PERGAMON

Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT®

Pattern Recognition 38 (2005) 787 – 798

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

Efficient 3D reconstruction for face recognition

Dalong Jiang[a, b, 1], Yuxiao Hu[c], Shuicheng Yan[d, 1], Lei Zhang[c, *], Hongjiang Zhang[c], Wen Gao[a, b]

[a] Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China
[b] Graduated School of Chinese Academy of Sciences, Beijing 100039, China
[c] Microsoft Research Asia, Beijing 100080, China
[d] School of Mathematical Sciences, Peking University, Beijing 100871, China

Received 9 August 2004

Abstract

Face recognition with variant pose, illumination and expression (PIE) is a challenging problem. In this paper, we propose an analysis-by-synthesis framework for face recognition with variant PIE. First, an efficient two-dimensional (2D)-to-three-dimensional (3D) integrated face reconstruction approach is introduced to reconstruct a personalized 3D face model from a single frontal face image with neutral expression and normal illumination. Then, realistic virtual faces with different PIE are synthesized based on the personalized 3D face to characterize the face subspace. Finally, face recognition is conducted based on these representative virtual faces. Compared with other related work, this framework has following advantages: (1) only one single frontal face is required for face recognition, which avoids the burdensome enrollment work; (2) the synthesized face samples provide the capability to conduct recognition under difficult conditions like complex PIE; and (3) compared with other 3D reconstruction approaches, our proposed 2D-to-3D integrated face reconstruction approach is fully automatic and more efficient. The extensive experimental results show that the synthesized virtual faces significantly improve the accuracy of face recognition with changing PIE.

Keywords: Face recognition; 3D face reconstruction; Multi-view; Illumination; Expression; Analysis by synthesis

1. Introduction

Human faces are one of the most important content in photograph, thus detecting and recognizing faces are extremely desirable in content understanding of digital photographs. However, robustly recognition of faces in digital photographs, especially family photographs, remains a challenging problem despite of over three decades of research efforts [1]. To evaluate the progress made both in theories and practices, face recognition vendor test in 2002 (FRVT2002) [2] evaluated the state of art algorithms and systems by large-scale, real-world test datasets. The results indicate that face recognition (verification) accuracy on frontal face with indoor lighting has reached about 90%, which is basically acceptable for general face recognition tasks. On the other hand, FRVT2002 also expose that face recognition among different pose, illumination and expression (PIE) is still far from satisfactory. The reason for the low face recognition accuracy on multi-view, un-constrained illumination and arbitrary expression samples is that two-dimension (2D) face images are greatly influenced by PIE besides the identity, i.e. unique head geometry and skin

[*] Corresponding author.
E-mail address: leizhang@microsoft.com (L. Zhang).
[1] This work was performed while the authors were visiting Microsoft Research Asia.

0031-3203/$30.00 © 2004 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2004.11.004

texture of a person. These differences between the gallery and probe samples should either be explicitly decoupled before classification or be implicitly described by the face model during recognition.

In order to deal with the aforementioned problems, two different strategies have been conducted in previous works, i.e. normalization based strategy and expansion based strategy. The first kind of methods either tries to normalize probe samples to a unified PIE which is the same or similar to the gallery samples to ensure the generalization capability of the classifier trained on the gallery samples [3–6], or tries to extract specific features which are invariant or insensitive to different PIE [7–10]. Besides these 2D methods, three-dimensional (3D) methods are also explored. In [11], face samples with out-of-plane rotation are warped to frontal faces according to a cylinder face model. Vetter et al. proposed a 3D alignment algorithm [12–14] to recover the shape and texture parameters of a 3D morphable model. In their solution, the shape parameters are computed from a shape error estimated by optical flow and the texture parameters are obtained from a texture error. Their algorithm uses linear equations to recover the shape and texture parameters irrespective of pose and lighting conditions of the face image. Face recognition is conducted by matching the recovered shape and texture parameters. In general, the aforementioned 2D based methods do not consider the specific structures of human faces, and thus frequently leads to the worse performance on profile pose face samples. 3D based methods overcome this problem, but they either require heavy manual labeling work or are time-consuming.

In contrast to the normalization based methods, the other kind of methods tries to utilize more samples which cover different PIE to enhance the representation capability of the face gallery. View-based method [15] has shown its efficiencies, but it needs sufficient gallery samples. While for typical face recognition systems, the quantity and quality of the training and testing samples are asymmetrical in most cases. Generally, it is cumbersome to collect sufficient face samples to represent the identities, but it is convenient to control the PIE of these face samples during acquisition or model these factors by sophisticated off-line analysis algorithms. On the other hand, face samples with variant PIE will appear in test sets, which are difficult to be predicted or controlled. Actually, such asymmetries are common in practical systems. For example, in public security applications such as security check in airports, there are generally two mug shots, one for frontal face and the other for profile face, being available to match a suspect. (Sometimes only one frontal face image is provided.) While the PIEs of the passengers' faces are frequently too different to be normalized, the asymmetry between training and testing samples requires the face recognition system to be able to characterize the face of each identity by as few training samples as possible, which may be achieved by analyzing training samples and generating more representative ones.

To enlarge the training set and improve its representative ability, variant analysis-by-synthesis methods are put forward, i.e., the labeled training samples are warped to cover different poses or re-lighted to simulate different illuminations [16–21]. Photometric stereo technologies such as illumination cones and quotation image are used to recover the illumination or relight the sample face images. Shape from shading [22–25] has been explored to extract 3D geometry information of a face and to generate virtual samples by rotating the result 3D face models.

The aforementioned expansion based algorithms have achieved improvement in face recognition; however, the intrinsic drawbacks limit their practice in real applications: (1) photometric methods assume that the faces have similar geometries; as a result, if the pose of an unknown face is not the same as that of the known face or it is not aligned well, the synthesized faces will not be realistic; (2) shape from shading algorithm requires that the face images are precisely aligned pixel-wise, which is difficult to be implemented in practice or even impossible for practical face recognition applications; and (3) the 3D face alignment [14] requires manual initialization and the speed (1 min for a face image) is not able to meet the requirement of most real face recognition systems.

