriptor and has been successfully used for face recognition. LVC is a method which choses K-means clustering to learn basic facial elements. K-fold cross validation was used on three methods. Because of the large face database, we did two groups of experiments. In the first, 10 images of each person consisting of 5 neutral images and 5 images with different expressions were divided into 10 groups and used for K-fold cross validation. In the second, all of the expression images were divided into 10 groups for K-fold cross validation. The procedure was repeated until all the images per subject have been used once as testing images. The final recognition rate was computed by averaging all the trails. The results are presented in Table 5. It shows that our method consistently outperforms other methods.

The size of the experiment database	group 1	group 2
LBP	83.32 %	85.29 %
LVC	91.87 %	92.35 %
3D GPSR	96.98 %	97.13 %

group 1: 5 images with neutral expression and 5 image with different expression
group 2: 10 images with different expression

Table 5. Comparison of recognition rates using K-fold cross validation

5.2 Verification Results

In this section, we consider the verification scenario. First, the test subject claimed an identity and the system matched the test identity with the corresponding claimed identity in the gallery. A confidence score was used to justify whether the test identity is a client or an impostor. By setting a threshold value, we can make a decision whether to accept or reject the test sample. We evaluated the verification performance of 3D face verification in terms of receive operating characteristic (ROC)

curves, which show the fluctuation of the false reject rate (FRR) versus the false accept rate (FAR) across a large range of thresholds. For the computation of the ROC curve we consider every possible subject and test sample pair.

We also illustrated the ROC curves on the three methods and used the dimensions that yield the best performance, which correspond to an optimal subspace in 3D GPSR. Six images with one neutral expression and different expressions were used for training while the rest were used for testing. Figure 5 shows our method had the lowest FRR and FAR. Observe again that 3D GPSR consistently outperforms LBP and LVC in all configurations and remains the best performer.

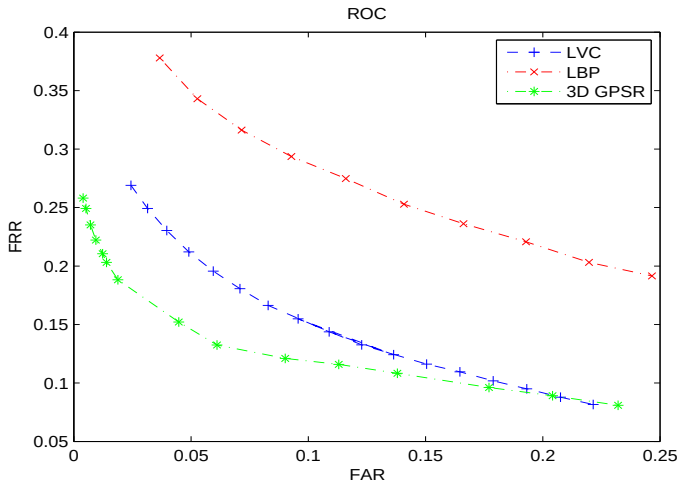


Fig. 5. ROC on FRGC

5.3 Future Work

In this paper, we connect three procedures directly into a 3D face recognition framework: preprocessing the original point clouds, extracting Gabor features and reducing dimensions. Intuitively, it is not the best way to fuse them. In the future, we will carry out further research and develop a more effective algorithm. Many areas need to be improved. Better preprocessing methods will reduce small registration errors retained from the current method. Innovative shape variation representation needs to be developed to better describe expression variations and encode relationships in neighboring points. More thorough testing using a larger database and practical applications with more pose and expression variations will give even better confirmation of test results.

6 CONCLUSION

In this paper, we propose a new 3D face recognition framework called 3D Gabor Patched Spectral Regression (3D GPSR). In our method we have introduced a multistage preprocess for extraction and registration of facial shape information in the original 3D point clouds. Our method handles variations in facial expressions and poses based on Gabor filters, which is invariant to rotations, scales and translations. In addition, our framework was developed from a graph embedding viewpoint of dimensionality reduction algorithms, which is directly related to the discriminating power. It combines spectral regression and local patches strategy to provide a more efficient and effective approach to the regularized subspace learning problem. Specifically, 3D GPSR only needs to solve a set of regularized least squares problems and efficiently minimize the influence of noise. Complexity shows huge savings in both time and memory. Finally, experimental results show that the performance of 3D GPSR is better than other popular approaches.

