Neural-Network-Based Adaptive Hybrid-Reflectance Model for 3-D Surface Reconstruction

Chin-Teng Lin, *Fellow, IEEE*, Wen-Chang Cheng, and Sheng-Fu Liang

Abstract—This paper proposes a novel neural-network-based adaptive hybrid-reflectance three-dimensional (3-D) surface reconstruction model. The neural network automatically combines the diffuse and specular components into a hybrid model. The proposed model considers the characteristics of each point and the variant albedo to prevent the reconstructed surface from being distorted. The neural network inputs are the pixel values of the two-dimensional images to be reconstructed. The normal vectors of the surface can then be obtained from the output of the neural network after supervised learning, where the illuminant direction does not have to be known in advance. Finally, the obtained normal vectors are applied to enforce integrability when reconstructing 3-D objects. Facial images and images of other general objects were used to test the proposed approach. The experimental results demonstrate that the proposed neural-network-based adaptive hybrid-reflectance model can be successfully applied to objects generally, and perform 3-D surface reconstruction better than some existing approaches.

Index Terms—Enforcing integrability, Lambertian model, neural network, reflectance model, shape from shading, surface normal.

I. INTRODUCTION

SHAPE recovery is a classical computer vision problem. The objective of shape recovery is to obtain a three-dimensional (3-D) scene description from one or more two-dimensional (2-D) images. The techniques used to recover the shape of an object are called shape-from-X techniques, where X denotes the specific information, such as shading, stereo, motion, and texture. Shape recovery from shading (SFS) is a major computer vision approach, which reconstructs the 3-D shape of an object from its gradual shading variation in 2-D images. When a point light source illuminates an object, they appear with different brightness, since the normal vectors corresponding to different parts of the object's surface are different. The spatial variation of brightness, referred to as *shading,* is used to estimate the orientation of surface and then calculate the depth map of the object. The recovered shape can be expressed in terms of depth, surface normal vector, surface gradient, or surface slant and tilt.

The SFS approach was first proposed by Horn [1] in the early 1970s and was further enhanced by himself and Brooks [2], [3]. Attempts have been made to enhance the recovery performance of SFS [4]–[10]. To resolve the SFS problem, image formation needs be studied. A straightforward image formation model is the Lambertian model, in which the gray intensity of a pixel in the image depends on the light source direction and the surface normal. The conventional approaches to solve the SFS problem are to employ the conventional Lambertian model and to minimize the cost function consisting of the brightness constraint, the smoothness constraint, the integrability constraint, the intensity gradient constraint, and the unit normal constraint [11], [12]. Since the nonlinear optimization problem takes a long computational time to solve and does not easily converge to the optimum solution, direct shape reconstruction approaches have been proposed [8], [13], [14]. These approaches require an extra smoothness constraint in the cost function to guarantee a smooth surface and stabilize the convergence to a unique solution.

Zhang [15] categorized the SFS techniques into four categories: minimization [5]–[7], [13], [16]–[21], propagation [1], [8], [22]–[24], local [9], [25], and linear approaches [10], [26]. Minimization approaches obtain the solution by minimizing an energy function. Propagation approaches propagate the shape information from a set of surface points to the entire image. Local approaches derive the shape by assuming of surface type. Linear approaches obtain the solution from linearizing the reflectance map. Zhang compared the computational time and accuracy of these four SFS approaches and indicated that none of the algorithms performed consistently for all images. The algorithms worked well for some images but performed poorly for others. Zhang concluded that minimization approaches are generally the most robust but are also the slowest. To obtain high 3-D reconstruction performance, this paper uses the minimization approach in the training stage of the proposed approach.

Recently, multilayer neural networks have also been adopted to handle the SFS problem [27], [28], [31]. However, these approaches are still restricted by the Lambertian model, which requires estimating the direction of the light source. Obviously, this restriction makes the algorithm impractical for many applications in which illumination information is not available. Additionally, the reflectance of objects does not always follow the Lambertian model, so a more general model is required. According to [28], a successful reflectance model for surface reconstruction of objects should combine two major components:

C.-T. Lin is with the Department of Electrical and Control Engineering and the Department of Computer Science, National Chiao-Tung University, Hsinchu, Taiwan. He is also with the Brain Research Center, University System of Taiwan, Taipei, Taiwan (e-mail: ctlin@mail.nctu.edu.tw).

S.-F. Liang is with the Department of Biological Science and Technology, National Chiao-Tung University, Hsinchu, Taiwan. He is also with the Brain Research Center, University System of Taiwan, Taipei, Taiwan (e-mail: sfliang@mail.nctu.edu.tw).

W.-C. Cheng is with the Department of Information Networking Technology, Hsiuping Institute of Technology, Taichung, Taiwan (e-mail: wccheng@mail.hit.edu.tw).

the diffuse and specular components. The Lambertian model describes the relationship between the surface normal and the light source direction by generally assuming that the surface reflection is due to diffuse reflection only, and assumes that the surface reflects the light equally in all directions. Thus, the Lambertian model ignores the specular component.

Some specular or non-Lambertian models have been proposed to model the specular component. Healy and Binford [29] used the Torrance–Sparrow model [30], which assumes that a surface is composed of small, randomly oriented, mirror-like facets to retrieve a local shape from specularity. Cho and Chow [31] proposed a novel hybrid approach using two self-learning neural networks to generalize the reflectance model by modeling the pure Lambertian surface and the specular component of the non-Lambertian surface, respectively. This model does not require the viewing direction and the light source direction and yields better shape recovery than previous approaches. However, the hybrid approach still has two drawbacks.

1) The albedo of the surface is disregarded or regarded as constant, distorting the recovered shape. Generally, the albedo is variant in different surface regions.
2) The combination ratio between the diffuse and specular components is regarded as a constant, which is determined by trial and error.

Therefore, the hybrid combination approach proposed in [31] is unsuitable for the surface reconstruction of human faces or general objects whose albedo and reflecting characteristic is not the same across the entire surface.

This paper proposes a novel adaptive hybrid-reflectance model to represent more general conditions. This model intelligently integrates both the diffuse and specular reflection components and does not require determining the hybrid ratio in advance. The pure diffuse and specular reflection components are both formed by similar feed-forward neural network structures. A supervised learning algorithm is applied to tune up the pointwise hybrid ratio automatically based on image intensities and to produce the normal vectors of the surface for reconstruction. The proposed approach estimates the illuminant direction, viewing direction, and normal vectors of object surfaces for reconstruction after training. Therefore, new shaded images could be generated under different illuminant conditions by controlling the above parameters. The 3-D surface can also be reconstructed according to these normal vectors using existing approaches such as enforcing integrability [17]. Additionally, the proposed approach considers the albedo and the reflecting characteristic of each surface point individually. According to the experimental results presented in Section VI, the shape recovery algorithm is the most robust for recovery of surfaces with the variant albedo and complex reflecting characteristic.

The remainder of this paper is organized as follows. Section II describes the proposed hybrid-reflectance model that includes the diffuse and specular components. The details of the neural-network-based hybrid-reflectance model and its learning rule derivations are presented in Sections III and IV. Section V presents the results of experiments performed to evaluate the performance of the proposed approach. Conclusions are drawn in the last section.

II. THE HYBRID REFLECTANCE MODELS

This study aims to recover the surface of a 3-D object from 2-D images by SFS techniques. To resolve the SFS problem, a mathematical reflectance model is needed to describe the relation between the reflection behavior and the image formation. A proper reflectance model could help reconstruct the surface shape accurately from the image intensity variation corresponding to the reflection characteristic of the surface. Computer vision has two light reflection components: diffuse reflection and specular reflection. Diffuse reflection is a uniform reflection of light with no directional dependence for the viewer and specular reflection is the reflected light visible only at the reflected direction. Considering only the diffuse (so-called Lambertian) component or the specular component singly is not enough for practical applications. Most surfaces are neither pure Lambertian nor pure specular, and their reflection characteristics are mixtures of these two reflection components. This paper derives a modified mathematical hybrid reflectance model based on [29]–[33] and uses it to develop a new neural-network-based adaptive hybrid-reflectance model.

In [31], a linear combination of the diffuse intensity and the specular intensity by a constant ratio was given as

$$R_{\text{hybrid}}(x,y) = \mu R_d(x,y) + (1-\mu)R_s(x,y) \quad \forall x,y \in D \quad (1)$$

where $R_{\text{hybrid}}(\cdot)$ denotes the total intensity of the hybrid reflectance model, which combines the diffuse intensity $R_d(\cdot)$ and the specular intensity $R_s(\cdot)$, and μ denotes the combination ratio for the hybrid reflectance model.

In (1), $R_d(\cdot)$ denotes the diffuse reflectance model (called the Lambertian model) which represents a surface illuminated by a single point light source given by

$$R_d\left(\mathbf{n}(x,y), \alpha(x,y)\right)$$
$$= \max\left\{L\alpha(x,y)\mathbf{n}(x,y)\mathbf{s}(x,y), 0\right\} \quad \forall x,y \in D \quad (2)$$

where $\alpha(x,y)$ denotes the surface's diffuse albedo at point (x,y) over the image plane domain D; $\mathbf{n}(x,y)$ denotes the surface normal on position (x,y); $\mathbf{s} \in R^3$ denotes the direction of point light at infinity; and L denotes the light strength. Assuming that the Lambertian surface of a convex object is given by the depth function $z(x,y)$, the surface normal $\mathbf{n}(x,y)$ on position (x,y) can be represented as

$$\mathbf{n}(x,y) = \frac{[-p(x,y) - q(x,y)\ 1]}{\sqrt{p^2(x,y) + q^2(x,y) + 1}} \quad (3)$$

where $p(x,y)$ and $q(x,y)$ denote the x- and y-partial derivatives of $z(x,y)$, respectively. In (2), the equation $\max\{\cdot\}$ sets all negative components corresponding to the surface points lying in attached shadow to zero, where a surface point (x,y) lies in an attached shadow iff $\mathbf{n}(x,y)\mathbf{s} < 0$ [11].