In this paper, we propose an efficient and fully automatic 2D-to-3D integrated face reconstruction method to provide a solution to the above problems in an analysis-by-synthesis manner. First, frontal face detection and alignment are utilized to locate a frontal face and the facial feature points within an image, such as the contour points of the face, left and right eyes, mouth and nose. Then, the 3D face shape is reconstructed according to the feature points and a 3D face database. After that, the face model is texture-mapped by projecting the input 2D image onto the 3D face shape. Based on this 3D face model, virtual samples with variant PIE are synthesized to represent the 2D face image space. Finally, face recognition is conducted in this enlarged face subspace after standard normalization of testing sample face images. The only input to this system is a frontal face image with normal illumination and neutral expression. The outputs are images with variant PIE for recognition. Compared with previous work, this framework has following advantages: (1) only one single frontal face is required for training, which avoids the burdensome enrollment work; (2) the synthesized face samples provide the capability of recognizing faces under complex conditions such as arbitrary PIE; (3) the proposed integrated 2D-to-3D face reconstruction approach is fully automatic and the speed is fast. It takes about 4 s per face image (512 × 512 pixels) on a P4 1.3 GHz, 256M RAM computer, which is about 15 times faster than the 3D face alignment processing [14].

The rest of this paper is organized as follows. The 2D-to-3D face reconstruction algorithm and the method of generating realistic virtual face sample images with variant IE is described in detail in Section 2. Face recognition experimental results are provided in Section 3 to justify the efficiency of the proposed algorithm. Section 4 gives conclusion remarks and discussions about the future directions.

2. Efficient 3D face reconstruction for face recognition

Previous works in face recognition have witnessed the efficiency of virtual faces and 3D face modeling. In this section, we present an efficient and fully automatic framework for face recognition by performing 3D face reconstruction and generating virtual faces from a single frontal face with normal illumination and neutral expression. The framework, as shown in Fig. 1, consists of two parts: (1) 2D-to-3D integrated face reconstruction; and (2) face recognition using the virtual faces with different PIEs. The following subsections will describe these two parts in detail.

2.1. 2D-to-3D integrated face reconstruction

The only required input to the system is a frontal face image of a subject with normal illumination and neutral expression. Based on our previous 2D alignment algorithm [26], 83 key feature points are automatically located. The feature points, as shown in Fig. 1, are accurate enough for face reconstruction in most cases. A general 3D face model is applied for personalized 3D face reconstruction. The 3D shapes have been compressed by the principal component analysis (PCA). After the 2D face alignment, the key feature points are used to compute the 3D shape coefficients of the eigenvectors. Then, the coefficients are used to reconstruct the 3D face shape. Finally, the face texture is extracted from the input image. By mapping the texture onto the 3D face geometry, the 3D face model for the input 2D face image is reconstructed.

2.1.1. Efficient 2D face alignment

Automatic alignment on multi-view face images is still an open problem. But face alignment on frontal face has been well studied [26]. In our work, the input 2D face images are in frontal pose with normal illumination and neutral expression, which is the most common case in face recognition. Under such a constraint, a fast and accurate 2D face alignment algorithm is deployed to locate key facial points such as face contour points, eye centers and nose tip. Eighty three feature points can be aligned, some of which are selected for face reconstruction, as shown in Fig. 2(a). The position of

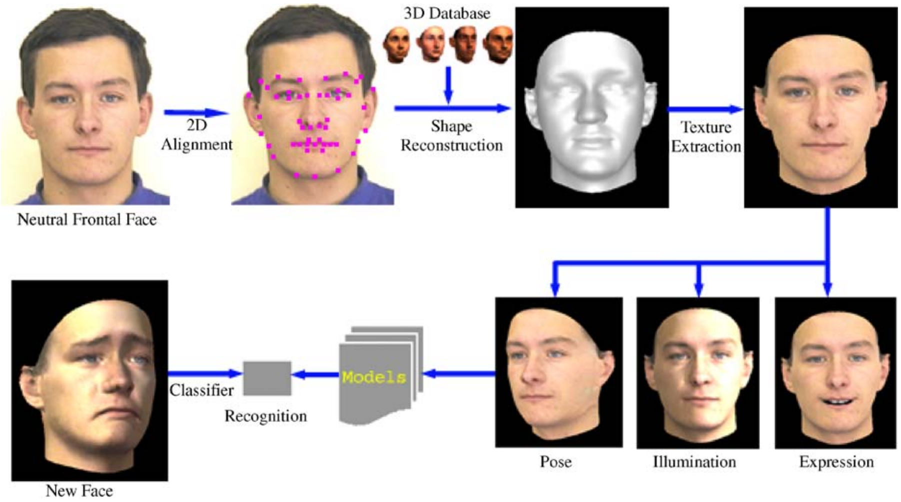

Fig. 1. System overview. (The input image is copied from AR face database [34].)

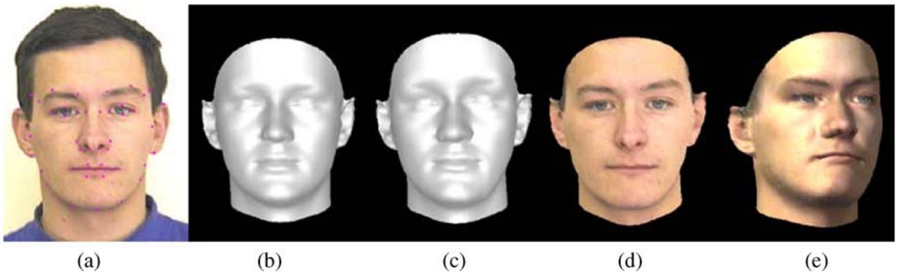

Fig. 2. 3D reconstruction. (a) 2D alignment; (b) 3D shape reconstructed by PCA coefficients of eigenvectors; (c) 3D shape after Kriging interpolation; (d) 3D model with texture; (e) a new view with PIE.

these feature points can be modified in case the alignment is not accurate, which rarely occurs.

2.1.2. 3D face geometry reconstruction

To reconstruct a 3D face model, we use the USF Human ID 3-D database which includes 100 laser-scanned heads [13]. Each face model in the database has approximately 70,000 vertices. In this paper, the number of the vertices is reduced to about 8900 for better performance. Then the vertices on the lip line are duplicated and the triangles around the lip line are reconstructed so that the mouth of the face model can be opened for lip motion and expression, which will be described later.