Acknowledgements

This work is supported by National Natural Science Foundation (60973060), the Research Fund for the Doctoral Program (20080004001) and Beijing Program (YB2008-1000401), and the Fundamental Research Funds for the Central Universities (2009-YJS025). A part of the research in this paper uses the CASIA 3D Face Database collected by The institute of Automatic, Chinese Academy of Science. The authors would like to thank Prof. Patrick Flynn for sharing with us the UND (University of Notre Dame) Biometrics database. The authors would also like to thank the Associate Editor and the anonymous reviewers for their helpful comments.

REFERENCES

[1] Zhao, W.—Chellappa,R.—Rosenfeld, A.—Phillips, P.: Face Recognition: A Literature Survey. CVL Technical Report, University of Maryland, Baltimore, 2000.

[2] Bowyer, K. W.—Chang, K.—Flynn, P.: A Survey of Approaches and Challenges in 3D and Multi-Modal 3D + 2D Face Recognition. Computer Vision and Image Understanding, Vol. 101, 2006, No. 1, pp. 1–15.

[3] Akaruan, L.—Gokberk, B.—Salah, A. A.: 3D Face Recognition for Biometric Applications. In Proceedings of the European Signal Processing Conference, Antalaya, September 2005.

[4] Lu, X.—Jain, A.: Deformation Modeling for Robust 3D Face Matching. In Proceedings of IEEE Computer Vision and Pattern Recognition, New York, June 2006, pp. 1391–1398.

[5] Xu, C.—Wang, Y.—Tan, T.—Quan, L.: A New Attempt to Face Recognition using 3D Eigenfaces. In Proceedings of the Asian Conference on Computer Vision, Jeju Island, January 2004, pp. 884–889.

[6] Al-Osaimi, F. R.—Bennamoun, M.—Mian, A.: Integration of Local and Global Geometrical Cues for 3D Face Recognition. Patter Recognition, Vol. 41, 2008, No. 3, pp. 1030–1040.

[7] Llonch, R. S.—Kokiopoulou, E.—Tosic, I.—Frossard, P.: 3D Face Recognition with Sparse Spherical Representations. Pattern Recognition, Vol. 43, 2010, No. 3, pp. 824–834.

[8] Faltemier, T. C.—Bowyer, K. W.—Flynn, P. J.: A Region Ensemble for 3D Face Recognition. IEEE Transactions on Information Forensics and Security, Vol. 3, 2008, No. 1, pp. 62–73.

[9] Gupta, S.—Aggarwal, J. K.—Markey, M. K.—Bovik, A. C.: 3D Face Recognition Founded on the Structural Diversity of Human Faces. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, June 2007.

[10] Gordon, G.: Face Recognition Based on Depth and Curvature Features. In SPIE Proceedings on Geometric Methods in Computer Vision, San Diego, July 1991, pp. 234–247.

[11] Wang, H.—Wei, H.—Wang, Y.: Face Representation under Different Illumination Conditions. In Proceedings of IEEE International Conference on Multimedia and Expo, Baltimore, July 2003, pp. 285–288.

[12] Ramamoorthi, R.: Analytic PCA Construction for Theoretical Analysis of Lighting Variability in Images of a Lambertian Object. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, 2002, No. 10, pp. 1322–1333.

[13] Yue, Z.—Zhao, W.—Chellappa, R.: Pose-Encoded Spherical Harmonics for Face Recognition and Synthesis Using a Single Image. EURASIP Journal on Advances in Signal Processing, 2008, No. 65, pp. 1–18.

[14] Kakadiaris, I. A.—Passalis, G.—Toderici, G.—Murtuza, N.—Lu, Y.—Karampatziakis, N.—Theoharis, T.: 3D Face Recognition in the Presence of Facial Expressions: An Annotated Deformable Model Approach. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 29, 2007, No. 4, pp. 640–649.

[15] Wang, Y.—Chua, C.—Ho, Y.: Facial Feature Detection and Face Recognition from 2D and 3D Images. Pattern Recognition Letters, Vol. 23, 2002, No. 10, pp. 1191–1202.

[16] Chang, K. I.—Bowyer, K. W.—Flynn, P. J.: An Evaluation of Multi-Model 2D+3D Biometrics. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 28, 2006, No. 4, pp. 1695–1700.

[17] Mian, A. S.—Bennamoun, M.—Owens, R.: An Efficient Multimodal 2D-3D Hybrid Approach to Automatic Face Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 29, 2007, No. 11, pp. 1927–1943.

[18] Xu, C.—Li, S.—Tan, T.—Quan, L.: Automatic 3D Face Recognition from Depth and Intensity Gabor Features. Pattern Recognition, Vol. 42, 2009, No. 9, pp. 1895–1905.

