# 國 立 交 通 大 學

## 電機與控制工程學系

## 博 士 論 文

# 基於影像之 3D 物體重建
# Image 3D Object Reconstruction

研 究 生：鄭 文 昌

指導教授：林 進 燈

中 華 民 國 九 十 四 年 一 月

# 基於影像之 3D 物體重建
# Image 3D Object Reconstruction


研 究 生：鄭文昌　　Student：Wen-Chang Cheng

指導教授：林進燈 博士　　Advisor：Dr. Chin-Teng Lin


國 立 交 通 大 學

電 機 與 控 制 工 程 學 系

博 士 論 文

A Dissertation

Submitted to Department of Electrical and Control Engineering

College of Electrical Engineering and Computer Science

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

in

Electrical and Control Engineering

January 2005

Hsinchu, Taiwan, Republic of China

中華民國九十四年一月

# 基於影像之 3D 物體重建

研究生：鄭文昌　　指導教授：林進燈 博士

國立交通大學電機與控制工程學系（研究所）博士班

## 摘　　要

在這篇論文中我們提出改善 3D 物體表面重建以及場景顏色重建的相關新技術。對於 3D 物體的表面重建，我們主要是採用 photometric stereo 的立體重建技術方法，這個方法是基於多張不同光源方向但是相同物體場景的影像來重建影像中物體的深度(depth)資訊，依據以往的研究，反射(specular)成分與散射(diffuse)成分的影響是必須被一起考慮，因此我們將改善的方法分成兩個階段來完成，第一階段我們先針對反射成分與散射成分的結合比例問題提出結合兩個對稱型的類神經網路來取代傳統利用固定比例結合的方法，透過我們提出的類神經網路學習機制，最後可以得到影像中每一個點對應到表面散射及反射成分最適當的混合比例及表面法線向量，進一步利用表面的法線向量，我們使用 Enforcing Integrability 方法重建出影像中物體表面的深度資訊，經實驗驗證我們所提出的類神經網路的方法比傳統固定比例結合的方法能更有效的重建 3D 表面的結果。

為了更有效的結合反射成分與散射成分，第二個階段我們提出一個單一非線性(nonlinear)混合模型同時用來表現物體表面的散射及反射成分，我們不需要從這個模型中將這兩個成分分開，透過非線性獨立成成分分析(independent component analysis)技術的計算，我們最後可以從這單一非線性混合模型中分解出對應到影像中每一點的表面法線向量，接著利用表面法線向量可以重建出影像中物體表面的深度資訊，經實驗驗證我們知道物體表面的反射成分與散射成分是可以用單一模型來表現，其次證實非線性獨立成分分析技術可以有效的應用在此一問題上。

對於 3D 物體的色彩重建，我們也提出一個基於類神經網路的影像中外在環境光源的估測方法，這個估測方法是依據影像上所有顏色在色域圖上的統計直條圖分佈來估測影像外在光源，我們以影像色域統計圖分佈的中心值當作類神經網路的輸入，透過 BP 學習法則學習過的類神經網路，我們可以得到相對應於光源參數的輸出，利用這些估測的光源參數，我們能完成色彩的重建，從實驗的結果可以驗證，不管是定量或定性的實驗顯示，我們所提出的方法確實能正確且有效快速的完成光源的估測及色彩的重建。

# Image 3D Object Reconstruction

Student：Wen-Chang Cheng　　Advisor：Dr. Chin-Teng Lin

Department of Electrical and Control Engineering
National Chiao-Tung University

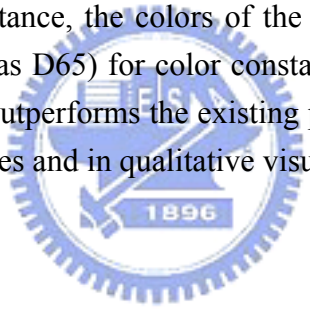
# Abstract

In this thesis, we propose three new techniques to improve the surface reconstruction and color reconstruction of 3D objects. For the surface reconstruction of 3D objects, photometric stereo is able to estimate local surface orientations by using several images of the same surface which are photographed from the same viewpoint but under the illuminations from different directions. According to previous researches, a successful reflectance model for surface reconstruction of 3D objects should combine two major components, the diffusion and specular components. As a result, in this thesis, we categorize the improvement by our methodology into two stages. In the first stage, a new neural-network-based adaptive hybrid-reflectance model is proposed for combining the diffusion and specular components automatically. The supervised learning algorithm is adopted and the hybrid ration for each point is updated in the learning iterations. After the learning process, the neural network can estimate the normal vector for each point on the surface of 3D objects in an image. The enforcing integrability method is applied to the reconstruction of 3D objects by using the obtained normal vectors. The experimental results demonstrate that the proposed network estimates the point-wisely adaptive combination ratio of the diffusion and specular intensities such that the different reflection properties of each point on the object surface are considered to achieve better performance on the surface reconstruction.

In the second stage, we further propose a new nonlinear reflectance model consisting of diffusion and specular components for modeling the surface reflectance of 3D objects in an image. Unlike the previous approaches, these two components are not separated and processed individually in the proposed model. An unsupervised learning adaptation algorithm is developed to estimate the reflectance model based on image intensities. In this algorithm, the post-nonlinear independent

component analysis (ICA) is used to obtain the surface normal on each point of an image. Then, the 3D surface model is reconstructed based on the estimated surface normal on each point of image by using the enforcing integrability method. The results clearly indicate the superiority of the proposed nonlinear reflectance model over the other linear hybrid reflectance model. The experimental results demonstrate that the post-nonlinear ICA method can be used in the problems of surface reconstruction.

For color recovering of 3D objects, a new neural-network-based algorithm for surrounding illumination estimation of image scenes is proposed. This estimation is based upon the chromaticity histogram of a color image, which is obtained by the accumulation of *CIE* chromaticity values corresponding to all the colors in the image. A neural network with a BP learning algorithm is used to model the nonlinearly functional relationship between the central values of the chromaticity histogram and the coefficients of illuminant functions. The trained BP network can then be used to estimate the spectral power distribution of the surrounding illuminant. By substituting this illuminant estimates into the finite-dimensional linear model of surface reflectance, the colors of the image can be recovered with the standard illuminant (such as D65) for color constancy. The experimental results show that the new algorithm outperforms the existing popular compared algorithms, both in quantitative error indices and in qualitative visual perception.



# Acknowledgment

感謝我的指導教授 林進燈博士，他讓我有很大的研究空間，完全不會束縛我的想法，讓我能依照自己的想法進行研究，並且在我遇到困難時，總是不厭其煩的給予指導，讓我順利完成很多研究工作。其次感謝學長 梁勝富博士及蒲賀章博士，有他們的指導我才能順利完成我的論文寫作。

感謝交大多媒體實驗室的全體同學，最後感謝我的家人，他們總是默默的在背後支持著我，讓我完全沒有其它顧慮下順利完成學位，沒有怨言、沒有抱怨，只有包容及鼓勵。

# Contents

# List of Figures

# List of Tables

# 1. Introduction

## 1.1. Motivation

When we use the camera to capture the images of 3D objects and scenes, we also loose the depth information of 3D objects and only obtain the 2D image information. However, the depth information of 3D objects plays an import role in many applications such as the 3D-object recognition and 3D-object display. Moreover, when we use the camera to capture images, the colors of objects in an image will change because of the various camera characteristics and environment illuminations. In order to show the original information of 3D objects, we must recover the true object colors. Hence, the problems of 3D object reconstruction from 2D images including the 3D surface reconstruction and the true color reconstruction. In this thesis, we propose thee new techniques to attack these two problems.

## 1.2. Surface Recovering Methods

Several candidate approaches exist for the recovery of surface topography, including binocular stereo, shape from shading, and photometric stereo.

### 1.2.1. Binocular stereo

*Binocular stereo* is a means of recovering depth by identifying corresponding points in two images developed from different viewpoints. Although binocular stereo has been used successfully in computer vision, it still has several drawbacks that should be overcome: (1) Additional hardware is necessary; as this method required

two cameras. (2) The difficulty in using binocular stereo arises from reliably determining the corresponding features between two separate images. It is essential that the position for us to observe remains fixed during the phase in image acquisition in order to prevent the so called correspondence problem. Implementing the matching algorithm also results in additional computation. (3) The missing depth information can be obtained by using stereoscopic imaging techniques. (4) The depth of surface is preferred to recovering than the surface orientation. All of above conditions introduce noises and artifacts.

### 1.2.2. Shape from shading from a single image

The image typically exhibits a smooth variation in brightness from one point to another, which is known as shading. This perception of shape from gradual changes in brightness is denoted as *shape from shading* (SfS). It was one of the first areas of studies in computer vision and was proposed by Horn [1] in the early 1970s.The topic of shape from shading (SFS) is concerned with determining the shape of an object solely from the intensity variation in the image plane [2]-[10]. Unfortunately, measurements of brightness at a single point in the image only provide one condition whereas describing surface orientation requires two variables. The problem is illposed unless further assumptions are made.

The single-image shape from shading algorithm is still limited even if the exact lighting condition and surface reflectivity are known. One extreme case is that the 3D surface information may be totally lost under certain lighting conditions, and so there is no way to recovery the surface orientation. The problems in extracting shapes from shading by a single image are as follows. (1) It is assumed that an approach relies on having a known reflectance function for a surface. (2) One constraint exists in terms of the mathematical solution for this method. The method relies on the continuity of

*I(x, y)*. This means there are no discontinuities on the surface, which is therefore unsuitable for 3D texture surface estimation. (3) We really need a starting point to grow a solution. The equations that we solved are not over-constrained so the method will be extremely susceptible to noises in the image.

### 1.2.3. Photometric stereo

*Photometric stereo* gives us the ability to estimate local surface orientation by using several images of the same surface taken from the same viewpoint but under the illumination from different directions. It was first introduced by Woodham [11]. The light sources are ideally point sources from some distance away in different directions, so that in each case there is a well-defined direction of light source from which we can measure the surface orientation. Therefore, the change of the intensities in the images depends on both local surface orientation and illumination direction.

Photometric stereo is a way in which the ill-posed problems in shading from shading can be resolved. It uses several images of the same surface under different illumination directions. The advantages of photometric stereo are as follows. (1) Unlike single image by shape from shading algorithms, photometric stereo makes no assumption of the smoothness of the surface. (2) Furthermore, it requires only additional lighting and can be easily implemented at a reasonable computational cost. (3) Each image brings along its own unique reflectance map, therefore we will define a unique set of possible orientations for each point in each image. (4) Photometric stereo can recover not only surface normal but also surface albedo.

## 1.3. Candidate Color Recovering Methods

Without light there is no color. Light sources therefore play a very important part in colorimetry. If the color is self-luminous, such as in the case of fireworks, for example, then the light source itself is the color. But, more often, colors are associated

with objects that, instead of being self-emitting, reflect or transmit light emitted by light sources. Thus, a variation of the illuminant in the scene changes the color of the surface as it appears in an image. As far as the human eyes are concerned, the color of the surface is almost different from each other. This phenomenon is called the *color constancy*. However, this function of color constancy does not exist for cameras, so we need to have an automatic color constancy function in order to obtain the correct color images from cameras.

The goal of color constancy can be defined as the transformation of a source image while being taken under an unknown illuminant for a target image which is identical to that which would have been obtained by the same camera in the same scene under a standard illuminant (such as D65). Thus, there are two stages in the process of color constancy. The block diagram in Fig. 1-1 illustrates this process. The first stage estimates the spectral power distribution of the surrounding illuminant of an input image, $\hat{E}(\lambda)$, and based on which we can correct the image pixel-wisely in the second stage with standard illuminant D65.

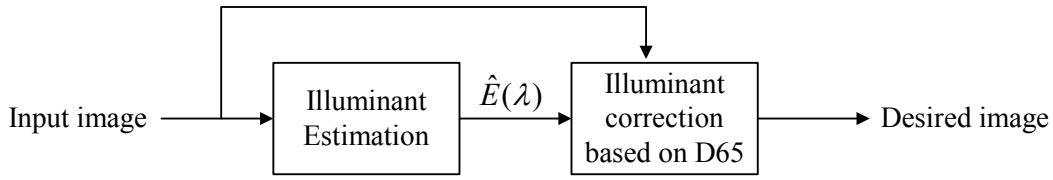



Figure 1-1. Block diagram of the color constancy process.

It is difficult to estimate the surrounding illumination from an image itself, because we totally have no idea about what color consists of an image. There are several methods that have been proposed for determining the surrounding illumination based on some strict assumptions [12]. These candidate approaches exist for the surrounding illumination estimation, including gray-world methods, color by correlation, and neural networks methods.

### 1.3.1. Gray World Methods

The "Gray-World" method is one well-known prior art technique, which estimates the color of the surrounding illumination as the average of color in the camera image [14]-[16]. This method is quite unreliable because the average is very unstable. For example, the color of the image of a large field of grass will be primarily green, so the average color in the image will be green. Color correction of such an image based on the gray-world technique produces readily perceptible color errors throughout the image. Another prior art method of estimating the color of the surrounding illumination is the brightest surface method [13], [17]-[20]. This method effectively uses the maximum value found within the image for each of the three RGB channels of color information as the estimate of the color of the surrounding illuminant. This method is also unstable because it depends on the assumption that there will be a surface that is maximally reflective in each of the three RGB channels of color information everywhere in the scene. This assumption is frequently violated. These methods are improper in general, though they are simple and easy.

### 1.3.2. Neural Network Methods

Usui [21] described a method of estimating the surrounding illumination by using a simple neural network as a decorrelator to minimize the correlation among the RGB information of an image. This simple neural network has three inputs and three outputs, but by itself does not accomplish satisfactory color constancy. In 1999, Funt et al. [22] proposed another neural network approach. The image colors are first mapped into a chromaticity space that is then divided into a plurality of separate regions. For each region, the binary value "1" is assigned to the region if the region contains chromaticity values, and the other binary value "0" is assigned to the region if it does not contain a chromaticity value. The assigned values are then used as inputs

to a pre-trained neural network, in which there are two output nodes representing the white point chromaticity of the surrounding illuminant. The drawbacks are that the input part is too complex and the weighting values are not easy to converge within the back-propagation (BP) algorithm.

### 1.3.3. Color by Correlation

The color correlation approach for the surrounding illumination estimation is proposed in [23]-[25]. A correlation matrix memory is built to correlate the data from any camera image to reference images under a range of illuminants. The vertical dimension of the matrix memory (the columns) is a rearrangement of the two-dimensional chromaticity space into a raster list of points. The horizontal dimension corresponds to a similar raster list that is the chromaticity for all possible illuminants from the view of the device. When a camera produces a picture image, the data are converted to chromaticity and a vector is created corresponding to the values existing in the scene. We can then multiply this vector by each column in the correlation matrix to create a new matrix. Each column is then summed, and the resulting values form a vector that represents the likelihood of each reference source being the surrounding illumination. Despite of this successful method, it suffers from some important drawbacks: (1) It needs to have enough memory space to save the correlation matrix, so it would make it difficult to implement. (2) This method uses a finite quantum of light sources to approximate all possible illuminants, and approximating errors exist since the majority determines the winner in estimating the light sources. Although increasing the number of light sources in the correlation matrix could solve this problem, it also results in the increase of the need of memory at the same time. (3) The color of the surrounding illuminant has two important attributes: chromaticity and luminance. The correlation method shows promise as a

means of selecting an appropriately adopted white chromaticity, but it does not provide information about the white point of the surrounding illuminant. Conversely, the previously described prior art methods provide a means of selecting the white point of the surrounding illuminant, but not by the chromaticity. (4) In some situations, some colors would occupy most range of the image. For examples, the green region would dominate the whole image if we got images from grass, or the images gotten from a whole blue sky or the ocean all belong to this kind of images. In these situations, we call most colors in the image as the *dominant color*. The dominant color usually leads to estimation error in the existing approaches.

## 1.4. Concluding Remarks

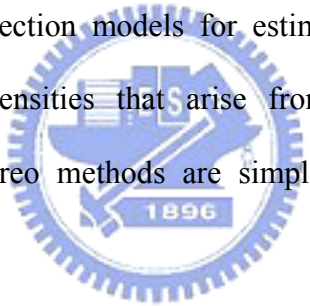
In this chapter, several traditional approaches for the recovery of surface topography were presented, where photometric stereo approach is able to estimate local surface orientation. Therefore, two approaches based on photometric stereo for improving the surface reconstruction were proposed in this thesis. First, a new neural-network-based adaptive hybrid-reflectance model is proposed to combine the diffusion and specular components automatically. The supervised learning algorithm is adopted and the hybrid ration for each point is updated in the learning iterations. Second, we further proposed a new nonlinear reflectance model that consists of the diffusion and specular components for modeling the surface reflectance of 3D objects in an image. Unlike the previous approaches, these two components are not separated and processed individually in the proposed model. An unsupervised learning adaptation algorithm is developed to estimate the reflectance model based on image intensities. In this algorithm, the post-nonlinear independent component analysis (ICA) is used to obtain the surface normal on each point of an image. Then, the 3D surface model is reconstructed based on the estimated surface normal on each point of the

image by using the enforcing integrability method.

For color recovering of 3D objects, we proposed a new neural-network-based algorithm for surrounding illumination estimation of the image scene. This estimation is based upon the chromaticity histogram of a color image, which is obtained by the accumulation of *CIE* chromaticity values corresponding to all colors in the image. The experimental results show that the new algorithm outperforms those existing popular algorithms in both the quantitative error index and the qualitative visual perception.

# 2. Photometric Stereo

The appearance of a surface in an image results from the effects of illumination, shape, and reflectance. Reflectance models have been developed to characterize the image radiance with respect to the illumination environment like viewing angles and material properties. These models provide a local description or reflection mechanism that can serve as a foundation on the representations of appearance. Photometric stereo approaches utilize reflection models for estimating surface properties from transformation of image intensities that arise from illumination changes [11]. Furthermore, photometric stereo methods are simple and elegant for Lambertian methods.



## 2.1. Introduction

Determining the shape of objects from an image in a scene is extremely difficult. The image typically exhibits a smooth variation in brightness from one point to another, which is known as shading. This perception of shape from gradual changes in brightness is denoted as *shape from shading* (SfS). It was one of the first areas of study in computer vision and was proposed by Horn [1] in the early 1970s. Unfortunately, measurements of brightness at a single point in the image only provide one constraint, whereas describing surface orientation requires two variables. Therefore, it is an ill-posed problem.