In (1), $R_s(\cdot)$ denotes the specular reflectance model. Healy and Binford [29] simplified the Torrance–Sparrow model [30], which uses a Gaussian distribution to model the facet orientation function and considers the other components as constants. Defining the halfway vector as $\mathbf{h} = (\mathbf{s} + \mathbf{v})/\|\mathbf{s} + \mathbf{v}\|$, which
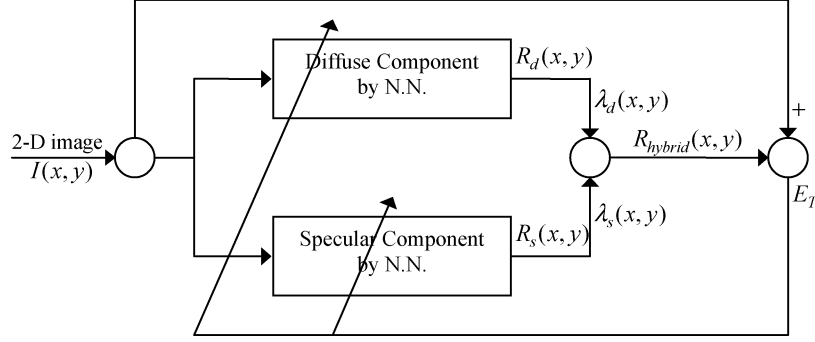
Fig. 1. Block diagram of the proposed adaptively hybrid reflectance model.

denotes the normalized vector sum between the light source direction $\mathbf{s}$ and the viewing direction $\mathbf{v}$, the model is represented as

$$R_s = \frac{1}{\sqrt{2\pi}\beta} \exp\left(-\frac{\varphi^2}{2\sigma^2}\right) \qquad (4)$$

where φ denotes the angle between the surface normal $\mathbf{n}(x,y)$ and the halfway vector $\mathbf{h}(x,y)$ at point (x,y) such that $\varphi = \cos^{-1}(\langle \mathbf{n}(x,y)\mathbf{h}(x,y)\rangle)$ and σ denotes the standard deviation, which can be interpreted as a measure of the surface roughness. The specular reflectance model in (4) is popular for the SFS problems but has many parameters to be determined. Therefore, this model cannot easily be integrated with the diffuse reflection component in the proposed approach.

The other well-known specular model is Phong's model [33]. For specular reflection, the amount of light seen by the viewer depends on the angle δ between the perfect reflected ray $\mathbf{r}$ and the direction of the viewer $\mathbf{v}$. Phong's model states that the light perceived by the viewer is proportional to $\cos \delta^r$ and can be represented as

$$R_s\left(\mathbf{n}(x,y), \mathbf{h}(x,y)\right) = \left(\langle \mathbf{n}(x,y), \mathbf{h}(x,y)\rangle\right)^r = (\cos \phi)^r \quad (5)$$

where r denotes a constant. Different values of the constant r denote different kinds of surfaces which are more or less mirror-like. Phong's model [33] is mathematically simpler than that of Healy and Binford [29], and hence is more appropriate for integrating the diffuse and specular reflections. Therefore, the proposed scheme employs a neural network used to describe the specular reflection based on Phong's model.

In this paper, the ratio between the diffuse and specular reflection components corresponding to each point is adjusted for each point on the object surface in the image. This ratio is regarded as an adaptive weight in the proposed neural network and is updated in the learning iterations to obtain a reasonable and appropriate value for each point on the 3-D surface.

III. NEURAL-NETWORK-BASED ADAPTIVE HYBRID-REFLECTANCE MODEL

This section proposes a novel neural-network-based hybrid-reflectance model in which the hybrid ratio of diffuse and specular components is regarded as the adaptive weight of the neural network. The supervised learning algorithm is adopted, and the hybrid ratio for each point is updated in the learning iterations. After the learning process, the neural network can be used to estimate the proper hybrid ratio for each point on the 3-D surface of any object in an image. Thus, the diffuse and specular components can be integrated intelligently and efficiently. Additionally, the proposed hybrid-reflectance model also considers the variant albedo effect. The variant albedo effect is sometimes claimed to influence the performance of 3-D surface reconstruction and cause distortion in conventional approaches [34]–[36].

Fig. 1 shows the schematic block diagram of the proposed adaptive hybrid-reflectance model, which consists of the diffuse and specular components. This diagram is used to describe the characteristics of the diffuse and specular components of the adaptive hybrid-reflectance model by two neural networks with similar structures. The composite intensity R_{hybrid} is obtained by combining the diffuse intensity R_d and the specular intensity R_s based on the adaptive weights $\lambda_d(x,y)$ and $\lambda_s(x,y)$. The system inputs are the 2-D image intensities of each point, and the outputs are the learned reflectance map. When solving the SFS problem by the proposed neural-network-based reflectance model, the cost function E_T is minimized to update the neural parameters. After training, the normal surface vectors can be obtained from the reflectance model to reconstruct the 3-D shape of the object. The model can also be combined with different light source directions and viewing directions to produce new shaded images.

Fig. 2 shows the framework of the proposed symmetric neural network which simulates the diffuse reflection model. The input/output pairs of the network are arranged like a mirror in the center layer, where the number of input nodes equals the number of output nodes, making it a symmetric neural network. The light source direction and the normal vector from the input 2-D images in the left side of the symmetric neural network are separated and then combined inversely to generate the reflectance map for diffuse reflection in the right side of the network. The function of each layer is discussed in detail below.

Assuming that an input image has m pixels in total, then the symmetric neural network includes m input variables. The 2-D image is rearranged to form an $m \times 1$ column vector represented as $\mathbf{I} = (I_1, I_2, \ldots, I_m)^T$ and fed into the symmetric neural network. Through the symmetric neural network, the reflectance map for diffuse reflection $(R_{d_1}, R_{d_2}, \ldots, R_{d_m})$ can be obtained in the output of the symmetric neural network.
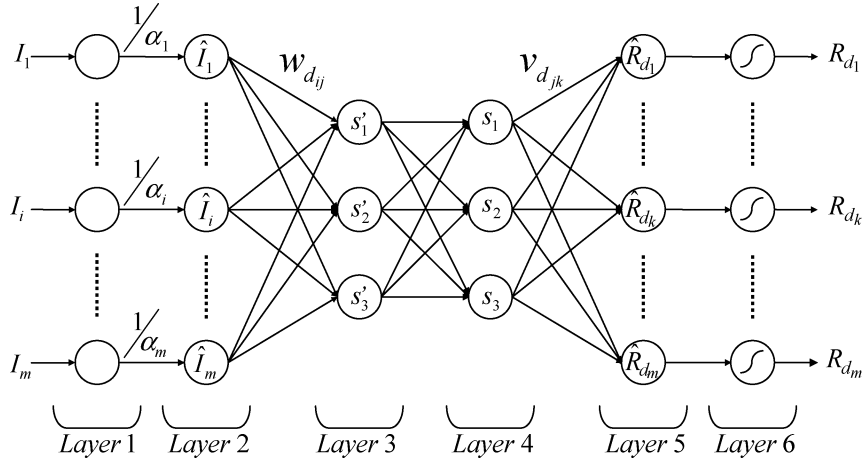
Fig. 2.   Framework of the symmetric neural network for diffuse reflection model.

Each node in the symmetric neural network has some finite "fan-in" of connections represented by weight values from the previous nodes and "fan-out" of connections to the next nodes. Associated with the fan-in of a node is an integration function f which combines information, activation, and evidence from other nodes and provides the net input for this node

$$\text{net input} = f(\text{inputs to this node}, \text{associated link weights}). \quad (6)$$

Each node also outputs an activation value as a function of its net input

$$\text{node output} = a^{(l)}(\text{net input}) = a^{(l)}(f) \quad (7)$$

where $a^{(l)}(\cdot)$ denotes the activation function and the superscript l denotes the layer number. The proposed symmetric neural network contains six layers. The functions of the nodes in each layer are described as follows.

Layer 1: This layer gathers the intensity values of the input images as the network inputs. Node I_i denotes the ith pixel of the 2-D image and m denotes the number of total pixels of the image. That is

$$f_i = I_i, \quad i = 1, \ldots, m$$
$$a_i^{(1)} = f_i, \quad i = 1, \ldots, m. \quad (8)$$

The following equation also uses this notation.

Layer 2: This layer adjusts the intensity of the input 2-D image with corresponding albedo value. Each node in this layer, corresponding to one input variable, divides the input intensity by the corresponding albedo and transmits it to the next layer. That is

$$f_i = \frac{I_i}{\alpha_i}, \quad i = 1, \ldots, m$$
$$\hat{I}_i = a_i^{(2)} = f_i, \quad i = 1, \ldots, m. \quad (9)$$

The output of this layer is the adjusted intensity value of the original 2-D image. The nodes of this layer are labeled as $\hat{I}_1, \hat{I}_2, \ldots, \hat{I}_m$. The term α_i denotes the ith albedo value corresponding to the ith pixel of the 2-D image and $1/\alpha_i$ denotes the weight between I_i and $\hat{I}_i$.

Layer 3: The purpose of Layer 3 is to separate the light source direction from the 2-D image. The light source directions of this layer are not normalized and are labeled as s_1', s_2', and s_3'. The link weight in layer 3 is represented as $w_{d_{ij}}$ for the connection between node i of layer 2 and node j of Layer 3

$$f_j = \sum_{i=1}^{m} \hat{I}_i w_{d_{ij}}, \quad i = 1, \ldots, m, \quad j = 1, 2, 3$$
$$s_j' = a_j^{(3)} = f_j, \quad j = 1, 2, 3. \quad (10)$$

Layer 4: The nodes of this layer represent the unit light source. Equation (11) is used to normalize the nonnormalized light source direction obtained in Layer 3. These nodes in Layer 4 are labeled as s_1, s_2, and s_3, respectively, and the light source direction is represented as $\mathbf{s} = (s_1, s_2, s_3)^T$. The output of s_j can be calculated from

$$f_j = \frac{1}{\sqrt{s_1'^2 + s_2'^2 + s_3'^2}}, \quad j = 1, 2, 3,$$
$$s_j = a_j^{(4)} = f_j \cdot s_j' = \frac{s_j'}{\sqrt{s_1'^2 + s_2'^2 + s_3'^2}}, \quad j = 1, 2, 3. \quad (11)$$

Layer 5: Layer 5 combines the light source direction $\mathbf{s}$ and normal vectors of the surface to generate the diffuse reflection reflectance map. The link weight connecting node j of Layer 4 and node k of Layer 5 is denoted as $v_{d_{jk}}$ and represents the normal vectors of the surface for the diffuse component. That is, $(v_{d_{1k}}, v_{d_{2k}}, v_{d_{3k}})^T$ denotes the normal vector of the surface for the diffuse component on the point k, where $k = 1, \ldots, m$. The outputs of the nodes in this layer are denoted as $\hat{R}_{d_k}$ and can be calculated as

$$f_k = \sum_{j=1}^{3} s_j v_{d_{jk}}, \quad k = 1, \ldots, m$$
$$\hat{R}_{d_k} = a_k^{(5)} = f_k, \quad k = 1, \ldots, m. \quad (12)$$

Significantly, $\hat{R}_{d_k}$ denotes the nonnormalized reflectance map of diffuse reflection, and therefore is normalized in Layer 6.

