

Recognizing Rotated Faces From Frontal and Side Views: An Approach Toward Effective Use of Mugshot Databases

Author

Zhang, Xiaozheng, Gao, Yongsheng, Leung, Maylor KH

Published

2008

Journal Title

IEEE Transactions on Information Forensics and Security

DOI

<https://doi.org/10.1109/TIFS.2008.2004286>

Copyright Statement

© 2008 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

Downloaded from

<http://hdl.handle.net/10072/22882>

Link to published version

<http://ieeexplore.ieee.org/servlet/opac?punumber=10206>

Griffith Research Online

<https://research-repository.griffith.edu.au>

Recognizing Rotated Faces From Frontal and Side Views: An Approach Toward Effective Use of Mugshot Databases

Xiaozheng Zhang, *Student Member, IEEE*, Yongsheng Gao, *Senior Member, IEEE*, and Maylor K. H. Leung, *Member, IEEE*

Abstract—Mug shot photography has been used to identify criminals by the police for more than a century. However, the common scenario of face recognition using frontal and side-view mug shots as gallery remains largely uninvestigated in computerized face recognition across pose. This paper presents a novel appearance-based approach using frontal and sideface images to handle pose variations in face recognition, which has great potential in forensic and security applications involving police mugshot databases. Virtual views in different poses are generated in two steps: 1) shape modelling and 2) texture synthesis. In the shape modelling step, a multilevel variation minimization approach is applied to generate personalized 3-D face shapes. In the texture synthesis step, face surface properties are analyzed and virtual views in arbitrary viewing conditions are rendered, taking diffuse and specular reflections into account. Appearance-based face recognition is performed with the augmentation of synthesized virtual views covering possible viewing angles to recognize probe views in arbitrary conditions. The encouraging experimental results demonstrated that the proposed approach by using frontal and side-view images is a feasible and effective solution to recognizing rotated faces, which can lead to a better and practical use of existing forensic databases in computerized human face-recognition applications.

Index Terms—Appearance-based recognition, face recognition, mug shot, police database, pose variation, virtual view synthesis, 3-D modelling.

I. INTRODUCTION

FACE recognition under pose variations is one of the key remaining problems in the research field of pattern recognition and computer vision [42], [62]. It is of great interest in many applications, most notably those dealing with indifferent or uncooperative subjects, for instance, in surveillance systems. Given the current state of technologies, however, computerized face recognition requires cooperative subjects who stay still in

a required pose (often frontal) to be checked [9]. The difficulty lies in the fact that the intraclass differences brought by pose variations are often larger than interclass differences for distinguishing different people. One of the most successful approaches for face recognition is the appearance-based approach [27]. In a pose-invariant face-recognition scenario, it compares an input (or probe) face in an arbitrary pose with a number of enrolled (or gallery) images per person in the database covering all possible viewing directions. The performance of appearance-based face recognition is largely dependent on the selection and availability of the gallery images.

Police mugshot databases¹ usually consist of a frontal view and a side view per person (Fig. 1), which are a primary source of gallery images for face recognition [55]. The police have used mug shots to identify criminals for more than a century [49] and there are many legacy mugshot databases which were routinely taken during the police booking process [31]. For computerized face recognition, however, frontal and side-view mug shots are not able to provide effective coverage of all possible conditions, when a probe image is in an arbitrary pose between the frontal and side-viewing angles. Although several pose-invariant face-recognition techniques [7], [24], [27] have been proposed recently, the common scenario of using frontal and side-view mug shots as a gallery remains largely uninvestigated. This research proposes a novel face-recognition approach which uses frontal and side-view face images as gallery and recognizes probe views in arbitrary poses, which has great potential in mugshot-related applications. The combination of frontal view and side-view balances the tractability and applicability of the proposed approach for face recognition across pose well. Compared to algorithms of face recognition from single images, multiple gallery views provide more information about the individual face so that the system can be more accurate and rely less on prior knowledge of an average human face. The proposed approach aims to make effective use of the frontal and side views of face pictures widely available in existing police mugshot databases and does not require additional acquisitions of gallery databases. It can easily fit into many face-recognition applications using frontal and side-view mug shots as a gallery for law enforcement and security surveillance, for example, in video-monitoring systems.

The proposed approach is appearance based, which synthesizes personalized virtual views in different poses from mugshot

Manuscript received October 09, 2007; revised April 21, 2008. Current version published November 19, 2008. This work was supported by the Australian Research Council (ARC) under Discovery Grant DP0451091. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Vijaya Kumar Bhagavatula.

X. Zhang and Y. Gao are with the Griffith School of Engineering, Griffith University, Nathan Qld 4111, Australia (e-mail: x.zhang@griffith.edu.au; yongsheng.gao@griffith.edu.au).

M. K. H. Leung is with the School of Computer Engineering, Nanyang Technological University, Singapore 639798 (e-mail: asmkleung@ntu.edu.sg).

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

Digital Object Identifier 10.1109/TIFS.2008.2004286

¹Mugshot databases in this paper refer to common police databases containing frontal and side views.



Fig. 1. Example of mugshot-style face pictures used as the gallery in the proposed approach.

gallery views and recognizes rotated probe views using face-recognition algorithms augmented with the synthesized virtual views. First, a hierarchical multilevel variation minimization approach reconstructs the personalized 3-D face shapes from two gallery views. Then, the facial textures are estimated considering the diffuse and specular reflections using the Phong reflection model. Virtual views are synthesized by rotating the reconstructed face shapes and performing pixel-wise texture synthesis to cover possible viewing angles. Appearance-based recognition of face views in arbitrary poses is finally performed using the synthesized virtual views as gallery images. The initial idea and early work of this research have been published in part as conference papers in [57], [58]. In this paper, we report the complete and improved formulation, thorough investigation and extended experimental evaluation of our system.

A. Related Works

Previous works show that it is realistic yet challenging to perform appearance-based face recognition across pose using synthesized virtual views [5], [22]. Two-dimensional techniques [5], [10], [14], [34], [48], [60] and 3-D approaches [7], [11], [22], [30] were used to predict the appearance of a human face in different poses. Beymer and Poggio [5] mapped a 2-D facial transformation obtained from a sample face rotation onto a gallery face view to generate virtual views under novel viewing conditions. Cootes *et al.* [14] proposed a 2-D statistical method to approximate facial appearances under novel viewing conditions from three gallery face images. González-Jiménez and Alba-Castro [24] extracted pose parameters from interpersonal variations of 2-D facial transformation in parallel deformation [5] using the point distribution model [36]. Despite these innovative attempts to synthesise virtual views using 2-D techniques, researchers argue that the pose variation is a 3-D transformation rather than a 2-D warping. In addition to the complex nonconvex face shape, virtual views from different viewpoints cannot be correctly synthesized by 2-D manipulations. This probably explains why only small rotation angles (15° in [5], 22.5° in [14], and 31° in [24]) can be properly handled in 2-D techniques for virtual view synthesis.

3-D approaches [7], [22], [23], [30] have advantages over 2-D techniques in synthesizing rotated virtual views under

large-angle rotations. In [30], a generic model was used as prior knowledge of the human face shape and facial features are detected as the constraints to deform the generic model to form an individualized human face shape. One of the most attractive approaches in this category is the 3-D morphable model proposed by Blanz and Vetter [7]. It successfully projected face shapes and textures onto eigenspaces to represent the identity using prior knowledge of human faces. The use of single gallery views makes their algorithm rely heavily on the pixel intensities, and consequently, the face modelling process is highly nonlinear and unstable [44], [45]. Georgiades *et al.* [22] introduced a generative 3-D illumination cone model trained from seven gallery images to synthesize virtual views under various illuminations and poses. Gross [27] proposed learning an eigenlight field from 2-D face images to perform face recognition. These two studies were based on assumptions that human face surfaces are Lambertian surfaces without considering specular reflections. In [23], the 3-D illumination cone model in [22] was patched with the consideration of a simplified Torrance-Sparrow model to resolve the issues associated with Lambertian assumption (e.g., bas-relief ambiguity [4]). It is assumed that the specular reflection component was homogeneous across the whole face surface, which is a better, but not as realistic modelling of real face surfaces.

In the virtual view synthesis of the proposed approach, the personalised 3-D face shape is constructed using the interrelations of facial features on the two gallery views, enabling a 3-D face modelling procedure without using a generic shape model. Since the shape modelling process does not directly use the pixel intensities to estimate face shapes, it does not require the large number of gallery views to resolve intractability as used in [22] and [23]. The estimation of the face surface properties is based on the Phong reflection model, which considers diffuse and specular reflectivities. The proposed face-recognition system consists of three steps, that is: 1) a hierarchical 3-D face modelling step which refines the shape reconstruction at multiple resolution levels; 2) a surface reflectance analysis and virtual texture synthesis step which estimates and refines facial textures from the pixel intensities of the two gallery images and synthesizes virtual images under arbitrary viewing conditions; and 3) an appearance-based face-recognition step augmented with synthesized virtual views in different viewing angles. The proposed approach is different from existing methods in the following aspects. First, it combines frontal view and side-view images for virtual face view synthesis and face recognition, which targets the effective use of existing police mugshot databases. Compared to other methods using a single gallery image per person, such as [7] and [32], the proposed approach is capable of estimating the shape and texture information more reliably and less dependent on prior knowledge of the 3-D human faces. Second, compared to 2-D approaches [5], [14], [24], the proposed 3-D modelling process is capable of handling large viewpoint changes (even vertical rotations). Third, this approach does not require any 3-D generic face model. In the process of 3-D shape modelling, the reconstructed face shape is personalized, which is free from prior knowledge of human faces belonging to other identities. Through this process, the discriminative shape information for face recognition can be well pre-

served. Last but not least, the texture synthesis considers spatially variable diffuse and specular reflectance using the Phong model, other than direct mapping neglecting specular reflection [5], [27], [32] or simple interpolation of pixel values [14], [30].

The main focus of this research is to handle pose variations in face recognition and this paper made no attempt to handle the variations of other conditions in face recognition, such as illumination variations and expression variations. For face-recognition techniques under illumination variations and expression variations, readers can refer to [2], [12], [20], [21], [35], [38], [45], [46], [63], as well as [8], [40], [41], [53], [56], and [59], respectively. The rest of this paper is organized as follows. In Section II, the proposed hierarchical 3-D shape modelling approach is elaborated. Texture analysis and virtual view synthesis considering spatially variable diffuse and specular reflectivities are described in Section III. In Section IV, virtual views in arbitrary viewing conditions are synthesized using the generated personal face models. Augmented face recognition using the synthesized virtual views is conducted on the CMU-PIE database and compared to benchmark algorithms. Finally, a conclusion is drawn in Section V.