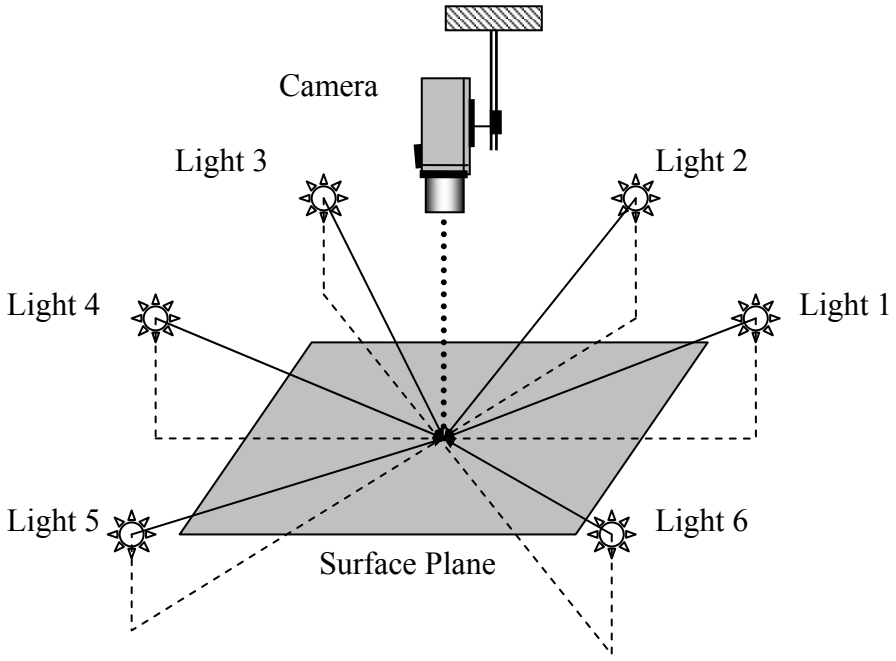



Figure 2-1 Illustration of photometric stereo geometry.

Photometric stereo approach is able to estimate local surface orientation by using several images of the same surface taken from the same viewpoint but under illuminations from different directions. It was first introduced based on the Lambertian reflectance model by R. J. Woodham [11]. It has received wide attention and several efforts have been made to improve the performance of recovery [26]-[44]. The main limitation of classical photometric stereo approach is that the light source positions must be accurately known and this necessitates a fixed, calibrated lighting rig. Hence, an improved photometric stereo method for estimating the surface normal and the surface reflectance of objects without a priori knowledge of the light source direction or the light source intensity is proposed by Hayakawa [43]. The method used the singular-value decomposition (SVD) method to factorize image data matrix of three different illuminations into surface reflectance matrix and light source matrix based on the Lambertian model. However, they still used one of the two added

constraints (i.e., at least 6 pixels in which relative value of the surface reflection is constant or known and at least 6 frames in which the relative value of the light-source intensity is constant or known) for finding the linear transformation between the surface reflectance matrix and the light source matrix. McGunnigle [30] introduced a simple photometric stereo scheme which only considered a Lambertian reflectance model, where the self and cast shadow as well as inter-reflections were ignored. Three images at tilt angle of 90° increments were captured. He suggested using his method as a first estimate for an iterative procedure. In fact, this method is a simplified version of Woodham's method in which the illumination directions are chosen by mathematics simplification. Belhumeur etc. [33] showed that a generalized bas-relief transformation is a transformation of both the surface shape and the surface albedo for an arbitrary Lambertian surface. The set of images of an object in fixed post but under all possible illumination conditions is a convex cone (illumination cone) in the space of images. When the surface reflectance can be approximated as Lambertian, this illumination cone can be constructed from a handful of images acquired under variable lighting. They used as few as seven images of a face seen in a fixed pose, but illumination by point light sources at varying, unknown position, to estimate its surface geometry and albedo map up to a generalized bas-relief transformation. Despite they announced their success under unknown light source directions, the estimation of surface methods still need to be assisted with some added constraints or more images.

Another more difficult problem is estimating a surface with an unknown reflectance map. Cho and Minamitani [27] tried to reduce 3D reconstruction errors due to specularities. Since the specular reflection produced incorrect surface normal by elevating the image intensity, they readjusted the pixel with greatest intensity by

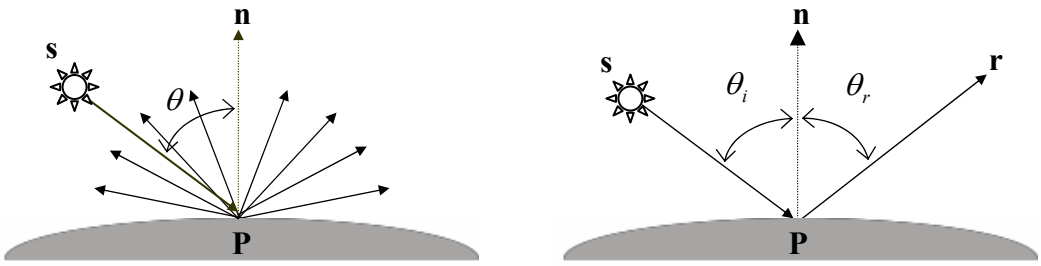
re-scaling with a modified reflectivity. Kay and Caelli [29] used the photometric stereo method to estimate not only the surface normal but also the roughness parameters associated with the Torrance-Sparrow (TS) reflectance model. The basis they used was to apply non-linear regression techniques to the photometric stereo method. Nayar [31] used a linear combination of Lambertian and an impulse specular component. He used distributed light sources for photometric stereo of surface whose reflection is a linear sum of specular and Lambertian components. A proper reflectance model could help us to reconstruct the surface shape accurately from the variation of image intensity corresponding to the reflection characteristic of the surface. In next section we shall discuss the basis reflectance models.

## 2.2. The Basic Reflectance Models

There are mainly two kinds of light reflection components considered in computer vision: diffuse reflection and specular reflection. Diffuse reflection is a uniform reflection of light with no directional dependence for the viewer. The phenomenon of diffuse reflection is illustrated in Fig. 2-2(a). **s** is a point light source, **n** is the normal vector of the surface on point P, $\theta$ is the angle between light source direction and the normal vector of the surface. When **s** illuminates straightly to the surface, the diffuse reflection scatters incoming light equally in all directions. Thus we have identical reflected energy for all viewing directions. The light reaching the surface is reflected in the reflected direction with the same angle. The phenomenon of specular reflection is illustrated in Fig. 2-2(b). It means if a point light source **s** illuminates to the surface, the reflected light is visible only at the reflected direction **r**, where $\theta_i = \theta_r$.

It is important that a good reflectance model should be able to describe the surface shape accurately from the image intensity [31]. Basically, the reflectance

surface can be categorized to be Lambertian or non-Lambertian. The Lambertian surfaces are surfaces that only have diffuse reflectance, which implies that the surface reflects light equally in all direction. On the other hand, the non-Lambertian model considers the specular component in addition to the diffuse component in the Lambertian model.




(a) (b)

Figure 2-2 Illustration of diffuse reflection. **s** is a point light source, **n** is the normal vector of surface on point P, $\theta$ is the angle between light source and the normal. The diffuse reflection scatters incoming light equally in all directions. (b) Illustration of specular reflection. It obeys Shell's law, that is, $\theta_i = \theta_r$ .

### 2.2.1. Lambertian Model

Suppose that the recovering of surface shape, denoted by $z(x, y)$, from shaded images depends upon the systematic variation of image brightness with surface orientation, where *z* is the depth field, and *x* and *y* form the 2D grid over the domain *D* of the image plane. Then, the Lambertian reflectance model used to represent a surface illuminated by a single point light source is written as:

$$R_d\left(\mathbf{n}(x, y), \alpha(x, y)\right) = \max\left\{L\,\alpha(x, y)\mathbf{s}^T\mathbf{n}(x, y), 0\right\}, \quad \forall x, y \in D\,, \tag{2-1}$$

where $R_d(\cdot)$ is diffuse component intensity, $\alpha(x, y)$ is diffuse albedo on position $(x, y)$ of surface, **s** is a column vector indicating the direction of point light, and *L* is light strength. The surface normal on position $(x, y)$, denoted by $\mathbf{n}(x, y)$, can be represented as

$$\mathbf{n}(x,y)=\frac{\begin{bmatrix}-p(x,y) & -q(x,y) & 1\end{bmatrix}^T}{\sqrt{p^2(x,y)+q^2(x,y)+1}}, \tag{2-2}$$

where $p(x,y)$ and $q(x,y)$ are the $x$- and $y$- partial derivatives of $z(x,y)$, respectively. In Eq. (2-1), $\max\{\cdot\}$ sets all negative components that correspond 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].

The Lambertian model describes a simple non shiny surface where any incident light is reflected evenly in all directions after modulation by the surface's reflectivity. It is a simple but useful reflectance model. It is commonly adopted in the field of computer vision as a model of the ideal surface. Despite the simplicity and the popularity of the Lambertian model, it is quite well known that this model is unable to generalize with strong specular components. However, in most cases, the surface does not often contain the Lambertian surface because the light source is often located at finite distance and at an unknown position. Therefore, as far as practical application in general is concerned, it is not enough for considering only diffuse component. As a result, the non-Lambertian surface model is considered.

### 2.2.2. Non-Lambertian Model

In order to effectively exhibit the reflectance model, both the diffuse component and specular component should be considered for reconstructing the surfaces of 3D objects. The kind of hybrid model is called non-Lambertian model. Specular component occurs when the incident angle of the light source is equal to the reflected angle and this component is formed by two terms: the specular spike and the lobe. The specular spike is zero in all directions except for very narrow range around the directions of specular reflectance. The specular lobe spreads around the direction of specular reflectance.

Healy and Binford [45] derived a more refined model by simplifying the Torrance-Sparrow model [46], in which the Gaussian distribution is used to model the facet orientation function and the other components are considered as constants. Defining the halfway-vector as $\mathbf{h} = (\mathbf{s} + \mathbf{v}) / \|\mathbf{s} + \mathbf{v}\|$ which represents the normalized vector sum between the light source direction $\mathbf{s}$ and the viewing direction $\mathbf{v}$, the specular component is represented as:

$$R_s = \frac{1}{\sqrt{2\pi\beta}} \exp(-\frac{\varphi^2}{2\sigma^2}), \tag{2-3}$$

where $\varphi$ is 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 $\sigma$ is the standard deviation, which can be interpreted as a measure of the roughness of the surface. The specular reflectance model in Eq. (2-3) is popular, but it has many parameters to be determined.

The other well-known specular model is Phong's model [47]. For specular reflection, the amount of light seen by the viewer depends on the angle $\delta$ between the perfect reflected ray $\mathbf{r}$ and the direction of the viewer $\mathbf{v}$. Phong's model says that the light perceived by the viewer is proportional to $\cos\delta^r$ and it can be represented as

$$R_s(\mathbf{n}(x, y), \mathbf{h}(x, y)) = (< \mathbf{n}(x, y), \mathbf{h}(x, y) >)^r = (\cos\phi)^r, \ \forall x, y \in D \tag{2-4}$$

where $r$ is a constant. Different values of the constant $r$ represent different kinds of surfaces which are more or less mirror-like. The mathematical model of Phong's model [47] is simpler than that of Healy and Binford [45].

With the specular component described in Eq. (2-4), the non-Lambertian model proposed in [31] can be represented as following linear combination equation:

$$R_{hybrid}(x,y) = \lambda R_d(x,y) + (1-\lambda) R_s(x,y), \quad \forall x, y \in D, \tag{2-5}$$

where $R_{hybrid}$ is the total intensity of the hybrid surface, and $R_d$ and $R_s$ are the diffuse intensity and the specular intensity, respectively, and $\lambda$ is the weight of the diffuse component. However, the existing approach considers only the linear hybrid combination as described by Eq. (2-5). It is not enough to model a nonlinear hybrid reflectance model.

## 2.3. Concluding Remarks

Photometric stereo allows us to estimate local surface orientations by using several images of the same surface taken from the same viewpoint but under illuminations from different directions. Unlike single image by the shape from shading algorithms, photometric stereo makes no assumption of the smoothness of the surface. Furthermore, it requires only additional lighting and can be easily implemented at a reasonable computational cost. In addition, photometric stereo can recover not only surface normal but also surface albedo.

On the other hand, we introduce the Lambertian reflectance model. The Lambertian model describes a simple non shiny surface where any incident light is reflected in all directions after the modulation by the reflectivity of surface. It is a simple but useful reflectance model. However, in most cases, the surface does not often contain the Lambertian surface. Next, the non-Lambertian surface model is considered.

In this thesis, we propose two novel nonlinear reflection models that consist of the diffusion and specular components for photometric stereo. In the next section, we shall give more discussions in the novel nonlinear reflection models.

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

In this chapter, a new neural-network-based adaptive hybrid-reflectance model is proposed for 3D surface reconstruction. The neural network combines the diffuse component and specular component into a hybrid model automatically. The characteristic of each point as well as variant albedo is also considered individually in the proposed model to avoid the distortion of surface reconstruction. The inputs of the neural network are the pixel values of the 2D images to be reconstructed, and then the normal vectors of the surface can be obtained from the output of the neural network after supervised learning, where the illuminant direction does not need to be known in advance. Finally, the enforcing integrability method is applied for the reconstruction of 3D objects by using the obtained normal vectors. In our experiments, facial images and images of other general objects are used to test the performance of the proposed method. The experimental results demonstrate that our neural-network-based adaptive hybrid-reflectance model can be successfully applied to more general objects and achieve better performance for 3D surface reconstruction as compared to some existing approaches.

## 3.1. Introduction

Recently, multi-layer neural networks have also been employed to deal with the

photometric stereo problem [48]-[50]. However, these approaches are still under the restriction of the Lambertian model in which the direction of light source must be given or reasonably estimated. Obviously, this restriction makes the algorithm impracticable for many applications in which information of the illumination is not available. In addition, the reflectance of objects do not always follow the Lambertian model, therefore, a more general model is required. According to the study in [49], a successful reflectance model for surface reconstruction of objects should combine two major components: the diffuse component and the specular component. The Lambertian model was established to describe the relationship between the surface normal and the light source direction by generally assuming the surface reflection is due to diffuse reflection only. It implies that the surface reflects the light equally in all directions. Thus the specular component is ignored in the Lambertian model.

In order to model the specular component, some specular models or non-Lambertian models have been proposed. Healy and Binford [45] employed the Torrance-Sparrow model [46], which assumes that a surface is composed of small, randomly oriented, mirror-like facets, to obtain local shape from specularity. Cho and Chow [50] 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. The viewing direction and the light source direction are no longer required for this model and the performance of shape recovery is more robust than that of former approaches. However, it still has two drawbacks that should be overcome: (1) The albedo of the surface is ignored or considered as constant; therefore, the recovered shape will be distorted. In general cases albedo is variant in different regions of the surface. (2) The combination ratio between diffuse component and specular component is regarded as

a constant which is determined by tried and error. Thus, the hybrid combination method proposed in [50] is not suitable for the surface reconstruction of human faces or some general objects whose albedo and reflecting characteristic is not the same for the whole surface.

In this chapter, a novel adaptive hybrid-reflectance model is proposed to represent more general conditions. This model intelligently combines both diffuse reflection component and specular reflection component, and the hybrid ratio does not need to be determined in advance. Both pure diffuse reflection component and pure specular reflection component are generated by the similar feed-forward neural network structures. A supervised learning algorithm is used to tune up the pointwise hybrid ratio automatically based on image intensities and to obtain the normal vectors of the surface for reconstruction. The proposed model will estimate the illuminant direction, viewing direction and normal vectors of surfaces of the object for reconstruction after training. Therefore, we could produce new shaded images under different illuminant conditions by controlling the above parameters. We can also reconstruct the 3D surface according to these normal vectors by employing the existing approaches such as the enforcing integrability method [51], etc. In addition, the albedo and the reflecting characteristic of each point of the surfaces are considered individually in our method. According to the experimental results presented in Section 3.4, our shape recovery algorithm is more robust for the recovery of the surfaces with variant albedo and complex reflecting characteristic.

The rest of this chapter is organized as follows. Section 3.2 describes the proposed hybrid-reflectance model that includes the diffuse component and the specular component. The details of the neural-network-based hybrid-reflectance model and its derivations of learning rules are presented in Section 3.2 and Section

3.3. Extensive experiments have been performed to evaluate the performance of the proposed approach, and parts of the results are presented in Section 3.4. Conclusions are summarized in the last section.

## 3.2. Neural-Network-Based Adaptive Hybrid-Reflectance Model

In this section, we propose a novel neural-network-based hybrid-reflectance model, in which the hybrid ratio of diffuse and specular components is regarded as adaptive weights of 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 estimate the proper hybrid ratio for each point on the 3D surface of any object in an image. In this manner, we can integrate diffuse component and specular component intelligently and efficiently. In addition, the variant albedo effect is also considered in our hybrid-reflectance model. It has been claimed that the variant albedo effect will influence the performance of 3D surface reconstruction and cause distortion in conventional methods [52]-[54].

The schematic block diagram of our proposed adaptive hybrid-reflectance model is shown in Fig. 3-1. The structure diagram consists of the diffuse part and the specular part. They are used to describe the characteristic of the diffuse component and specular component of our adaptive hybrid-reflectance model, respectively, by two neural networks with similar structure. The composite intensity $R_{hybrid}$ is obtained by using the adaptive weights $\lambda_d(x,y)$ and $\lambda_s(x,y)$ to combine the diffuse intensity $R_d$ and the specular intensity $R_s$. The inputs of the system are 2D image intensity of each point and the outputs are the learned reflectance map. In solving the photometric stereo problem by our neural-network-based reflectance

model, the cost function $E_T$ is minimized to update the neural parameters. After training, the normal vectors of the surface can be obtained from the reflectance model to reconstruct the 3D shape of the object for reconstruction and we can also combine it with different light source directions and viewing directions to produce new shaded images.

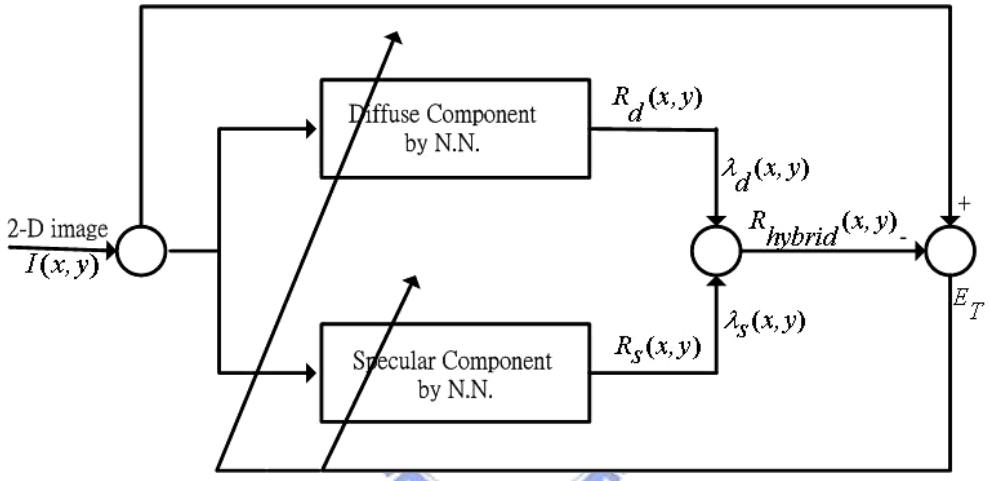



Figure 3-1. Block diagram of the proposed adaptively hybrid reflectance model.

The structure of the proposed symmetric neural network used to simulate the diffuse reflection model is shown in Fig. 3-2. The input/output pairs of the network are arranged in a form like a mirror in the center layer and the number of input nodes is equal to the number of output nodes; therefore, we call it a symmetric neural network. We separate the light source direction and the normal vector from the input 2D images in the left part of the symmetric neural network and then we combine them inversely to generate the reflectance map for diffuse reflection in the right part of the network. In the following, we will discuss the function of each layer in details.