In general, matching a 3D geometry to a 2D image is an ill-posed problem. Fortunately, the differences between the 3D shapes of different face models are not dramatic. In this paper, the geometry of a 3D face model is represented with a shape-vector $S=(X_1, Y_1, Z_1, X_2, \ldots, Y_n, Z_n)^{\mathrm{T}} \in \Re^{3n}$, which contains the X, Y, Z coordinates of its n vertices. Since all facial feature points such as corners of eyes and tips of nose of different faces are fully corresponded by means of their semantic position, PCA is appropriate to be conducted to get a more compact shape representation of face by the primary components. Let $\overline{S}$ be the average shape, $P \in \Re^{3n\times m}$ be the matrix of the first m eigenvectors (in descending order according to their eigenvalues). A new face shape S' can be expressed as

$$S' = \overline{S} + P\vec{\alpha}, \tag{1}$$

where $\vec{\alpha}=(\alpha_1, \alpha_2, \ldots, \alpha_m)^{\mathrm{T}} \in \Re^m$ is the coefficients of the shape eigenvectors.

In the alignment step, it is assumed that t 2D facial feature points are selected for 3D reconstruction. t vertices, corresponding to the feature points, are also chosen on the face geometry. Let $S_f=(X_1, Y_1, X_2, \ldots, X_t, Y_t)^{\mathrm{T}} \in \Re^{2t}$ be the set of X, Y coordinates of the feature vertices on the face. Thus, S_f is the sub shape-vector of S. According to Eq. (1), the X, Y coordinates of those feature vertices of a new face shape S'_f, assumed zero centered, can be expressed as

$$S'_f = \overline{S_f} + P_f\vec{\alpha}, \tag{2}$$

where $\overline{S}_f \in \Re^{2t}$ and $P_f \in \Re^{2t\times m}$ are the X, Y coordinates of the feature vertices on $\overline{S}$ and P, respectively. To transform face coordinate to image coordinate, let S''_f be the transformed shape, which can be obtained from the face alignment result, then

$$S''_f = cS'_f + T, \tag{3}$$

where $T \in \Re^{2t}$ is the translation vector and $c \in \Re$ is the scale coefficient. Note that since the 2D face image and 3D face model are both frontal, the rotation matrix is not required. Since P_f is an orthogonal matrix, $\vec{\alpha}$ can be derived from Eq. (2) as

$$\vec{\alpha} = (P_f^{\mathrm{T}} P_f)^{-1} P_f^{\mathrm{T}} (S'_f - \overline{S_f}). \tag{4}$$

To avoid the outliers, the priors are applied to constrain $\vec{\alpha}$, and Eq. (4) is changed to

$$\vec{\alpha} = (P_f^{\mathrm{T}} P_f + \lambda\Lambda^{-1})^{-1} P_f^{\mathrm{T}} (S'_f - \overline{S_f}), \tag{5}$$

where $\Lambda = diag(v_1, v_2, \ldots, v_m)$, λ is the weighting factor, and v_i is the ith eigenvalue.

In Eqs. (2) and (3), there are five variables ($\vec{\alpha}$, S'_f, S''_f, T, c). To compute the face geometry coefficient $\vec{\alpha}$, an iterative procedure is applied as outlined below.

Before the first iteration, let $\overline{S_f}$ be the initial value of S'_f.

Step 1. Let T_x and T_y be the average offsets of all t feature points of S''_f to the origin along X, Y axes, respectively, then

$$(T_x, T_y)^{\mathrm{T}} = \frac{1}{t}\sum_{i=1}^{t} S''_{fi}.$$

Then $T=(T_x, T_y, \ldots, T_x, T_y)^{\mathrm{T}}$ and

$$c = \frac{\sum_{i=1}^{t}\left\langle S''_f - (T_x, T_y)^{\mathrm{T}}, S'_f\right\rangle}{\sum_{i=1}^{t}\|S'_f\|^2},$$

S'_f can then be computed from Eq. (3).

Step 2: The face geometry coefficient $\vec{\alpha}$ can be computed using Eq. (5); and then a new S'_f can be obtained by applying $\vec{\alpha}$ to Eq. (2).

The geometry coefficient $\vec{\alpha}$ generally converges to a fixed value after repeating step 1 and step 2 for mostly 10 iterations. Then we apply $\vec{\alpha}$ to Eq. (1) to get the whole 3D face geometry S'. The reconstructed face shape is shown as Fig. 2(b). The face geometry looks quite well, but the X, Y coordinates of the feature vertices on the face are somewhat different from the X, Y coordinates of the feature points on the 2D image. The reason is that the shape space is limited by the 3D face database. To ensure that the feature vertices are exactly correct, the X, Y coordinates of the feature vertices on the face are forced to be aligned to the X, Y coordinates of the feature points on the 2D image. According to the displacements of the feature vertices, the Kriging interpolation method [27] is used to compute the displacement of non-feature vertices. For interpolation purpose, radius base function (RBF) is a good alternative. By using the method we described above, the final 3D face geometry is reconstructed with the accurate feature vertices. The final 3D face shape is shown as Fig. 2(c).

2.1.3. Texture extraction

In this paper, the input image is projected orthogonally to the 3D geometry to generate the face texture. Compared

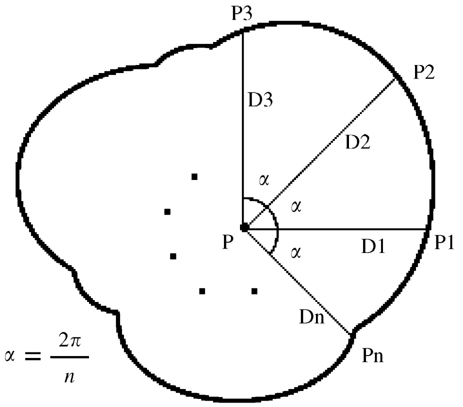

Fig. 3. Texture interpolation.

with Vetter's method [12], our texture extraction is real time, and the extracted texture has detailed and realistic features, which is important for face recognition. On the other hand, the face texture can also be reconstructed by PCA or other subspace methods. But the insufficient training data compared to the high dimensional face texture space, i.e. curse of dimension, limits the reconstruction effect so that it is not practical here, which has been indicated in our experiments.