II. 3-D SHAPE MODELLING

The goal of 3-D face-shape modelling is to achieve an accurate and smooth 3-D face model from the frontal and side-view gallery images. The surface smoothness reduces the texture estimation errors by providing sound normal directions of the 3-D shape for the texture analysis in Section III. The proposed approach hierarchically estimates the 3-D shape of the personalized face from a set of constraining points on the gallery views. In each level of 3-D shape estimation, it simultaneously models and smoothes the face surfaces by minimizing the surface variations between constraining points and nonconstraining points. Table I lists the notions and symbols used in the proposed approaches for 3-D face-shape modelling and reflectance analysis and synthesis.

A. Single-Level 3-D Shape Reconstruction

Suppose a 3-D face shape is formulated as a depth graph $\vec{z} = z(x, y)$, where (x, y) represents a pixel position on the frontal image plane (x - y plane). The side view is on the y - z plane. Given a set of q constraining points specified on frontal and side-view face images, the coordinates of the i th constraining points on the frontal image and side image are denoted as $(x_{c,i}, y_{c,i})$ and $(z_{c,i}, y_{c,i})$, respectively. The location of the i th constraining point in the 3-D space is then $(x_{c,i}, y_{c,i}, z_{c,i})$. Denote the constraining point set as $\Pi = \{(x_{c,i}, y_{c,i}) | 1 \leq i \leq q\}$, and the depth information of the i th constraining point is expressed as $c(x_{c,i}, y_{c,i}) = z_{c,i}$. The 3-D shape reconstruction is performed by minimizing the surface roughness

$$E(\vec{z}) = \int \int_{(x,y)} \left\{ \left(\frac{\partial^2 \vec{z}}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 \vec{z}}{\partial x \partial y} \right) + \left(\frac{\partial^2 \vec{z}}{\partial y^2} \right)^2 \right\} dx dy \quad (1)$$

subject to $z(x, y) = c(x, y), \forall (x, y) \in \Pi$, [26], [57]. If the initial surface is

$$z(x, y) = \begin{cases} c(x, y), & \text{if } (x, y) \in \Pi \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

TABLE I
NOTIONS AND SYMBOLS USED IN THE PROPOSED ALGORITHMS

3D Shape modelling (Section II)	
I_f, I_s	Frontal and side view image intensities
n	Image width (=height)
$z(x, y)$	Height function of the 3D face model
(x_c, y_c, z_c)	Constraining point
$c(x_c, y_c)$	Height function of constraining point
m	Number of shape modelling levels
h_l	Spacing at level l
$z_{l,0}(x, y)$	Initial shape at level l
$z_l^*(x, y)$	Optimised shape at level l
Φ_l	Control lattice at level l
Π_l	Constraining set at level l
E_s	Cost function of shape modelling
Reflectance analysis (Section III)	
\vec{N}	Normal unit vector
\vec{R}	Reflecting unit vector
\vec{L}	Lighting unit vector
H_a, H_d, H_s	Ambient, diffuse, and specular reflection components
γ_f, γ_s	Frontal and side view texture refinement coefficients
E_t	Cost function of texture analysis
I_N	Image intensity of the novel view

the disconnection around Π results in a large E . The discrete form of (1) is expressed as

$$E_s(\vec{z}) = \sum_{x=1}^{n-2} \sum_{y=0}^{n-1} (z(x+1, y) - 2z(x, y) + z(x-1, y))^2 + \sum_{x=0}^{n-1} \sum_{y=1}^{n-2} (z(x, y+1) - 2z(x, y) + z(x, y-1))^2 + 2 \sum_{x=0}^{n-2} \sum_{y=1}^{n-2} (z(x, y) - z(x-1, y) - z(x, y-1) + z(x-1, y-1))^2 \quad (3)$$

where n is the width (and height) of the frontal image. To perform 3-D shape reconstruction based on the constraining points, (3) is minimized under the constraint of $z(x, y) - c(x, y) = 0$. The optimization is a quadratic nonlinear programming problem and the optimal set of z can be approximated using an iterative gradient projection method [39].

The process of improving the approximation from an initial vector is as follows.

- 1) At the k th iteration, initial \vec{z}_k is the result of the $(k-1)$ th iteration.
- 2) Construct the constraint matrix A_c , where A_c is a $q \times n^2$ matrix with rows consisting of a 1 in the position corresponding to the location in the frontal image (x, y) for $(x, y) \in \Pi$.
- 3) Calculate the projection matrix P_c as $P_c = I - A_c^T A_c$, which consists of all 1s except for diagonal elements in those rows corresponding to a constraining point, with such elements being 0s.
- 4) Calculate the steepest gradient direction \vec{d} as $\vec{d} = -P \nabla E(\vec{z}_k)^T$, along which a better estimate is found and fastest minimization of the E is achieved.

- 5) Find the optimal solution along \vec{d} by calculating α_1 and α_2 as follows, respectively

$$\max_{\alpha_1} \left\{ \alpha | \vec{z}_k + \alpha \vec{d} \text{ is feasible} \right\} \quad (4)$$

$$\min_{\alpha_2} \left\{ E(\vec{z}_k + \alpha \vec{d}) | 0 \leq \alpha \leq \alpha_1 \right\} \quad (5)$$

- 6) Update the estimate \vec{z}_{k+1} by $\vec{z}_{k+1} = \vec{z}_k + \alpha_2 \vec{d}$.

The initial \vec{z}_0 can be any shape which satisfies the constraint. Without loss of generality, (2) is used to provide an initial state of the surface \vec{z} , in which all surface points are at 0 heights except for the constraining points with their heights being $c(x, y)$, where $(x, y) \in \Pi$. The iteration continues to improve the reconstructed surface, until the change is smaller than the threshold ε_s as

$$E_s(\vec{z}_k) - E_s(\vec{z}_{k+1}) < \varepsilon_s. \quad (6)$$

B. Multilevel 3-D Shape Reconstruction

In the single-level 3-D shape reconstruction process, the neighboring points of a constraining point will move gradually toward the constraining point along the z -axis by minimizing surface roughness E . This minimization requires a dense set of constraining points which is infeasible for 3-D face-shape reconstruction, because the facial features (facial components) used as constraining points are, at best, sparse points on the relatively large images of human faces.

Under this circumstance, a coarse-to-fine hierarchy is introduced to resolve the problem with sparse features in the single-level 3-D surface reconstruction. Its hierarchically refines the estimations of the 3-D shape by inheriting more accurate initial data from the previous levels of reconstruction. Suppose the resolution of the frontal gallery image (x - y plane) is $n \times n$, where n is of the powers of 2. A coarse-to-fine hierarchy of control lattices $\Phi_0, \Phi_1, \dots, \Phi_m$ is applied onto the frontal image to generate a sequence of surface reconstructions at different resolution levels, where Φ_0 is the coarsest control lattice and Φ_m is the finest control lattice (i.e., the actual image resolution). As shown in Fig. 2, the l th lattice Φ_l has a resolution of $(2 \cdot 2^l) \times (2 \cdot 2^l)$ grids, and each grid consists of $h_l \times h_l$ pixels in the frontal image, where $h_l = (n/2 \cdot 2^l)$.

For the coarsest lattice Φ_0 , the resolution is 2×2 and $h_0 = (n/2)$, while for the finest lattice Φ_m , the resolution is $n \times n$ and $h_m = 1$. m can be determined from the size of the frontal image as $m = \log_2 n - 1$. At control lattice Φ_l , where $l \in [0, m]$, pixel (x, y) on the frontal image is mapped to grid (u_l, v_l) , where $u_l = \lfloor x/h_l \rfloor$ and $v_l = \lfloor y/h_l \rfloor$. ($\lfloor a \rfloor$ denotes the floor function that returns the largest integer less than or equal to a .) The constraining pointset Π_l defined on Φ_l becomes

$$\Pi_l = \left\{ \left(\left\lfloor \frac{x}{h_l} \right\rfloor, \left\lfloor \frac{y}{h_l} \right\rfloor \right) \mid (x, y) \in \Pi \right\} \quad (7)$$

and the heights are converted into the new coordinate (u_l, v_l) as $c_l(u_l, v_l) = c(x, y)$, if $(x, y) \in \Pi$, where $(u_l, v_l) = (\lfloor x/h_l \rfloor, \lfloor y/h_l \rfloor)$.

Based on the converted constraining points, a single-level 3-D shape reconstruction can be performed from an initial estimate

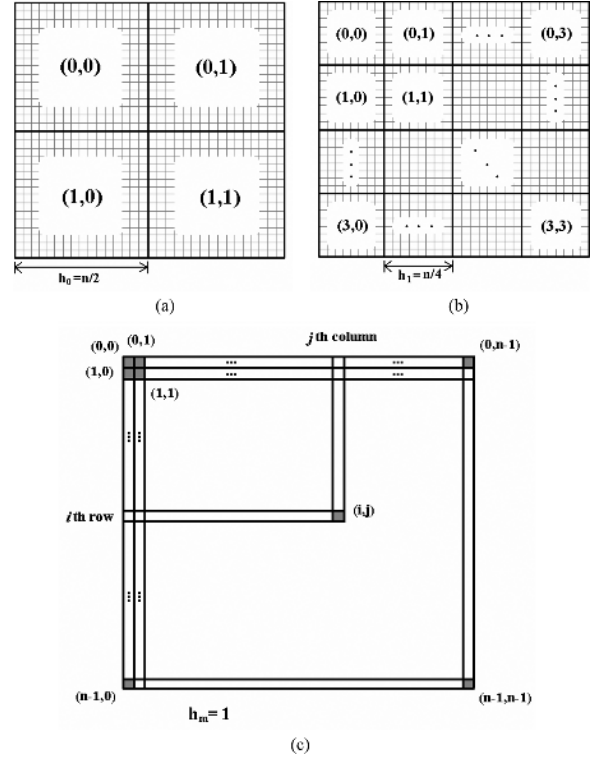


Fig. 2. Control lattices in the proposed multilevel variation minimization for 3-D shape reconstruction. (a) The coarsest lattice Φ_0 , (b) Φ_1 , and (c) the finest lattice Φ_m , where the lattice is the actual frontal image.