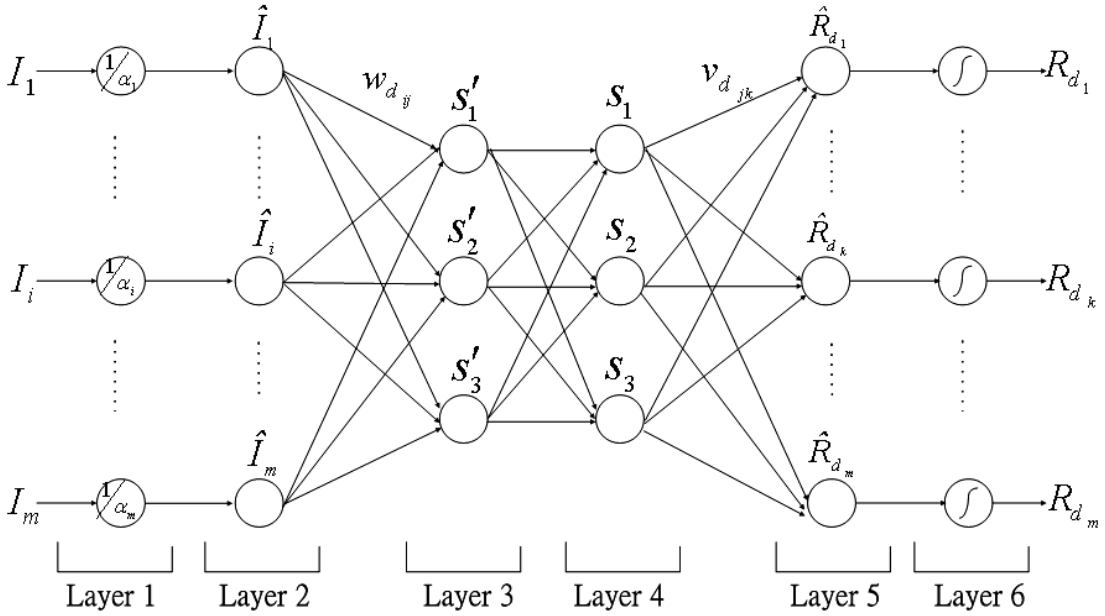



Figure 3-2. Structure of the symmetric neural network for diffuse reflection model.

Assuming an input image has $m$ pixels totally, therefore, there are $m$ input variables in the symmetric neural network. The 2D image is rearranged to a $m \times 1$ column vector denoted as $\mathrm{I} = (I_1, I_2, ..., I_m)^T$ and fed into the symmetric neural network. Through the symmetric neural network, the reflectance map for diffuse reflection $(R_{d_1}, R_{d_2}, ..., R_{d_m})$ can be obtained in the output of the symmetric neural network.

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 serves to combine information, activation, or evidence from other nodes. The function provides the net input for this node,

$$net\text{-}input = f(\text{inputs to this node; associated link weights}). \quad (3\text{-}1)$$

A second action of each node is to output an activation value as a function of its net-input,

$$node\text{-}output = a^{(l)}(\text{net-input}) = a^{(l)}(f), \tag{3-2}$$

where $a^{(l)}(\cdot)$ denotes the activation function and the superscript $l$ indicates the layer number. There are 6 layers in the proposed symmetric neural network and the functions of the nodes in each layer are described in the following.

**Layer 1**: This layer collects the intensity values of the input images as the inputs of the network. Node $I_i$ represents the $i$th pixel of the 2D image and $m$ is the number of total pixels of the image. That is

$$\begin{aligned} f_i &= I_i, \qquad i = 1, ..., m, \\ a_i^{(1)} &= f_i, \qquad i = 1, ..., m. \end{aligned} \tag{3-3}$$

This notation will also be used in the following equation.

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

$$\begin{aligned} f_i &= \frac{I_i}{\alpha_i}, \qquad i = 1, ..., m, \\ \hat{I}_i &= a_i^{(2)} = f_i, \qquad i = 1, ..., m. \end{aligned} \tag{3-4}$$

The output of this layer is the adjusted intensity value of the original 2D image and we label the nodes of this layer as $\hat{I}_1, \hat{I}_2, ..., \hat{I}_m$. The $\alpha_i$ is the $i$th albedo value corresponding to the $i$th pixel of the 2D image and $1/\alpha_i$ is the weight between $I_i$ and $\hat{I}_i$.

**Layer 3**: The function of layer 3 is to separate the light source direction from the

2D image. The light source direction of this layer is un-normalized, and we label them as $s_1'$, $s_2'$, and $s_3'$. The link weight in layer 3 is denoted as $w_{d_{ij}}$ for the connection between node *i* of layer 2 and node *j* of layer 3.

$$\begin{aligned} f_j &= \sum_{i=1}^{m} \hat{I}_i w_{d_{ij}}, \qquad i = 1, ..., m, \;\; j = 1, 2, 3, \\ s_j' &= a_j^{(3)} = f_j, \qquad j = 1, 2, 3. \end{aligned} \tag{3-5}$$

**Layer 4**: The nodes of this layer represent the unit light source and we normalize the un-normalized light source direction obtained in layer 3 by Eq. (3-6). 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 by:

$$\begin{aligned} f_j &= \frac{1}{\sqrt{s_1'^2 + s_2'^2 + s_3'^2}}, j = 1, 2, 3, \\ s_j &= a_j^{(4)} = f_j \cdot s_j' = \frac{s_j'}{\sqrt{s_1'^2 + s_2'^2 + s_3'^2}}, \quad j = 1, 2, 3. \end{aligned} \tag{3-6}$$

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

$$\begin{aligned} f_k &= \sum_{j=1}^{3} s_j v_{d_{jk}}, \qquad k = 1, ..., m, \\ \hat{R}_{d_k} &= a_k^{(5)} = f_k, \qquad k = 1, ..., m. \end{aligned} \tag{3-7}$$

It is noted that $\hat{R}_{d_k}$ represents the un-normalized reflectance map of diffuse

reflection; therefore, we will normalize these values in layer 6.

**Layer 6**: The function of this layer is to transfer the un-normalized reflectance map of diffuse reflection obtained in layer 5 into the interval [0, 255]. These nodes, $R_{d_1}, R_{d_2}, ..., R_{d_m}$, represent the normalized reflectance map of diffuse reflection, and the output of these nodes can be calculated by:

$$\begin{aligned} f_k &= \hat{R}_{d_k}, \qquad k = 1, ..., m, \\ R_{d_k} &= a_k^{(6)} = \frac{255\left(f_k - \min\left(\hat{\mathbf{R}}_d\right)\right)}{\max\left(\hat{\mathbf{R}}_d\right) - \min\left(\hat{\mathbf{R}}_d\right)} = \frac{255\left(\hat{R}_{d_k} - \min\left(\hat{\mathbf{R}}_d\right)\right)}{\max\left(\hat{\mathbf{R}}_d\right) - \min\left(\hat{\mathbf{R}}_d\right)}, \quad k = 1, ..., m, \end{aligned} \tag{3-8}$$

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

Similar to the diffuse reflection model, we also use a symmetric neural network as Fig. 3-2 to simulate the specular component in our hybrid-reflectance model. The major differences between these two networks are the representation of nodes in layer 3 and layer 4, and the active function of layer 5. The nodes of layer 3 represent the un-normalized half-way vector, labeled as $\mathbf{h}' = (h_1', h_2', h_3')^T$, and the nodes of layer 4 represent the normalized half-way vector labeled as $\mathbf{h} = (h_1, h_2, h_3)^T$. According to Eq. (2-5), layer 5 of the symmetric neural network for specular component combines the half-way vector **h** and normal vectors of the surface to generate the reflectance map of specular reflection. Let the link weight connects node *j* of layer 4 and node *k* of layer 5 be denoted as $v_{s_{jk}}$ and it 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:

$$\begin{aligned} f_k &= \sum_{j=1}^{3} h_j v_{s_{jk}}, \qquad k = 1, ..., m, \\ \hat{R}_{s_k} &= a_k^{(5)} = \left(f_k\right)^r, \qquad k = 1, ..., m, \end{aligned} \tag{3-9}$$

where the active function $a_k^{(5)}$ in this layer is the $r$ degree of the net-input of this layer. It is noted that $\hat{R}_{S_k}$ represents the un-normalized reflectance map of specular reflection and it will also be normalize in layer 6.

Through the supervised learning algorithm derived in the following section, we can get the normal vectors of the surface automatically and then we can use the enforcing integrability approach [51] to obtain the depth information for reconstructing the 3D surface of an object by the obtained normal vectors. In our approach, the reflectance characteristic of the hybrid surfaces can be determined without a priori information of the relative strengths of the diffuse and specular components. This is an important improvement of the conventional algorithms. The hybrid intensity of each point on the surfaces is considered individually such that we can reduce the distortion that the conventional methods met in the recovery process. In addition, by the symmetric neural network for diffuse reflection, we can get the light source direction $\mathbf{s}$ in the hidden nodes of the symmetric neural network and we can solve the photometric stereo problem without specifying illuminant positions in advance. This also relaxes the constraint in the conventional approaches and is more suitable for practical applications to 3D surface reconstruction.

## 3.3. Training Algorithm of the Proposed Model

The back-propagation learning method is used for the supervised training of the proposed model and the goal is to minimize the error function defined as

$$E_T = \sum_{i=1}^{m} \left( R_{hybrid_i} - D_i \right)^2, \tag{3-10}$$

where $m$ is the number of total pixels of the 2D image, $R_{hybrid_i}$ is the $i$th output of the neural network, and $D_i$ is the $i$th desired output that is equal to the $i$th intensity of

the original 2D image. For each 2D image, starting at the input nodes, a forward pass is used to compute 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 compute $\partial E_T / \partial \omega$, where $\omega$ represents the adjustable parameters in the network. The general parameter update rule is

$$\omega(t+1) = \omega(t) + \Delta\omega(t) = \omega(t) + \eta\left(-\frac{\partial E_T}{\partial \omega(t)}\right), \tag{3-11}$$

where $\eta$ is the learning rate.

In the following, we will show the details of the learning rules corresponding to each adjustable parameter.

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

$$\begin{aligned}\lambda_{d_k}(t+1) &= \lambda_{d_k}(t) + \Delta\lambda_{d_k}(t) \\ &= \lambda_{d_k}(t) + 2\eta\left(D_k(t) - R_{hybrid_k}(t)\right)R_{d_k}(t),\ k = 1, \ldots, m,\end{aligned} \tag{3-12}$$

$$\begin{aligned}\lambda_{s_k}(t+1) &= \lambda_{s_k}(t) + \Delta\lambda_{s_k}(t) \\ &= \lambda_{s_k}(t) + 2\eta\left(D_k(t) - R_{hybrid_k}(t)\right)R_{s_k}(t),\ k = 1, \ldots, m,\end{aligned} \tag{3-13}$$

where $D_k(t)$ is the $k$th desired output, $R_{hybrid_k}(t)$ is the $k$th system output, $R_{d_k}(t)$ is the $k$th diffuse intensity obtained from the up sub-network, $R_{s_k}(t)$ is the $k$th specular intensity obtained from the low sub-network (as shown in Fig. 3-1), $m$ is the number of total pixel of an 2D image, and $\eta$ is the learning rate of the neural network.

For a gray image, the intensity value of a pixel is in the interval of [0, 255]. If we want to avoid the intensity value of $R_{hybrid_k}$ exceeding the interval [0, 255], we must force

$$\lambda_d + \lambda_s = 1, \tag{3-14}$$

where $\lambda_d \geq 0$ and $\lambda_s \geq 0$. Therefore, we normalize the combination ratio $\lambda_{d_k}$ and $\lambda_{s_k}$ by

$$\begin{aligned} \lambda_{d_k}(t+1) &= \frac{\lambda_{d_k}(t+1)}{\lambda_{d_k}(t+1)+\lambda_{s_k}(t+1)}, \quad k=1,...,m, \\ \lambda_{s_k}(t+1) &= \frac{\lambda_{s_k}(t+1)}{\lambda_{d_k}(t+1)+\lambda_{s_k}(t+1)}, \quad k=1,...,m. \end{aligned} \tag{3-15}$$

**Sub-networks**: The normal vector calculated from the sub-network corresponding to the diffuse part is represented as $\mathbf{n}_{d_k} = (v_{d_{1k}}, v_{d_{2k}}, v_{d_{3k}})$ for *k*th point on the surface; the normal vector calculated from the sub-network corresponding to the specular part is represented as $\mathbf{n}_{s_k} = (v_{s_{1k}}, v_{s_{2k}}, v_{s_{3k}})$ for *k*th point. The normal vectors $\mathbf{n}_{d_k}$ and $\mathbf{n}_{s_k}$ are updated iteratively by the gradient method as:

$$\begin{aligned} v_{d_{jk}}(t+1) &= v_{d_{jk}}(t) + \Delta v_{d_{jk}} \\ &= v_{d_{jk}}(t) + 2\eta\, s_j(t)\left(D_k(t) - R_{hybrid_k}(t)\right), \quad j=1,2,3, \\ v_{s_{jk}}(t+1) &= v_{s_{jk}}(t) + \Delta v_{s_{jk}} \\ &= v_{s_{jk}}(t) + 2\eta\, r\, h_j(t)\left(D_k(t) - R_{hybrid_k}(t)\right), \quad j=1,2,3, \end{aligned} \tag{3-16}$$

where $s_j(t)$ is the *j*th element of illuminant direction **s**, $h_j(t)$ is the *j*th element of the halfway-vector **h**, *r* is the degree of the specular equation shown in Eq. (3-9). The updated $v_{d_{j,k}}$ and $v_{s_{j,k}}$ should be normalized as follows:

$$v_{d_{jk}}(t+1) = \frac{v_{d_{jk}}(t+1)}{\left\|\mathbf{n}_{d_k}(t+1)\right\|}, \quad v_{s_{jk}}(t+1) = \frac{v_{s_{jk}}(t+1)}{\left\|\mathbf{n}_{s_k}(t+1)\right\|}, \quad j=1,2,3. \tag{3-17}$$

In order to get the reasonable normal vectors of the surface from our adaptive hybrid-reflectance model, we compose $\mathbf{n}_{d_k}$ and $\mathbf{n}_{s_k}$ as the hybrid normal vector, $\mathbf{n}_k$, of the surface on the *k*th point by

$$\mathbf{n}_k(t+1) = \lambda_{d_k}(t)\,\mathbf{n}_{d_k}(t) + \lambda_{s_k}(t)\,\mathbf{n}_{s_k}(t), \tag{3-18}$$

where $\lambda_d(t)$ and $\lambda_s(t)$ are 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 and thus the update rule for the weights between layer 2 and layer 3 of the two sub-networks denoted as $\mathbf{W}_d$ and $\mathbf{W}_s$ (see Fig. 3-2) can be calculated by least square method. Hence, $\mathbf{W}_d$ and $\mathbf{W}_s$ at time $t+1$ can be obtained by

$$\mathbf{W}_d(t+1) = \left(\mathbf{V}_d(t+1)^T \mathbf{V}_d(t+1)\right)^{-1} \mathbf{V}_d(t+1)^T, \tag{3-19}$$

$$\mathbf{W}_s(t+1) = \left(\mathbf{V}_s(t+1)^T \mathbf{V}_s(t+1)\right)^{-1} \mathbf{V}_s(t+1)^T, \tag{3-20}$$

where $\mathbf{V}_d(t+1)$ and $\mathbf{V}_s(t+1)$ are weights betweens the output layer and the center layer of the two sub-networks for diffuse and specular components, respectively.

In addition, for fast convergence, the learning rate $\eta$ of the neural network is adaptive in the updating process. If the current error is smaller than the errors of the previous two iterations, it indicates that the current direction of adjustment is correct. Thus we should maintain the current direction and increase the step size to speed up convergence. On the contrary, if the current error is larger than the errors of the previous two iterations, we must decrease the step size because the current adjustment is wrong. Otherwise, learning rate $\eta$ will not change. In this manner, the cost function $E_T$ could reach minimum quickly and avoid oscillation around the local minimum. The adjustment rule of the learning rate is shown as follows:

If (Err(*t-1*) > Err(*t*) and Err(*t*-2)> Err(*t*))

$\eta(t+1) = \eta(t) + \xi$,

Else If (Err(*t-1*) < Err(*t*) and Err(*t-2*) < Err(*t*))

$\eta(t+1) = \eta(t) - \xi$, where $\xi$ is a given scalar.

Else $\eta(t+1) = \eta(t)$.

We also used the prior knowledge as the initial values of the proposed neural network for specific object classes to improve the results of 3D 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.

## 3.4. Experimental Results and Discussions

In this section, four experiments are performed to demonstrate the proposed method. In these experiments, both direction of the light and the observer's viewing direction are unknown. It is used to test whether our algorithm can reconstruct the objects well even if we don't have the information about lighting and viewing directions in advance. In the first experiment, images of the synthetic objects are used for testing. The estimated depth map is compared with the true depth map to examine the performance of reconstruction. In the second experiment, several images corresponding to real surfaces of human faces are used for testing. These images are downloaded from the Yale Face Database B [55]. They are 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 are used to show the generality of our method.

### 3.4.1. Experiment on Images of Synthetic Objects

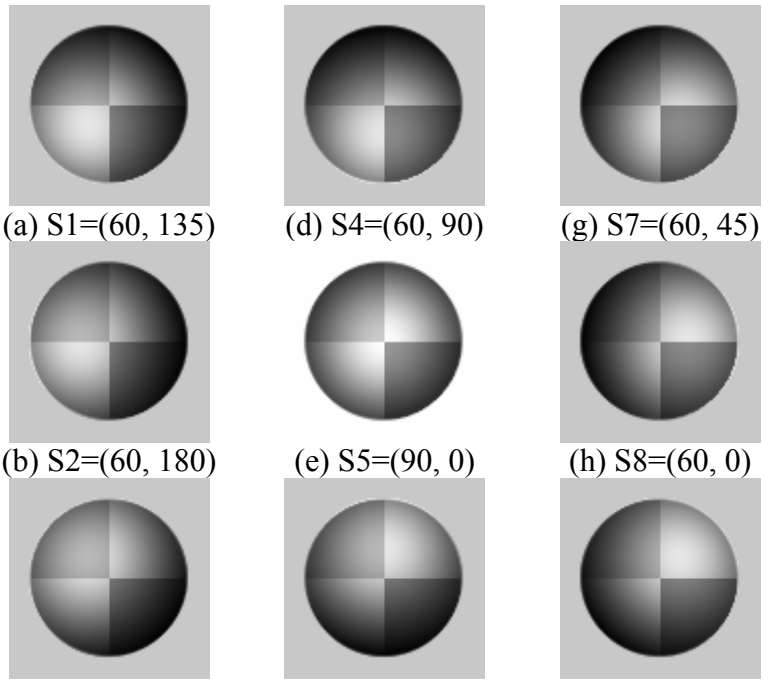
In this experiment, quantitative results of synthetic-object reconstruction are presented. The results of the proposed method are compared with three existing methods including the diffuse model [56], the specular model [57], and the hybrid model [50]. Three synthetic objects, sphere, sombrero, and vase mathematically generated by Eqs. (3-21), (3-22), and (3-23), respectively, are used for testing.

$$z(x,y)=\begin{cases}\left|\sqrt{r^2-x^2-y^2}\right|, & \text{if } x^2+y^2\le r^2\\ 0, & \text{otherwise}\end{cases}, \tag{3-21}$$

$$z(x, y) = 15 + 15\cos(\frac{\pi\sqrt{x^2 + y^2}}{17}), \tag{3-22}$$

$$z(x, y) = \sqrt{f(y)^2 - x^2}\ . \tag{3-23}$$

In Eq. (3-21), $r = 45$, $0 \le x, y \le 100$, and the center is located at $(x, y) = (50, 50)$. In Eq. (3-23), $f(y) = 0.6 - 0.3y(6y+1)^2(y-1)^2(3y-2)$. The shaded images of the sphere are synthesized with variant albedo and different directions as shown in Fig 3-3. The different albedos are, 0.6 for right-bottom of the sphere, 0.8 for left-top of the sphere, and 1 for the rest part. The locations of light sources in Figs. 3-3(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. The experimental results are shown in Fig. 3-4 and Table 3-1. In Table 3-1, we take 5 groups of images with different illuminant angles from the left, the right, and the front for 3D reconstruction. Both estimated surface and synthetic one are normalized within the interval [0, 1]. According to the first row (sphere object) of Fig. 3-4, the surface with variant albedo is hard to handle by the conventional methods and our proposed method performs superior reconstruction result. According to Table 3-1, it is found that the proposed method can achieve the lowest mean errors compared with the other methods in all illumination conditions.