Layer 6: This layer transfers the nonnormalized reflectance map of diffuse reflection obtained in Layer 5 into the interval [0, 255]. These nodes $R_{d_1}, R_{d_2}, \ldots, R_{d_m}$ denote the normalized

reflectance map of diffuse reflection, and their output can be calculated as

$$f_k = \hat{R}_{d_k}, \quad k = 1, \ldots, m$$
$$R_{d_k} = a_k^{(6)}$$
$$= \frac{255 \left(f_k - \min(\hat{\mathbf{R}}_d) \right)}{\max(\hat{\mathbf{R}}_d) - \min(\hat{\mathbf{R}}_d)}$$
$$= \frac{255 \left(\hat{R}_{d_k} - \min(\hat{\mathbf{R}}_d) \right)}{\max(\hat{\mathbf{R}}_d) - \min(\hat{\mathbf{R}}_d)}, \quad k = 1, \ldots, m \quad (13)$$

where $\hat{\mathbf{R}}_d = (\hat{R}_{d_1}, \hat{R}_{d_2}, \ldots, \hat{R}_{d_k})^T$ and the link weights between Layers 5 and 6 are unity.

Similar to the diffuse reflection model, a symmetric neural network given by Fig. 2 is used to simulate the specular component in the hybrid-reflectance model. The major differences between these two networks are the node representation in Layers 3 and 4 and the active function of Layer 5. The nodes of Layer 3 represent the nonnormalized halfway vector, labeled as $\mathbf{h}' = (h'_1, h'_2, h'_3)^T$, and the nodes of Layer 4 represent the normalized halfway vector labeled as $\mathbf{h} = (h_1, h_2, h_3)^T$. According to (5), Layer 5 of the specular component of the symmetric neural network combines the halfway vector $\mathbf{h}$ and normal surface vectors to generate the reflectance map of specular reflection. The link weight connecting node j of Layer 4 and node k of Layer 5 is denoted as $v_{s_{jk}}$ and represents the normal vectors of the surface for the specular component. Then, the outputs of the nodes in Layer 5, denoted as $\hat{R}_{s_k}$, can be calculated as

$$f_k = \sum_{j=1}^{3} h_j v_{s_{jk}}, \quad k = 1, \ldots, m$$
$$\hat{R}_{s_k} = a_k^{(5)} = (f_k)^r, \quad k = 1, \ldots, m \quad (14)$$

where the active function $a_k^{(5)}$ in this layer represents the r degree of its net input. Significantly, $\hat{R}_{s_k}$ denotes the nonnormalized reflectance map of specular reflection, and is also normalized in Layer 6.

Through the supervised learning algorithm derived in the following section, the normal surface vectors can be obtained automatically. Then, the enforcing integrability method can be used [17] to obtain the depth information for reconstructing the 3-D surface of an object by the obtained normal vectors. In the proposed approach, the reflectance characteristic of the hybrid surfaces can be decided without a priori information of the relative strengths of the diffuse and specular components. This feature is a significant improvement of conventional algorithms. The hybrid intensity of each point on the surface is considered individually to reduce the distortion met by conventional approaches in the recovery process. Additionally, using the symmetric neural network for diffuse reflection, the light source direction $\mathbf{s}$ in the hidden nodes of the symmetric neural network can be obtained, and the SFS problem can be resolved without specifying illuminant positions in advance. This approach also relaxes the constraint in conventional approaches and is appropriate for practical 3-D surface reconstruction applications.
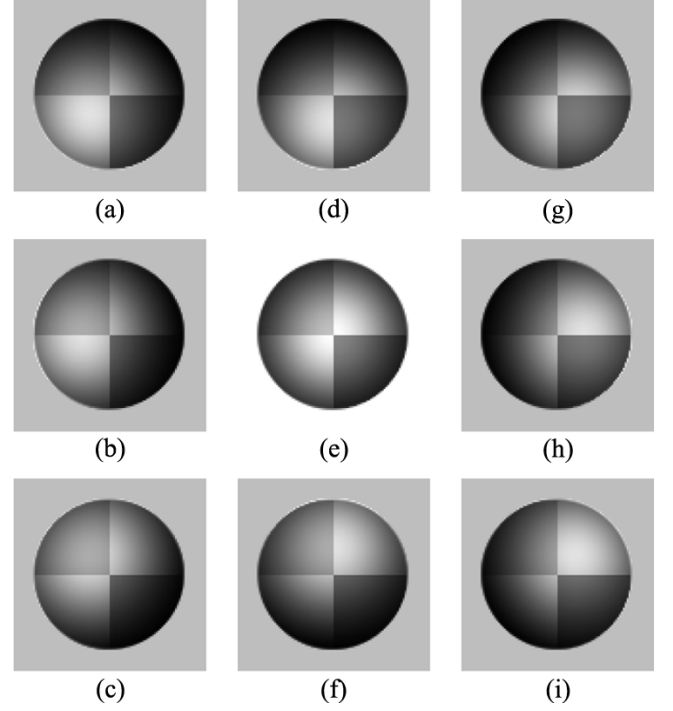

Fig. 3. 2-D sphere images generated with varying albedo and lighting directions (degree of tilt angle and pan angle). (a) $S1 = (60, 135)$; (b) $S2 = (60, 180)$; (c) $S3 = (60, -135)$; (d) $S4 = (60, 90)$; (e) $S5 = (90, 0)$; (f) $S6 = (60, -90)$; (g) $S7 = (60, 45)$; (h) $S8 = (60, 0)$; (i) $S9 = (60, -45)$.

IV. TRAINING ALGORITHM OF THE PROPOSED MODEL

Back-propagation learning is employed for supervised training of the proposed model to minimize the error function defined as

$$E_T = \sum_{i=1}^{m} \left(R_{\text{hybrid}_i} - D_i \right)^2 \quad (15)$$

where m denotes the number of total pixels of the 2-D image, R_{hybrid_i} denotes the ith output of the neural network, and D_i denotes the ith desired output equal to the ith intensity of the original 2-D image. For each 2-D image, starting at the input nodes, a forward pass is used to calculate the activity levels of all the nodes in the network to obtain the output. Then, starting at the output nodes, a backward pass is used to calculate $\partial E_T / \partial \omega$, where ω denotes the adjustable parameters in the network. The general parameter update rule is given by

$$\omega(t+1) = \omega(t) + \Delta\omega(t) = \omega(t) + \eta \left(-\frac{\partial E_T}{\partial \omega(t)} \right) \quad (16)$$

where η denotes the learning rate.

The details of the learning rules corresponding to each adjustable parameter are given below.

Output layer: The combination ratio for each point $\lambda_{d_k}(t)$ and $\lambda_{s_k}(t)$ is calculated iteratively by

$$\lambda_{d_k}(t+1) = \lambda_{d_k}(t) + \Delta\lambda_{d_k}(t)$$
$$= \lambda_{d_k}(t) + 2\eta \left(D_k(t) - R_{\text{hybrid}_k}(t) \right) R_{d_k}(t),$$
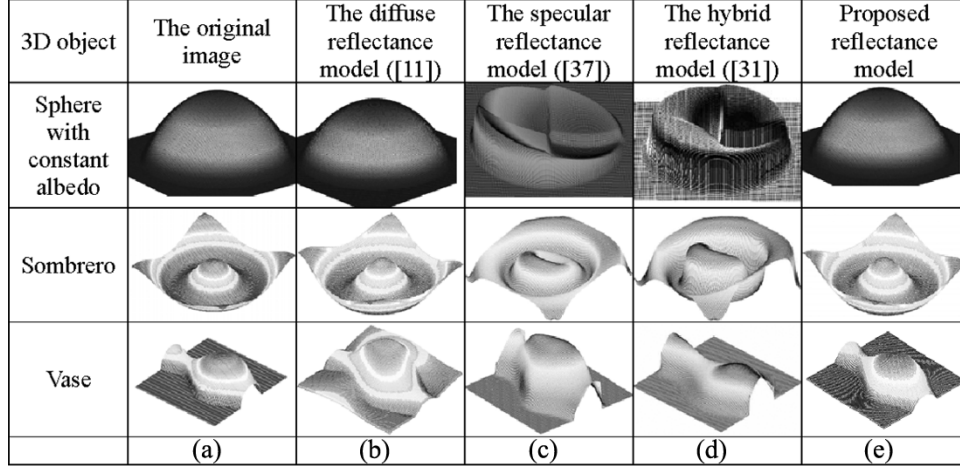$$k = 1, \ldots, m \quad (17)$$

Fig. 4. Comparisons of synthetic images and recovered surfaces of sphere, sombrero and vase. (a) The depth map of the objects. (b) The recovered result by the diffuse reflectance model [11]. (c) The recovered result by the specular reflectance model [37]. (d) The recovered result by the hybrid reflectance model [31]. (e) The recovered result by the proposed approach.