by minimizing a quadratic variation of the surface \vec{z}_l as

$$\begin{aligned} E_s(\vec{z}_l) = & \sum_{u_l=1}^{2 \cdot 2^l - 2} \sum_{v_l=0}^{2 \cdot 2^l - 1} (z_l(u_l + 1, v_l) - 2z_l(u_l, v_l) \\ & + z_l(u_l - 1, v_l))^2 \\ & + \sum_{u_l=0}^{2 \cdot 2^l - 1} \sum_{v_l=1}^{2 \cdot 2^l - 2} (z_l(u_l, v_l + 1) - 2z_l(u_l, v_l) \\ & + z_l(u_l, v_l - 1))^2 \\ & + 2 \sum_{u_l=0}^{2 \cdot 2^l - 2} \sum_{v_l=1}^{2 \cdot 2^l - 2} (z_l(u_l, v_l) - z_l(u_l - 1, v_l) \\ & - z_l(u_l, v_l - 1) \\ & + z_l(u_l - 1, v_l - 1))^2 \end{aligned} \quad (8)$$

subject to $z_l(u_l, v_l) = c_l(u_l, v_l)$. The initial $\vec{z}_{l,0}$ is inherited from the estimated result \vec{z}_{l-1}^* on the previous control lattice Φ_{l-1} as

$$z_{l,0}(u_l, v_l) = z_{l-1}^* \left(\left\lfloor \frac{u_l}{2} \right\rfloor, \left\lfloor \frac{v_l}{2} \right\rfloor \right). \quad (9)$$

The initial $\vec{z}_0^* = \{z_0^*(u_0, v_0)\}$ for the coarsest control lattice Φ_0 is expressed as

$$z_0^*(u_0, v_0) = \begin{cases} c_0(u_0, v_0), & \text{if } (u_0, v_0) \in \Pi_0 \\ 0, & \text{otherwise.} \end{cases} \quad (10)$$

Fig. 3 shows a sequence of shape reconstruction results from the frontal and side views of Fig. 1 with different control lattices.

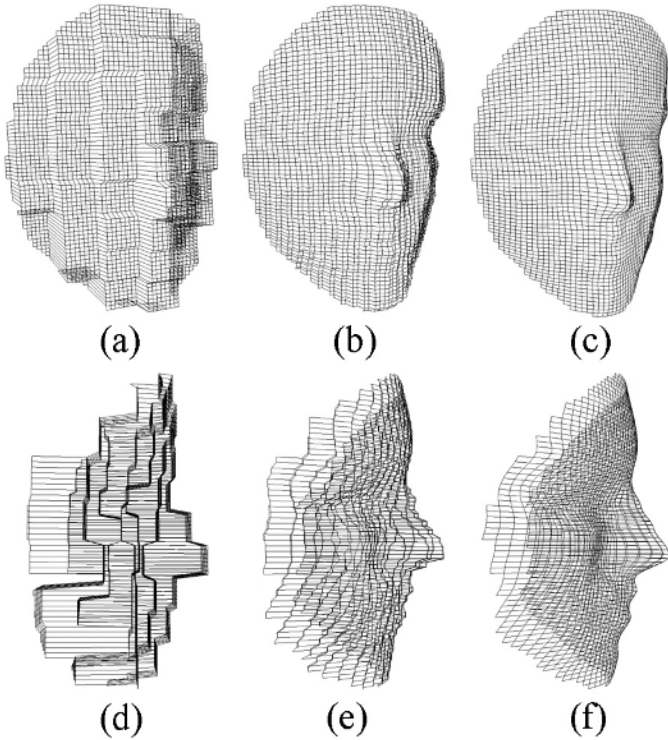


Fig. 3. The 3-D face shape reconstructed with a coarse-to-fine hierarchy. Reconstructed face shapes with a 16×16 control lattice in 30° rotation (a) and 90° rotation (d), with a 64×64 control lattice in 30° rotation (b) and 90° rotation (e), and with a 512×512 control lattice (the finest) in 30° rotation (c), and 90° rotation (f).

The proposed 3-D shape modelling approach requires a set of facial features specified on frontal and side views to be used as constraining points. A number of automatic algorithms for features location or detection have been proposed recently, such as the face vectorizing algorithm [6]; active shape model (ASM) [13], [36], [52]; Bayesian shape model (BSM) [54]; active contours [25], [29]; morphable models [33], [51]; and 2-D shape clustering [17], which could be incorporated into the proposed framework to automatically provide the facial features on face images. On the other hand, manual facial feature specifications are also widely used in various face modelling and recognition research [15], [24], [37]. In this research, facial features are specified interactively on the gallery views (Fig. 4). They are eyes, eyebrows, nose profiles, lips, and ears, which are ASM control points (red in Fig. 4) and interpolated points (yellow in Fig. 4) which are points between control points [61]. Since only half of the face is visible in the side view, facial features can only be located on one side of the face. Based on the assumption that a human face is bilateral symmetric as in [5] and [19], the feature locations can be mirrored to the other half. In our experiments, 80 corresponding facial features were manually labelled in frontal and side views for each person as constraining points in the proposed 3-D face modelling step.

III. REFLECTANCE ANALYSIS AND RECONSTRUCTION OF FACE SURFACES

The textural information of human faces is critical to the performance of face-recognition systems. A realistic and accurate

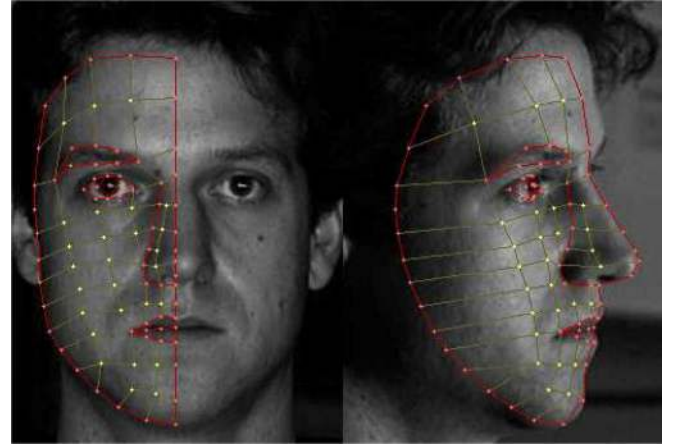


Fig. 4. Facial feature specification on the frontal and side gallery views.

texture mapping of virtual face views in arbitrary viewing conditions can greatly improve the recognition accuracy. From the two gallery images of the frontal and side views, the surface textural properties need to be estimated and are then used to synthesize virtual textures of human faces in the virtual face views. In this section, a stable reflectance reconstruction approach based on Phong reflection model is proposed to estimate the human face textural properties and to synthesize virtual views in arbitrary conditions. First, a uniform set of textural properties for Phong reflection model is estimated from the textures of the two gallery views. Then, corresponding to specific viewing conditions, spatially varying textural property coefficients are refined and estimated from the actual pixel intensities of the gallery images to represent the texture details of the human faces for recognition.

A. Phong Reflection Model for Human Skins

In the proposed reflectance analysis step, facial textures are analyzed from the image intensities of gallery views based on the Phong reflection model [43], since human faces exhibit diffuse and specular reflections [16]. In the Phong model, the image intensity (reflected intensity) is expressed as

$$I = k_a I_a + k_d I_{po}(\vec{L} \cdot \vec{N}) + k_s I_{po}(\vec{R} \cdot \vec{V})^n \quad (11)$$

where k_a , k_d , and k_s are the ambient, diffuse, and specular reflectance, respectively; I_a and I_{po} are the ambient and directional incident light intensities; \vec{L} is the direction of I_{po} ; \vec{N} is the direction of the surface normal; \vec{V} is the viewing direction; n is the specular exponent; and \vec{R} is the reflecting direction corresponding to \vec{L} which is calculated as

$$\vec{R} = \begin{cases} 2(\vec{N} \cdot \vec{L})\vec{N} - \vec{L}, & (\vec{N} \cdot \vec{L}) > 0 \\ 0, & (\vec{N} \cdot \vec{L}) \leq 0. \end{cases} \quad (12)$$

In this paper, the light source is limited to be a single point light from infinity. This assumption is often used (e.g., [7], [22], [23]) and considered as a reasonable approximation, because in taking mug shots, the major light source is the camera flash. If the camera is far enough from the face compared to the size of the face, it can be approximated as a light source from infinity.

Thus, the lighting intensities I_a and I_{po} and the lighting direction \vec{L} in (11) are constant for all of the surface points. n is fixed [7], [14], [28] as 2 [30]. The normal \vec{N} of a given point on the face surface is calculated from the results of the previous step in 3-D shape reconstruction using B-spline surface approximation on 16 neighboring points on the face as [18, Ch. 11]

$$\vec{N} = \frac{\partial z}{\partial x} \times \frac{\partial z}{\partial y}. \quad (13)$$

Though the lighting intensities I_a and I_{po} are unknown, they are constant values at every point on the face surface. By denoting $I_a k_a$ as H_a , $I_{po} k_d$ as H_d and $I_{po} k_s$ as H_s , (11) becomes

$$I = H_a + H_d(\vec{L} \cdot \vec{N}) + H_s(\vec{R} \cdot \vec{V})^2. \quad (14)$$

B. Reflection Component Estimation from the Two Gallery Views

Given a set of frontal and side views under a parallel flashlight \vec{L} , we have $\vec{L} = \vec{V}_F$ for the frontal view and $\vec{L} = \vec{V}_S$ for the side view. To estimate the three reflection components, the frontal face view and the sideface view are both used. For the i th point, (11) becomes

$$I_F^i = H_a^i + H_d^i(\vec{V}_F \cdot \vec{N}^i) + H_s^i(\vec{R}_F^i \cdot \vec{V})^2 \quad (15)$$

for the corresponding pixel on the frontal view, and

$$I_S^i = H_a^i + H_d^i(\vec{V}_S \cdot \vec{N}^i) + H_s^i(\vec{R}_S^i \cdot \vec{V})^2 \quad (16)$$

for the corresponding pixel on the side view. The reflecting direction \vec{R}_F^i and \vec{R}_S^i are calculated as

$$\vec{R}_F^i = \begin{cases} 2(\vec{N}^i \cdot \vec{V}_F) \vec{N}^i - \vec{V}_F, & (\vec{N}^i \cdot \vec{V}_F) > 0 \\ 0, & (\vec{N}^i \cdot \vec{V}_F) \leq 0 \end{cases} \quad (17)$$

and

$$\vec{R}_S^i = \begin{cases} 2(\vec{N}^i \cdot \vec{V}_S) \vec{N}^i - \vec{V}_S, & (\vec{N}^i \cdot \vec{V}_S) > 0 \\ 0, & (\vec{N}^i \cdot \vec{V}_S) \leq 0 \end{cases} \quad (18)$$

respectively.