(a) S1=(60, 135) (d) S4=(60, 90) (g) S7=(60, 45)

(b) S2=(60, 180) (e) S5=(90, 0) (h) S8=(60, 0)

(c) S3=(60, -135) (f) S6=(60, -90) (i) S9=(60, -45)

Figure 3-3. The 2D sphere images generated with varying albedo and different lighting directions (the degree of tilt angle, the degree of pan angle).

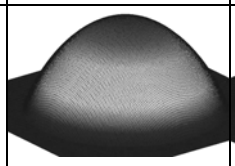
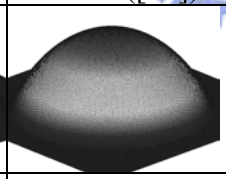
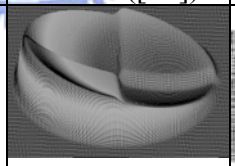
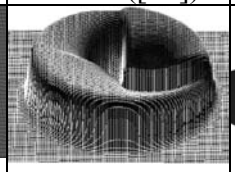
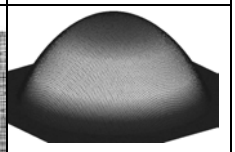
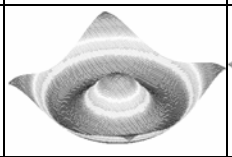
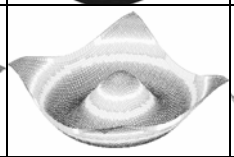
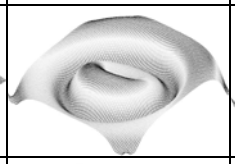
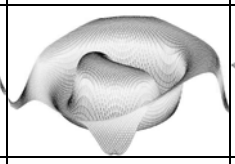
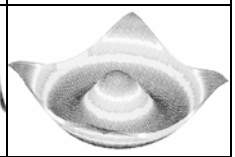
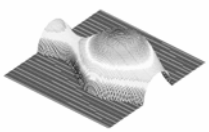
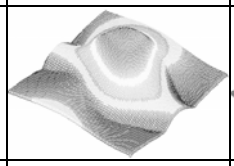
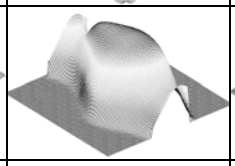
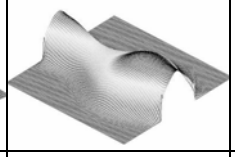
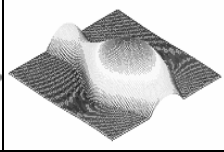
| 3D Object | The original image | The diffuse reflectance model ([56]) | The specular reflectance model ([57]) | The hybrid reflectance model ([50]) | Proposed reflectance model |
|---|---|---|---|---|---|
| Sphere with constant albedo |  |  |  |  |  |
| Sombrero |  |  |  |  |  |
| Vase |  |  |  |  |  |
| | (a) | (b) | (c) | (d) | (e) |

Figure 3-4. Comparisons of synthetic images and the recovered surfaces of sphere, sombrero and vase. (a) The depth map of the objects. (b) The recovered result by the diffuse reflectance model [56]. (c) The recovered result by the specular reflectance model [57]. (d) The recovered result by the hybrid reflectance model [50]. (e) The recovered result by the proposed method.

Table 3-1. The absolute mean errors between estimated depths and desired depths of synthetic object's 3D surfaces. (Both light and viewing directions are unknown in the experiment.) (Iterations = 10)

| Mean absolute depth error and CPU time | Lights | The diffuse reflectance model ([56]) | The specular reflectance model ([57]) | The hybrid reflectance model ([50]) | 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 |

For the sombrero object, the results of the diffuse reflectance model and our method are very similar to the original shape, and the results of the specular reflectance model and the hybrid reflectance model are not very well. The shape of sombrero object is very sharp and their images have many shadows (both cast and attached shadows). It is very likely to cause distortion by Cho's methods (the specular reflectance model [57] and the hybrid reflectance model [50]) because they use single image to recover the shape. The image may have invalid pixel values because of saturated pixels in shadows; therefore, the information of single image may be not enough.

For the vase object, our method reconstructs the synthetic vase very successfully, but the result of the diffuse reflectance model has obviously distortion. This may be

caused by the convex of the vase. When the vase is illuminated, the convex of the vase will be shiny and the Lambertian assumption could not approximate it well.

In addition, the CPU time used by each method for shape reconstructions is also included in Table 3-1. We implemented each method in Matlab 6.1 software on a 1.2GHz Pentium III-based PC with 256 MB RAM. According to the results, the CPU time used by the proposed method is close to that used by the Lambertian method [56] and it is greatly reduced compared with the specular reflectance method [57] and the hybrid reflectance method [50].

The data set in the University of Notre Dame Biometrics Database [58] is also used in our experiments for objective comparison. The database consists of 3D face coordinate data and their corresponding 2D front view. Fig. 2-5 shows 6 individuals in the database with 160*160 image size. Unlike the synthetic objects, both the estimated surface and 3D face surface are not normalized within the interval [0, 1]. In order to evaluate the performance of our approach and other existing methods, we define a function that use the $x$- and $y$- partial derivatives of $z(x, y)$ instead of absolute mean error to compute the error between the estimated depths and desired depths of 3D face surfaces as:

$$Err = \frac{\sum_{y=1}^{M}\sum_{x=1}^{N}\sqrt{\left(\partial z(x,y)/\partial x - \partial \hat{z}(x,y)/\partial x\right)^2 + \left(\partial z(x,y)/\partial y - \partial \hat{z}(x,y)/\partial y\right)^2}}{M \cdot N}, \qquad (3\text{-}24)$$

where $\partial z(x,y)/\partial x$ and $\partial \hat{z}(x,y)/\partial x$ are the $x$-partial derivatives of $z(x,y)$ and $\hat{z}(x,y)$, respectively, $\partial z(x,y)/\partial y$ and $\partial \hat{z}(x,y)/\partial y$ are the $y$-partial derivatives of $z(x, y)$ and $\hat{z}(x,y)$, respectively, and $M$ and $N$ are $x$ and $y$ dimensions. Table 3-2 shows the mean errors between the desired depths of 3D face surfaces and the estimated depths by using different methods. According to the experiment on the

images of the three synthetic objects and the dataset in the University of Notre Dame Biometrics Database, our approach can be applied to more general objects and can achieve better performance than the existing methods.

In order to evaluate the impact of adaptive hybrid ratio on the performance of the proposed network, the performance of the proposed neural network with constant hybrid ratio is also provided in the experiments. According to Table 3-2, it is obvious that the mean errors of the proposed network with constant hybrid ratio are less than those of other approaches and can be further reduced if the hybrid ratio is adjusted in the learning process.

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

Figure 3-5. Six individuals in the Notre Dame Biometrics Database D [58] used to test our algorithm (these images include both males and females.).

Table 3-2. The mean errors between the estimated depths and desired depths of 3D face surfaces shown in Fig. 3-5. (Iterations = 20)

| Mean error (Eq. 3-24) | The diffuse reflectance model ([56]) | The specular reflectance model ([57]) | The hybrid reflectance model ([50]) | The proposed reflectance model with const of hybrid ratio | The proposed reflectance model |
|---|---|---|---|---|---|
| Fig. 3-5(a) | 1.5524 | 1.5268 | 1.5259 | 1.1586 | 1.0992 |
| Fig. 3-5(b) | 1.3517 | 1.3550 | 1.3533 | 1.0358 | 0.9644 |
| Fig. 3-5(c) | 1.9476 | 1.9542 | 1.9546 | 1.5963 | 1.5348 |
| Fig. 3-5(d) | 1.6007 | 1.6018 | 1.6039 | 1.2913 | 1.2243 |
| Fig. 3-5(e) | 1.2396 | 1.2457 | 1.2443 | 0.9576 | 0.9060 |
| Fig. 3-5(f) | 1.9240 | 1.9155 | 1.9160 | 1.5613 | 1.5224 |
| Average | 1.6027 | 1.5998 | 1.5997 | 1.2668 | 1.2085 |

### 3.4.2. Experiment on Yale Face Database B

In this experiment, face images downloaded from the Yale Face Database B [55] are used for testing. For each person, we take three images in which their pose is fixed and they are illuminated by three different light directions. After processing by our algorithm, we can get the diffuse intensity, the specular intensity, and the hybrid intensity. Besides, we can also get the surface normal vectors and use them to reconstruct the surface. Fig. 3-5 shows an example that uses the proposed method to estimate different reflection components and normal vectors of a human face in the Yale Face Database B [55]. Fig. 3-6 shows the comparison between three existing approaches and the proposed method for human face reconstruction. Figs. 3-6(b), (c), and (d) are the reconstructed results of the diffuse reflectance model [56], the specular reflectance model [57], the hybrid reflectance model [50], respectively. The result of our method is shown in Fig. 3-6(e). The results clearly indicate that the performance of our proposed algorithm is better than that of these three approaches. From the comparison between Fig. 3-6(b) and Fig. 3-6(e), the reconstructed shape by our method is sharper and more apparent, especially on the part of nose. The reconstructed results of the specular reflectance model by [57] and the hybrid reflectance model by [50] as shown in Figs. 3-6(c) and (d) have serious distortions.

From the above reconstructed results, the reconstructed performance of the specular model [57] and the hybrid reflectance model [50] are obviously not good for human faces. Therefore, we will only compare the diffuse reflectance model [56] and the proposed method in the following experiments. Fig. 3-7 shows the comparison of the reconstructed results between the diffuse reflectance model [56] and the proposed method. According to the experimental results, the reconstructed results of the two methods are similar. Finally, more reconstructed results of human faces in Yale Face

Database B by using the proposed method are shown in Fig. 3-8. In order to compare them easily, the angles of the faces are set to be equal. The reconstructed results demonstrate that our algorithm performs well on different human faces in Yale Face Database B.

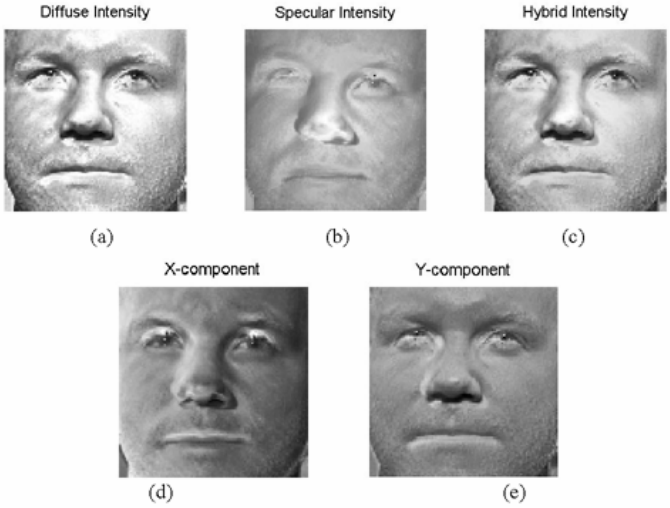



Figure 3-6. Estimated reflection components and normal vectors of a human face in the Yale Face Database B [55] by the proposed method. (a) The diffuse intensity. (b) The specular intensity. (c) The hybrid intensity. These three images are the estimated results of the proposed method. (d) The X-component of the normal vector. (e) The Y-component of the normal vector.

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

Figure 3-7. Reconstructed surfaces of the proposed algorithm compared with three existing approaches. (a) The original 2D facial image. (b) The recovered result by the diffuse reflectance model [56]. (c) The recovered result by the specular reflectance model by [57]. (d) The recovered result by the hybrid reflectance model [50]. (e) The recovered result by the proposed method.

(a) (b) (c)

Figure 3-8. Reconstructed results of two different methods. (a) The original 2D image of the human face. (b) Reconstructed result of the diffuse reflectance model [56]. (c) Reconstructed result of the proposed method.

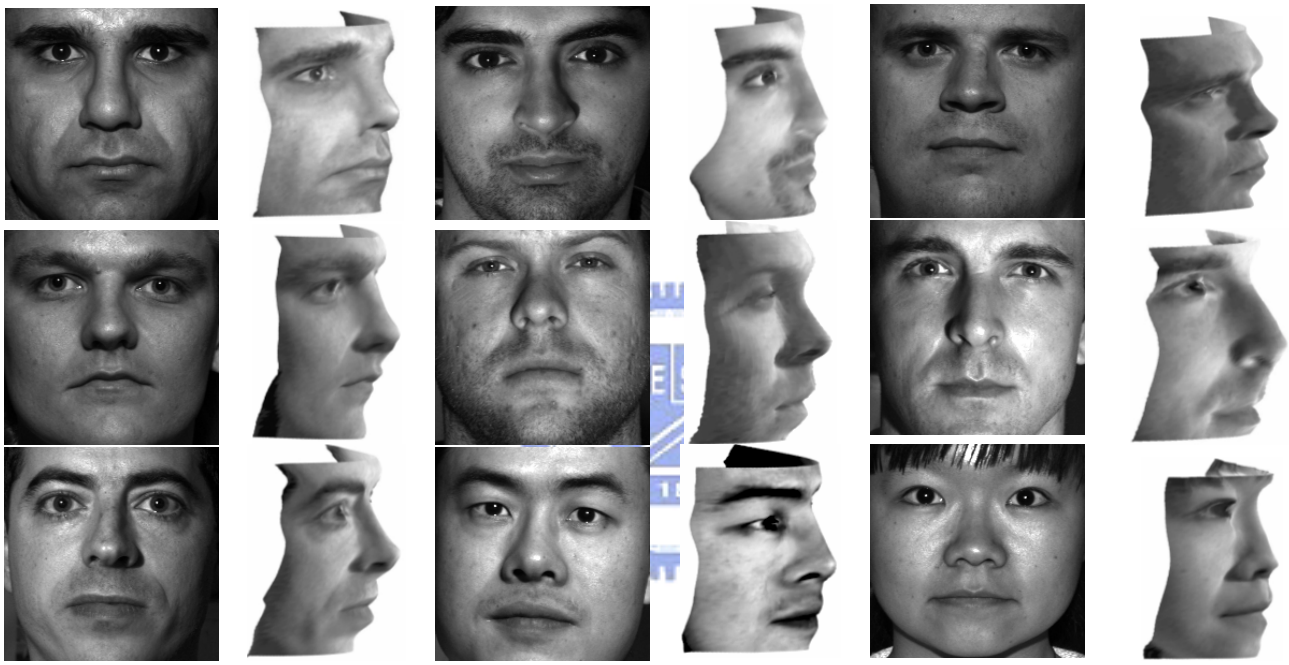


Figure 3-9. More reconstructed results of the human faces in Yale Face Database B by using the proposed method.

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

In order to test and verify the performance of our proposed algorithm on the facial images of our laboratory members and other images of general objects, we design and construct a photographing environment as shown in Fig. 3-10. To make the strength of different light sources to photographed objects equal, the photographing environment is constructed as a hemisphere. The radius of the hemisphere is 2 meters and there are eight computer-controlled electronic flashes (Mikona MV-328) placed

on the hemisphere and the positions of these eight light sources are $\mathbf{s}_1$ = (53, 30), $\mathbf{s}_2$ = (30, 63), $\mathbf{s}_3$ = (20, 30), $\mathbf{s}_4$ = (15, 45), $\mathbf{s}_5$ = (-20, 30), $\mathbf{s}_6$ = (-30, 63), $\mathbf{s}_7$ = (-53, 60), and $\mathbf{s}_8$ = (-53, 30), respectively. The representation of light position is (degree of pan angle, degree of tilt angle) where the center of the hemisphere is as the origin of the coordinate. The captured images of a bear pottery that are illuminated by the 8 light sources are shown in Fig. 3-10. The images of Figs. 3-10(a)-(d) are illuminated by the light sources at the right-hand side of the object; Figs. 3-10(e)-(h) are illuminated by the light sources at the left-hand side of the object, and their angles and positions are a little different.

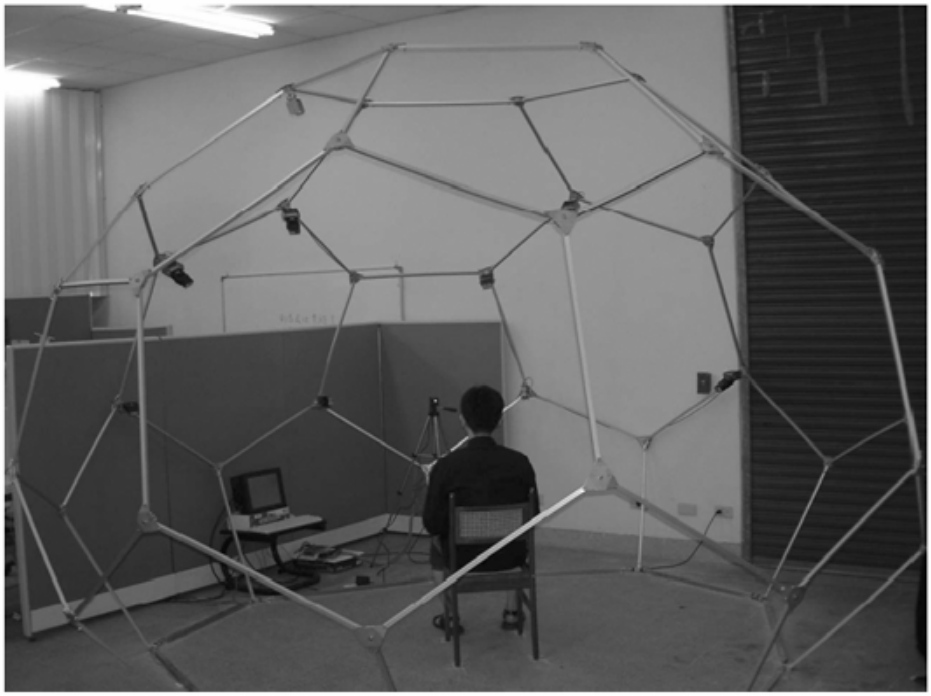
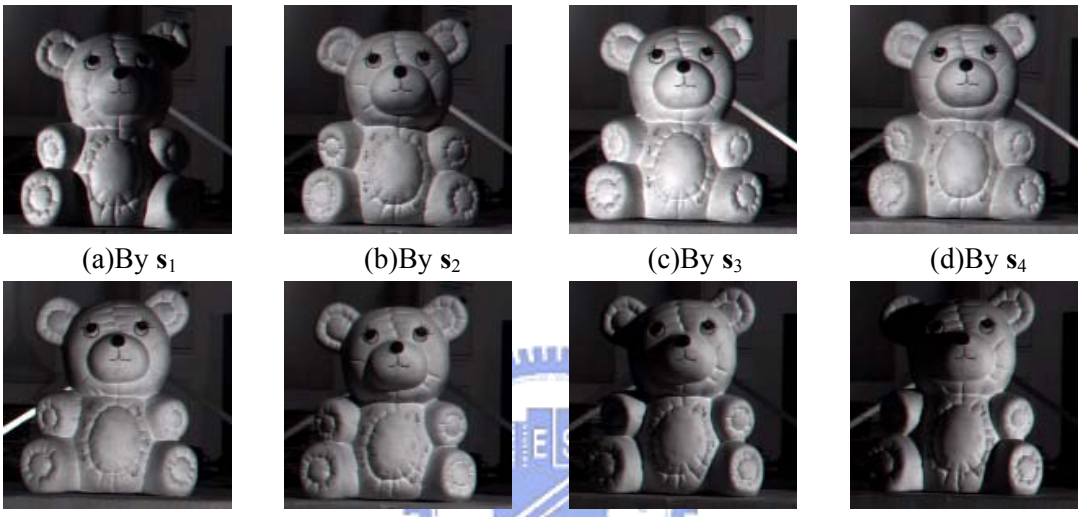


Figure 3-10. The photographing environment with eight electronic flashes used to capture images under variations in illumination set up in our lab.