TABLE I
THE ABSOLUTE MEAN ERRORS BETWEEN ESTIMATED AND DESIRED DEPTHS OF A SYNTHETIC OBJECT'S 3-D SURFACES, WHERE BOTH LIGHT AND VIEWING DIRECTIONS ARE UNKNOWN. (Iterations $= 10$)

Mean absolute depth error and CPU time	Lights	The diffuse reflectance model ([11])	The specular reflectance model ([37])	The hybrid reflectance model ([31])	The proposed reflectance model
Sphere with Variant albedo	S1, S4, S7	0.1738	0.5542	0.5270	0.153
	S2, S5, S8	0.1699	0.6079	0.6300	0.151
	S3, S6, S9	0.1730	0.5555	0.5337	0.154
	S1, S5, S9	0.1669	0.5584	0.6031	0.148
	S3, S5, S7	0.1669	0.5944	0.6073	0.148
	CPU time (Avg.)	50.102 sec	309.325 sec	405.283 sec	58.885 sec
Sombrero	S1, S4, S7	0.4125	0.4319	0.4034	0.1396
	S2, S5, S8	0.4125	0.4446	0.4334	0.1395
	S3, S6, S9	0.4124	0.4421	0.4086	0.1399
	S1, S5, S9	0.5126	0.4320	0.4161	0.1514
	S3, S5, S7	0.5124	0.4319	0.4130	0.1516
	CPU time (Avg.)	50.953 sec	310.667 sec	410.490 sec	60.067 sec
Vase	S1, S4, S7	1.3145	0.982	0.854	0.1808
	S2, S5, S8	1.3077	1.018	0.920	0.1859
	S3, S6, S9	1.3103	1.010	0.887	0.1886
	S1, S5, S9	1.3055	1.004	0.887	0.1861
	S3, S5, S7	1.3207	0.725	0.595	0.1877
	CPU time (Avg.)	52.465 sec	316.055 sec	412.783 sec	60.798 sec

$$\lambda_{s_k}(t+1) = \lambda_{s_k}(t) + \Delta\lambda_{s_k}(t)$$
$$= \lambda_{s_k}(t) + 2\eta \left(D_k(t) - R_{\mathrm{hybrid}_k}(t) \right) R_{s_k}(t),$$
$$k = 1, \ldots, m, \qquad (18)$$

where $D_k(t)$ denotes the kth desired output; $R_{\mathrm{hybrid}_k}(t)$ denotes the kth system output; $R_{d_k}(t)$ denotes the kth diffuse intensity obtained from the up subnetwork; $R_{s_k}(t)$ denotes the kth specular intensity obtained from the low subnetwork (as shown in Fig. 1); m denotes the total number of pixels in a 2-D image, and η denotes the learning rate of the neural network.

For a gray image, the intensity value of a pixel is in the interval [0, 255]. To prevent the intensity value of R_{hybrid_k} from exceeding the interval [0, 255], then the rule

$$\lambda_d + \lambda_s = 1 \qquad (19)$$

where $\lambda_d \geq 0$ and $\lambda_s \geq 0$, must be enforced. Therefore, the combination ratio λ_{d_k} and λ_{s_k} is normalized by

$$\lambda_{d_k}(t+1) = \frac{\lambda_{d_k}(t+1)}{\lambda_{d_k}(t+1) + \lambda_{s_k}(t+1)}, \quad k = 1, \ldots, m$$
$$\lambda_{s_k}(t+1) = \frac{\lambda_{s_k}(t+1)}{\lambda_{d_k}(t+1) + \lambda_{s_k}(t+1)}, \quad k = 1, \ldots, m. \qquad (20)$$

Subnetworks: The normal vector calculated from the subnetwork corresponding to the diffuse component is denoted as $\mathbf{n}_{d_k} = (v_{d_{1k}}, v_{d_{2k}}, v_{d_{3k}})$ for the kth point on the surface, and the normal vector calculated from the subnetwork corresponding to the specular component is denoted as $\mathbf{n}_{s_k} = (v_{s_{1k}}, v_{s_{2k}}, v_{s_{3k}})$

TABLE II
THE MEAN ERRORS BETWEEN THE ESTIMATED DEPTHS AND DESIRED DEPTHS OF THE 3-D FACE SURFACES IN FIG. 5. (Iterations= 20)

| Mean error (Eq. 29)	The diffuse reflectance model ([11])	The specular reflectance model ([37])	The hybrid reflectance model ([31])	The proposed reflectance model with const of hybrid ratio	The proposed reflectance model
Fig. 5(a)	1.5524	1.5268	1.5259	1.1586	1.0992
Fig. 5(b)	1.3517	1.3550	1.3533	1.0358	0.9644
Fig. 5(c)	1.9476	1.9542	1.9546	1.5963	1.5348
Fig. 5(d)	1.6007	1.6018	1.6039	1.2913	1.2243
Fig. 5(e)	1.2396	1.2457	1.2443	0.9576	0.9060
Fig. 5(f)	1.9240	1.9155	1.9160	1.5613	1.5224
Average	1.6027	1.5998	1.5997	1.2668	1.2085

(a) (b) (c) (d) (e) (f)
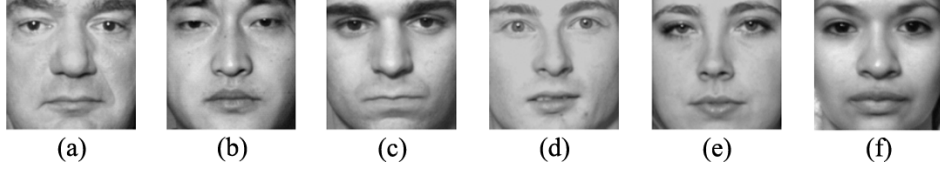
Fig. 5. Six individuals in the Notre Dame Biometrics Database D used to test the proposed approach (these images include both males and females).
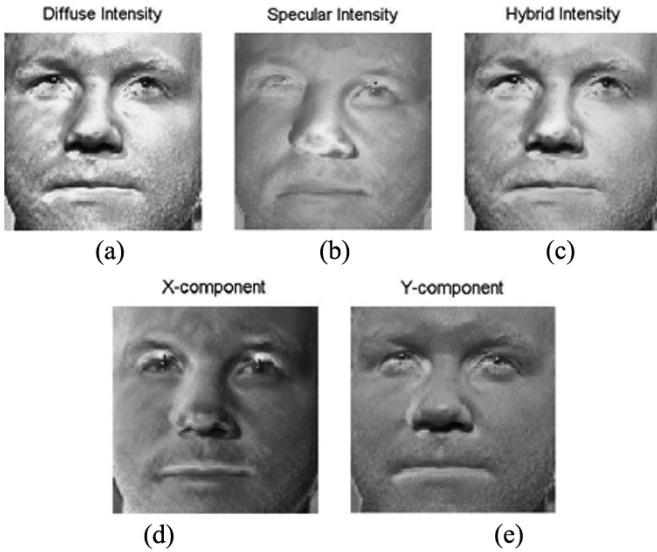

Fig. 6. Estimated reflection components and normal vectors of a human face in the Yale Face Database B by the proposed approach. (a) The diffuse intensity. (b) The specular intensity. (c) The hybrid intensity. These three images are the estimated results of the proposed approach. (d) The X-component of the normal vector. (e) The Y-component of the normal vector.

for the kth point. The normal vectors $\mathbf{n}_{d_k}$ and $\mathbf{n}_{s_k}$ are updated iteratively using the gradient method as

$$
\begin{aligned}
v_{d_{jk}}(t+1) &= v_{d_{jk}}(t) + \Delta v_{d_{jk}} \\
&= v_{d_{jk}}(t) + 2\eta s_j(t)\left(D_k(t) - R_{\text{hybrid}_k}(t)\right), \\
&\quad j = 1, 2, 3, \\
v_{s_{jk}}(t+1) &= v_{s_{jk}}(t) + \Delta v_{s_{jk}} \\
&= v_{s_{jk}}(t) + 2\eta\, r\, h_j(t)\left(D_k(t) - R_{\text{hybrid}_k}(t)\right), \\
&\quad j = 1, 2, 3
\end{aligned}
\tag{21}
$$

where $s_j(t)$ denotes the jth element of illuminant direction $\mathbf{s}$; $h_j(t)$ denotes the jth element of the halfway vector $\mathbf{h}$, and r

denotes the degree of the specular equation in (14). The updated $v_{d_{j,k}}$ and $v_{s_{j,k}}$ should be normalized as follows:

$$
\begin{aligned}
v_{d_{jk}}(t+1) &= \frac{v_{d_{jk}}(t+1)}{\|\mathbf{n}_{d_k}(t+1)\|} \\
v_{s_{jk}}(t+1) &= \frac{v_{s_{jk}}(t+1)}{\|\mathbf{n}_{s_k}(t+1)\|}, \quad j = 1, 2, 3.
\end{aligned}
\tag{22}
$$

To obtain the reasonable normal vectors of the surface from the adaptive hybrid-reflectance model, $\mathbf{n}_{d_k}$ and $\mathbf{n}_{s_k}$ are composed from the hybrid normal vector $\mathbf{n}_k$ of the surface on the kth point by

$$
\mathbf{n}_k(t+1) = \lambda_{d_k}(t+1)\mathbf{n}_{d_k}(t+1) + \lambda_{s_k}(t+1)\mathbf{n}_{s_k}(t+1) \tag{23}
$$

where $\lambda_d(t+1)$ and $\lambda_s(t+1)$ denote the combination ratios for the diffuse and specular components.

Since the structure of the proposed neural networks is like a mirror in the center layer, the update rule for the weights between Layers 2 and 3 of the two subnetworks denoted as $\mathbf{W}_d$ and $\mathbf{W}_s$ (see Fig. 2) can be calculated by the least square method. Hence, $\mathbf{W}_d$ and $\mathbf{W}_s$ at time $t+1$ can be calculated by

$$
\mathbf{W}_d(t+1) = \left(\mathbf{V}_d(t+1)^T\mathbf{V}_d(t+1)\right)^{-1}\mathbf{V}_d(t+1)^T \tag{24}
$$

$$
\mathbf{W}_s(t+1) = \left(\mathbf{V}_s(t+1)^T\mathbf{V}_s(t+1)\right)^{-1}\mathbf{V}_s(t+1)^T \tag{25}
$$

where $\mathbf{V}_d(t+1)$ and $\mathbf{V}_s(t+1)$ denote the weights betweens the output and central layers of the two subnetworks for the diffuse and specular components, respectively.