All surface points are initialized homogeneously and the set of reflection components are denoted as H_a , H_d , and H_s . For the whole images of the frontal and side gallery views, we have

$$\begin{cases} I_F^1 = H_a + H_d(\vec{V}_F \cdot \vec{N}^1) + H_s(\vec{R}_F^1 \cdot \vec{V})^2 \\ \vdots \\ I_F^p = H_a + H_d(\vec{V}_F \cdot \vec{N}^p) + H_s(\vec{R}_F^p \cdot \vec{V})^2 \\ I_S^1 = H_a + H_d(\vec{V}_S \cdot \vec{N}^1) + H_s(\vec{R}_S^1 \cdot \vec{V})^2 \\ \vdots \\ I_S^p = H_a + H_d(\vec{V}_S \cdot \vec{N}^p) + H_s(\vec{R}_S^p \cdot \vec{V})^2 \end{cases} \quad (19)$$

where p is the number of points. Equation (19) is solved in a least square sense and the objective function to be minimized is expressed as

$$\begin{aligned} E_t(H_a, H_d, H_s) &= \sum_{i=1}^p \left[H_a + H_d(\vec{V}_F \cdot \vec{N}^i) + H_s(\vec{R}_F^i \cdot \vec{V})^2 - I_F^i \right]^2 \\ &+ \sum_{i=1}^p \left[H_a + H_d(\vec{V}_S \cdot \vec{N}^i) + H_s(\vec{R}_S^i \cdot \vec{V})^2 - I_S^i \right]^2. \end{aligned} \quad (20)$$

To ensure that all of the reflection components H_a , H_d , and H_s are semipositive, (20) is minimized subject to

$$\begin{cases} H_a \geq 0 \\ H_d \geq 0 \\ H_s \geq 0 \end{cases} \quad (21)$$

The optimization is again a quadratic nonlinear programming problem and the optimal set of (H_a^*, H_d^*, H_s^*) can be estimated by using the same iterative gradient projection method as used in the step of 3-D shape reconstruction. Compared to the nonlinear programming problem in the previous 3-D shape reconstruction step, the dimension of the variables is reduced to 3 and the constraints are no longer constantly active (e.g., when $H_a > 0$, the constraint $H_a \geq 0$ is inactive). Accordingly, the process of the reflection component estimation is as follows.

- Step 1) At the k th iteration, initial $(H_{a,k}, H_{d,k}, H_{s,k})$ is obtained from the $(k-1)$ th iteration.
- Step 2) From $(H_{a,k}, H_{d,k}, H_{s,k})$, determine the active constraint set as $M_t = \{H_{\tau,k} | H_{\tau,k} = 0, (\tau = a, d, s)\}$, which is the subset of the constraint set defined only by the active constraints.
- Step 3) Construct the constraint matrix A_t , where A_t is a $q_t \times 3$ matrix with rows consisting of a 1 in the position corresponding to the location of an active constraint.
- Step 4) Calculate the projection matrix P_t as $P_t = I - A_t^T (A_t A_t^T)^{-1} A_t$, which consists of all 1s except for diagonal elements in those rows corresponding to an active constraint, with such elements being 0s.
- Step 5) Calculate the steepest gradient direction \vec{d}_t as $\vec{d}_t = -P_t \nabla E_t(H_a, H_d, H_s)^T$, along which a better estimate can be found and fastest minimization of the E_t can be achieved.
- Step 6) If $\vec{d}_t \neq \vec{0}$, go to Step 7). Otherwise, go to Step 9).
- Step 7) Find the optimal solution along \vec{d}_t by calculating $\alpha_{t,1}$ and $\alpha_{t,2}$ as follows, respectively

$$\max_{\alpha_{t,1}} \left\{ \alpha_t | (H_{a,k}, H_{d,k}, H_{s,k})^T + \alpha_t \vec{d}_t \text{ is feasible} \right\} \quad (22)$$

$$\min_{\alpha_{t,2}} \left\{ E_t[(H_{a,k}, H_{d,k}, H_{s,k})^T + \alpha \vec{d}_t] | 0 \leq \alpha \leq \alpha_{t,1} \right\}. \quad (23)$$

- Step 8) Update the estimate as $(H_{a,k+1}, H_{d,k+1}, H_{s,k+1})^T = (H_{a,k}, H_{d,k}, H_{s,k})^T + \beta_2 \vec{d}_t$, and return to Step 2).
- Step 9) Calculate the constraint projection direction vector as

$$\vec{\beta}_{t,k} = - (A_q A_q^T)^{-1} A_q \nabla E_t(H_a, H_d, H_s)^T \quad (24)$$

which indicates the potential descent direction of the active constraint.

- If $\vec{\beta}_{t,k} \geq \vec{0}$, the iterations reach their optimal estimate and stop.
- Otherwise, find the row i_β corresponding to the inequality with the most negative component of $\vec{\beta}_{t,k}$ as $i_\beta = \min_{i_\beta} \{ \beta_{t,k}^i \}$ and delete the i_β th row from A_q as

$$A_{q,k+1} = \begin{bmatrix} I_{(i_\beta-1) \times (i_\beta-1)} & 0 & 0 \\ 0 & 0 & I_{(q_t-i_\beta-1) \times (q_t-i_\beta-1)} \end{bmatrix} A_{q,k} \quad (25)$$

and return to Step 4) to search for a new descent direction.

The initial parameter vector is chosen as

$$(H_{a,0}, H_{d,0}, H_{s,0})^T = (1, 1, 1)^T \quad (26)$$

where all three constraints are inactive. The optimization is terminated when the progress is smaller than the threshold ε_t as $E_{t,k} - E_{t,k+1} < \varepsilon_t$ or it reaches Step 9(a). The optimal estimate is denoted as (H_a^*, H_d^*, H_s^*) .

C. Reflection Component Refinement and Virtual Texture Synthesis

Since face surfaces have spatially variable reflectance properties, each surface point has a unique set of reflection components, i.e., pixel-wise reflection components (H_a^i, H_d^i, H_s^i) are expected from the uniform estimate (H_a^*, H_d^*, H_s^*) . To estimate the pixel-wise reflection components, the difference of the generated pixel intensity from (H_a^*, H_d^*, H_s^*) and the real pixel intensity from the gallery views is used. From (H_a^*, H_d^*, H_s^*) and the reconstructed 3-D face shape, the pixel intensities of the i th pixel on the frontal and side views are rendered as

$$I_F^i = H_a^* + H_d^* (\vec{V}_F \cdot \vec{N}^i) + H_s^* (\vec{R}_F^i \cdot \vec{V})^2 \quad (27)$$

and

$$I_S^i = H_a^* + H_d^* (\vec{V}_S \cdot \vec{N}^i) + H_s^* (\vec{R}_S^i \cdot \vec{V})^2 \quad (28)$$

respectively. The corresponding pixel intensities from the real gallery frontal and side views are I_F^i and I_S^i , respectively. The refinement coefficients γ_F^i and γ_S^i for the frontal view and the side view are defined as $\gamma_F^i = I_F^i / I_F^i$ and $\gamma_S^i = I_S^i / I_S^i$.

By multiplying the frontal-view refinement coefficient γ_F^i to the optimized uniform reflection components (H_a^*, H_d^*, H_s^*) , the refined reflection components can synthesize the exact pixel intensity in the same condition of the frontal gallery view, which are calculated as

$$\begin{aligned} H_{a,F}^i &= \gamma_F^i H_a^*, & H_{d,F}^i &= \gamma_F^i H_d^*, \\ H_{s,F}^i &= \gamma_F^i H_s^*. \end{aligned} \quad (29)$$

Similarly, using the side-view refinement coefficient γ_S^i , a set of pixel-wise reflection components can be refined as $H_{a,S}^i = \gamma_S^i H_a^*$, $H_{d,S}^i = \gamma_S^i H_d^*$, $H_{s,S}^i = \gamma_S^i H_s^*$, to achieve exact pixel intensity on the side gallery view. Ideally, the two sets of parameters are equal, because they are the intrinsic parameters of the face surface of the same point. In practice, however, a number of imperfections may cause the two sets to be different. For example, the inaccuracy of the 3-D shape reconstruction can result in inaccurate pixel correspondence where the corresponding pixels on the frontal and side views are actually two different points on the face surface. Given a viewing and lighting condition $(\vec{V}_n$ and $\vec{L}_n)$, the novel virtual textures are rendered from the estimated reflection components. If $(\vec{V}_F \cdot \vec{V}_N) \geq (\vec{V}_S \cdot \vec{V}_N)$, the frontal-view-based refined reflection components are applied in the Phong reflectance model as

$$I_N^i = H_{a,F}^i + H_{d,F}^i (\vec{L}_N \cdot \vec{N}^i) + H_{s,F}^i (\vec{R}_N^i \cdot \vec{V}_N)^2 \quad (30)$$

where

$$\vec{R}_N^i = \begin{cases} 2(\vec{N}^i \cdot \vec{L}_N) \vec{N}^i - \vec{L}_N, & (\vec{N}^i \cdot \vec{L}_N) > 0 \\ 0, & (\vec{N}^i \cdot \vec{L}_N) \leq 0. \end{cases} \quad (31)$$

If $(\vec{V}_F \cdot \vec{V}_N) < (\vec{V}_S \cdot \vec{V}_N)$, the side-view-based refined reflection components are used as

$$I_N^i = H_{a,S}^i + H_{d,S}^i (\vec{L}_N \cdot \vec{N}^i) + H_{s,S}^i (\vec{R}_N^i \cdot \vec{V}_N)^2. \quad (32)$$

The virtual textures are then mapped to the rotated 3-D face shape and the virtual face views are synthesized. The entire process for the virtual view synthesis from the frontal and side view mugshot-style gallery images is illustrated in Fig. 5.

IV. EXPERIMENTS AND RESULTS

The proposed approach for recognizing rotated faces using mug shots was tested on the subset of the CMU-PIE database [47] consisting of 884 face images of 68 persons under 13 different poses (labelled as pose 02, 05, 07, 09, 11, 14, 22, 25, 27, 29, 31, 34, and 37, refer to Fig. 6). The viewing directions of these poses are listed in Table II. The frontal view (pose 27) and side view (pose 22) of each person were selected as the mug shots to provide gallery views for the proposed system. The personalized face models were reconstructed and the virtual views under different viewing conditions were synthesized. For performance evaluation, the most widely used holistic matching method, principle component analysis (PCA) [50], and the most appealing local matching approach, local binary patterns (LBP) [1], were selected as the classifiers to build face-recognition systems. It is interesting to further test the performance of the proposed method on larger databases in the future when they are available.