After the 2D image is directly mapped to the 3D geometry, no corresponding color information is available for some vertices because they are occluded in the frontal face image. Therefore, it is possible that there are still some blank areas on the generated texture map, which need to be corrected, as shown in Fig. 4(a). In this paper, a linear interpolation algorithm is employed to interpolate the blank areas by known colors. As shown in Fig. 3, let P be one of the points in the blank area. It is assumed that n radials, which averagely divide the 360°, come from P. Let P_i $(i = 1, 2, \ldots, n)$ be the points at which the radials intersect the known color areas. Then, $C(R, G, B)$, the *RGB* color of P, is computed by using the following equation:

$$C(R, G, B) = \frac{\sum_{i=1}^{n}(\lambda_i C_i(R, G, B))}{\sum_{i=1}^{n} \lambda_i}, \tag{6}$$

where $\lambda_i = 1/D_i$ and D_i is the distance between P and P_i. If P_i is at the edge of the image, then λ_i is set to 0. The non-interpolated texture image in Fig. 4(a) can be compared with the interpolated texture image in Fig. 4(b) and (c), where the numbers of radials n are 4 and 16, respectively. The interpolated areas may be not accurate enough and lose some details, but most of them are near the neck and ears, which are not crucial to face recognition. The texture mapped face model is also shown in Fig. 2(d) and (e). For interpolation purpose, the thin-plate relaxation algorithm [28] may be a good alternative.

2.1.4. Discussion

Only 100 3D heads in the USF Human ID 3D database are used for PCA in this paper. The face space spanned by these models is quite limited. For reconstruction purpose, more 3D heads are needed. If there are more 3D head samples, the reconstruction results should be more accurate, especially the vertices's positions along the *Z*-axis.

Another promising method to improve the 3D reconstruction accuracy is to add one additional input face image with the profile pose. Alignment on the side view acquires the accurate *Z* coordinates of the feature vertices. Combined with the alignment on the frontal image, the accurate positions of the feature vertices can be obtained. The reconstructed 3D model will be more accurate by using the accurate feature vertices. However, the automated alignment algorithms on profile faces are not as robust as those on frontal faces, thus manual alignment will be required.

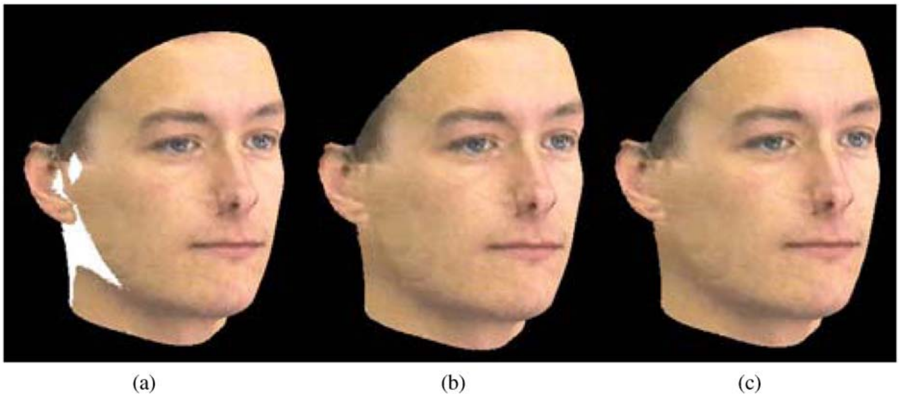

Fig. 4. Fill the blank areas in the texture. (a) Texture before filling the blank; (b) texture after filling the blank, $n = 4$; (c) texture after filling the blank, $n = 16$.

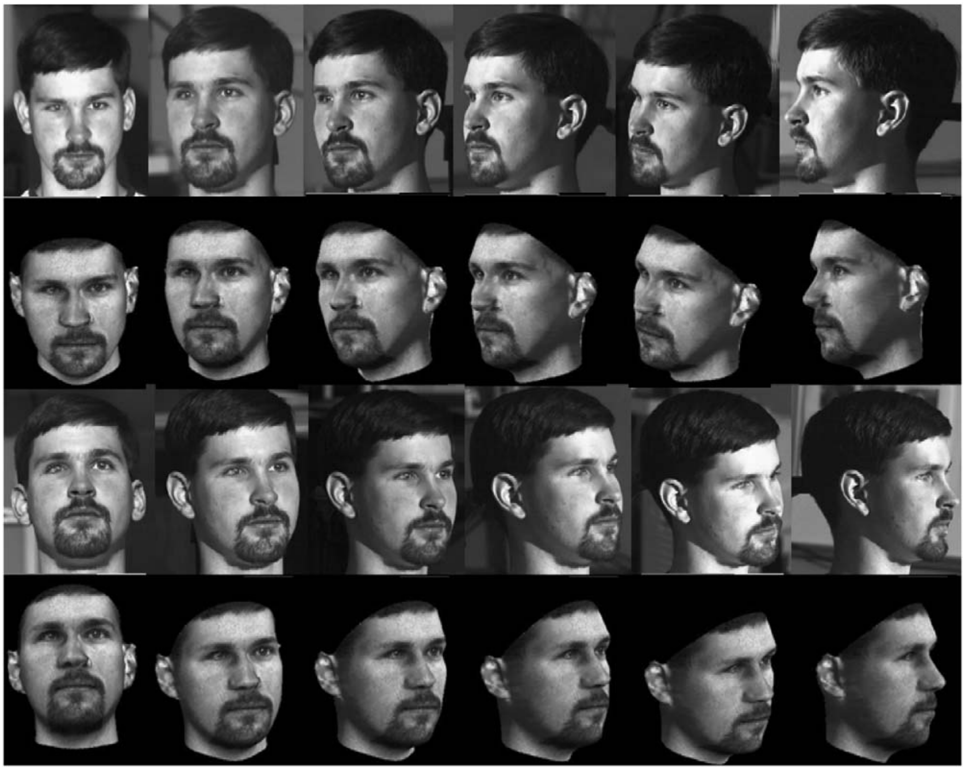

Fig. 5. Poses. The first and third lines are poses in CMU-PIE. The second and fourth lines are the corresponding poses generated by rotating the reconstructed model.

2.2. Synthesis with different pose, illumination and expression

In natural environments, PIE remains a critical and challenging issue in face recognition algorithms. To increase the accuracy of face recognition, acquiring sample face images with variant PIE are necessary. However, it's difficult to generate new face images with different PIE from a frontal image using any existing 2D-to-2D methods. The problem could be solved by proposed approach to reconstructing the 3D face model from the given 2D face image. The reconstructed 3D face model is then rotated to generate images with variant poses. By applying different lights, variant illuminations are created. Finally, a MPEG-4 based facial animation technique is used to generate expressions, which are also an important factor in face recognition but are not considered in most researches.