Additionally, for fast convergence, the learning rate η of the neural network is adaptive in the updating process. If the current error is smaller than the errors of the previous two iterations, then the current direction of adjustment is correct. Thus, the current direction should be maintained, and the step size should be increased, to speed up convergence. By contrast, if the current error is larger than the errors of the previous two iterations, then
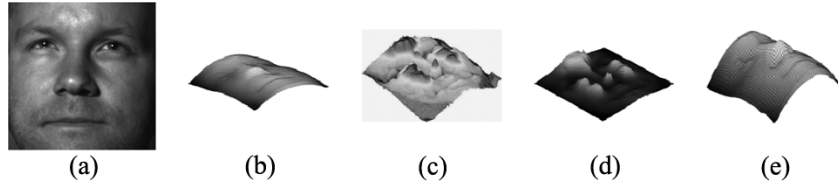
Fig. 7. Reconstructed surfaces of the proposed algorithm compared with three existing approaches. (a) The original 2-D facial image. (b) The recovered result by the diffuse reflectance model [11]. (c) The recovered result by the specular reflectance model by [37]. (d) The recovered result by the hybrid reflectance model [31]. (e) The recovered result by the proposed approach.
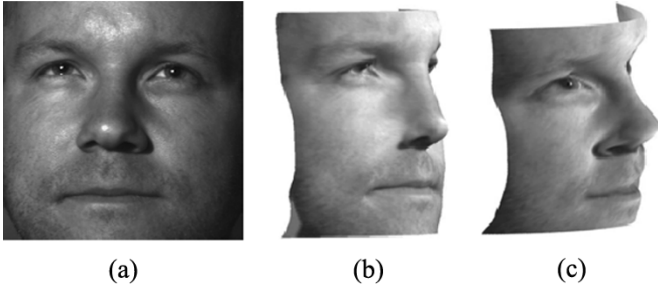

Fig. 8. Reconstructed results from two different approaches. (a) The original 2-D image of the human face. (b) The reconstructed result of the diffuse reflectance model [11]. (c) The reconstructed result of the proposed approach.

the step size must be decreased because the current adjustment is wrong. Otherwise, the learning rate η does not change. Thus, the cost function E_T could reach the minimum quickly and avoid oscillation around the local minimum. The adjustment rule of the learning rate is given as follows:

```
If (Err(t−1) > Err(t) and Err(t−2) > Err(t))
   η(t + 1) = η(t) + ξ,
Else If (Err(t − 1) < Err(t) and Err(t − 2) <
Err(t))
     η(t + 1) = η(t) − ξ, where ξ denotes a
given scalar.
   Else η(t + 1) = η(t).
```

Additionally, the prior knowledge was used as the initial values of the proposed neural network for specific object classes to enhance the results of 3-D surface reconstruction and reduce the learning time. For example, for the face surface reconstruction problem, the normal vectors of a sphere's surface were used as the initial values of the proposed neural network due to their similar structures.

V. EXPERIMENTAL RESULTS AND DISCUSSIONS

In this section, four experiments were performed to demonstrate the proposed approach. In these experiments, both the direction of the light and the observer's viewing direction were unknown. The experiments were intended to test whether the proposed algorithm can reconstruct the objects well even when the lighting and viewing directions are not known in advance. The first experiment used images of the synthetic objects used for testing. The estimated depth map was compared with the true depth map to examine the reconstruction performance. The second experiment used several images corresponding to real

surfaces of human faces for testing. These images were downloaded from the Yale Face Database B[1] under different lighting conditions with variant albedos. In the third and fourth experiments, images of human faces and general objects captured in our photographing environment were used to demonstrate the generality of the proposed approach.

A. Experiment on Images of Synthetic Objects

This section presents quantitative results of synthetic-object reconstruction. The results of the proposed approach were compared with those of three existing approaches: the diffuse [11], specular [37], and hybrid models [31]. Three synthetic objects—sphere, sombrero, and vase, which were mathematically generated by (26)–(28), respectively—were used for testing

$$z(x,y) = \begin{cases} |\sqrt{r^2 - x^2 - y^2}|, & \text{if } x^2 + y^2 \leq r^2 \\ 0, & \text{otherwise} \end{cases} \quad (26)$$

$$z(x,y) = 15 + 15\cos\left(\frac{\pi\sqrt{x^2 + y^2}}{17}\right) \quad (27)$$

$$z(x,y) = \sqrt{f(y)^2 - x^2}. \quad (28)$$

In (26), $r = 45$, $0 \leq x, y \leq 100$, and the center is located at $(x, y) = (50, 50)$. In (28), $f(y) = 0.6 - 0.3y(6y + 1)^2(y - 1)^2(3y - 2)$. The shaded images of the sphere were synthesized with different albedo values and directions as shown in Fig. 3. The different albedos were 0.6 for the bottom right of the sphere, 0.8 for the top left of the sphere, and 1 for the rest part. The locations of light sources in Fig. 3(a)–(i) are S1 = $(60, 135)$, S2 = $(60, 180)$, S3 = $(60, -135)$, S4 = $(60, 90)$, S5 = $(90, 0)$, S6 = $(60, -90)$, S7 = $(60, 45)$, S8 = $(60, 0)$, and S9 = $(60, -45)$, where the first component is the degree of tilt angle and the second component is the degree of pan angle. The center of image is set as the origin of the coordination. The x-y plane is parallel to the image plane. The z axis is perpendicular to the image plane.

Fig. 4 and Table I show the experimental results. Table I presents five groups of images with different illuminant angles from the left, right, and front for 3-D reconstruction. Both the estimated and synthetic surfaces were normalized within the interval [0, 1]. The first row (sphere object) of Fig. 4 indicates that the surface with variant albedo is difficult to handle by conventional approaches and that the proposed approach performs best. Table I indicates that the proposed approach achieves the lowest mean errors in all illumination conditions.

[1]http://cvc.yale.edu/projects/yalefacesB/yalefacesB.html.
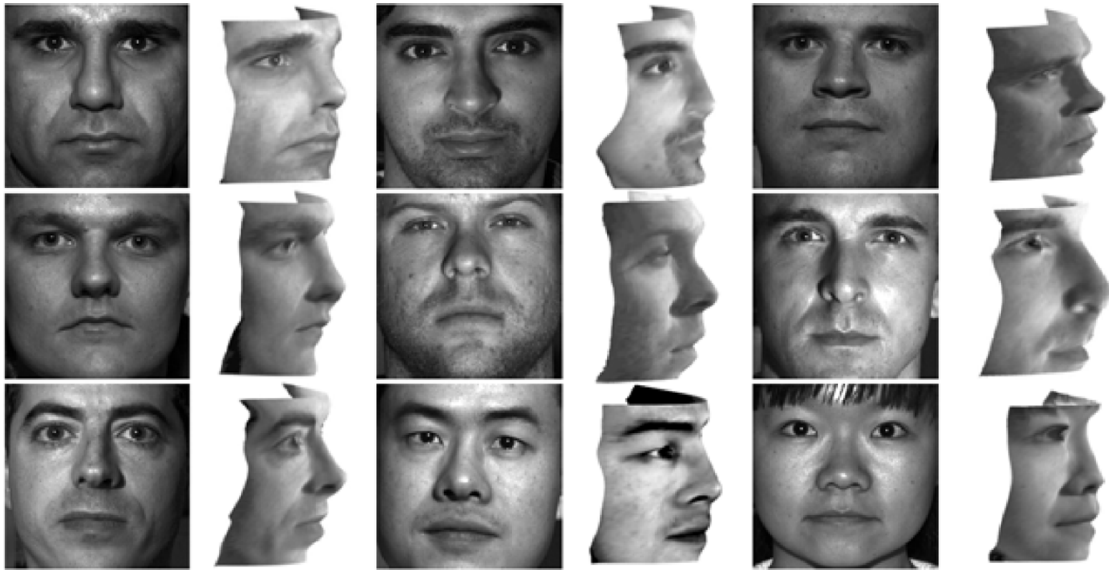
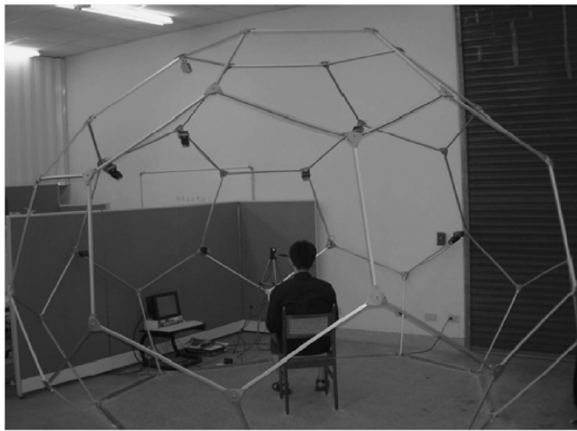Fig. 9. More reconstructed results of human faces in Yale Face Database B by using the proposed approach.

Fig. 10. Photographing environment with eight electronic flashes used to capture images under variations in illumination set up in our lab.

For the sombrero object, the experimental results from the diffuse reflectance model and the proposed approach are very similar to the results for original shape, and the specular and hybrid reflectance models did not yield very good results. The sombrero object had a very sharp shape and many shadows (both cast and attached). Cho's approaches (the specular [37] and hybrid reflectance models [31]) are very likely to cause distortion because they use a single image to recover the shape. The image may have invalid pixel values because of saturated pixels in shadows, and therefore the information from single image may not be sufficient.

For the vase object, the proposed approach reconstructs the synthetic vase very successfully, but the result of the diffuse reflectance model caused clear distortion, which may be due to the convex of the vase. When the vase was illuminated, the convex of the vase was shiny, and the Lambertian assumption could not approximate it well.

Additionally, Table I includes the CPU time used by each approach for shape reconstructions. Each approach was implemented using Matlab 6.1 software on a 1.2 GHz Pentium III-based PC with 256 MB RAM. The analytical results demonstrate that the CPU time used by the proposed approach is similar to that used by the Lambertian model [11] and is much lower than that used by the specular [37] and hybrid reflectance approaches [31].

These experiments also employ the data set in the University of Notre Dame Biometrics Database[2] for objective comparison. The database comprises 3-D face coordinate data and their corresponding 2-D front view. Fig. 5 shows six images of size 160 × 160 from the database. Unlike with the synthetic objects, neither the estimated surface nor the 3-D face surface of these images are normalized within the interval [0, 1]. To evaluate the performance of the proposed approach and other approaches, a function was defined to use the x- and y-partial derivatives of $z(x, y)$ instead of the absolute mean error to calculate the error between the estimated depths and desired depths of 3-D face surfaces as

$$Err = \frac{\sum_{y=1}^{M} \sum_{x=1}^{N} \sqrt{\left(\frac{\partial z(x,y)}{\partial x} - \frac{\partial \hat{z}(x,y)}{\partial x}\right)^2 + \left(\frac{\partial z(x,y)}{\partial y} - \frac{\partial \hat{z}(x,y)}{\partial y}\right)^2}}{M \cdot N}$$

(29)

where $\partial z(x,y)/\partial x$ and $\partial \hat{z}(x,y)/\partial x$ denote the x-partial derivatives of $z(x,y)$ and $\hat{z}(x,y)$, respectively; $\partial z(x,y)/\partial y$ and $\partial \hat{z}(x,y)/\partial y$ denote the y-partial derivatives of $z(x,y)$ and $\hat{z}(x,y)$, respectively; and M and N denote the x and y lengths, respectively. Table II shows the mean errors between the desired depths of 3-D face surfaces and the estimated depths using different approaches. According to the experiment on the images of the three synthetic objects and the dataset in the University of Notre Dame Biometrics Database, the proposed approach can be applied to more general objects and can perform better than existing approaches.