A. Virtual View Synthesis

Given a set of mug shots, the 3-D shape and texture information of a person's face can be estimated by using the proposed approach from the mug shots. Personalized virtual views under novel variations in pose can be synthesized from the estimated

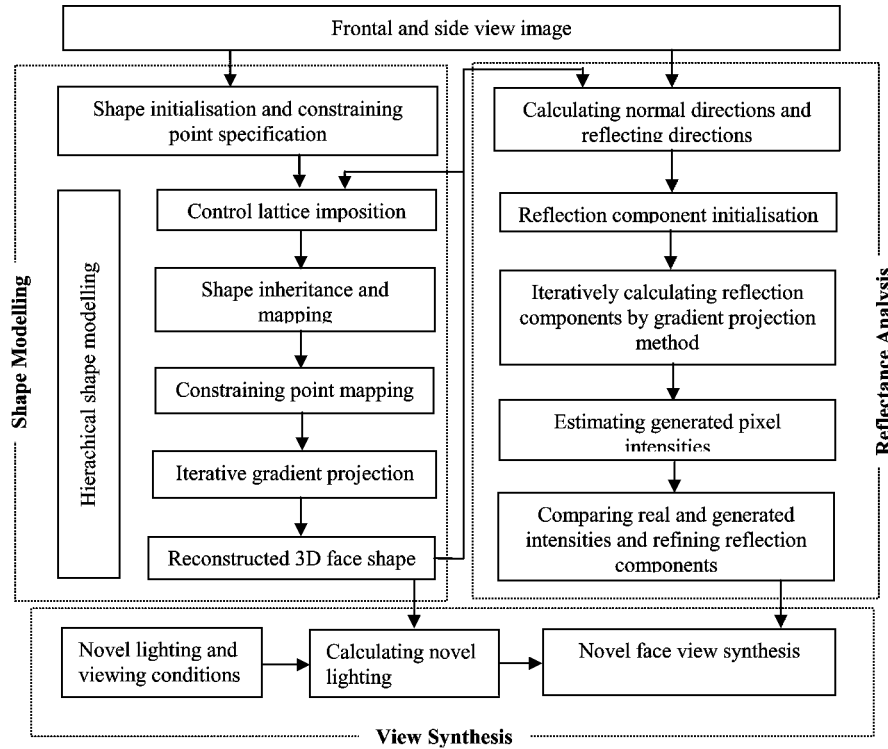


Fig. 5. Process of shape modelling, reflectance analysis, and view synthesis.

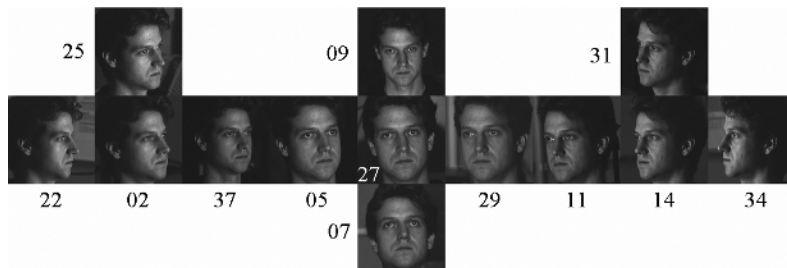


Fig. 6. Example of the 13 poses from the CMU-PIE database [47] used in the experiments.

parameters using Algorithm 2. Fig. 7 shows examples of virtual face views synthesized in our experiments. A comparison between the real probe view of the face and a virtual view generated by our algorithm is shown in Fig. 8. Note that these images were synthesized from the frontal and side views in the mugshot databases.

B. Face Recognition (Identification) Experiments

Using principal component analysis (PCA) [50] and local binary patterns (LBP) [1] as classifiers representing two major schemes of 2-D appearance-based face recognition (i.e., holistic and local matching schemes), the following six strategies were employed in our identification experiments. The first three strategies were used as baselines for comparison purposes, while the last three strategies used the proposed approach to augment the face recognition under pose variations. For PCA, the available gallery views were used for training to construct the eigenfeature space and the dimensionality of the feature space was chosen to be 60. The image resolution was 160×160

and the three most significant principal components in projected eigenvectors were discarded to yield better recognition performance [3]. For LBP, the image resolution was 80×80 and the standard (8,1) neighborhood was chosen. The face images were divided into 20×20 regions, the histograms were quantized to 32 bins, and the classification was based on the unweighted chi-square method. For PCA and LBP, the nearest neighbor (NN) has been used as the classification scheme when the gallery includes more than one view per person (e.g., B2, A2, and AA).

- 1) **B2**: The two real views per person [Fig. 9(a)] were used as gallery views and the other rotated real views [Fig. 9(c)], labelled as 02, 05, 07, 09, 11, 14, 25, 29, 31, 34, and 37, were used as probe views.
- 2) **BF**: The frontal real views [Fig. 9(a)-“27”] were used as gallery views and the rotated real views [Fig. 9(c)] were used as probe views.
- 3) **BS**: The side real views [Fig. 9(a)-“22”] were used as gallery views and the rotated real views [Fig. 9(c)] were used as probe views.

TABLE II
ROTATION ANGLES OF THE DIFFERENT POSES IN THE CMU-PIE DATABASE

Pose	02	05	07	09	11	14	22	25	27	29	31	34	37
Horizontal	-44°	-16°	0°	0°	32°	47°	-62°	-44°	0°	17°	47°	66°	-31°
Vertical	0°	0°	-13°	13°	0°	0°	1°	11°	0°	0°	11°	1°	0°

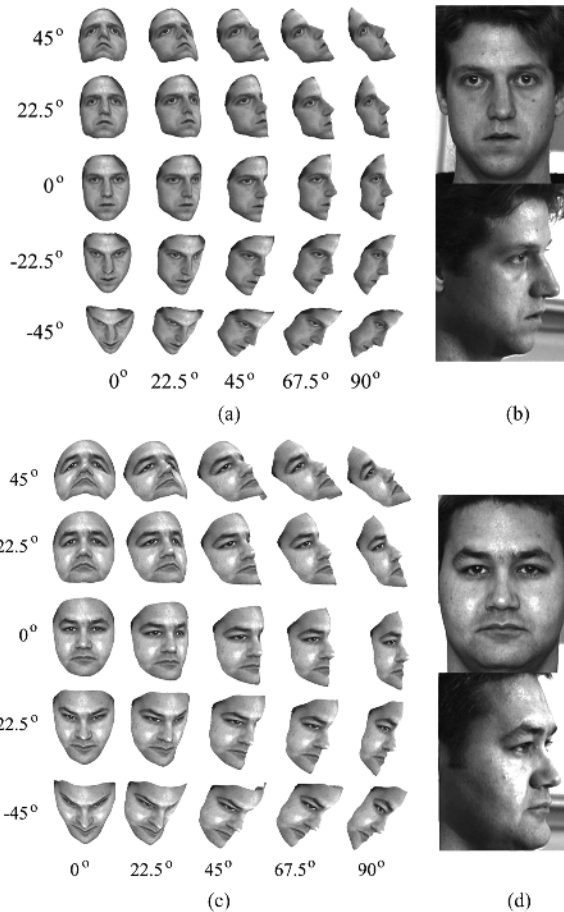


Fig. 7. Examples of virtual views. (a) and (c) Virtual views generated from gallery views by the proposed method. (b) and (d) frontal and side gallery views in the CMU-PIE database [47].

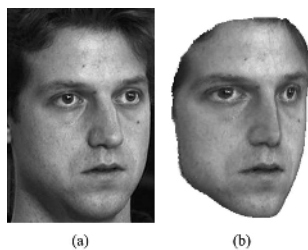


Fig. 8. Comparison between (b) the virtual view synthesized using the proposed approach and (a) its corresponding real view.

- 4) **AC**: The virtual views in a single pose [e.g., Fig. 9(b)-“02”] synthesized from the mug shots [Fig. 9(a)] were used as gallery views and the real views in the corresponding pose [e.g., Fig. 9(c)-“02”] were used as probe views.
- 5) **A2**: The virtual views in a single pose [e.g., Fig. 8(b)-“02”] plus the real frontal and side views [Fig. 8(a)] were used as

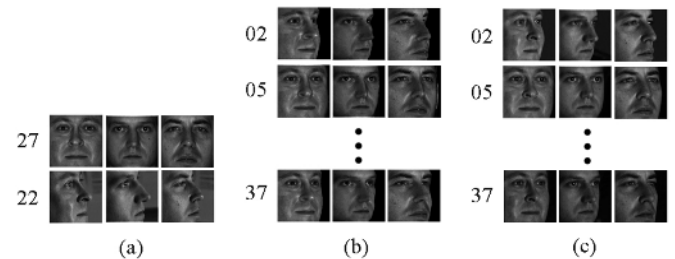


Fig. 9. Examples of the face images used in the experiments. (a) The mug shots in frontal and side viewing directions. (b) The synthesized virtual views in rotated poses. (c) The rotated real views.

gallery views and the real views in the corresponding pose [e.g., Fig. 8(c)-“02”] were used as probe views.

- 6) **AA**: All of the synthesized virtual views plus the two mug shots [i.e., Fig. 8(a) and (b)-“02”, 8b-“05”,..., 8b-“37”] were used as gallery views and the real views [e.g., Fig. 8(c)-“02”] were used as probe views.

The algorithms were run in Microsoft Visual C++ on a Pentium IV 2.8-GHz desktop computer with 1-GB memory. On average, it takes 16.43 min for face-shape modelling, 3.76 min for reflectance analysis, and 0.256 min for synthesizing 11 virtual views. Note that all of these are virtual view database processes in real applications, which are done offline. In the recognition experiment, the AA-PCA algorithm (the most time-consuming strategy) takes 24.04 s for eigenspace training and projection which are offline processes and 0.11 s for online probe feature extraction and recognition. The AA-LBP algorithm takes 4.16 s for the offline gallery view LBP feature extraction and 0.635 s for the online probe feature extraction and recognition. Though the steps of virtual view synthesis are offline processes, the computational times of the algorithms should and can be reduced. For instance, the lattice resolution can be reduced from 512×512 to 256×256 as the virtual views used in recognition are of lower resolution (160×160 for PCA and 80×80 for LBP). We will work on reducing the processing time in a future study.