[19] Phillips, P. J.—Flynn, P. J.—Scuggs, T.—Bowyer, K. W.: Overview of the Face Recognition Grand Challenge. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, San Diego, June 2005, pp. 947–954.

[20] Cai, D.—He, X.—Han, J.: Spectral Regression for Efficient Regularized Subspace Learning. In Proceedings of IEEE International Conference on Computer Vision, Rio de Janeiro, Brazil, October 2007.

[21] Xu, C.—Wang, Y.—Tan, T.—Quan, L.: A Robust Method for Detecting Noise on 3D Point Cloud. Pattern Recognition Letters, Vol. 27, 2006, No. 9, pp. 1687–1497.

[22] Siarry, P.—Berthiau, G.—Durbin, F.—Haussy, J.: Enhanced Simulated Annealing for Globally Minimizing Functions of Many-Continuous Variables. ACM Transactions on Mathematical Software, Vol. 23, 1997, No. 2, pp. 209–228.

[23] Cai, D.—He, X.—Han, J.: Using Graph Model for Face Analysis. Technical Report, UIUCDCS-R-2005-2636, UIUC, September 2005.

[24] Golub, G. H.—Loan, C. F. V.: Matrix computations. Johns Hopkins University Press, 3rd edition, 1996.

[25] Paige, C. C.—Saunders, M. A.: Algorithm 583 LSQR: Sparse Linear Equations and Least Squares Problems. ACM Transactions on Mathematical Software, Vol. 8, 1982, No. 2, pp. 195–209.

[26] Chua, C.—Han, F.—Ho, Y.: 3D Human Face Recognition using Point Signature. In Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition, Grenoble, Washington D.C., March 2000, pp. 233–238.

[27] Beumier, C.—Acheroy, M.: Automatic 3D Face Authentication. Image and Vision Computing, Vol. 18, 2000, No. 4, pp. 315–321.

[28] Turk, M.—Pentland, A.: Eigenfaces for Recognition. Journal of Cognitive Neuroscience, Vol. 3, 1991, No. 1, pp. 71–86.

[29] Cai, D.—He, X.—Han J.—Zhang, H.: Orthogonal Laplacianfaces for Face Recognition. IEEE Transactions on Image Processing, Vol. 15, 2006, No. 11, pp. 3608–3614.

[30] Zhong, C.—Sun, Z.—Tan, T.: Robust 3D Face Recognition Using Learned Visual Codebook. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, June 2007, pp. 1–6.

[31] Ahonen, T.—Hadid, A.—Pietikainen, M.: Face Description with Local Binary Patterns: Application to Face Recognition. IEEE Trasactions on Pattern Analysis and Machine Intelligence, Vol. 28, 2008, No. 12, pp. 2037–2041.

[32] Jones, J. P.—Palmer, L. A.: An Evaluation of the Two-Dimensional Gabor Filter Model of Simple Receptive Fields in Cat Striate Cortex. Journal of Neurophysiology, Vol. 27, 1987, No. 1, pp. 1233–1258.

[33] Liu, C.—Wechsler, H.: Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition. IEEE Transactions on Image Processing, Vol. 4, 2002, No. 4, pp. 467–476.

[34] Jin, M.: A Research on Military Target Recognition and Tracking based on Gabor Filter. Ph. D. thesis of Chinese Academy of Science, 2005 (in Chinese).

[35] Zhong, C.—Sun, Z.—Tan T.—He, Z.: Robust 3D Face Recognition in Uncontrolled Environments. in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, June 2008.

[36] Holmes, M. P.—Gary, A. G.—Isbell, C. L.: Fast SVD for Large-Scale Matrix. `http://bigml.wikispaces.com/file/view/Holmes.pdf`.

Yue Ming is currently working towards her Ph. D. degree at the Institute of Information Science, Beijing Jiaotong University. Her research interests include image processing, pattern recognition, 3D face recognition and reconstruction, etc.

Qiuqi Ruan received his B. Sc. and M. Sc. degrees from Beijing Jiaotong University, China in 1969 and 1981, respectively. From January 1987 to May 1990, he was a Visiting Scholar at the University of Pittsburgh, and at the University of Cincinnati. Subsequently, he has been a Visiting Professor in USA for several times. He has published 2 books and more than 100 papers, and achieved a national patent. Now he is a Professor, Doctorate Supervisor, at Northern Jiaotong University. He is a senior member of IEEE. His main research interests include digital signal processing, computer vision, pattern recognition, virtual reality, etc.

Xueqiao Wang is currently working towards her Ph. D. degree at the Institute of Information Science, Beijing Jiaotong University. Her research interests include image processing, pattern recognition, data compression, etc.