To evaluate the impact of adaptive hybrid ratio on the performance of the proposed network, the performance of the pro-

(a) By s_1 (b) By s_2 (c) By s_3 (d) By s_4

(e) By s_5 (f) By s_6 (g) By s_7 (h) By s_8
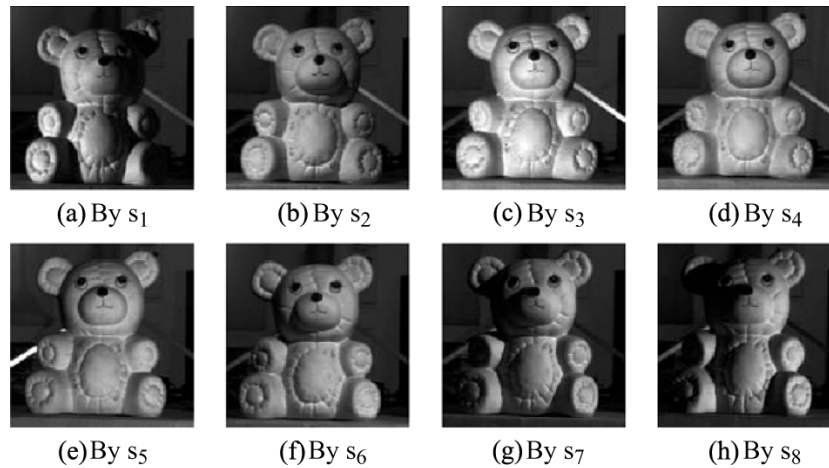
Fig. 11.   Images of an object illuminated by the eight different light sources, respectively. These images can be categorized into three groups: left-hand-side illuminated images (by $s1, s2, s3$), right-hand-side illuminated images (by $s6, s7, e8$), and front illuminated images (by $s3, s4, s5$).

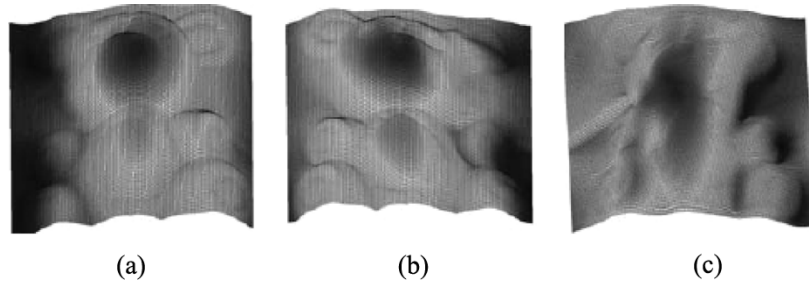(a)   (b)   (c)

Fig. 12.   Reconstructed results of the object obtained from images in Fig. 11. (a) From the images illuminated by s_1, s_2, and s_3. (b) From the images illuminated by s_6, s_7, and s_8. (c) From the images illuminated by s_3, s_4, and s_5.

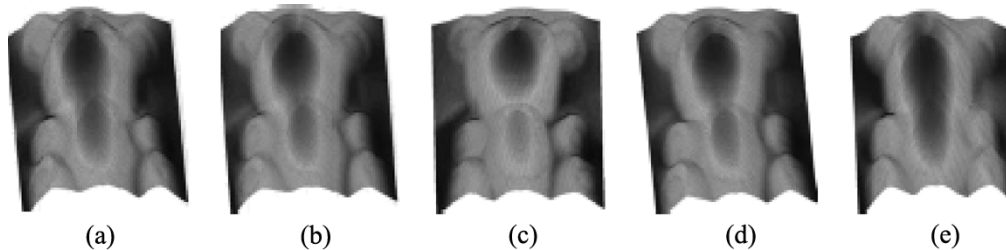(a)   (b)   (c)   (d)   (e)

Fig. 13.   Better reconstructed results of the object from the images in Fig. 11. (a) From the images illuminated by s_2, s_4, and s_7. (b) From the images illuminated by s_2, s_3, and s_7. (c) From the images illuminated by s_2, s_5, and s_6. (d) From the images illuminated by s_2, s_3, and s_8. (e) From the images illuminated by s_1, s_3, and s_8.

posed neural network with constant hybrid ratio was also determined. Table II clearly indicates that the mean errors of the proposed network with constant hybrid ratio are less than those of other approaches and can be further decreased if the hybrid ratio is adjusted in the learning process.

B. Experiment on Yale Face Database B

This experiment employed face images downloaded from the Yale Face Database B for testing. For each person, three images were taken in which their pose is fixed and they are illuminated by three different light directions. After processing with the proposed algorithm, the diffuse, specular, and hybrid intensity values. Additionally, the surface can be reconstructed using the surface normal vectors. Fig. 6 shows an example using the proposed approach to estimate different reflection components and normal vectors of a human face

in the Yale Face Database B. Fig. 7 shows the comparison between three existing approaches and the proposed human face reconstruction approach. Fig. 7(b)–(d) illustrates the reconstructed results of the diffuse [11], specular [37], and hybrid reflectance models [31], respectively. Fig. 7(e) illustrates the reconstructed results of the proposed approach, which clearly indicate that the proposed algorithm performs better than the three conventional approaches. Fig.7 (b) and (e) shows that the reconstructed shape from the proposed approach is sharper and more apparent, especially on part of the nose, than that from the specular reflectance model. The reconstructed results of the specular reflectance model by [37] and the hybrid reflectance model by [31], as shown in Fig. 7(c) and (d), exhibit serious distortions.

The above reconstructed results clearly show that the reconstructed performance of the specular model [37] and the hybrid

(a) By s_1 (b) By s_1,s_8 (c) By s_1,s_5,s_6

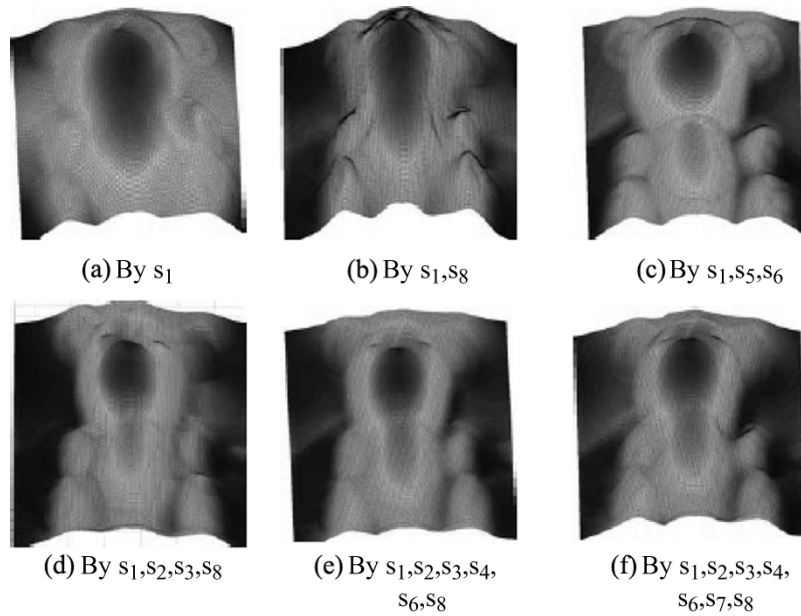(d) By s_1,s_2,s_3,s_8 (e) By $s_1,s_2,s_3,s_4,$ s_6,s_8 (f) By $s_1,s_2,s_3,s_4,$ s_6,s_7,s_8

Fig. 14. Results of 3-D object reconstruction from Fig. 11 using different numbers of images (light sources). (a) The reconstruction uses one 2-D image. (b) The reconstruction uses two 2-D images. (c) The reconstruction uses three 2-D images. (d) The reconstruction uses four 2-D images. (e) The reconstruction uses seven 2-D images. (f) The reconstruction uses eight 2-D images.
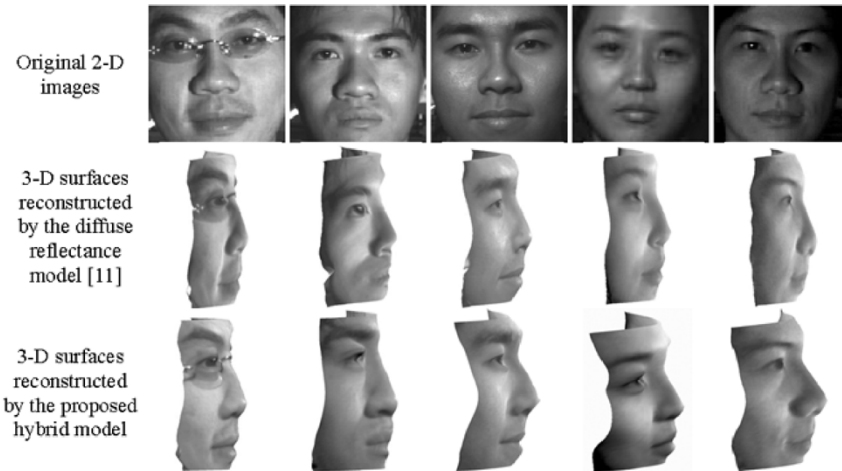

Fig. 15. Reconstructed 3-D facial surfaces from the 2-D pictures of our laboratory team members by the diffuse reflectance model [11] and the proposed approach.

reflectance model [31] do not work well on human faces. Therefore, the following experiment only compared the diffuse reflectance model [11] with the proposed approach. Fig. 8 compares the reconstructed results between the diffuse reflectance model [11] and the proposed approach. The experiments indicate that the reconstructed results of the two approaches are similar. Finally, Fig. 9 shows more reconstructed results of human faces from Yale Face Database B. To compare the results easily, the angles of the faces were set to be equal. The reconstructed results demonstrate that the proposed approach performs well on different human faces in Yale Face Database B.