The recognition rates of the aforementioned identification experiments using PCA [50] as the classifier are summarized in Table III. The results show that the proposed face-recognition approaches dramatically improve the face-recognition rates against pose variations. The improvements of the average recognition rates range from 35.45% to 77.91%. Under every viewing direction, the proposed approaches outperformed the baseline face-recognition approaches. Four plots of cumulative match curves for four representative poses were shown in Fig. 10 to compare the performances of the six different methods. These four poses (i.e., pose “05,” “37,” “02,” and “25”) were in an ascending series of rotations (i.e., pose “05” is under a 16° horizontal rotation, “37” is under a 31° horizontal rotation, “02” is under a 44° horizontal rotation, and “25” is

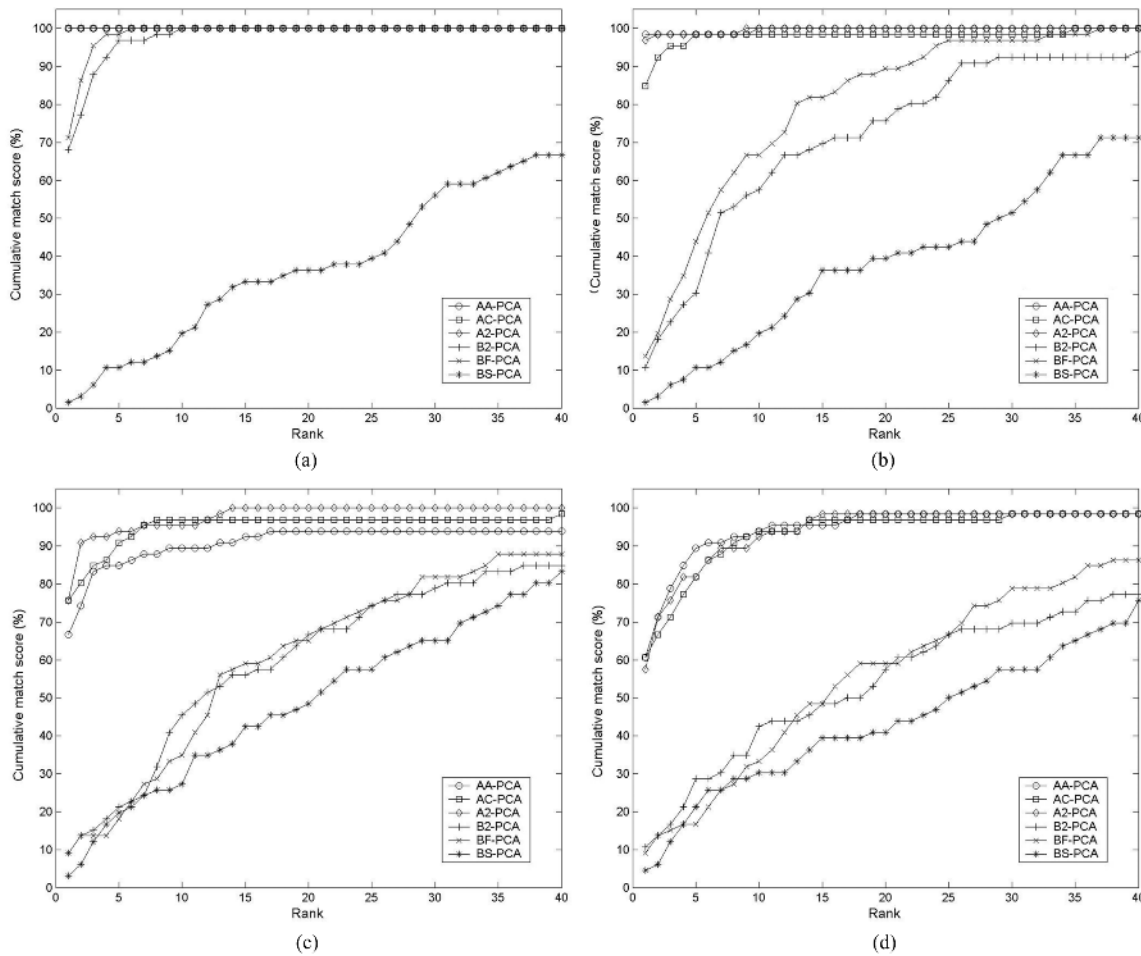


Fig. 10. Cumulative match curves of the proposed approaches and the baseline approaches using PCA as the classifier for (a) pose 05, (b) 37, (c) 02, and (d) 25.

under a 44° horizontal, and 11° vertical rotations. In all poses, the proposed approaches using view synthesis and augmentation outperformed the three baseline approaches in every rank. For small rotations (e.g., pose “05”), the proposed approaches reached 100% much earlier than the baseline approaches (B2 and BF only). For larger rotations (e.g., pose “02” and “25”), the proposed approaches improved the performances significantly over the baseline approaches. Moreover, these three proposed approaches (i.e., AC-PCA, A2-PCA, and AA-PCA) performed similarly, which indicates that the selections of recognition strategies are not decisive to performance.

The recognition rates of the face identification experiments using LBP [1] as the classifier are summarized in Table IV. It shows that the proposed approaches again significantly improved the face-recognition performance. The improvements on the average accuracy range from 14.82% to 81.81%. Four plots of cumulative match curves for the poses “05,” “37,” “02,” and “25” were shown in Fig. 11. Fig. 11 showed that by using the local LBP classifier, the proposed approaches outperformed the baseline approaches in large pose variations, while in small pose variations, the two approaches performed perfectly. Compared to Fig. 10, the LBP classifier is more robust to pose variations than the PCA classifier using the baseline recognition scheme. Consistently, switching from

TABLE III
FACE RECOGNITION RATES (%) OF THE BASELINE RECOGNITION STRATEGIES AND THE PROPOSED APPROACHES USING PCA AS THE CLASSIFIER

Pose	Recognition without the proposed approach			Recognition with the proposed approach		
	B2-PCA	BF-PCA	BS-PCA	AC-PCA	A2-PCA	AA-PCA
02	9	9	3	76	76	67
05	68	71	2	100	100	100
07	83	71	3	100	100	100
09	92	80	3	100	100	100
11	36	26	2	82	85	95
14	5	9	3	67	71	79
25	11	9	5	61	58	61
29	80	73	5	97	100	100
31	5	8	2	52	68	67
34	47	2	45	17	71	65
37	11	14	2	85	97	98
Average	40.64	33.82	6.82	76.09	84.18	84.73

PCA to LBP also provides more robustness for the proposed face-recognition approaches using view synthesis and augmentation to pose variations. For instance, AA-LBP outperformed AA-PCA under pose “02” and “25.” The improvements shown in cumulative match curves contributed to the synthesis of virtual views in the same (or similar) viewing conditions to the probe images. By performing the proposed approaches, the image differences between gallery images and probe images

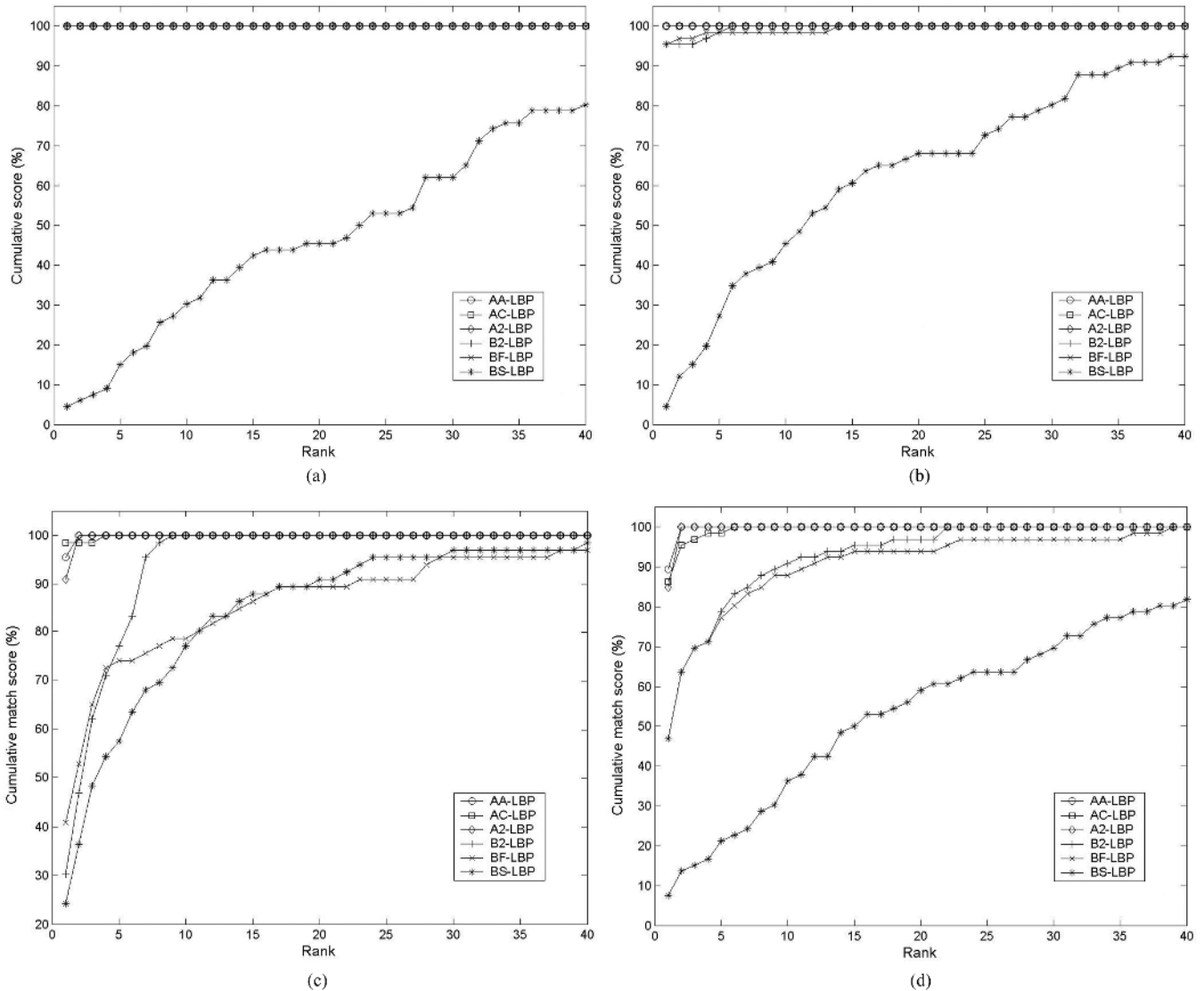


Fig. 11. Cumulative match curves of the proposed approaches and the baseline approaches using LBP as the classifier for (a) pose 05, (b) 37, (c) 02, and (d) 25.