In order to understand the influence of illuminant positions and angles on the performance of the proposed method, we test our approach on three groups of images: left-hand-side illuminated images (by **s**1, **s**2, **s**3), right-hand-side illuminated images (by **s**6, **s**7, **s**8), and front illuminated images (by **s**3, **s**4, **s**5). The illuminant direction is

determined based on the photographed objects' viewpoint. Fig. 3-12 shows the reconstructed results that are calculated from the images corresponding to these three groups, respectively. The reconstruction in Fig. 3-12(a) is calculated from the images of Figs. 3-11(a), (b), and (c); the reconstruction in Fig. 3-12(b) is calculated from the images of Figs. 3-11(f), (g), and (h); and the reconstruction in Fig. 3-12(c) is calculated from the images of Figs. 3-11(c), (d), and (e).



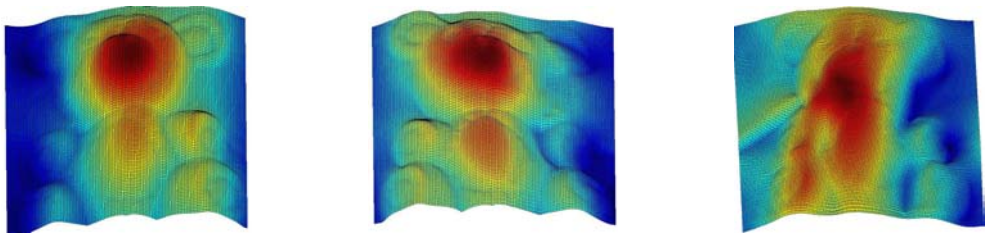
(a)By $\mathbf{s}_1$ (b)By $\mathbf{s}_2$ (c)By $\mathbf{s}_3$ (d)By $\mathbf{s}_4$

(e)By $\mathbf{s}_5$ (f)By $\mathbf{s}_6$ (g)By $\mathbf{s}_7$ (h)By $\mathbf{s}_8$

Figure 3-11. (a)~(h) The images of an object illuminated by the eight different light sources, respectively. They can be separated to three groups: left-hand-side illuminated images (by **s**1, **s**2, **s**3), right-hand-side illuminated images (by **s**6, **s**7, **s**8), and front illuminated images (by **s**3, **s**4, **s**5).

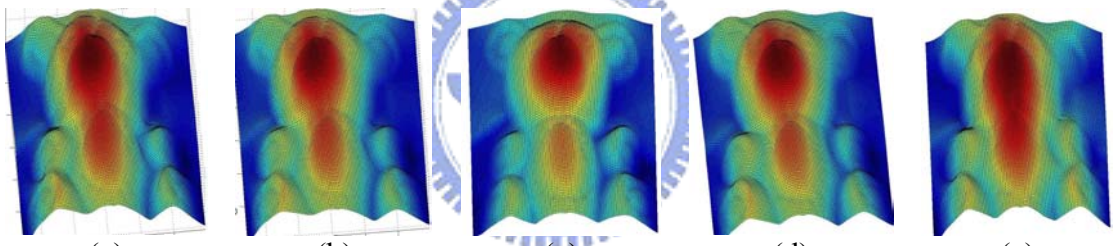
For the reconstructed result in Fig. 3-12(a), the object is illuminated from the right-hand-side, so its variant intensities are more obvious on the left part of the object's surface. Therefore, the left part of the reconstructed object is better than its right part. Similarly, in Fig. 2-12(b), the object is illuminated from the left-hand-side and the reconstructed result is better in the right part of the object. Besides, Fig. 2-12(c) is calculated from the images in Figs. 2-11(c), (d), and (e). These images are illuminated from the front and they are very similar, and the variant intensities in images are not obvious. Thus the provided information is not enough to reconstruct the 3D surface well. If the input images are too similar, it is hard to solve the least

square problem for the unique solution of the irradiance equation. The experimental results indicate that the information of only one-side-illuminated images is not enough for 3D reconstruction. Therefore, to perform better recovery, we should take more different illuminated conditions to obtain more information of the object's surface.



(a) (b) (c)

Figure 3-12. The reconstructed results of the object that are calculated from images in Fig. 3-11. (a) From the images illuminated by $\mathbf{s}_1$, $\mathbf{s}_2$, and $\mathbf{s}_3$. (b) From the images illuminated by $\mathbf{s}_6$, $\mathbf{s}_7$, and $\mathbf{s}_8$. (c) From the images illuminated by $\mathbf{s}_3$, $\mathbf{s}_4$, and $\mathbf{s}_5$.



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

Figure 3-13. The better reconstructed results of the object from the images in Fig. 3-11. (a) From the images illuminated by $\mathbf{s}_2$, $\mathbf{s}_4$, and $\mathbf{s}_7$. (b) From the images illuminated by $\mathbf{s}_2$, $\mathbf{s}_3$, and $\mathbf{s}_7$. (c) From the images illuminated by $\mathbf{s}_2$, $\mathbf{s}_5$, and $\mathbf{s}_6$. (d) From the images illuminated by $\mathbf{s}_2$, $\mathbf{s}_3$, and $\mathbf{s}_8$. (e) From the images illuminated by $\mathbf{s}_1$, $\mathbf{s}_3$, and $\mathbf{s}_8$.

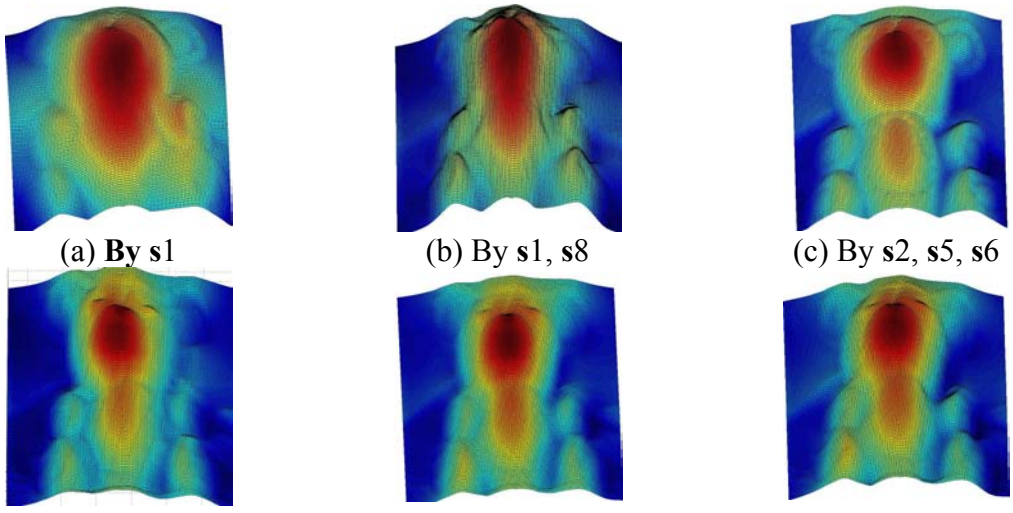
Figure 3-13 shows some better reconstructed results of the objects and they are calculated by different combinations of the left-hand-side illuminated, right-hand-side illuminated, and front illuminated images. Therefore, in order to perform better reconstructed results, we should take images with different illuminant angles from the left, the right, and the front to provide sufficient information of the surface for 3D reconstruction.

In order to check the performance of the proposed reconstruction method under

different numbers of light sources, the experimental results by using 1, 2, 3, 4, 7, and 8 images (light sources) are shown in Fig. 3-14. Obviously, the results in Figs. 3-14(a) and (b) that use 1 and 2 images are not good, and many features of the bear are not recovered. On the contrary, the results in Figs. 3-14(c)-(f) that use more than 2 images are better than the results in Figs. 3-14(a) and (b). The key features of the bear are obvious and the details of the shape are not lost. According to the experimental results shown in Fig 3-14, it is concluded that we should use three images at least for fine reconstruction. However, the reconstructed results by using more than three images are not necessarily better than the reconstructed result by using three images. This experimental result is consistent with the theoretical basis of the proposed scheme, in which three sets of variables are to be determined, so two images only make it a under-determined problem, and more than three images make it a over-determined problem. Therefore, we use three 2D images to reconstruct the surface of a 3D object in the proposed method and the unnecessary calculation could also be avoided.

As a result, we know the influence of the illuminant angles and positions is very important. Because our approach is based on the shape from shading method, the reconstructed information from 2D images will determine the performance of results. From the experimental results, it is concluded that the illuminant positions and angles should not be too close.

Fig. 3-15 shows the reconstructed results of human faces of our laboratory members by the diffuse reflectance model [56] and by our method. For the purpose of comparison, all angles of the faces are set to be similar. The reconstructed faces by our method are sharper and the face features are more apparent. For example, the noses of the reconstructed results by our method are more conspicuous.



(a) **By s**1 (b) By **s**1, **s**8 (c) By **s**2, **s**5, **s**6

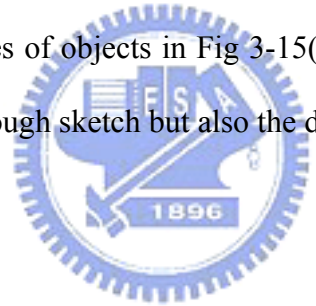
(d) By **s**1,**s**2,**s**3,**s**8 ( e ) By **s**1,**s**2,**s**3,**s**4,**s**5,**s**6,**s**8 ( f ) By **s**1,**s**2,**s**3,**s**4,**s**5, **s**6,**s**7,**s**8

Figure 3-14. Results of 3D object reconstruction using different numbers of images (light sources) in Fig. 3-11. (a) The reconstruction uses one 2D image. (b) The reconstruction uses two 2D images. (c) The reconstruction uses 3 2D images. (d) The reconstruction uses 4 2D images. (e) The reconstruction uses 7 2D images. (f) The reconstruction uses 8 2D images.


Figure 3-15. The reconstructed 3D facial surfaces from the 2D pictures of our laboratory members by the diffuse reflectance model [56] and our proposed method.

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

In that last experiment, images of a pottery bear, a dummy head, a toy figurine, a basketball, and an octagon iron box captured in our photographing environment are used for testing. The 2D images and the results of 3D reconstructions by using the diffuse reflectance model [56] and the proposed method are shown in Fig. 3-12. The reconstructed results by the proposed method seem well. In Fig. 3-12(d), the imprint of the English words of the basketball clearly appears on the reconstructed surface. In Fig. 3-12(e), the details of the box such as the ridge and the edge are also well reconstructed. However, the reconstructed results by the diffuse reflectance model [56] are not as well as the results of our method and the diffuse reflectance model even cannot reconstruct the surfaces of objects in Fig 3-15(e). It indicates that our method can reconstruct not only the rough sketch but also the details of the surface.



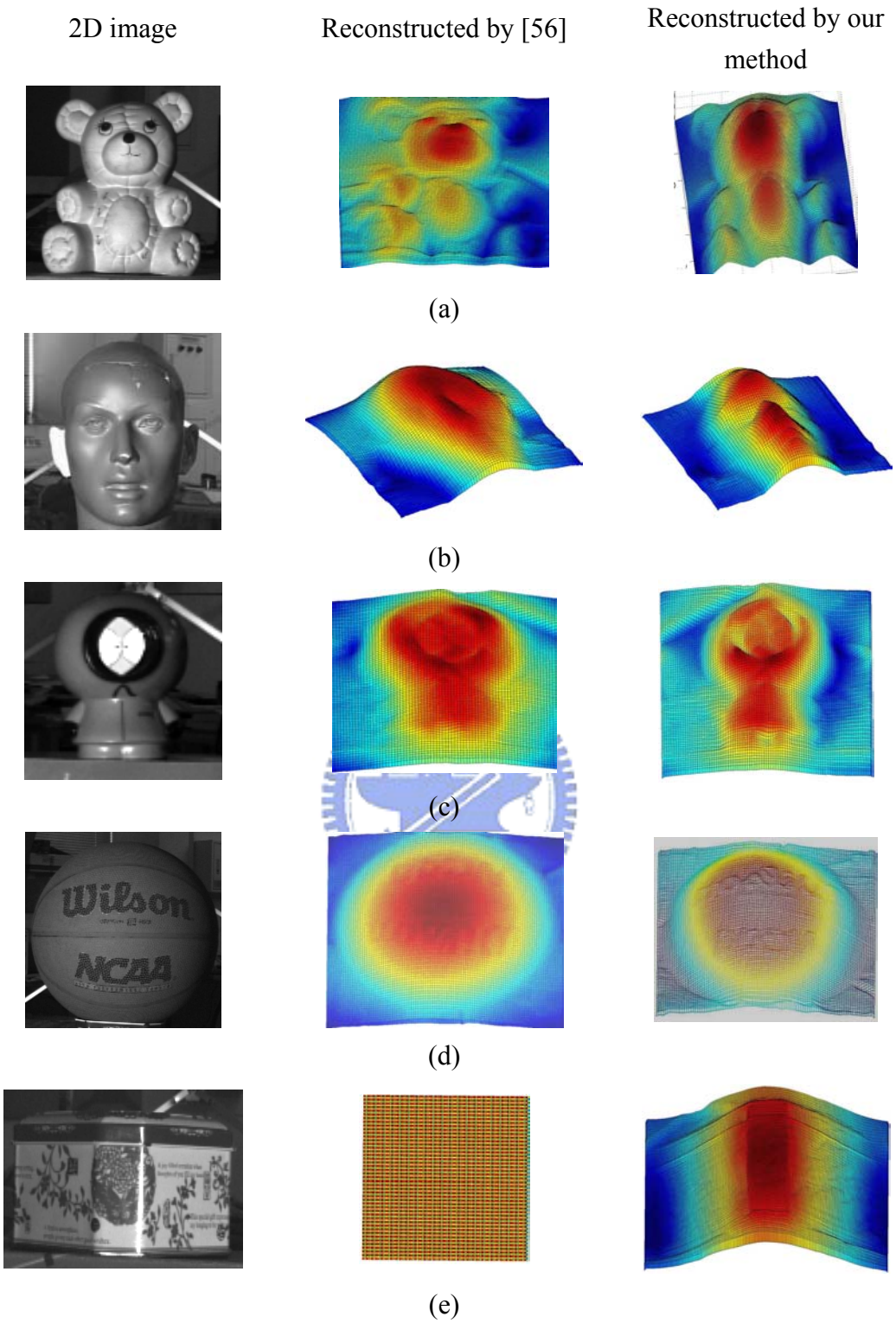



Figure 3-16. The reconstructed 3D surfaces of general objects by the diffuse reflectance model [56] and the proposed method. The left side of each raw is the 2D image of the object, the center part is the reconstructed surface by the diffuse reflectance model [56], and the right side is the reconstructed surface by the proposed method. (a) A pottery bear. (b) A dummy head. (c) A toy figurine. (d) A basketball. (e) An octagon iron box.

## 3.5. Concluding Remarks

In this chapter, a novel 3D image reconstruction method was proposed. This method considers both the diffuse and specular reflection components of the reflectance model simultaneously. We used two neural networks with symmetric structure to estimate these two reflection components separately and combine them with an adaptive ratio for each point on the object surface. We also tried to reduce the distortion due to variable albedo variation by adjusting the intensity value for each pixel by dividing the pixel's intensity by the corresponding rough-albedo value. Then these intensities 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 and so on, are also obtained from the learning process of the neural network. The obtained normal vectors of the surface can then be applied to 3D surface reconstruction by enforcing integrability approach. Extensive experimental results based on public image database and the images captured in the photographing environment built in our lab have demonstrated that the proposed technique can reconstruct the 3D surfaces of more general and real-world objects better as compared to several exiting approaches.

The contributions of this chapter can be summarized as follows. (1) We used the images caught under three different light sources to solve the photometric stereo problem without the information of exact light source locations. (2) The proposed method considers the changes of the albedo on the object surface, so we can obtain good reconstruction results not only for human faces but also for general objects with variant albedo. (3) The proposed symmetric neural network structure with adaptive learning procedure dose not need any special parameter setting and the smoothing conditions. It is also easier to converge and makes the system stable. (4) The proposed

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

# 4. A Post-Nonlinear ICA Reflectance Model for 3D Surface Reconstruction

In this chapter, we propose a new photometric stereo scheme based on a new reflectance model and the post-nonlinear independent components analysis (ICA) method. The proposed *nonlinear reflectance model* consists of diffuse components and specular components for modeling the surface reflectance of a stereo object in an image. Unlike the previous approaches, these two components are not separated and processed individually in the proposed model. An unsupervised learning adaptation algorithm is developed to estimate the reflectance model based on image intensities. In this algorithm, the post-nonlinear ICA method is used to obtain the surface normal on each point of an image. Then, the 3D surface model is reconstructed based on the estimated surface normal on each point of image by using the enforcing integrability method. Two experiments are performed to assess the performance of the proposed approach. We test our algorithm on synthetically generated images for the reconstruction of surface of objects and on a number of real images captured from the Yale Face Database B. These testing images contain variability due to illumination and varying albedo in each point of surface of human faces. All the experimental results are compared to those of the existing photometric stereo approaches tested on the same images. The results clearly indicate the superiority of the proposed nonlinear reflectance model over the conventional Lambertian model and the other linear hybrid reflectance model.

## 4.1. Introduction

In this chapter, we propose a novel post-nonlinear ICA-based reflection model that consists of the diffuse components and the specular components. We do not need to separate the two components from the novel nonlinear reflection model. An unsupervised learning adaptation algorithm is used to tune up the proportion of hybrid automatically based on image intensities. The technique of the post-nonlinear independent components analysis (ICA) model [59]-[61] is used to solve the surface normal on each point of an image. The goal of post-nonlinear ICA is to nonlinearly transform the data such that the transformed variables are as statistically independent from each other as possible. Finally, the 3D surface model is reconstructed from the surface normal on each point of an image, obtained by the post-nonlinear ICA technique, using the method of enforcing integrability [51]. The reason is that it is easy to implement.