C. Experiment on Images of Human Faces Captured in Our Photographing Environment

To test and verify the performance of the proposed approach on the facial images of our laboratory members and other images of general objects, a photographing environment was designed as shown in Fig. 10. To equalize the strength of different light sources to photographed objects, the photographic environment was constructed as a hemisphere. The radius of the hemisphere was 2 m. Eight computer-controlled electronic flashes (Mikona MV-328) were placed on the hemisphere at positions $s_1 = (53,30)$, $s_2 = (30,63)$, $s_3 = (20,30)$, $s_4 = (15,45)$, $s_5 = (-20,30)$, $s_6 = (-30,63)$, $s_7 = (-53,60)$, and $s_8 = (-53,30)$, respectively. The light position is represented as (degree of pan angle, degree of tilt angle), where the center of the hemisphere is the same as the origin of the coordinate. Fig. 11 shows the captured images of a bear pottery illuminated by the eight light sources. The images of Fig. 11(a)–(d) were illuminated by the light sources on the right-hand side of the object, while those of Fig. 11(e)–(h) were illuminated by the light sources on the left-hand side of the object with slightly different angles and positions.
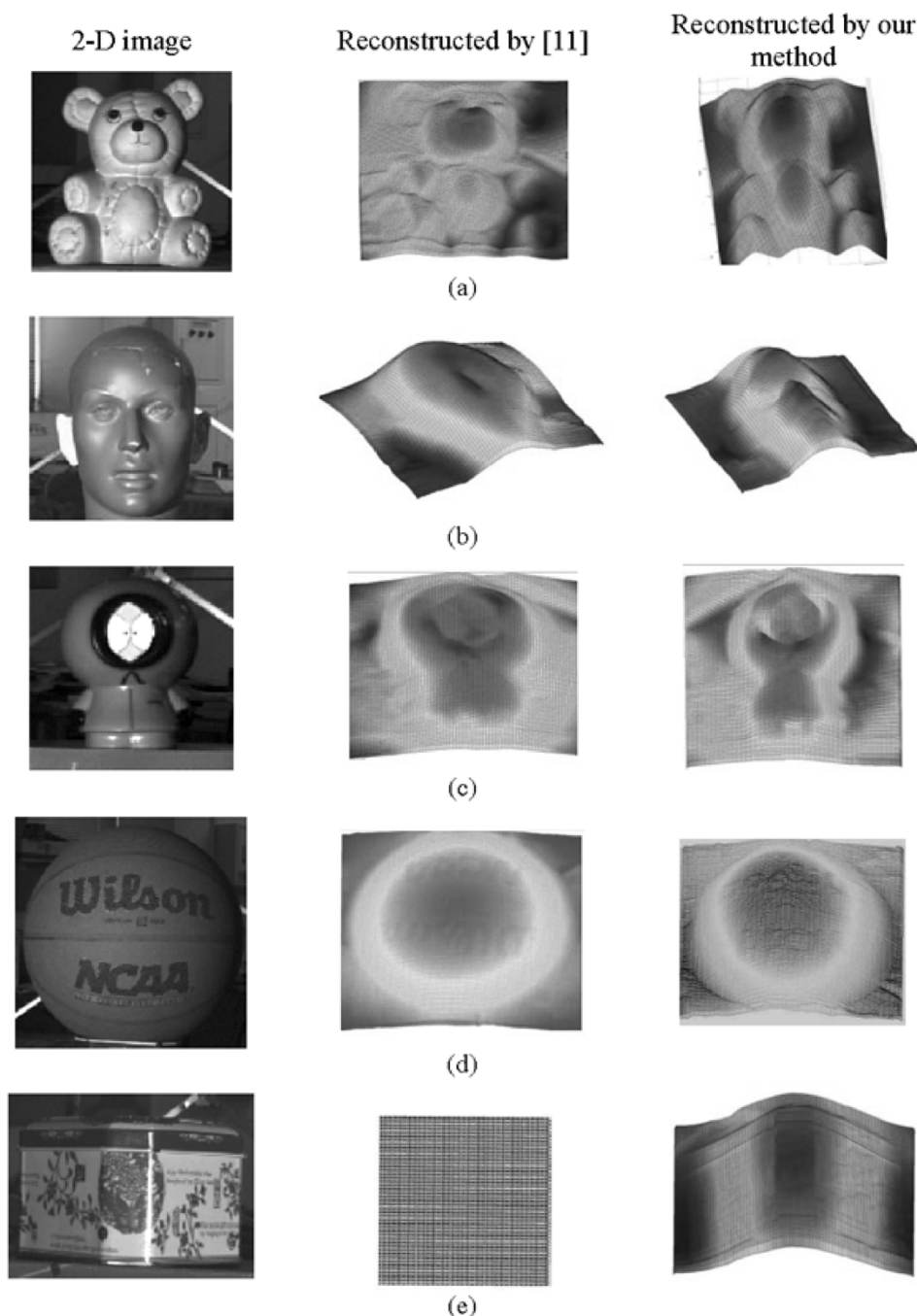
Fig. 16. Reconstructed 3-D surfaces of general objects by the diffuse reflectance model [11] and the proposed approach. The left side of each raw image denotes the 2-D image of the object; the center part denotes the reconstructed surface by the diffuse reflectance model [11]; and the right side denotes the reconstructed surface by the proposed approach. (a) A pottery bear. (b) A dummy head. (c) A toy figurine. (d) A basketball. (e) An octagon iron box.

To understand the influence of illuminant positions and angles on the performance of the proposed approach, the proposed approach was tested on three groups of images: left-hand-side (by $s1$, $s2$, $s3$), right-hand-side (by $s6$, $s7$, $s8$), and front illuminated images (by $s3$, $s4$, $s5$). The illuminant direction was based on the viewpoint of the photographed objects. Fig. 12 shows the reconstructed results calculated from the images corresponding to these three groups, respectively. The reconstruction in Fig. 12(a) was calculated from the images of Fig. 11(a)–(c); the reconstruction in Fig. 12(b) was calculated from the images of Fig. 11(f)–(h), and the reconstruction in Fig. 12(c) was calculated from the images of Fig. 11(c)–(e).

In Fig. 12(a), the object is illuminated from the right-hand side, so its variant intensities are most obvious on the left part of the object's surface. Therefore, the left side of the reconstructed object is better than its right part. Similarly, in Fig. 12(b), the object is illuminated from the left-hand side, so the reconstructed result is better in the right side of the object. The result in Fig. 12(c) is calculated from the images in Fig. 11(c)–(e). These images are illuminated from the front and have very similar image intensities. Therefore, the image information is insufficient to reconstruct the 3-D surface well. If the input images are too similar, then the least square problem in the irradiance equation is difficult to solve. The experimental

results indicate that the information of images illuminated on only one side is (not enough *or* insufficient) for 3-D reconstruction. Therefore, for better recovery, more different illumination conditions should be used to obtain more data on the object's surface.

Fig. 13 shows better reconstructed results of the objects, calculated using different combinations of the left-hand-side illuminated, right-hand-side illuminated, and front illuminated images. Therefore, to obtain better reconstructed results, images should be taken with different illuminant angles from the left, right, and front to provide sufficient surface data for 3-D reconstruction.

To verify the performance of the proposed reconstruction approach under different numbers of light sources, Fig. 14 shows the experimental results using one, two, three, four, seven, and eight images (light sources). Clearly, the results in Fig. 14(a) and (b), using one and two images, are not good, since they overlook many features of the bear. By contrast, the results in Fig. 14(c)–(f) using more than two images are better than those in Fig. 14(a) and (b). The bear retains its key features and shape details. The experimental results in Fig. 14 show that at least three images should be used for fine reconstruction. However, the reconstructed results using more than three images are not necessarily better than the reconstructed result using three images. This experimental result is consistent with the theoretical basis of the proposed approach, which requires evaluating three sets of variables. Hence, using two images only leads to an underdetermined problem, while using more than three images leads to an overdetermined problem. Therefore, three 2-D images were used to reconstruct the surface of a 3-D object using the proposed approach and avoid unnecessary calculation.

Consequently, the influence of the illuminant angles and positions is very significant. Since the proposed approach is based on the shape from the shading approach, the results depend on the reconstructed information from 2-D images. The experimental results demonstrate that the illuminant positions and angles should not be too close.

Fig. 15 shows the reconstructed results of human faces of the laboratory members using the diffuse reflectance model [11] and the proposed approach. For comparison, all angles of the faces were set to be similar. The reconstructed faces from the proposed approach are sharpest and have the strongest facial features. For example, the noses of the reconstructed results are more conspicuous by the proposed approach than by the diffuse reflectance model.

D. Experiment on Images of General Objects Captured in Our Photographing Environment

In the final experiment, images of a pottery bear, a dummy head, a toy figurine, a basketball, and an octagon iron box captured in the photographing environment were used for testing. Fig. 16 shows the 2-D images and the results of 3-D reconstructions using the diffuse reflectance model [11] and the proposed approach. The reconstructed results by the proposed approach seem good. In Fig. 16(d), the imprint of the English words of the basketball clearly appears on the reconstructed surface. In Fig. 16(e), the details of the box such as the ridge and the edge

are also reconstructed well. However, the reconstructed results by the diffuse reflectance model [11] are not as good as those by the proposed approach, and the diffuse reflectance model cannot reconstruct the surfaces of objects, as shown in Fig. 16(e). The experimental results indicate that the proposed approach can reconstruct not only the rough sketch but also the surface detail.

VI. CONCLUSION

This paper proposed a novel 3-D image reconstruction approach, which considers both the diffuse and specular reflection components of the reflectance model simultaneously. Two neural networks with symmetric structures were used to estimate these two reflection components separately and to combine them with an adaptive ratio for each point on the object surface. Additionally, this paper attempted to reduce the distortion caused by variable albedo variation by dividing each pixel's intensity by the corresponding rough-albedo value. Then, these intensity values were fed into the neural network to learn the normal vectors of the surface by the back-propagation learning algorithm. The critical parameters, such as the light source and the viewing direction, were also obtained from the learning process of the neural network. The normal surface vectors thus obtained can then be applied to 3-D surface reconstruction by enforcing integrability method. Extensive experimental results based on a public image database, and the images captured in the photographing environment built in the lab, have demonstrated that the proposed approach can reconstruct the 3-D surfaces of more general and real-world objects better than several existing approaches.

The contributions of this paper can be summarized as follows.

1) Images caught under three different light sources were used to solve the SFS problem without exact light source locations.

2) The proposed approach considers the changes in albedo on the object surface, so that good reconstruction results could be obtained not only for human faces but also for general objects with variant albedo.

3) The proposed symmetric neural network structure with adaptive learning procedure does not require any special parameter setting or smoothing conditions. The system also converges easily and is stable.

4) The proposed network estimates the point-wise adaptive combination ratio of diffuse and specular intensities such that the different reflecting properties of each point on the object surface are considered to achieve better surface reconstruction performance.