TABLE IV
FACE RECOGNITION RATES (%) OF THE BASELINE RECOGNITION STRATEGIES
AND THE PROPOSED APPROACHES USING LBP AS THE CLASSIFIER

Pose	Recognition without the proposed approach			Recognition with the proposed approach		
	B2-LBP	BF-LBP	BS-LBP	AC-LBP	A2-LBP	AA-LBP
02	30	41	24	98	91	95
05	100	100	5	100	100	100
07	100	100	0	100	100	100
09	100	100	3	100	100	100
11	100	97	3	100	100	100
14	33	33	3	89	89	91
25	47	47	8	86	84	89
29	100	100	2	100	100	100
31	39	39	2	83	82	80
34	73	2	73	24	73	73
37	95	95	5	100	100	100
Average	74.27	68.55	11.64	89.09	92.64	93.45

caused by pose changes were minimized, which helped face classifiers consistently handle rotated probe images.

Tables III and IV show that holistic (PCA) and local (LBP) classifiers suffer from pose variations in face recognition, especially when the rotation angle is larger than 30° . Without using the proposed approaches, the recognition accuracies of both classifiers were below 47%, when the probe views were in poses 02 (44° rotation), 14 (47° rotation), 25 (44° yaw, and 11° tilt rotations), and 31 (47° yaw and 11° tilt rotations). The proposed approach significantly improved the performance of face recognition under large pose variations by increasing the recognition rates by at least 37%. The only exception is pose 34 (a side view opposite that is opposite pose 22), where the recognition rates of the proposed approaches remain as limited improvement (Table III) and no improvement (Table IV). This may be because the virtual views were synthesized from the opposite side view of the face and, consequently, the asymmetry of human faces affected the quality of the virtual view synthesis. Comparing different strategies using the proposed approach, the recognition performance using frontal, side, and corresponding rotated views (A2) is the best for PCA and the recognition performance by using the corresponding rotated views only (AC) is the best

TABLE V
ACCURACY (%) COMPARISON OF THE PROPOSED APPROACHES (AA-PCA, AA-LBP) AND THE REPORTED APPROACHES OF THE EIGENLIGHT FIELD (ELF-F, ELF-S)[27] AND POINT DISTRIBUTION MODEL WITH FACIAL SYMMETRY (PDM)[24]

Pose	Reported recognition rates			The proposed approach	
	ELF-F	ELF-S	PDM	AA-PCA	AA-LBP
02	58	68	72	67	95
05	94	53	100	100	100
07	89	42	N/A	100	100
09	94	62	N/A	100	100
11	88	50	94	95	100
14	70	42	62	79	91
25	56	62	N/A	61	89
29	57	41	98	100	100
31	56	47	N/A	67	80
34	47	58	20	65	73
37	89	62	89	98	100
Average	72.55	53.36	N/A	84.73	93.45

for LBP except for pose 34. The recognition of AC-PCA is lower probably because of the small training dataset for the eigenspace establishment in PCA, and the use of a large training dataset may improve AC-PCA performance. LBP outperformed PCA in almost every aspect even in the baseline recognition strategies, which indicates that local classifiers are more suitable for face recognition under pose variations than holistic classifiers.

To compare the proposed approach with existing face-recognition methods considering pose variations, face recognition using eigenlight fields [27], and face recognition using point distribution models and facial symmetry [24], two of the best methods using single images per person as gallery views with the most thorough experiment tests on CMU-PIE database, were chosen as the benchmark techniques. The experimental results of our proposed approach were tabulated in Table V, together with the reported results in [27] and [24]. In [27], either the frontal view or the side view per person was used as gallery view to recognize images under other viewing conditions. The recognition accuracy using the frontal views as gallery images is denoted as ELF-F and that using the side views as gallery images is denoted as ELF-P. Our proposed face-recognition approaches achieved better performances in every pose. The average recognition rates of ELF using frontal gallery views and using side gallery views are 72.55% and 53.36%, respectively, while the proposed AA-PCA and AA-LBP achieving accuracy increases, ranging from 12.18% to 40.09%. Note the recognition rates in [27] and [24] were achieved using single views per person as gallery views. These improvements of AA-PCA and AA-LBP over ELF-F, ELF-P [27], and PDM [24] were achieved by introducing an additional view as gallery views for virtual view synthesis in the proposed approach.

In [24], frontal views were used as gallery views and the rotated views labelled as 02, 05, 11, 14, 29, 34, and 37 were used as probe views (denoted as PDM in Table V). Since their method is based on 2-D transformations, the recognition rates drop as the rotation angle increases. Under $\sim 15^\circ$ rotation (pose 05 and 29), $\sim 30^\circ$ rotation (pose 11 and 37), and $\sim 45^\circ$ rotation (pose 02 and 14), the recognition rates were 99%, 92%, and 67.5%, respectively. Our proposed approach achieved 100%, 100%, and

93% of accuracies in these three poses, respectively. The recognition rates under the poses associated with large vertical rotations (i.e., for pose 07, 09, 25, and 31, refer to Fig. 6) were not available in [24]. The challenges for face recognition under different poses are different. For instance, it is more difficult to recognize a face in pose 25 than that in pose 02, because pose 25 contains an additional vertical rotation than that in pose 02. Comparing the available recognition rates in [24], the proposed approach achieved higher recognition rates in all poses when using LBP as the classifier.

V. CONCLUSIONS

In this paper, we have presented a novel approach to recognize faces in rotated poses using frontal and side view mug shots as gallery images. Targeting this common application scenario in law enforcement and security surveillance, which appears overlooked by the research community of face recognition, this research investigated the feasibility and effectiveness of building a face-recognition framework toward the use of widely available frontal and side view police mugshot databases. The proposed approach is appearance based, which recognizes input face views in arbitrary poses by comparing them against the virtual views synthesized from gallery views. Personalized face models are reconstructed using hierarchical multilevel variation minimization for 3-D shape modelling and pixel-wise texture analysis considering diffuse and specular reflections of human face surfaces. Virtual face views synthesized in arbitrary poses containing horizontal and vertical in-depth rotations have qualitatively justified the benefit of using frontal and side views as gallery images, especially when the rotation angle is larger than 30° .

Quantitatively, we have tested the proposed face-recognition approach on the pose subset of the CMU-PIE face database, consisting of 884 images in 13 different poses. The proposed face-recognition approach significantly improved the recognition accuracies over the baseline holistic [50] and local [1] face-recognition algorithms and achieved much higher recognition rates compared to state-of-the-art pose-invariant face-recognition techniques using single gallery views [24], [27]. Specifically, the experimental results reveal that:

- using two mugshot face images as gallery views achieved higher recognition rates over face-recognition systems using single face images as gallery views;
- the accuracy improvements of the proposed approach over the baseline strategies were consistently achieved in holistic and local face-matching schemes;
- the proposed approach is capable of handling large horizontal and/or vertical rotation angles between gallery views and probe views, which is a challenging task for face recognition;
- the proposed approach using frontal and side view mugshot gallery images outperformed the state-of-the-art techniques of recognizing rotated faces [24], [27] using single gallery images.

The experimental results demonstrated that the use of mugshot face images as gallery views is a feasible and beneficial solution, which provides an effective means toward pose-invariant face recognition. It shows the possibility of a

better and practical use of existing forensic information by incorporating frontal and side views in existing police mugshot databases into computerized face recognition, leading to an effective solution for face-recognition applications dealing with uncooperative subjects, for instance, in security surveillance systems.

ACKNOWLEDGMENT

The authors would like to thank W. Wang, Dr. B. Zhang, and S. Zhao for their help in the experiments of this research, and the reviewers for helpful comments that improved the quality of this paper.