The rest of this chapter is organized as follows. Section 4.2 describes the basic reflectance models, including the Lambertian model and non-Lambertian. The details of the proposed post-nonlinear ICA-based reflectance model and its derivations are presented in Section 4.3. Extensive experiments have been performed to evaluate the performance of the proposed approach, and parts of the results are presented in Section 4.4. Conclusions are summarized in the last section.

## 4.2. The Proposed Non-linear Reflectance Model

In this chapter, we propose a new nonlinear reflectance model; it can model both the diffuse components and specular components into a single model. This model is described by

$$R_{nonlinear}\left(\mathbf{n}(x,y),\sigma(x,y),\gamma(x,y)\right)=L\gamma(x,y)\exp\left(-\frac{\left(\cos^{-1}\left(\mathbf{a}^{T}\mathbf{n}(x,y)\right)\right)^{2}}{2\sigma^{2}(x,y)}\right), \tag{4-1}$$

where $R_{nonlinear}(\cdot)$ denotes the nonlinear reflectance intensities, **a** is a 3×1 column vector and it represents to the light direction and viewing direction, $L$ is light strength, and $\gamma(x,y)$ is composite albedo on position $(x,y)$ of surface. $\sigma(x,y)$ is the variance of exponential function. Fig. 4-1 shows the normalized exponential functions with different sigma values, where angle is from $-\pi/2$ to $\pi/2$. When $\sigma(x,y)$ is large, the $R_{nonlinear}(\cdot)$ models more the diffuse component intensity. When $\sigma(x,y)$ gets smaller, $R_{nonlinear}(\cdot)$ models more the specular component intensity. So, we can obtain the best approximation by the adjustment of $\sigma(x,y)$. However, the following task is to solve the surface normal, $\mathbf{n}(x,y)$ for all $x$ and $y$, of Eq. (4-1) from 2-D intensities images. Since the $\mathbf{n}(x,y)$ vector is a 3×1 column vector, it is a limit that we need at least three images under different light directions. If the location of light sources were given, we could solve the normal vector on surfaces of every location $(x,y)$. But unfortunately, light sources could not be known in the general applications. Because the problem of solving Eq. (4-1) is a blind separation problem, an unsupervised learning adaptation algorithm based on images intensities can be used in solving Eq. (4-1). The technique of the nonlinear independent components analysis (ICA) model is used to solve the surface normal on each point of image. The post-nonlinear ICA is a technique that exploits higher-order statistical structure in data. This method has recently gained attention due to its applications to signal processing problems including speech enhancement, discrete signal processing and image processing. The goal of post-nonlinear ICA is to nonlinearly transform the data such that the transformed variables are as statistically independent from each other as possible. The detail for how to find $\mathbf{n}(x,y)$ of each point in image is showed in the following post-nonlinear ICA model.

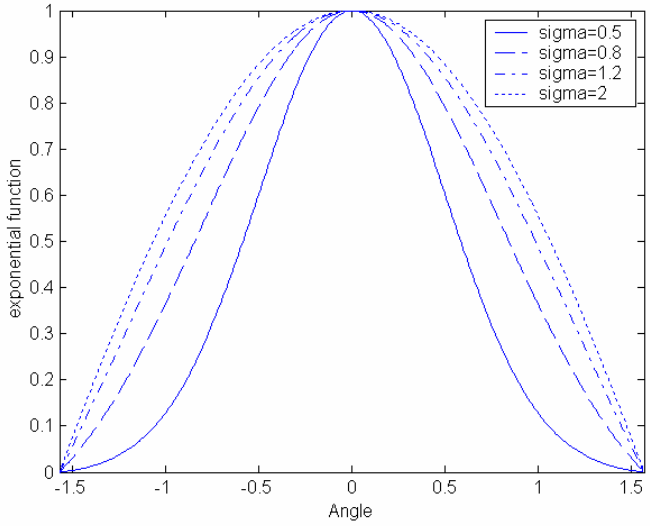



Figure 4-1. Normalized exponential functions with different sigma values (Angle: $-\frac{\pi}{2} \sim \frac{\pi}{2}$).

## 4.3. Post-Nonlinear ICA Model

In this section, we introduce the particular nonlinear mixtures, which can be considered to be a hybrid structure consisting of a linear stage followed by a nonlinear stage. It is shown in Fig. 4-2. This structure, which was introduced by Taleb and Jutten [59], provides the observation $\mathbf{x}(t) = (x_1(t), x_2(t), \ldots, x_n(t))^T$, which is the unknown nonlinear mixture of the unknown statistically independent source $\mathbf{s}(t) = (s_1(t), s_2(t), \ldots, s_n(t))^T$ :

$$x_i(t) = f_i\left(\sum_{j=1}^{n} a_{ij} s_j(t)\right),\ i = 1,\ 2,\ ...,\ n, \tag{4-2}$$

where $f_i(\cdot)$ are unknown invertible derivable nonlinear functions, and $a_{ij}$ $(i, j = 1, 2, \ldots, n)$ denote the scalar elements of a regular mixing matrix **A**. In the following, the mixture vector $\mathbf{x}(t)$, and by extension the pair (**A**, **f**), will be called a post-nonlinear (PNL) model.

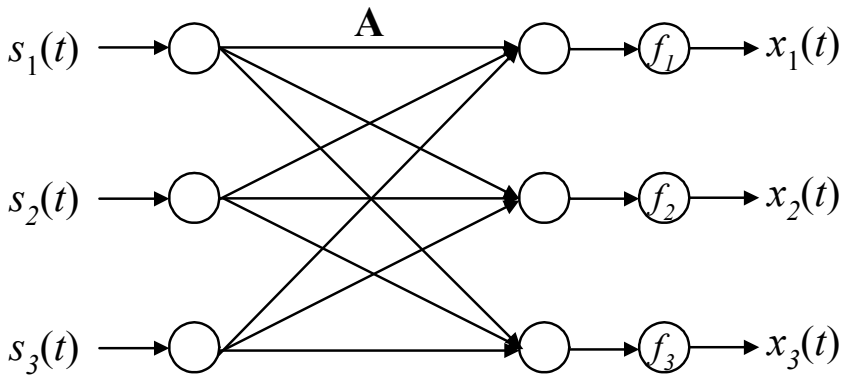



Figure 4-2. Post-nonlinear mixing ICA model ($n$ = 3).

Contrary to general nonlinear mixtures, the PNL model has a favorable separability property. That is, using the separation structure (**g**, **B**) shown in Fig. 4-3, it can be demonstrated, under weak conditions on the mixing matrix **A** and on the source distribution, that the output independence can be obtained if and only if $f_i \bullet g_i$ are linear for all index $i$ from $1$ to $n$. This means that the sources $\mathbf{y}(t) = (y_1(t), y_2(t), \ldots, y_n(t))^T$, which was estimated using an independence criterion on the outputs, are equal to the unknown sources with the same indeterminacies noted in linear mixture model.

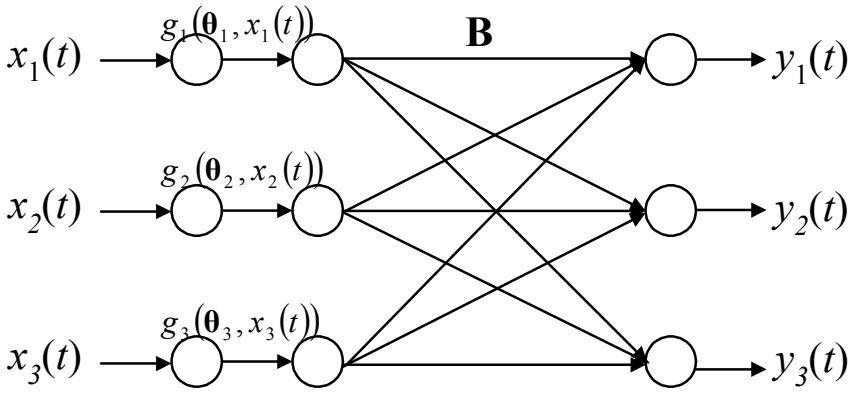



Figure 4-3. Separation architecture of the post nonlinear ICA model ($n$ = 3).

A very popular approach to estimating the ICA model is the maximum likelihood (ML) estimation. Maximum likelihood estimation is a fundamental method of statistical estimation. One interpretation of ML estimation is that we take those

parameter values as estimates that give the highest probability for the observations. In following section, we show how to apply ML estimation technique to post-nonlinear ICA estimation. The similar derivations of equations (4-3)-(4-13) based on the mutual information as a cost function is shown in the chapter by Taleb [59], [60].

### 4.3.1. Independence Criterion and Deriving the Likelihood

The statistical independence of the sources is the main assumption. Then, any separation architecture is tuned so that the components of its output **y** become statistically independent. This is achieved if the joint density factorizes as the product of the marginal densities

$$p(\mathbf{y}) = \prod_{i=1}^{n} p_i(y_i). \tag{4-3}$$

According to this result, the density $p_x(\mathbf{x})$ of the mixture vector $\mathbf{x} = \mathbf{f}(\mathbf{As})$ can be formulated as

$$p_x(\mathbf{x}) = |\det \mathbf{B}| \prod_{i=1}^{n} |g'(\boldsymbol{\theta}_i, x_i)| p(\mathbf{y}) = |\det \mathbf{B}| \prod_{i=1}^{n} |g'(\boldsymbol{\theta}_i, x_i)| \prod_{j=1}^{n} p_j(y_j), \tag{4-4}$$

where $\mathbf{B} = \mathbf{A}^{-1}$, $g(\cdot)$ is the inverse function of $f(\cdot)$, the parameters $\boldsymbol{\theta}_i$ are adjusted to cancel the effect of nonlinear function $f(\cdot)$, and the $p_j(y_j)$ denote the densities of the independent components. Equation (4-4) can be expressed as a function of $\mathbf{B} = (\mathbf{b}_1, \mathbf{b}_2, ..., \mathbf{b}_n)^T$ and **x**, giving

$$p_x(\mathbf{x}) = |\det \mathbf{B}| \prod_{i=1}^{n} |g'(\boldsymbol{\theta}_i, x_i)| \prod_{j=1}^{n} p_j\left(\mathbf{b}_j^T \mathbf{g}(\boldsymbol{\theta}, \mathbf{x})\right), \tag{4-5}$$

where $\mathbf{g}(\boldsymbol{\theta}, \mathbf{x}) = (g_1(\boldsymbol{\theta}_1, x_1), g_2(\boldsymbol{\theta}_2, x_2), ..., g_n(\boldsymbol{\theta}_n, x_n))^T$.

Assume that we have *T* observations of **x**, denoted by **x**(*1*), **x**(*2*), …, **x**(*t*), …, **x**(*T*). Then the likelihood can be obtained as the product of this density evaluated at the *T* points. This is denoted by $L(\mathbf{B}, \boldsymbol{\theta})$ and we have

$$L(\mathbf{B},\boldsymbol{\theta}) = \prod_{t=1}^{T} |\det \mathbf{B}| \prod_{i=1}^{n} |g'(\boldsymbol{\theta}_i, x_i)| \prod_{j=1}^{n} p_j\left(\mathbf{b}_j^T \mathbf{g}(\boldsymbol{\theta}, \mathbf{x}(t))\right). \tag{4-6}$$

In general, it is more practical to use the logarithm of the likelihood, since it is algebraically simpler. This does not make any difference here since the maximum of the logarithm is obtained at the same point as the maximum of the likelihood. The log-likelihood is given by

$$\log L(\mathbf{B},\boldsymbol{\theta}) = \sum_{t=1}^{T} \log\left\{ \prod_{i=1}^{n} |g'(\boldsymbol{\theta}_i, x_i)| \prod_{j=1}^{n} p_j\left(\mathbf{b}_j^T \mathbf{g}(\boldsymbol{\theta}, \mathbf{x}(t))\right) |\det \mathbf{B}| \right\}$$

$$= \sum_{t=1}^{T} \left\{ \sum_{i=1}^{n} \log|g'(\boldsymbol{\theta}_i, x_i)| + \sum_{j=1}^{n} \log\left[p_j\left(\mathbf{b}_j^T \mathbf{g}(\boldsymbol{\theta}, \mathbf{x}(t))\right)\right] + \log|\det \mathbf{B}| \right\}. \tag{4-7}$$

To simplify notation and to make it consistent to what can denote the sum over the sample index *t* by an expectation operator, thus we have

$$\frac{1}{T}\log L(\mathbf{B},\boldsymbol{\theta}) = E\left\{ \sum_{i=1}^{n} \log|g'(\boldsymbol{\theta}_i, x_i)| \right\} + E\left\{ \sum_{j=1}^{n} \log\left[p_j\left(\mathbf{b}_j^T \mathbf{g}(\boldsymbol{\theta}, \mathbf{x}(t))\right)\right] \right\} + \log|\det \mathbf{B}|, \tag{4-8}$$

where the expectation operator here is an average computed from the observed samples.

### 4.3.2. The Derivation of Adaptation Rules with Maximum Likelihood (ML) Estimation

To perform maximum log-likelihood estimation in practice, we need an algorithm to perform the numerical maximization of log-likelihood. In this section, we perform the numerical maximization of log-likelihood by gradient methods. First, the maximization of the log-likelihood requires the computation of its gradient with respect to the separation architecture parameters **B** and $\boldsymbol{\theta}_i, i = 1, 2, ..., n$.

*The first layer*: To estimate the linear stage parameters, we must compute the gradient of log-likelihood of Eq. (4-8) with respect to the separation architecture parameters **B**. Therefore, we have

$$\frac{1}{T}\frac{\partial \log L(\mathbf{B},\boldsymbol{\theta})}{\partial \mathbf{B}} = E\left\{\sum_{j=1}^{n} h_j\left(\mathbf{b}_j^T \mathbf{g}(\boldsymbol{\theta}, \mathbf{x}(t))\right)\mathbf{g}(\boldsymbol{\theta}, \mathbf{x}(t))^T\right\} + \left(\mathbf{B}^T\right)^{-1}, \tag{4-9}$$

where

$$h_j(y_j) = \frac{\partial \log p_{y_j}(y_j)}{\partial y_j} = \frac{p'_{y_j}(y_j)}{p_{y_j}(y_j)} \quad \text{and} \quad y_j = \mathbf{b}_j^T \mathbf{g}(\boldsymbol{\theta}, \mathbf{x}(t)).$$

Therefore, this immediately gives the following adaptation rule for ML estimation:

$$\Delta\mathbf{B} \propto E\left\{\sum_{j=1}^{n} h_j\left(\mathbf{b}_j^T \mathbf{g}(\boldsymbol{\theta}, \mathbf{x}(t))\right)\mathbf{g}(\boldsymbol{\theta}, \mathbf{x}(t))^T\right\} + \left(\mathbf{B}^T\right)^{-1} \quad \text{and} \tag{4-10}$$

$$\mathbf{B}^{(k+1)} = \mathbf{B}^{(k)} + \eta_{\mathbf{B}}(\Delta\mathbf{B}), \tag{4-11}$$

where $\eta_{\mathbf{B}}$ is the learning rate for adapting **B**. This result has the same expression as in the linear source separation. This algorithm is often called the Bell-Sejnowski algorithm [61]. It is the simplest algorithm for maximizing likelihood by gradient methods. However, due to the inversion of the matrix **B** in Eq. (4-10) is needed in every step, it converges very slowly. The convergence can be improved by whitening the data, and especially by using the natural gradient [63] that is based on the geometrical structure of the parameter space. Therefore, Eq. (4-12) is used to estimate the linear stage instead of the Eq. (4-10).

$$\Delta\mathbf{B} \propto \left(E\left\{\sum_{j=1}^{n} h_j(\mathbf{y}(t))\mathbf{y}(t)^T\right\} + \mathbf{I}\right)\mathbf{B} \tag{4-12}$$

where $\mathbf{y}(t) = \mathbf{b}_j^T \mathbf{g}(\boldsymbol{\theta}, \mathbf{x}(t))$.

*The second layer*: The derivation of the log-likelihood with respect to parameters $\boldsymbol{\theta}_i$ of the nonlinear function $g_i(\boldsymbol{\theta}_i, x_i)$ is

$$\frac{1}{T}\frac{\partial \log L(\mathbf{B},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i} = E\left\{\frac{\partial \log\left|g_i'(\boldsymbol{\theta}_i, x_i)\right|}{\partial \boldsymbol{\theta}_i}\right\} + E\left\{\left[\sum_{j=1}^{n} h_j\left(\mathbf{b}_j^T \mathbf{g}(\boldsymbol{\theta}, \mathbf{x}(t))\right) b_{ji}\right]\frac{\partial g_i(\boldsymbol{\theta}_i, x_i)}{\partial \boldsymbol{\theta}_i}\right\}. \tag{4-13}$$

From the derivation of the log-likelihood with respect to parameters $\boldsymbol{\theta}_i$, we update the parameters $\boldsymbol{\theta}_i$ of the $g_i(\boldsymbol{\theta}_i, \mathbf{x}(t))$ function by the following adaptation rule:

$$\Delta\boldsymbol{\theta}_i \propto E\left\{\frac{\partial \log\left|g_i'(\boldsymbol{\theta}_i, x_i)\right|}{\partial \boldsymbol{\theta}_i}\right\} + E\left\{\left[\sum_{j=1}^{n} h_j\left(\mathbf{b}_j^T \mathbf{g}(\boldsymbol{\theta}, \mathbf{x}(t))\right) b_{ji}\right]\frac{\partial g_i(\boldsymbol{\theta}_i, x_i)}{\partial \boldsymbol{\theta}_i}\right\} \quad \text{and} \tag{4-14}$$

$$\boldsymbol{\theta}_i^{(k+1)} = \boldsymbol{\theta}_i^{(k)} + \eta_{\boldsymbol{\theta}_i}\left(\Delta\boldsymbol{\theta}_i\right) \tag{4-15}$$

where $\eta_{\boldsymbol{\theta}_i}$ is the learning rate of adapting $\boldsymbol{\theta}_i$.

### 4.3.3. Estimation of the Source Densities

Denoted by $p_{\mathbf{y}}(\mathbf{y})$ the assumed densities of the independent components, and

$$h_j(y_j) = \frac{\partial \log p_{y_j}(y_j)}{\partial y_j} = \frac{p_{y_j}'(y_j)}{p_{y_j}(y_j)}. \tag{4-16}$$

Constrain the estimates of independent components $y_j = \mathbf{b}_j^T \mathbf{g}(\boldsymbol{\theta}, \mathbf{x}(t))$ to be uncorrelated and to have unit variance. Then the ML estimator is locally consistent, if the assumed densities $p_{y_j}(y_j)$ fulfill

$$E\{y_j h_j(y_j) - h_j'(y_j)\} > 0, \ \forall\ j. \tag{4-17}$$

The proof can be found in the [64]. Therefore, the limitation shows how to construct families consisting of only two densities, so that the condition in Eq. (4-17) is true for one of these densities. For example, consider the following log-densities:

$$\log p^{+}(s) = \alpha_1 - 2\log\cosh(s) \tag{4-18}$$

$$\log p^{-}(s) = \alpha_2 - \left(\frac{s^2}{2} - \log\cosh(s)\right), \tag{4-19}$$

where $\alpha_1, \alpha_2$ are positive parameters that are fixed so as to make these two functions logarithms of probability densities. Actually, these constants can be ignored in the following. Then, for super-Gaussian independent components, the pdf defined by Eq. (4-18) is usually used. This means that the nonlinear function $h(\cdot)$ is the tanh function:

$$h^{+}(y) = -2\tanh(y). \tag{4-20}$$

For sub-Gaussian independent components, the other pdf defined by Eq. (4-19) is used. Then the nonlinear function $h(\cdot)$ can be written as:

$$h^{-}(y) = \tanh(y) - y. \tag{4-21}$$

Finally, the choice between the two nonlinearities in Eq. (4-20) and Eq. (4-21) can be made by computing the nonpolynomial moment:

$$k_j = sign\left\{E\left[-\tanh(y_j)y_j + \left(1-\tanh(y_j)^2\right)\right]\right\},\ j = 1, 2, ..., n, \tag{4-22}$$

using some estimates of the independent components. Then, the source distribution is super-Gaussian when $k_j = 1$ and sub-Gaussian when $k_j = -1$, where the expectation value in the formulas is for all *t*, *t* = *1*, *2*, …, *T*.