2.2.1. Pose

Pose variation is the primary source of difficulties for face recognition. The difficulties have been documented in the FERET test report and suggested as a major research issue [29]. The performance of face recognition systems drops dramatically, when large pose variations are presented in the input images, especially when the system's training data have few non-frontal images. A reasonable way to improve multi-view recognition is to use multiple training views. In our work, it is very easy to generate any views by rotating the 3D model to the right pose. The poses in CMU-PIE [30] and those generated by our approach are compared in Fig. 5.

For face recognition training purpose, the positions of feature points on the multi-view face images are needed. In general, face alignment on arbitrary multi-view face images is quite difficult, and no technique is able to automatically solve this problem with high accuracy so far. Most multi-view face recognition methods require manually labeling these feature points on large number of training and testing sample images to align them, which is inaccurate and time-consuming.

In the proposed method, since the multi-view images are generated by rotating the 3D face model, the alignment on the new face images is no longer a problem. When a multi-view face image is projected after rotating the 3D model, the positions of facial feature points are obtained by projecting the corresponding feature vertices on the 3D model to 2D image, i.e., no more alignment is required on the generated multi-view images. The acquisition of the feature point positions on the multi-view face images is thus automatic and accurate.

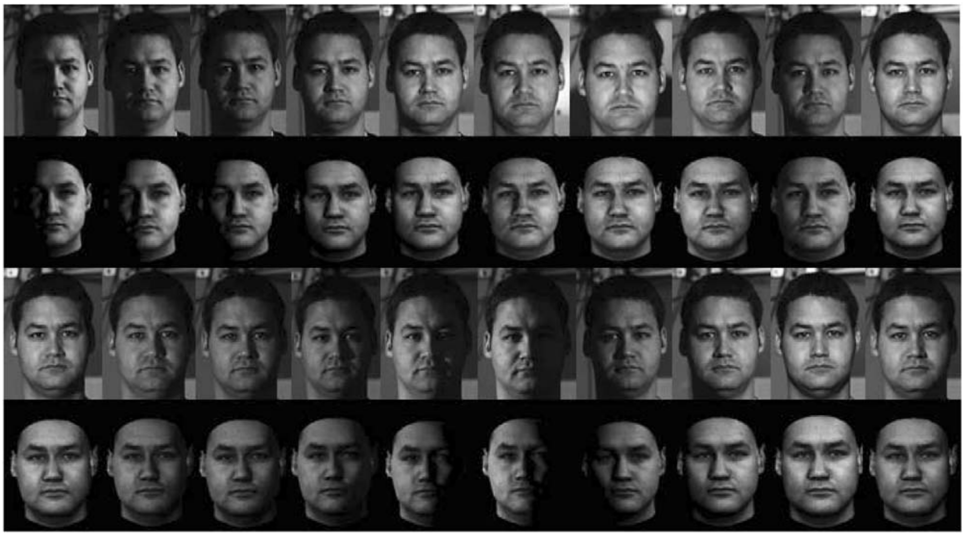

Fig. 6. Illuminations. The first and third lines are CMU-PIE images. The second and fourth lines are the corresponding generated images.

2.2.2. Illumination

Illumination is another important issue for face recognition. The same face appears different due to changes in lighting. The changes induced by illumination are often larger than the differences between individuals.

In our work, in order to generate variant illumination images of a 3D face model, two lights are applied to the 3D models. One is an environment light; the other is a movable spot light. The whole face model illumination is controlled by the environment light. The specular areas and shadows of the face model are generated by the spot light. Several attributes of the spot light can be controlled, including ambient, diffuse, specular and position. Some illumination images are imitated with CMU-PIE. Fig. 6 shows that the generated illuminations of the images are quite similar to CMU-PIE. If enough lights are applied, most common illumination conditions can be generated.

In this paper, it is assumed that all the face surfaces are in the same material. Actually, the materials in different face are different. Even for the same face, the materials are also different in the different areas. For example, eyeballs should reflect more lights than skins. In the future work, the complex face material should be considered to simulate more real illuminations.

2.2.3. Expression

In general, expression changes are not as important as pose and illumination changes for face recognition. But, expression changes are still a problem to be solved in order to achieve robust face recognition.

In our work, the MEPG-4 based animation framework is used to drive the 3D face model and generate different expressions [31,32]. In the MPEG-4 standard, there are altogether 68 facial animation parameters (FAPs), each of which expresses the motion in a specific direction on specific region of the face. Complex facial expressions are generated by combining all the FAPs. FAPs are a set of general parameters. One specific set of FAPs denotes one expression. It is independent of the facial model being used.

In the MPEG-4 based facial animation system, the motion trajectory of each mesh vertex is piece-wise linearly approximated. Precise results can be acquired by increasing the segments. Each vertex within the control region of each FAP has a 3D motion factors in each segment of a specific FAP. Facial animation table (FAT) is combined by these 3D motion factors, which depicts how the FAP influences the facial model to perform the desired animation. Getting the value of an FAP, FAT provides corresponding information to convert the FAP to facial animation. FAT depends on facial models being used. The FAT for our general facial model has been built up manually. But, it can be only applied to those facial models that have the same topology as the general model. In this paper, the 3D facial models have different topology. To drive these 3D facial models, a novel approach is proposed to build the FAT for the 3D models based on the known FAT of the general facial model.

Since the face muscle motions are mainly within the X, Y plane, we mainly compute the motion factors of mesh vertices in the X, Y plane when constructing FAT for a new model. First, the Kriging interpolation method [27] is used to transform the general model according to the extracted feature points on the new model. In this way, the vertices on the general model have the same physical meaning as those on the new model. Second, each vertex on the new model is projected orthographically onto the X, Y plane of