REFERENCES

- [1] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006.
- [2] A. U. Batur and M. H. Hayes, III, "Segmented linear subspaces for illumination-robust face recognition," *Int. J. Comput. Vis.*, vol. 57, no. 1, pp. 49–66, 2004.
- [3] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.
- [4] P. N. Belhumeur, D. J. Kriegman, and A. L. Yuille, "The bas-relief ambiguity," *Int. J. Comput. Vis.*, vol. 35, no. 1, pp. 33–44, 1999.
- [5] D. Beymer and T. Poggio, "Face recognition from one example view," in *Proc. Int. Conf. Comput. Vis.*, 1995, pp. 500–507.
- [6] D. Beymer, "Feature correspondence by interleaving shape and texture computations," in *Proc. IEEE Conf. CVPR*, 1996, pp. 921–928.
- [7] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1063–1074, Sep. 2003.
- [8] A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Robust expression-invariant face recognition from partially missing data," in *Proc. Eur. Conf. Comput. Vis.*, 2006, vol. 3953 NCS, pp. 396–408.
- [9] S. Cass and M. Riesenman, "Improving security, preserving privacy," *IEEE Spectr.*, vol. 39, no. 1, pp. 44–49, Jan. 2002.
- [10] X. Chai, S. Shan, X. Chen, and W. Gao, "Locally linear regression for pose-invariant face recognition," *IEEE Trans. Image Process.*, vol. 16, no. 7, pp. 1716–1725, Jul. 2007.
- [11] Q. Chen, J. Yao, and W. K. Cham, "3D model-based pose invariant face recognition from multiple views," *Inst. Eng. Technol. Comput. Vis.*, vol. 1, no. 1, pp. 25–34, 2007.
- [12] W. Chen, E. M. Joo, and W. Shiqian, "Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 36, no. 2, pp. 458–466, Apr. 2006.
- [13] T. F. Cootes, D. Cooper, C. J. Taylor, and J. Graham, "Active shape models—Their training and application," *Comput. Vis. Image Understand.*, vol. 61, no. 1, pp. 38–59, 1995.
- [14] T. F. Cootes, G. V. Wheeler, K. N. Walker, and C. J. Taylor, "View-based active appearance models," *Image Vis. Comput.*, vol. 20, pp. 657–64, 2002.
- [15] I. C. Cameron, "Parameterizing images for recognition and reconstruction," *Proc. Brit. Mach. Vis. Conf.*, pp. 367–370, 1991.
- [16] P. Debevec, T. Hawkins, C. Tchou, H. P. Duiker, W. Sarokin, and M. Sagar, "Acquiring the reflectance field of a human face," in *Proc. SIGGRAPH*, 2000, pp. 145–156.
- [17] N. Duta, A. K. Jain, and M.-P. Dubuisson-Jolly, "Automatic construction of 2D shape models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 5, pp. 433–446, May 2001.
- [18] J. D. Foley, A. van Dam, S. K. Feiner, and J. F. Hughes, *Computer Graphics: Principles and Practice*, 2nd ed. Reading, MA: Addison-Wesley, 1995.
- [19] Y. Gao, M. K. H. Leung, W. Wang, and S. C. Hui, "Fast face identification under varying pose from a single 2-D model view," *Proc. Inst. Elect. Eng., Vis. Image Signal Process.*, vol. 148, no. 4, pp. 248–253, 2001.
- [20] Y. Gao and M. K. H. Leung, "Face recognition using line edge map," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 6, pp. 764–779, Jun. 2002.
- [21] Y. Gao and Y. Qi, "Robust visual similarity retrieval in single model face databases," *Pattern Recogn.*, vol. 38, pp. 1009–1020, 2005.
- [22] A. S. Georghades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 643–660, Jun. 2001.
- [23] A. S. Georghades, "Incorporating the Torrance and Sparrow model of reflectance in uncalibrated photometric stereo," in *Proc. ICCV*, 2003, vol. 2, pp. 816–823.
- [24] D. González-Jiménez and J. L. Alba-Castro, "Toward pose-invariant 2-D face recognition through point distribution models and facial symmetry," *IEEE Trans. Inf. Forensics Security*, vol. 2, no. 3, pp. 413–429, Jul. 2007.
- [25] T. Goto, W.-S. Lee, and N. Magnenat-Thalmann, "Facial feature extraction for quick 3D face modeling," *Signal Process.: Image Commun.*, vol. 17, pp. 243–259, 2002.
- [26] W. E. L. Grimson, "A computational theory of visual surface interpolation," *Phil. Trans. Roy. Soc. London*, vol. B, no. 298, pp. 395–427, 1982.
- [27] R. Gross, I. Matthews, and S. Baker, "Appearance-based face recognition and light-fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 4, pp. 449–465, Apr. 2004.
- [28] K. Hara, K. Nishino, and K. Ikeuchi, "Light source position and reflectance estimation from a single view without the distant illumination assumption," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 4, pp. 493–505, Apr. 2005.
- [29] R.-L. Hsu and A. K. Jain, "Generating discriminating cartoon faces using interacting snakes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 11, pp. 1388–1398, Nov. 2003.
- [30] H. H. S. Ip and L. Yin, "Constructing a 3D individualized head model from two orthogonal views," *Vis. Comput.*, vol. 12, no. 5, pp. 254–266, 1996.
- [31] J. J. Woodward, "Using biometrics to achieve identity, dominance in the global war on terrorism," *Military Rev.*, pp. 30–34, 2005.
- [32] D. Jiang, Y. Hu, S. Yan, L. Zhang, H. Zhang, and W. Gao, "Efficient 3D reconstruction for face recognition," *Pattern Recogn.*, vol. 38, no. 6, pp. 787–98, 2005.
- [33] M. J. Jones and T. Poggio, "Multidimensional morphable models: A framework for representing and matching object classes," *Int. J. Comput. Vis.*, vol. 29, no. 2, pp. 107–131, 1998.
- [34] T. K. Kim and J. Kittler, "Design and fusion of pose-invariant face-identification experts," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 9, pp. 1096–1106, Sep. 2006.
- [35] S. G. Kong, H. Jingu, F. Boughorbel, Z. Yue, B. R. Abidi, A. Koschan, Y. Mingzhong, and M. A. Abidi, "Multiscale fusion of visible and thermal IR images for illumination-invariant face recognition," *Int. J. Comput. Vis.*, vol. 71, no. 2, pp. 215–233, 2007.
- [36] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Automatic interpretation and coding of face images using flexible models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 743–756, Jul. 1997.
- [37] B. Lévy, "Constrained texture mapping for polygonal meshes," in *Proc. SIGGRAPH*, 2001, pp. 417–424.
- [38] S. Z. Li, C. RuFeng, L. ShengCai, and Z. Lun, "Illumination invariant face recognition using near-infrared images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 627–639, Apr. 2007.
- [39] D. G. Luenberger, *Introduction to Linear and Nonlinear Programming*. Reading, MA: Addison-Wesley, 1973.
- [40] A. M. Martínez, "Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 6, pp. 748–763, Jun. 2002.
- [41] F. Perronnin, J.-L. Dugelay, and K. Rose, "A probabilistic model of face mapping with local transformations and its application to person recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 7, pp. 1157–1171, Jul. 2005.
- [42] P. J. Phillips, P. Grother, R. Micheals, D. M. Blackburn, E. Tabassi, and M. Bone, "Face recognition vendor test 2002," in *Proc. IEEE Int. Workshop Analysis Modeling of Faces Gestures*, 2003, p. 44.
- [43] B. T. Phong, "Illumination for computer generated pictures," *Commun. ACM*, vol. 18, no. 6, pp. 311–317, 1975.
- [44] S. Romdhani and T. Vetter, "Estimating 3D shape and texture using pixel intensity, edges, specular highlights, texture constraints and a prior," in *Proc. IEEE Conf. CVPR*, 2005, vol. 2, pp. 986–993.

- [45] S. Romdhani, J. Ho, T. Vetter, and D. J. Kriegman, "Face recognition using 3-D models: Pose and illumination," *Proc IEEE*, vol. 94, no. 11, pp. 1977–1999, Nov. 2006.
- [46] L. Seong-Whan, S. W. Lee, and M. Song-Hyang, "Face recognition under arbitrary illumination using illuminated exemplars," *Pattern Recogn.*, vol. 40, no. 5, pp. 1605–1620, 2007.
- [47] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression database," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1615–1618, Dec. 2003.
- [48] R. Singh, M. Vatsa, A. Ross, and A. Noore, "A mosaicing scheme for pose-invariant face recognition," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 37, no. 5, pp. 1212–1225, Oct. 2007.
- [49] R. R. Thomas, *Detective Fiction and the Rise of Forensic Science*. Cambridge, U.K.: Cambridge University Press, 2000.
- [50] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proc. IEEE Conf. CVPR*, 1991, pp. 586–591.
- [51] T. Vetter, M. Jones, and T. Poggio, "A bootstrapping algorithm for learning linear models of object classes," in *Proc. IEEE Conf. CVPR*, 1997, pp. 40–46.
- [52] K.-W. Wan, K.-M. Lam, and K.-C. Ng, "An accurate active shape model for facial feature extraction," *Pattern Recogn. Lett.*, vol. 26, no. 15, pp. 2409–2423, 2005.
- [53] Z. Xu and J. Luo, "Face recognition by expression-driven sketch graph matching," *Proc. ICPR*, vol. 3, pp. 1119–1122, 2006.
- [54] Z. Xue, S. Z. Li, and E. K. Teoh, "Bayesian shape model for facial feature extraction and recognition," *Pattern Recogn.*, vol. 36, no. 12, pp. 2819–2833, 2003.
- [55] P. C. Yuen and C. H. Man, "Human face image searching system using sketches," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 37, no. 4, pp. 493–504, Jul. 2007.
- [56] L. Zhang, A. Razdan, G. Farin, J. Femiani, M. Bae, and C. Lockwood, "3D face authentication and recognition based on bilateral symmetry analysis," *Vis. Comput.*, vol. 22, no. 1, pp. 43–55, 2006.
- [57] X. Zhang, Y. Gao, and M. K. H. Leung, "Multilevel quadratic variation minimization for 3D face modeling and virtual view synthesis," in *Proc. Int. Multimedia Model. Conf.*, 2005, pp. 132–138.
- [58] X. Zhang, Y. Gao, and B.-L. Zhang, "Recognising rotated faces from two orthogonal views in mugshot databases," in *Proc. ICPR*, 2006, vol. 1, pp. 195–198.
- [59] Y. Zhang and A. M. Martínez, "Recognition of expression variant faces using weighted subspaces," in *Proc. ICPR*, 2004, vol. 3, pp. 149–152.
- [60] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 915–928, Jun. 2007.
- [61] S. Zhao, W. Gao, S. Shan, and B. Yin, "Enhance the alignment accuracy of active shape models using elastic graph matching," in *Proc. Int. Conf. Biometric Auth.*, 2004, pp. 52–58.
- [62] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surv.*, vol. 35, no. 4, pp. 399–459, 2003.

- [63] S. K. Zhou, G. Aggarwal, R. Chellappa, and D. W. Jacobs, "Appearance characterization of linear Lambertian objects, generalized photometric stereo, and illumination-invariant face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 2, pp. 230–245, Feb. 2007.



Xiaozheng Zhang (S'07) received the B.Eng. degree in mechanical engineering from Tsinghua University, Beijing, China, in 2001 and the Ph.D. degree in computer science from Griffith University, Nathan, Australia, in 2008.

Currently he is a Postdoctoral Research Fellow with the Institute for Integrated and Intelligent Systems, Griffith University. His research interests include computer vision, image processing, and pattern recognition. His particular interests are in the fields of face recognition, 3-D face modelling, and surface reflectivity for 3-D object rendering and reconstruction.



Yongsheng Gao (SM'02), received the B.Sc. and M.Sc. degrees in electronic engineering from Zhejiang University, China, in 1985 and 1988, respectively, and the Ph.D. degree in computer engineering from Nanyang Technological University, Singapore.

Currently, he is an Associate Professor with the School of Engineering, Griffith University, Nathan, Australia. His research interests include face recognition, biometrics, image retrieval, computer vision, and pattern recognition.



Maylor K. H. Leung (M'92), received the B.Sc. degree in physics from the National Taiwan University in 1979, and the B.Sc., M.Sc., and Ph.D. degrees in computer science from the the University of Saskatchewan, Saskatoon, SK, Canada, in 1983, 1985, and 1992, respectively.

Currently, he is an Associate Professor with Nanyang Technological University, Singapore. His research interest is in the area of computer vision, pattern recognition, and image processing. His particular interests are object recognition, video behavior detection, robot navigation, line pattern analysis (e.g., Hausdorff distances of line and curve), and computer aids for the visually impaired.

surveillance for human analysis (e.g., Hausdorff distances of line and curve), and computer aids for the visually impaired.