## 4.4. Solving the Proposed Nonlinear Reflectance Model by Post-Nonlinear ICA Model

In this section, we shall describe the way of applying the post-nonlinear ICA model to estimate the normal vector $\mathbf{n}(x, y)$ on the object surface corresponding to each pixel in the image. Since the $\mathbf{n}(x, y)$ vector is a 3×1 column vector, it is required that we need at least three images under different light directions for its estimation. Hence, to reconstruct the 3D surface of an object through its images, we have to take three gray-value images under three different illuminants. Assuming an image contains *T* pixels in total, then we can rearrange all the gray values of the three

images into a $3 \times T$ matrix, with each row representing an image, and each column the gray values of a single pixel under three different illuminants. Putting this matrix into Eq. (4-1), and comparing Eq. (4-1) with Eq. (4-2), we can define the nonlinear function in the post-nonlinear ICA model as:

$$f_j\left(o_j(t)\right) = L\gamma(t)\exp\left(-\frac{\left(\cos^{-1}\left(o_j(t)\right)\right)^2}{2\sigma^2(t)}\right), t = 1,\ 2,\ ...,\ T, \text{and } j = 1,\ 2,\ 3, \qquad (4\text{-}23)$$

where $o(t) = \sum_{i=1}^{3} a_i s_i(t)$ and $\mathbf{s}(t)$ is the $\mathbf{n}(x, y)$ vector that we are looking for. From Eq. (4-23), we can obtain the inverse nonlinear $f(\cdot)$ function as

$$g_j\left(\sigma(t), I_j(t)\right) = \cos\left(\sqrt{-2\sigma^2(t)\ln(\frac{I_j(t)}{\gamma(t)})}\right), t = 1,\ 2,\ ...,\ T, \text{and } j = 1,\ 2,\ 3, \qquad (4\text{-}24)$$

where **I**(t) is the input vector, i.e., the three gray values of the *t*th pixel of the three images with different illuminants, and $\sigma(t)$ is the variant of the *t*th pixel in the exp() function in Eq. (4-23). We shall feed the input vectors, **I**(t), *t=1, 2, ..., T,* to the input of the network shown in Fig. 4-3. Because all these input vectors come from the images belonging to the same object, the estimated reflectance model should be exactly the same for each of the three images; i.e., all the $g_i(\cdot)$, *i=1, 2, 3* are the same. With such setting, we can obtain the final outputs through the network computations shown in Fig. 4-3. To ensure the final outputs to be independent components, we apply the unsupervised adaptation rules derived in Subsection 4.3.2 to tune $\sigma(t)$ and **B** matrix. Upon convergence, the final output is the estimated normal vector $\mathbf{n}(x, y)$ on the object surface corresponding to the *t*th pixel in the image for *t*=1, 2, …, *T*. The complete algorithm for the above computation is shown in Fig. 4-4, which consists of 9 steps.

**Algorithm:**

1. Set $k = 1$ and arbitrarily assign the initial values of $\sigma(t)$ and $\gamma(t)$, $t = 1, \ldots, T$, and $\mathbf{B}$。
2. Set sampling index $t = 1$.
3. Compute $\mathbf{e}(t)$ from Eq. (4-23),
$$e_i(t) = \cos\left(\sqrt{-2\sigma^2(t)\ln(I_i(t)/\gamma(t))}\right), \quad i = 1, 2, 3.$$
4. Compute $\mathbf{y}(t)$ as follows:
$$\mathbf{y}(t) = \mathbf{B}\mathbf{e}(t).$$
5. Calculate the normal vector of surface $\mathbf{n}(t) = \dfrac{\mathbf{y}(t)}{\|\mathbf{y}(t)\|}$ and the surface albedo $\gamma(t) = \|\mathbf{y}(t)\|, \ \forall t.$
6. Update **B** matrix by
$$\mathbf{B}^{(k+1)} = \mathbf{B}^{(k)} + \eta_{\mathbf{B}}\left\{\mathbf{I} + \left[\mathbf{h}(\mathbf{y}(t))\mathbf{y}(t)^T\right]\right\}\mathbf{B}^{(k)},$$
where $\eta_{\mathbf{B}}$ is the learning rate of **B**.
7. Update $\sigma(t)$ value using the following equation:
$$\sigma_i(t)^{(k+1)} = \sigma_i(t)^{(k)} + \eta_\sigma\left\{\frac{2}{\sigma_i(t)} + 4\sigma_i(t)\log(x_i(t))\left[\sum_{j=1}^{3} h_j(y_j)b_{ji}\right]\right\},$$
where $\eta_\sigma$ is the learning rate of $\sigma$.
8. Repeat Step 3~Step 7, until *t* equal to *T*.
9. Set $k = k + 1$ and repeat Step 2~Step 8, until convergence.

Figure 4-4. Unsupervised updating rules for the proposed 3D surface reconstruction scheme.

The separation architecture of the post-nonlinear ICA model can be considered to be a hybrid structure consisting of a nonlinear stage followed by a linear stage. Therefore, after compensating for the post-nonlinearities, the problem is essentially reduced to a linear mixture of the form [matrix depending on lighting and viewing directions] * [surface normal vector]. Using the ICA decomposition, we rewrite the equation in Step 4 in Fig. 4-4 as

$$\mathbf{y}(t) = \mathbf{B}\mathbf{e}(t), \text{ or } \mathbf{e}(t) = \mathbf{B}^{-1}\mathbf{y}(t) = \gamma(t)\hat{\mathbf{A}}^T\hat{\mathbf{n}}(t), \tag{4-25}$$

where $\hat{\mathbf{A}}^T = [\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3]^T = \mathbf{B}^{-1}$ is the matrix depending on lighting and viewing

directions and has unit length, $\hat{\mathbf{n}}(t)$ is the estimated normal vector corresponding to the $t$th pixel, $t$ = 1, 2, …, $T$, and $\gamma(t)$ is albedo of the $t$th pixel. However, the decomposition in Eq. (4-25) is not unique. If there is an invertible matrix **G**, which satisfies

$$\mathbf{A}^T = \hat{\mathbf{A}}^T\mathbf{G} \quad \text{and} \quad \mathbf{n} = \mathbf{G}^{-1}\hat{\mathbf{n}}, \tag{4-26}$$

where **A** and $\mathbf{n}(t)$ are, respectively, the true matrix depending on lighting and viewing directions of images **I,** and the normal vector of the $t$th pixel in the standard XYZ coordinates, then the linear ambiguity is belonging to subset of GBR [33]-[35]. On the one hand, according to Georghiades's [35] studies, if the surface of an object is seen under variable light direction, but with fixed viewpoint, then the linear ambiguity can be reduced to three GBR parameters. As far as the surface normal vectors are concerned, we can only recover

$$\mathbf{n} \cong \mathbf{G}^{-1}\hat{\mathbf{n}},$$

and

$$\mathbf{G}^{-1} = \frac{1}{g_3}\begin{bmatrix} g_3 & 0 & 0 \\ 0 & g_3 & 0 \\ -g_1 & -g_2 & 1 \end{bmatrix}, \tag{4-27}$$

where $g_i$ are the three GBR parameters. On the other hand, the three light sources corresponding to the three images do not lie in the same plane (non-coplanar), so the columns of matrix **A** are linearly independent. In addition, using the ICA decomposition in Eq. (4-25), we can obtain an independent basis matrix $\hat{\mathbf{A}}$, so the ambiguity can further be denoted a diagonal matrix, i.e., $g_1 = 0$ and $g_2 = 0$. So, the relation between the normals in the standard XYZ coordinates and those in the most independent coordinates system is only by $g_3$ factor. For the performance evaluation of 3D image reconstruction, both estimated surface and synthetic one are normalized

within the interval [0, 1]. Therefore, the influence of $g_3$ factor on the estimated 3D surface can be removed.

Fig. 4-5 shows a simple example of a sphere object. The first row shows a set of shadow images and the second row shows the other set of shadow images. This synthetic image was generated using the depth function of a sphere object with different light directions. So the content of the images is different. Fig. 4-6(a) shows the true normal vectors of the sphere. The estimated normal vectors in Fig. 4-6(b) and 4-6(c) were generated by using our approach corresponding to the two sets of shadow images in the first row and second row of Fig. 4-5. According to the estimated normal vectors in Fig. 4-6(b) and 4-6(c), it is obvious that the waveforms are similar to the true normal vectors. So, the estimated normal vectors do not depend on the content of the images. Furthermore, as far as the order of the sources being concerned, the similarity between human face and sphere is adopted in the supervised ICA algorithm to find the order of sources in the proposed scheme. We compute the correlation between the estimated normal vectors of surface of faces and the normal vectors of a sphere due to their similar structure, so the order of normal components can be identified.

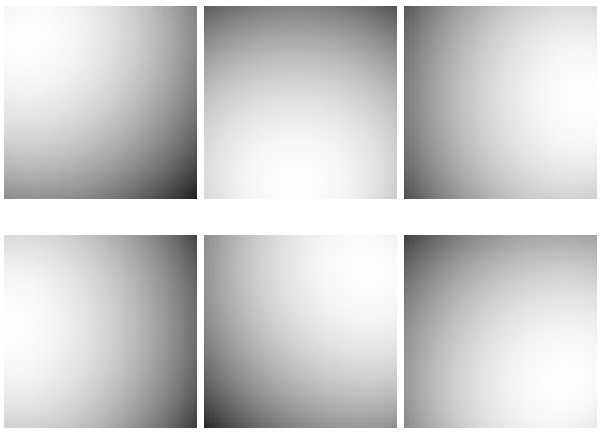
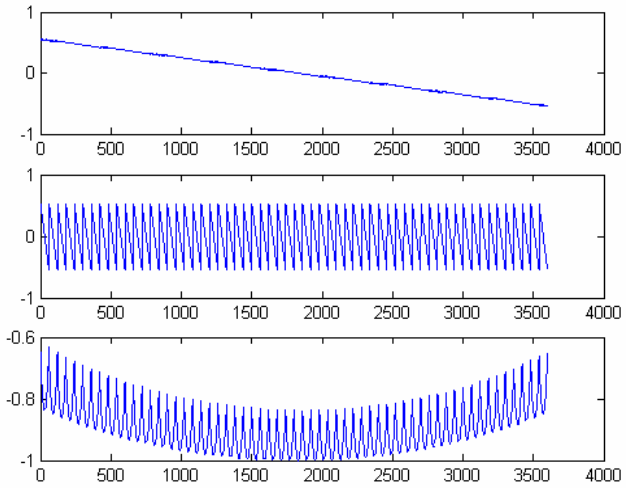
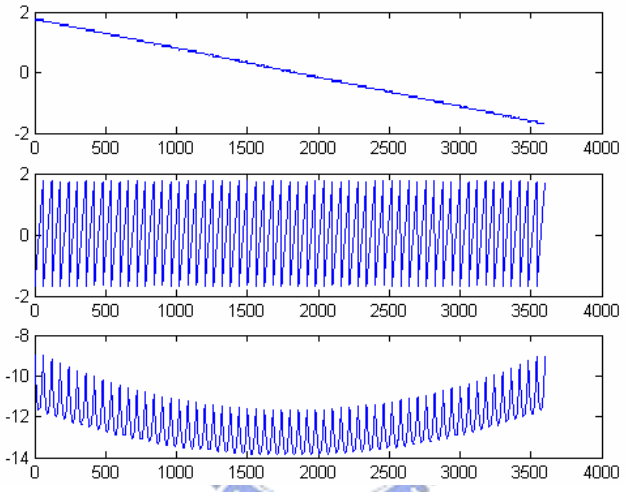


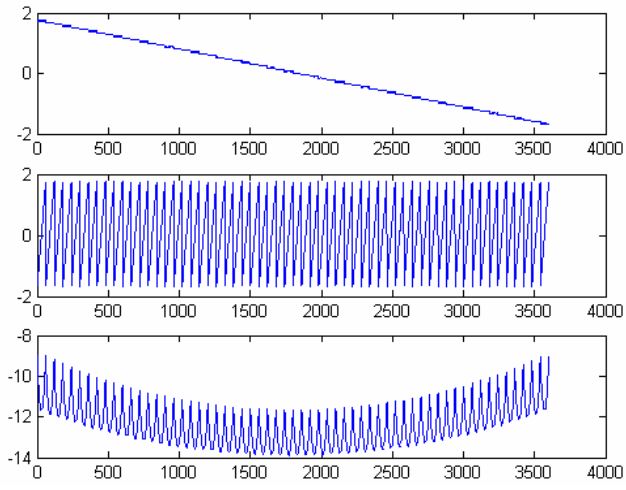
Figure 4-5. Shadow images of sphere object with different light directions.




(a)




(b)




(c)

Figure 4-6. The estimated normal vectors of sphere object by our approach. (a) The true normal vectors. (b) The estimated normal vectors from the first row of Figure 4-5. (c) The estimated normal vectors from the second row of Figure 4-5.

## 4.5. 3D Surface Reconstruction from the Surface Normal by Enforcing Integrability

In this section, we use the enforcing integrability approach to obtain the deeper information for reconstructing of the surface of an object by its normal vectors. This approach was proposed in the earliest stage by R. T. Frankot and R. Chellappa in 1988. Suppose that we represent the surface $z(x, y)$ by a finite set of integrable basis functions $\phi(x, y, \boldsymbol{\omega})$ so that

$$z(x, y) = \sum_{\boldsymbol{\omega} \in \Omega} c(\boldsymbol{\omega}) \phi(x, y, \boldsymbol{\omega}), \tag{4-28}$$

where $\boldsymbol{\omega} = (u, v)$ is a two-dimensional index, $\Omega$ is a finite set of indexes, and $\{\phi(x, y, \boldsymbol{\omega})\}$ is a finite set of integrable basis functions which are not necessarily mutually orthogonal. We chose the discrete cosine basis so that $\{c(\boldsymbol{\omega})\}$ is exactly the full set of discrete cosine transform (DCT) coefficients of $z(x, y)$. Since the partial derivatives of the basis functions, $\phi_x(x, y, \boldsymbol{\omega})$ and $\phi_x(x, y, \boldsymbol{\omega})$, are integrable, the partial derivatives of $z(x, y)$ are guaranteed to be integrable as well; that is, $z_{xy}(x, y) = z_{yx}(x, y)$. Note that the partial derivatives of $z(x, y)$ can also be expressed in terms of this expansion, giving

$$z_x(x, y) = \sum_{\boldsymbol{\omega} \in \Omega} c(\boldsymbol{\omega}) \phi_x(x, y, \boldsymbol{\omega}) \tag{4-29}$$

$$z_y(x, y) = \sum_{\boldsymbol{\omega} \in \Omega} c(\boldsymbol{\omega}) \phi_y(x, y, \boldsymbol{\omega}), \tag{4-30}$$

where $\phi_x(x, y, \boldsymbol{\omega}) = \partial \phi(\cdot) / \partial x$ and $\phi_y(x, y, \boldsymbol{\omega}) = \partial \phi(\cdot) / \partial y$.

Suppose we now have the possibly nonintegrable estimate $\mathbf{n}(x, y)$ from which we can easily deduce from Eq. (2-2) the possibly nonintegrable partial derivatives $\hat{z}_x(x, y)$ and $\hat{z}_y(x, y)$. These partial derivatives can also be expressed as a series,

giving

$$\hat{z}_x(x,y)=\sum_{\boldsymbol{\omega}\in\Omega}\hat{c}_1(\boldsymbol{\omega})\phi_x(x,y,\boldsymbol{\omega}) \tag{4-31}$$

$$\hat{z}_y(x,y)=\sum_{\boldsymbol{\omega}\in\Omega}\hat{c}_2(\boldsymbol{\omega})\phi_y(x,y,\boldsymbol{\omega}). \tag{4-32}$$

A method has been proposed for finding the expansion coefficients $c(\boldsymbol{\omega})$ given a possibly nonintegrable estimate of surface slopes $\hat{z}_x(x,y)$ and $\hat{z}_y(x,y)$:

$$c(\boldsymbol{\omega})=\frac{p_x(\boldsymbol{\omega})\hat{c}_1(\boldsymbol{\omega})+p_y(\boldsymbol{\omega})\hat{c}_2(\boldsymbol{\omega})}{p_x(\boldsymbol{\omega})+p_y(\boldsymbol{\omega})}, \text{ for } \boldsymbol{\omega}=(u,v)\in\Omega, \tag{4-33}$$

where $p_x(\boldsymbol{\omega})=\iint|\phi_x(x,y,\boldsymbol{\omega})|^2dxdy$ and $p_y(\boldsymbol{\omega})=\iint|\phi_y(x,y,\boldsymbol{\omega})|^2dxdy$. Finally, we can reconstruct the object's surface by performing the inverse 2-D DCT on the coefficients $c(\boldsymbol{\omega})$.

## 4.6. Experimental Results and Discussions

In this section, two experiments are performed to assess the performance of the proposed approach. In the first experiment, we test the algorithm on synthetically generated images for the reconstruction of surface of objects. The light direction and viewing direction are unknown. In the second experiment, we test the algorithm on a number of real images captured from the Yale Face Database B showing the variability due to illumination and there is varying albedo in each point of surface of human faces. All the experimental results are compared to those of the Georghiades's approach in [34] and the Hayakawa's approach in [43] tested on the same images.

### 4.6.1. Quantitative Experimental Results by Reconstructing a Synthetic Sphere Object

Quantitative experimental results have been obtained by reconstructing a synthetic sphere object. The true depth map of the synthetic sphere object is generated

mathematically as

$$z(x,y)=\begin{cases}\left|\sqrt{r^2-x^2-y^2}\right|, & \text{if } x^2+y^2\le r^2\\ 0, & \text{otherwise}\end{cases}, \tag{4-34}$$

where $r$=48, $0<x,y\le 100$, and the center is located at ($x$, $y$)=(51, 51). The sphere object is showed in Fig. 4-7. This synthetic image was generated using the depth function in Eq. (4-34) and the surface gradients were computed using the discrete approximation. Fig. 4-8 shows the synthetic images generated according to the non-Lambertian model with varying albedo and different directions. The different albedos are, 0.6 for right-bottom of the sphere, 0.8 for left-top of the sphere, and 1 for the rest part. The locations of light sources in Figs. 4-9(a)-(i) are S1=(30, 140), S2=(30, 90), S3=(30, 40), S4=(30, 180), S5=(0, 0), S6=(30, 0), S7=(30, -140), S8=(30, -90), and S9=(30, -40), 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. The experimental results are shown in Table 4-1 and the proposed method is compared with two photometric stereo algorithms, Hayakawa's method and Georghiades's method. In Table 4-1, we take 5 groups of images with different illuminant angles from the left, the right, and the front for 3D reconstruction. Both estimated surface and synthetic one are normalized within the interval [0, 1]. According to Table 4-1, it is found that the proposed method can achieve the lowest mean errors compared with the other methods in all illumination conditions. In addition, we implemented each method in Matlab 6.1 software on a 1.2GHz Pentium III-based PC with 256 MB RAM. According to the results, the CPU time used by the proposed method is close to that used by the Georghiades's method ([34]) and it is greatly reduced compared with the Hayakawa's method ([43]).