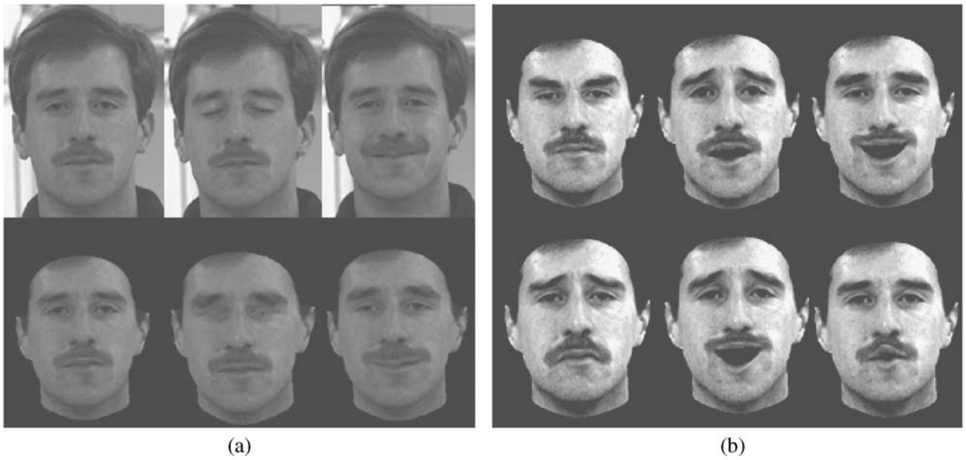

Fig. 7. Expressions. (a) The first line is the expressions in CMU-PIE; the second line is the generated expressions corresponding to the first line; (b) other generated expressions.

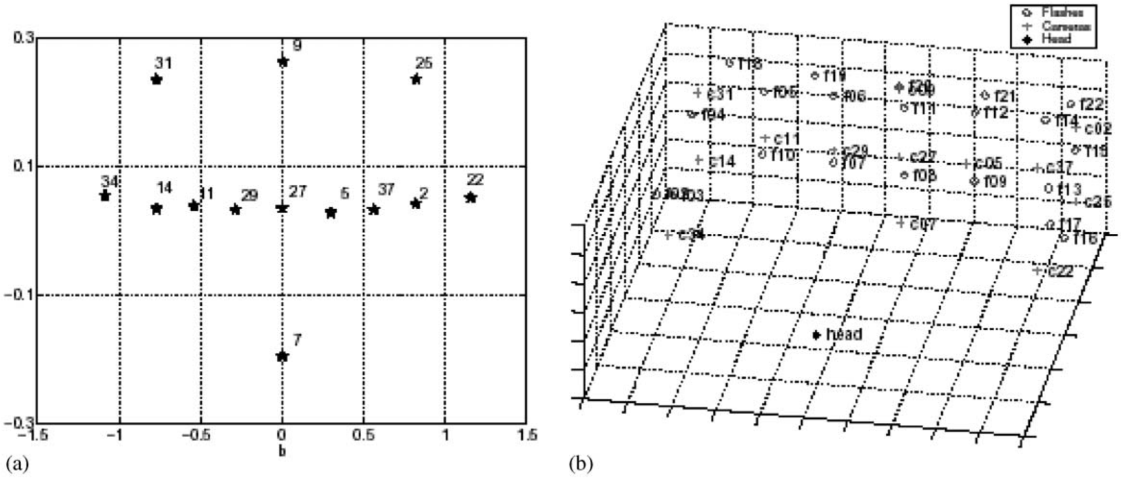

Fig. 8. CMU-PIE database camera and light positions. (a) A plot of the azimuth and altitude angles of the cameras; (b) 3D locations of the cameras, the flashes and the head. (The pictures are copied from [35] and [30].)

the general model. From the projection of each vertex, it is determined that which FAP controls this vertex and what the motion factors should be for this vertex. After all the mesh vertices are computed, the FAT for the new model can be obtained. The detail to construct FAT for arbitrary 3D face models has been thoroughly investigated in earlier works [33]. By the FAT built for the reconstructed 3D face model in this paper, variant expressions are generated. As Fig. 7(a) shows, the expressions in the first line are images in CMU-PIE. The expressions in the second line, which corresponds to the first line, are generated by our MPEG-4 animation system. Fig. 7(b) shows some other generated expressions.

3. Experiments

In this work, we aim at exploring face recognition performance across variant PIE. We systematically evaluated the performance of our algorithm compared with the conventional algorithm that do not use the virtual faces synthesized from the personalized 3D face models. The CMU-PIE database is used in the evaluation since it takes into account all the three factors. The CMU-PIE database contains 68 subjects with 41,368 face images, captured by 13 synchronized cameras and 21 flashes, under varying PIE. The CMU-PIE database camera and light positions are shown in Fig. 8.

Table 1
Recognition accuracy comparison between face recognition with/without virtual face using PCA

Pose.	22	25	02	37	05	09	07	29	11	14	31	34
Con (%)	3.9	5.3	4.5	6.5	65.9	82.7	79.8	48.5	6.0	4.4	3.8	4.3
Vir (%)	12.0	34.3	28.4	44.7	67.4	83.8	83.0	46.5	42.3	25.2	14.3	6.3
Vir+ (%)	13.6	34.7	29.8	45.2	66.6	83.2	82.5	46.7	43.0	17.0	10.7	6.4

(Con: conventional algorithms; Vir: using virtual faces; Vir+: using virtual face for special pose.)

Table 2
Recognition accuracy comparison between face recognition with/without virtual face using LDA

Pose.	22	25	02	37	05	09	07	29	11	14	31	34
Con (%)	4.4	4.3	3.8	4.7	64.7	75.9	79.6	48.6	6.3	6.7	4.7	4.5
Vir (%)	12.6	38.3	36.3	53.0	77.4	78.2	81.4	54.0	45.2	22.9	17.5	8.4
Vir+ (%)	15.6	35.1	28.6	53.3	84.9	85.0	92.1	68.0	48.4	22.1	17.6	9.2

(Con: conventional algorithms; Vir: using virtual faces; Vir+: using virtual face for special pose.)

We used the frontal face at pose-27 with neutral expression and environment light to automatically construct the personalized 3D faces. All the 68 3D faces are constructed and the virtual faces with different PIE combinations are synthesized; the comparisons with the real faces are illustrated at Figs. 5–7. Note that all the reconstruction is fully automatic. Only one frontal face of a subject with normal illumination and neutral expression is required to construct the personalized 3D face model of the subject, which can be easily satisfied in real application.