REFERENCES

[1] K. P. Horn, "Shape from shading: A smooth opaque object from one view," Ph. D. dissertation, Massachusetts Inst. of Technology, Cambridge, MA, 1970.

[2] K. P. Horn and M. J. Brooks, *Shape From Shading*. Cambridge, MA: MIT Press, 1989.

[3] K. P. Horn and R. W. Sjoberg, "Calculating the reflectance map," *Appl. Opt*, vol. 18, no. 11, pp. 37–75, 1990.

[4] R. Zhang, P. S. Tsai, J. E. Cryer, and M. Shah, "Analysis of shape from shading techniques," in *Proc. IEEE Computer Vision Pattern Recognition*, 1994, pp. 377–384.

[5] M. J. Brooks and K. P. Horn, "Shape and source from shading," in *Proc. Int. Joint Conf. Artificial Intelligence*, 1985, pp. 932–936.

[6] Q. Zheng and R. Chellappa, "Estimation of illuminant direction, albedo, and shape from shading," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, no. 7, pp. 680–702, 1991.

[7] K. M. Lee and C. C. J. Kuo, "Shape from shading with a linear triangular element surface model," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, no. 8, pp. 522–815, 1993.

[8] M. Bichsel and A. Pentland, "A simple algorithm for shape from shading," in *Proc. IEEE Computer Vision Pattern Recognition*, 1992, pp. 459–464.

[9] C. H. Lee and A. Rosenfeld, "Improved methods of estimating shape form shading using the light source coordinate system," *Artif. Intell.*, vol. 26, pp. 125–143, 1985.

[10] A. Pentland, "Shape information from shading: a theory about human perception," in *Proc. IEEE 2nd Int. Conf. Computer Vision*, 1988, pp. 404–413.

[11] S. Georghiades, 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. Machine Intell.*, vol. 23, pp. 643–660, Jun. 2001.

[12] K. P. Horn and M. J. Brooks, "The variational approach to shape from shading," *Comput. Vision Graphics Image Process.*, vol. 33, pp. 174–208, 1986.

[13] R. Szeliski, "Fast shape from shading," *Comput. Vision, Graph., Image Process.: Image Understand.*, vol. 53, no. 2, pp. 129–153, 1991.

[14] W. P. Cheung, C. L. Lee, and K. C. Li, "Direct shape from shading with improved rate of convergence," *Pattern Recognit.*, vol. 30, no. 3, pp. 353–365, 1997.

[15] R. Zhang, P. S. Tsai, J. E. Cryer, and M. Shah, "Shape from shading: a survey," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 21, pp. 690–706, Aug. 1999.

[16] K. Ikeuchi and K. P. Horn, "Numerical shape from shading and occluding boundaries," *Artif. Intell.*, vol. 17, no. 1–3, pp. 141–184, 1981.

[17] R. T. Frankot and R. Chellappa, "A method for enforcing integrability in shape from shading algorithms," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 10, pp. 439–451, Jul. 1988.

[18] K. P. Horn, "Height and gradient from shading," *Int. J. Comput. Vision*, pp. 37–75, 1989.

[19] J. Malik and D. Maydan, "Recovering three dimensional shape from a single image of curved objects," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 555–566, Jun. 1989.

[20] Y. G. Leclerc and A. F. Bobick, "The direct computation of height from shading," in *Proc. IEEE Computer Vision Pattern Recognition*, 1991, pp. 552–558.

[21] O. E. Vega and Y. H. Yang, "Shading logic: a heuristic approach to recover shape from shading," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, pp. 592–597, Jun. 1993.

[22] E. Rouy and A. Tourin, "A viscosity solution approach to shape from shading," *SIAM J. Numer. Analy.*, vol. 29, no. 3, pp. 867–884, 1992.

[23] J. Oliensis, "Shape from shading as a partially well-constrained problem," *Comput. Vision, Graph., Image Process.: Image Understand.*, vol. 54, pp. 163–183, 1991.

[24] R. Kimmel and A. M. Bruckstein, "Shape from shading via level sets," Israel Inst. of Technology, CIS Rep. 9209, 1992.

[25] A. P. Pentland, "Local shading analysis," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 170–187, 1984.

[26] P. S. Tsai and M. Shah, "Shape from shading using linear approximation," *Image Vision Comput. J.*, vol. 12, no. 8, pp. 487–498, 1994.

[27] G. Q. Wei and G. Hirzinger, "Learning shape from shading by a multilayer network," *IEEE Trans. Neural Netw.*, vol. 17, pp. 985–995, 1996.

[28] S. Y. Cho and T. W. S. Chow, "Shape recovery from shading by a new neural-based reflectance model," *IEEE Trans. Neural Netw.*, vol. 10, pp. 1536–1541, 1999.

[29] G. Healey and T. O. Binford, "Local shape from specularity," *Comput. Vision, Graph., Image Process.*, vol. 42, pp. 62–86, 1988.

[30] E. Torrance and E. M. Sparrow, "Theory for off-specular reflection from roughened surfaces," *J. Opt. Amer.*, vol. 57, pp. 1105–1114, 1967.

[31] S. Y. Cho and T. W. S. Chow, "Neural computation approach for developing a 3-D shape reconstruction model," *IEEE Trans. Neural Netw.*, vol. 12, pp. 1204–1214, Sep. 2001.

[32] S. K. Nayar, K. Keuchi, and T. Kanade, "Determining shape and reflectance of hybrid surface by photometric sampling," *IEEE Trans. Robot. Automat.*, vol. 6, pp. 418–431, Aug. 1990.

[33] B. T. Phong, "Illumination for computer generated pictures," *Commun. ACM*, vol. 18, no. 6, pp. 311–317, Jun. 1975.

[34] P. S. Tsai and M. Shah, "Shape from shading with variable albedo," *Opt. Eng.*, vol. 37, no. 4, Apr. 1998.

[35] C. H. Lee and A. Rosenfeld, "Albedo estimation for scene segmentation," *Pattern Recognit. Lett.*, vol. 1, pp. 155–160, 1983.

[36] K. P. Horn, "Obtaining shape from shading from shading information," in *The Psychology of Machine Vision*, P. H. Winston, Ed. New York: McGraw-Hill, 1975, pp. 115–155.

[37] S. Y. Cho and T. W. S. Chow, "Learning parametric specular reflectance model by radial basis function network," *IEEE Trans. Neural Netw.*, vol. 11, pp. 1498–1503, Nov. 2000.

Chin-Teng (CT) Lin (F'05) received the B.S. degree from National Chiao-Tung University (NCTU), Taiwan, in 1986, and the Ph.D. degree in electrical engineering from Purdue University, West Lafayette, IN, in 1992.

He is currently the Chair Professor of Electrical and Control Engineering, Dean of Computer Science College, and Director of Brain Research Center at NCTU. He was Director of the Research and Development Office of NCTU from 1998 to 2000, Chairman of Electrical and Control Engineering Department of NCTU from 2000 to 2003, and Associate Dean of the College of Electrical Engineering and Computer Science from 2003 to 2005. His current research interests are fuzzy neural networks, neural networks, fuzzy systems, cellular neural networks, neural engineering, algorithms and VLSI design for pattern recognition, intelligent control and multimedia (including image/video and speech/audio) signal processing, and intelligent transportation system (ITS). He is the coauthor of *Neural Fuzzy Systems—A Neuro-Fuzzy Synergism to Intelligent Systems* (Englewood Cliffs, NJ: Prentice Hall, 1996) and the author of *Neural Fuzzy Control Systems with Structure and Parameter Learning* (New York: World Scientific, 1994). He has published more than 90 journal papers in the areas of neural networks, fuzzy systems, multimedia hardware/software, and soft computing, including about 60 IEEE JOURNAL papers.

Dr. Lin was elected the IEEE Fellow for his contributions to biologically inspired information systems. He is a Member of Tau Beta Pi, Eta Kappa Nu, and Phi Kappa Phi. He is on Board of Governors of the IEEE Circuits and Systems (CAS) Society for 2005 and IEEE Systems, Man, Cybernetics (SMC) Society for 2003–2005. He is the Distinguished Lecturer of the IEEE CAS Society for 2003 to 2005. He is the International Liaison of the International Symposium of Circuits and Systems (ISCAS) 2005 in Japan, Special Session Cochair of ISCAS 2006 in Greece, and the Program Cochair of the IEEE International Conference on SMC 2006, in Taiwan. He has been the President of the Asia Pacific Neural Network Assembly since 2004. He received the Outstanding Research Award from the National Science Council, Taiwan, from 1997 to present, the Outstanding Electrical Engineering Professor Award from the Chinese Institute of Electrical Engineering (CIEE) in 1997, the Outstanding Engineering Professor Award from the Chinese Institute of Engineering (CIE) in 2000, and the 2002 Taiwan Outstanding Information-Technology Expert Award. He was elected to be one of the *38th Ten Outstanding Rising Stars in Taiwan* (2000). He is an Associate Editor of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS, PART I and PART II, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS, and IEEE TRANSACTIONS ON FUZZY SYSTEMS and *International Journal of Speech Technology*.

Wen-Chang Cheng received the B.S. degree from National Cheng-Kung University, Tainan, Taiwan, R.O.C., in 1997 and the M.S. degree from National Chung-Cheng University, Chiayi, Taiwan, in 1999, both in electrical engineering. He received the Ph.D. degree in electrical and control engineering from the National Chiao-Tung University, Hsinchu, Taiwan, in 2005.

Currently, he is an Assistant Professor in the Information Networking Technology Department, Hsiuping Institute of Technology, Taichung, Taiwan. His current research interests include neural networks, fuzzy systems, neurofuzzy systems, image processing, machine learning, artificial intelligence, and intelligent transportation systems.

Sheng-Fu Liang was born in Tainan, Taiwan, in 1971. He received the B.S. and M.S. degrees in control engineering and the Ph.D. degree in electrical and control engineering from National Chiao-Tung University (NCTU), Taiwan, in 1994, 1996, and 2000, respectively.

From 2001 to 2005, he was a Research Assistant Professor in Electrical and Control Engineering, NCTU. In 2005, he joined the Department of Biological Science and Technology, NCTU, where he is an Assistant Professor. He has been the Chief Executive of the Brain Research Center, NCTU Branch, University System of Taiwan, since September 2003. His current research interests are biomedical signal/image processing, machine learning, fuzzy neural networks (FNN), the development of brain–computer interface (BCI), and multimedia signal processing.