In all the experiments, the conventional method used only the frontal faces at pose-27 for training and the other faces are all used for testing. The comparison experiments have been conducted to evaluate the effectiveness of the virtual faces from constructed 3D face model for face recognition with arbitrary PIE. We used two traditional methods, principal component analysis (PCA) and linear discriminant analysis (LDA), to perform dimensionality reduction, and we used the nearest neighbors (NN) as similarity matching approach for classification. The results using PCA and LDA are listed in Tables 1 and 2. From the listed results, we have the following observations:

(1) In general, face recognition accuracy using virtual faces from reconstructed 3D faces is higher than conventional algorithms, especially for the experiment using LDA and with the pose information ahead.
(2) Our proposed algorithm significantly improved the performance in half-profile views, like pose 37 and 11; while for the profile views, the improvements are limited. This is because that the reconstructed texture for the unseen points in frontal view is not accurate enough. We are exploring new methods for realistic missing data reconstruction, like using the 3D texture models.
(3) With prior pose information, the performance was improved than that using all the virtual faces and one single global model, when we constructed separated models for each view.

4. Conclusions

Experimental evaluation of face reconstruction for face recognition have illustrated that the proposed fully automatic system is efficient and of high accuracy and robustness. Compared to other related works, this framework has following highlights: (1) only one single frontal face is required for face recognition and the outputs are realistic images with variant PIE for the individual of the input image, which avoids the burdensome enrollment work; (2) the synthesized face samples provide the capability to conduct recognition under difficult conditions of complex PIE; and (3) the proposed 2D-to-3D integrated face reconstruction approach is fully automatic and faster than other 3D reconstruction approaches.

In order to compare with the most related work by Blanz et al. [12], Table 3 lists the differences between their method and ours. In this table, the two methods are compared in brief on input requirements, initialization, shape and texture reconstruction methods and overall system performance. As a result, our approach is fully automatic and much faster than their method. Although the input requirements in our method are stricter than theirs, it is not difficult to satisfy such requirements, which are quite general in most scenarios.

In the work presented in this paper, the CMU-PIE image database was used in the experiments. The images captured by camera c27 with only environment light are used as the input images to reconstruct 3D face models. The images captured by camera c27 should be frontal in CMU-PIE. Actually, many of these “frontal” images are not really in frontal. Such input images contribute to the errors in the

Table 3
Comparison with Vetter's work

	Vetter's method	Our method
Input	Single face image with arbitrary pose, illumination	Single frontal face image with homogeneous illumination, neutral expression
Initialization	Manually initialization is required	Fully automatic
Shape	Shape parameters are computed from a shape error estimated by optical flow	The 3D shape is recovered by the correspondence between the 2D–3D fiducial feature points and a statistical model
Texture	Texture parameters are obtained from the texture error	The 3D face texture is directly mapped from the 2D input image
Speed	About 1 min per face image	Less than 5 s for a 512 × 512 face image on P4 1.3 GHz CPU

3D reconstruction and the generated images. If all the input images are constrained strictly with frontal pose, normal illumination and neutral expression, the experiment results should be more encouraging.

In the future works, the 3D alignment work which reconstructs 3D face model based on non-frontal face images will be investigated based on the method introduced in this paper. Reconstruction based on multiple face images with in-plane or out-plane variance will also be considered to improve the precision of the reconstructed shape and texture.

Acknowledgements

The authors thank the University of South Florida for providing the USF Human ID 3D Database. The first author is supported partly by Natural Science Foundation of China (No. 60303018).

References

[1] T. Kanade, Picture processing system by computer complex and recognition of human faces, doctoral dissertation, Kyoto University, November, 1973.

[2] P.J. Phillips, P. Grother, R.J. Micheals, D.M. Blackburn, E. Tabassi, M. Bone, Face Recognition Vendor Test 2002: Evaluation Report, 2002.

[3] T.S. Jebara, 3D Pose Estimation and Normalization for Face Recognition, Centre for Intelligent Machines, McGill University, 1995.

[4] H. Imaoka, S. Sakamoto. Pose-independent face recognition method, in: Proceedings of IEICE Workshop of Pattern Recognition and Media Understanding, June 1999, pp. 51–58.

[5] M. Lando, S. Edelman, Generalization from a single view in face recognition, in: Proceedings of the International Workshop on Automatic Face and Gesture Recognition, Zurich, 1995, pp. 80–85.

[6] T. Maurer, C. von der Malsburg, Single-view based recognition of faces rotated in depth, in: Proceedings of the International Workshop on Automatic Face and Gesture Recognition, Zurich, 1995, pp. 248–253.

[7] JianHuang Lai, Pong C Yuen, GuoCan Feng, Face recognition using holistic Fourier invariant features, Pattern Recognition 34 (1) (2001) 95–109.

[8] Laurenz Wiskott, Jean-Marc Fellous, Norbert Krüger, et al., Face recognition by elastic bunch graph matching. Seventh International Conference on Computer Analysis of Images and Patterns, CAIP'97, Kiel.

[9] P.S. Penev, Reducing the dimensionality of face space in a sparse distributed local-features representation, FG'2000.

[10] Z.M. Hafed, M.D. Levine, Face recognition using the discrete cosine transform, Int. J. Comput. Vision 43 (3) (2001) 167–188.

[11] Kin-Man Lam, Hong Yan, An analytic-to-holistic approach for face recognition based on a single frontal view, PAMI98, vol. 2(7), pp. 673–686.

[12] V. Blanz, S. Romdhani, T. Vetter, Face-identification across different poses and illuminations with a 3D morphable model, Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, 2002.

[13] V. Blanz, T. Vetter, A morphable model for the synthesis of 3D-faces, in: SIGGRAPH 99 Conference3 Proceedings, Los Angeles, 1999, pp. 187–194.

[14] S. Romdhani, V. Blanz, T. Vetter, Face identification by fitting a 3D morphable model using linear shape and texture error functions, in: Computer Vision—ECCV'02, vol. 4, 2002, pp. 3–19.

[15] A. Pentland, B. Moghaddam, T. Starner, O. Oliyide, M. Turk, View-based and modular eigenspaces for face recognition, Technical Report 245, M.I.T Media Lab, 1993.

[16] T. Riklin-Raviv, A. ShaShua, The quotient image: class based re-rendering and recognition with varying illuminations, Pattern Anal. Mach. Intell. 23 (2) (2001) 129–139.

[17] Zicheng Liu, Ying Shan, Zhengyou Zhang, Expressive expression mapping with ratio images, SIGGRAPH 2001.

[18] A.S. Georghiades, P.N. Belhumeur, D.J. Kriegman, From few to many: illumination cone models for face recognition under variable lighting and pose, IEEE Trans. Pattern Anal. Mach. Intell. (2001) 643–660.

[19] A. Talukder, Nonlinear feature extraction for pattern recognition applications, Dissertation of CMU, Pittsburg, 1999.

[20] A. Talukder, D. Casasent, Pose-invariant recognition of faces at unknown aspect views, IJCNN 1999, Washington, DC.

[21] T. Vetter, T. Poggio, Linear object classes and image synthesis from a single example image, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 733–741.

[22] Ruo Zhang, Ping-Sing Tai, James Edwin Cryer, Mubarak Sha, Shape from shading: a survey, IEEE Trans. Pattern Anal. Mach. Intell. 21 (8) (1999) 690–706.

[23] J. Atick, P. Griffin, N. Redlich, Statistical approach to shape from shading: reconstruction of three dimensional face surfaces from single two dimensional image, Neural Comput. 8 (1996) 1321–1340.

[24] Wenyi Zhao, Rama Chellappa, SFS based view synthesis for robust face recognition, Proceedings of the Fourth International Conference on Face and Gesture Recognition, Grenoble, France, 2000, pp. 285–292.

[25] T. Sim, T. Kanade, Combining models and exemplars for face recognition: an illuminating example, Proceedings of the CVPR 2001 Workshop on Models versus Exemplars in Computer Vision, December, 2001.

[26] S.C. Yan, M.J. Li, H.J. Zhang, Q.S. Cheng, Ranking prior likelihood distributions for Bayesian shape localization framework, in: Proceedings of the Ninth International Conference on Computer Vision, France, Nice, ICCV'03.

[27] M.A. Oliver, R. Webster, Kriging: a method of interpolation for geographical information system, Int. J. Geogr. Inf. Syst. 4 (3) (1990) 313–332.

[28] D. Terzopoulos, The computation of visible-surface representations, IEEE Trans. Pattern Anal. Mach. Intell. 10 (4) (1988) 417–438.

[29] P.J. Phillips, P. Rauss, S. Der, Feret (face recognition technology) recognition algorithm development and test report, ARL-TR 995, US Army Research Laboratory, 1996.

[30] T. Sim, S. Baker, M. Bsat, The CMU pose, illumination, and expression (PIE) database, The 2002 International Conference on Automatic Face and Gesture Recognition.

[31] ISO/IEC 14496-1:2001, Coding of Audio-Visual Objects: Systems.

[32] ISO/IEC 14496-2:2001, Coding of Audio-Visual Objects: Visual.

[33] D. Jiang, W. Gao, Z. Li, Z. Wang, Animating arbitrary topology 3D facial model using the MPEG-4 FaceDefTables, The Fourth International Conference on Multi-modal Interface, IEEE ICMI'2002, Pittsburgh, USA, 14–16 October, 2002, pp. 517–522.

[34] A.R. Martinez, R. Benavente, The ar face database, Technical Report 24, Computer Vision Center (CVC) Technical Report, Barcelona, Spain, June 1998.

[35] R. Gross, J. Shi, J. Cohn, Quo vadis face recognition? in: Third Workshop on Empirical Evaluation Methods in Computer Vision, 2001.

About the Author—DALONG JIANG received the B.S. degree in computer science from Tsinghua University, Beijing, China, in 1999. He is currently working toward the Ph.D. degree in computer science at the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China.

He has been a Research Assistant at the Joint R&D Lab (JDL), Chinese Academy of Sciences, since 1999. His research interests include virtual reality, computer graphics and animation.

About the Author—YUXIAO HU received the Master degree in computer science in 2001, a Bachelors degree in computer science in 1999, both from the Tsinghua University, Beijing, China.

He is an assistant researcher in Media Computing Group, Microsoft Research Asia. His current research interests are in multimedia processing, pattern recognition and human head tracking and pose estimation.

About the Author—SHUICHENG YAN received the B.S. and Ph.D. from Applied Mathematics Department, School of Mathematical Sciences from Peking University, China in 1999 and 2004, respectively.

His research interests include computer vision, machine learning, and pattern recognition.

About the Author—LEI ZHANG received his B.S. and M.S. degrees in Computer Science from Tsinghua University in 1993 and 1995, respectively. After 2 years of working in industry, he returned to Tsinghua University and received his Ph.D. degree in Computer Science in 2001. Then he joined Media Computing group in Microsoft Research Asia as an associate researcher. His research interests include machine learning, content-based image retrieval and classification, image processing and computer vision.

About the Author—HONGJIANG ZHANG (F'03) received his Ph.D. from the Technical University of Denmark and his BS from Zhengzhou University, China, both in Electrical Engineering, in 1982 and 1991, respectively.

From 1992 to 1995, he was with the Institute of Systems Science, National University of Singapore, where he led several projects in video and image content analysis and retrieval and computer vision. From 1995 to 1999, he was a research manager at Hewlett–Packard Labs, where he was responsible for research and technology transfers in the areas of multimedia management; intelligent image processing and Internet media. In 1999, he joined Microsoft Research Asia, where he is currently a Senior Researcher and Assistant Managing Director in charge of media computing and information processing research.

Dr. Zhang is a member of ACM and a Senior Member of IEEE. He has authored 3 books, over 260 referred papers and book chapters, 7 special issues of international journals on image and video processing, content-based media retrieval, and computer vision, as well as over 45 patents or pending applications. He currently serves on the editorial boards of five IEEE/ACM journals and a dozen committees of international conferences.

About the Author—WEN GAO (M'99) received the M.S. degree and the Ph.D. degree in computer science from Harbin Institute of Technology, Harbin, China, in 1985 and 1988, respectively, and the Ph.D. degree in electronics engineering from the University of Tokyo, Tokyo, Japan, in 1991.

He was a Research Fellow at the Institute of Medical Electronics Engineering, the University of Tokyo, in 1992, and a Visiting Professor at the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, in 1993. From 1994 to 1995 he was a Visiting Professor at MIT Artificial Intelligence Laboratories. Currently, he is the Vice President of the University of Science and Technology of China, the Deputy President of Graduate School of Chinese Academy of Sciences, Professor in Computer Science at Harbin Institute of Technology, and Honor Professor in Computer Science at City University of Hong Kong. He has published 7 books and over 200 scientific papers. His research interests are in the areas of signal processing, image and video communication, computer vision and artificial intelligence.

Dr. Gao is the head of Chinese National Delegation to MPEG working group (ISO/SC29/WG11). He is the Editor-in-Chief of the Chinese Journal of Computer, and the general co-chair of the IEEE International Conference on Multi-model Interface in 2002.