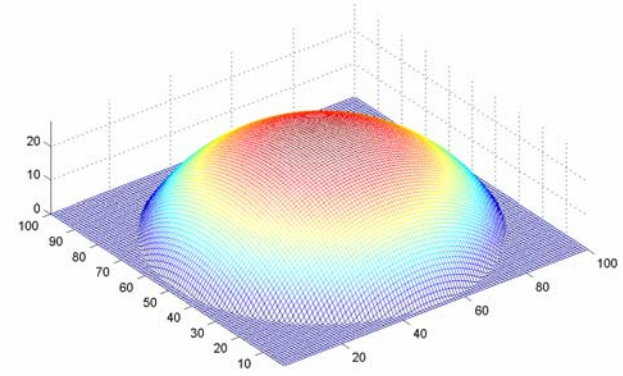
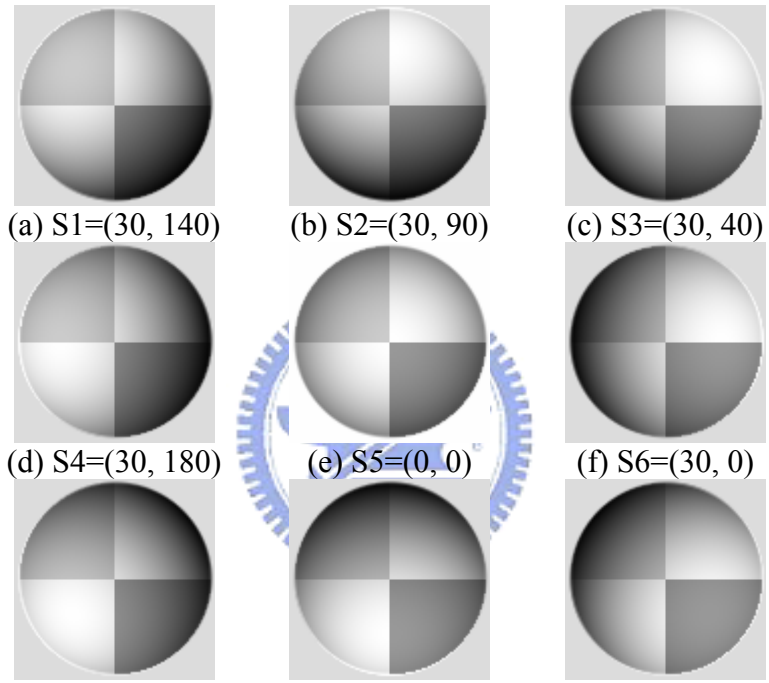



Figure 4-7. Synthetic sphere surface object.



(a) S1=(30, 140) (b) S2=(30, 90) (c) S3=(30, 40)

(d) S4=(30, 180) (e) S5=(0, 0) (f) S6=(30, 0)

(g) S7=(30, -140) (h) S8=(30, -90) (i) S9=(30, -40)

Figure 4-8. The 2D sphere images generated with varying albedo and different lighting directions (the degree of tilt angle, the degree of pan angle).

Table 4-1 The absolute mean errors between estimated depths and desired depth of synthetic object's 3D surfaces. (Both light and viewing directions are unknown in the experiment.)

| Mean absolute depth error | Lights | Georghiades's method ([34]) | Hayakawa's method ([43]) | Our proposed method |
|---|---|---|---|---|
| Sphere with Variant albedo | S1, S2, S3 | 0.048875 | 0.07688 | 0.02025 |
| | S7, S8, S9 | 0.050316 | 0.07924 | 0.02687 |
| | S1, S5, S3 | 0.033133 | 0.07279 | 0.02055 |
| | S1, S8, S6 | 0.033529 | 0.07869 | 0.01837 |
| | S1, S5, S7 | 0.031729 | 0.07621 | 0.01829 |
| | CPU time(Avg.) | 27.38 sec (Iterations = 10) | 56.85 sec | 25.9 sec (Iterations = 10) |

### 4.6.2. Experimental Results on Real Images with Varying Albedo

In the second experiment, we test the algorithm on a number of real images from the Yale Face Database B [55] showing the variability due to illumination and there is varying albedo in each point of surface of human faces. This subset contains 444 viewing conditions (1 pose × 37 illumination conditions, where these illumination conditions contain Subset1 (12°) and Subset2 (25°) in the Yale Face Database B.) for 10 individuals. Fig. 4-9 shows the 10 individuals from the Yale Face Database B used to test our algorithm, where each image size is $100 \times 100$ in pixels.

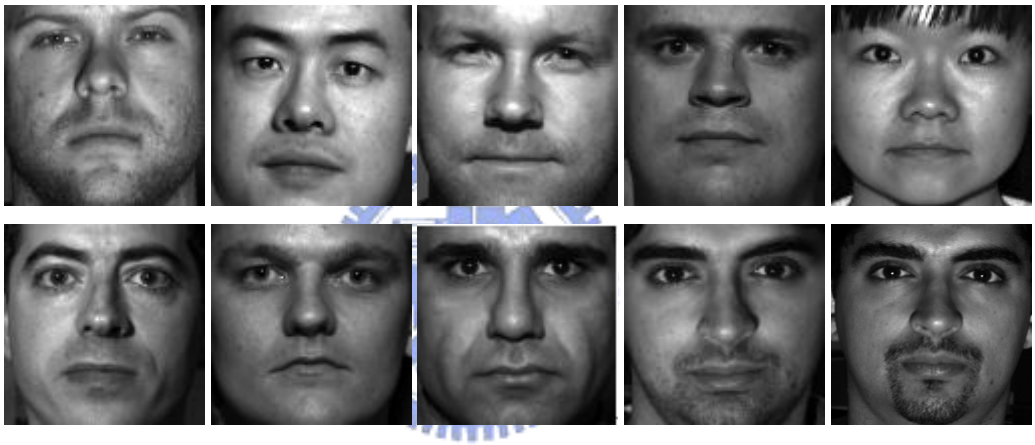


Figure 4-9. 10 individuals from the Yale Face Database B used to test our algorithm.

First, we take the images of the same person that was photographed under three different light sources from these testing images arbitrarily shown in Fig. 4-10. We feed the normalized images into our algorithm. After updating the parameters by several iterations, we can get the normal vector of the surfaces of human faces corresponding to each pixel in an image in the output nodes. The results are shown in the second row in Fig. 4-10, which are the first component, the second component, and the third component of the surface normal vector in order. Fig. 4-11 presents the results of 3D human face reconstruction. Fig. 4-11(a) shows the surface albedo of human face in Fig. 4-10. Fig. 4-11(b) shows the result with our proposed algorithm.

By using the Georghiades's approach [34] and the Hayakawa's approach [43], the reconstructed results are demonstrated in Fig. 4-11(c) and 4-11(d), respectively. The results clearly indicate that the performance of our proposed nonlinear reflectance model is better than that of the Georghiades's approach and the Hayakawa's approach. Comparing to the results obtained by the Georghiades's approach, the reconstructed surfaces with the consideration of specular components in our algorithm, are obviously better in high-gradient parts such as the nose. Besides, the Hayakawa's approach did need added constraints, it could reconstruct the 3D model of human face similar as our approach, but when the constraints is unavailable, then it could not reconstruct the 3D model of human face. Finally, the reconstructed results for the testing patterns are shown in Fig. 4-12 and Fig. 4-13.

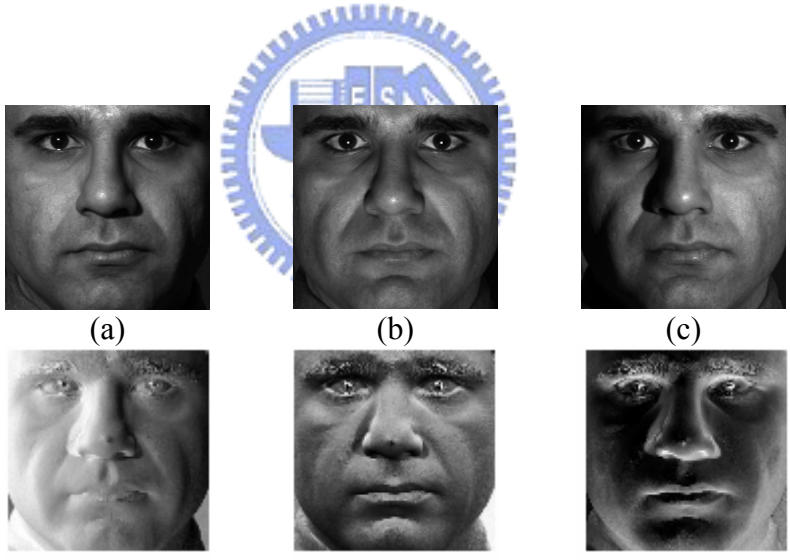
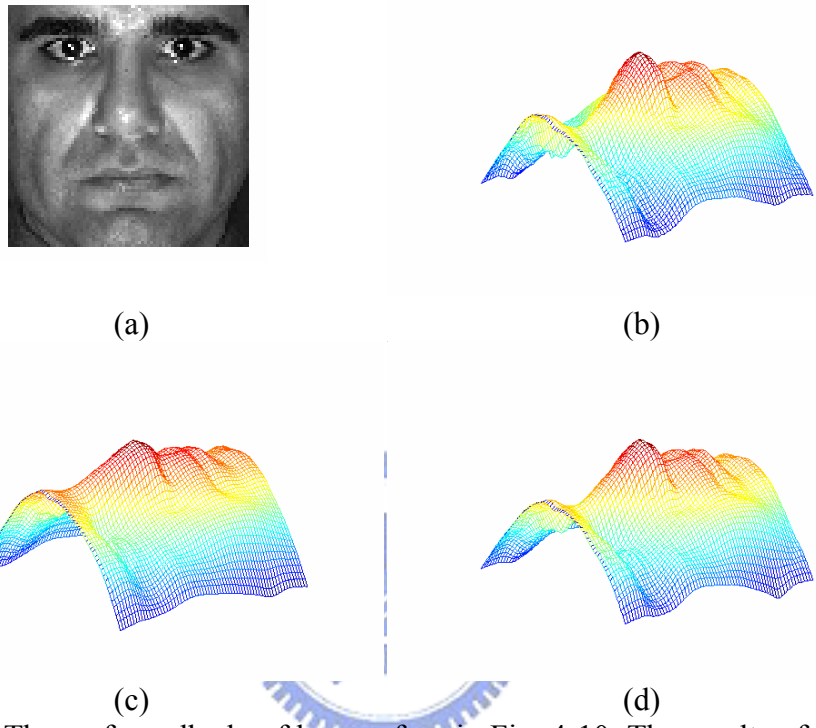


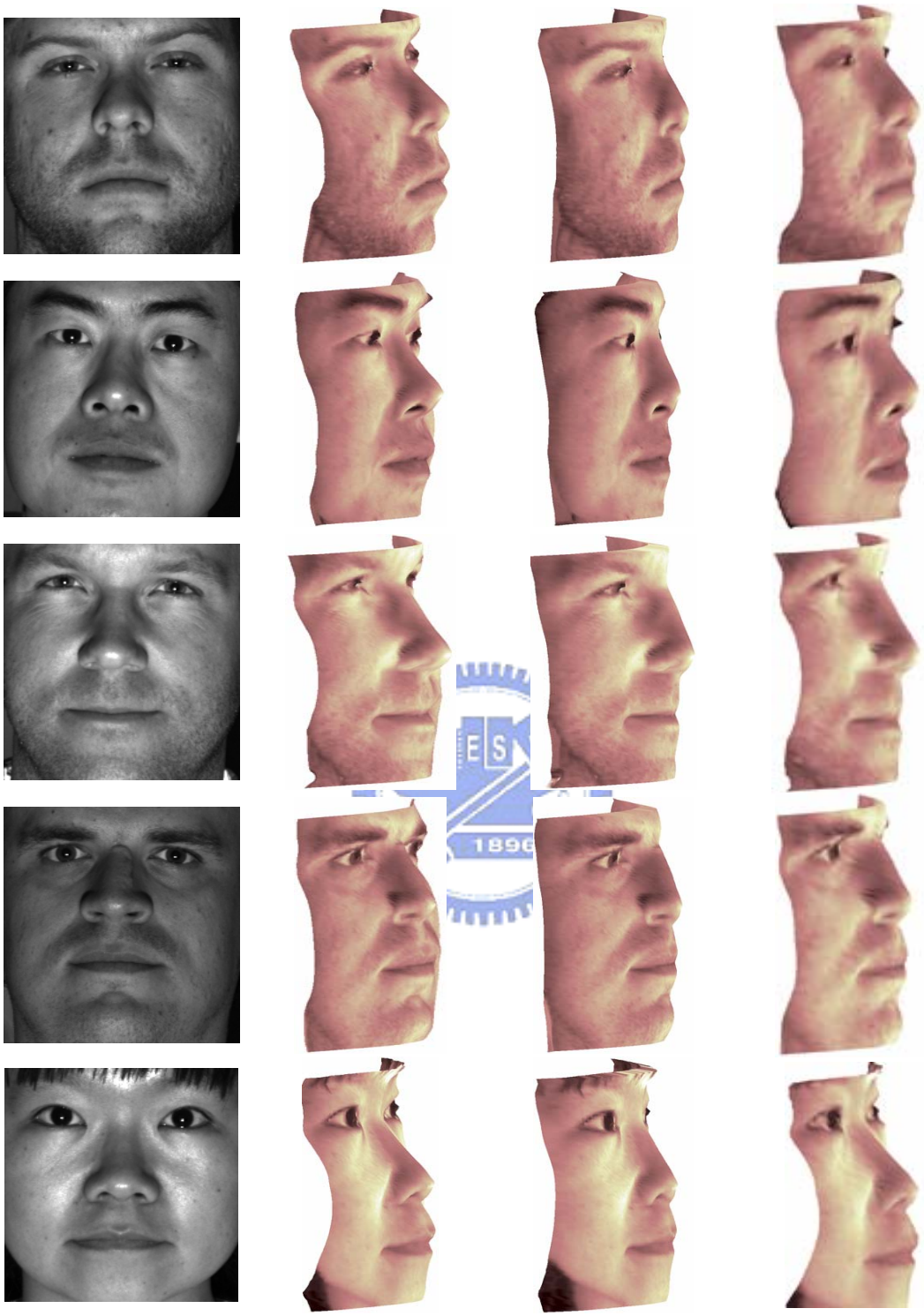

Figure 4-10. Three training images with differ light source positions from Yale Face Database B in frontal. (b) Surface normal corresponding to the three source images.



(a) (b)

(c) (d)

Figure 4-11. The surface albedo of human face in Fig. 4-10. The results of 3D model reconstruction by (b) our proposed algorithm, (c) Georghiades's approach in [34], and (d) Hayakawa's approach in [43].



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

Figure 4-12.The results of 3D model reconstruction by (b) our proposed algorithm, (c) Georghiades's approach in [34], and (d) Hayakawa's approach in [43].

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

Figure 4-13.The results of 3D model reconstruction by (b) our proposed algorithm, (c) Georghiades's approach in [34], and (d) Hayakawa's approach in [43].

## 4.7. Concluding Remarks

It has been claimed that methods based on reconstructing 3D face model for face recognition are quite successful. When we are estimating the surface shape, the success of the reflectance model for surface reconstruction of objects depends on two major components: the diffuse component and the specular component. Therefore, in this chapter, we proposed a new nonlinear reflection model consisting of the diffuse components and the specular components. The past researches only considered the linear combination of the diffuse components and the specular components. We do not need to separate the two components in the proposed novel nonlinear reflection model. In addition to this major contribution, several contributions of the proposed algorithm are listed below:

(a). In the past, we have to know the locations of light sources first for solving the photometric stereo problems. But this is not practical in the real situations. In this chapter, we used the images under three different light source locations to solve this problem. In our method, we can still obtain a very good result even if the locations of light sources are not given.

(b). Using the unsupervised non-linear ICA network for solving photometric stereo problems does not need any desired output value and the smoothing conditions. It is easier to converge and make the system stable.

The performance comparisons of our proposed nonlinear reflectance model to the Georghiades's approach in [34] and the Hayakawa's approach in [43] were made. In the first experiment, we test the algorithm on synthetically generated images for the reconstruction of surface of objects. The results clearly indicate that the performances of our proposed nonlinear reflectance model are better than that of the Georghiades's approach in [34] and the Hayakawa's approach in [43]. In the second experiment, we

test the algorithm on a number of real images from the Yale Face Database B containing the variability due to illumination and varying albedo in each point of surface of human faces. All the experimental results showed that the performance of the proposed nonlinear reflectance model is better than those of the two proposed existing photometric stereo methods.

# 5. An Illumination Estimation Scheme for Color Reconstruction

In this chapter a new algorithm for surrounding illumination estimation of image scene for color constancy was proposed. This estimation is based upon the chromaticity histogram of a color image, which is obtained by the accumulation of *CIE* chromaticity values corresponding to all colors in the image. Unlike the existing approaches, the proposed scheme estimates the white-point values of the surrounding illuminant by detecting the central value of all color's distribution in the chromaticity histogram. The white point estimation of a color image based on the chromaticity histogram has the advantages of high efficiency, good robustness, and no strict assumptions. After the illumination estimation, a neural network with back-propagation (BP) learning algorithm is used to model the nonlinear functional relationship between the central value of chromaticity histogram and coefficients of illuminant functions. The trained BP network can then be used to estimate the spectral power distribution of surrounding illuminant. Substituting this illuminant estimate into the finite-dimensional linear model of surface reflectance, the colors of the image can be corrected and recovered with the standard illuminant (D65) for color constancy. For performance evaluation, two sets of color-recovery experiments are performed in this chapter based on synthetic images and nature images captured from a still digital camera, respectively. All the results are compared to those of two existing popular algorithms (Max-RGB and Gray-World algorithms) on the same sets of images. The

experimental results show that the new algorithm outperforms these two popular compared algorithms, both in quantitative error index and in qualitative visual perception.

## 5.1. Introduction

In this chapter, we propose a new illumination estimation scheme based on chromaticity histogram and neural network. At first, the central values of the chromaticity histogram characterize the distribution of image colors with an unknown surrounding illuminant is detected. When the color of the surrounding light source of the sampling image approaches to red, the image colors in the chromaticity histogram would be obviously distributed in the region of lower color temperatures. Conversely, the image color in the chromaticity histogram will be obviously distributed in the region of higher color temperatures, when the color of the surrounding light source of the sampling image approaches to blue. That is why we use the chromaticity histogram to detect the central values of the main distribution of image colors. Next, the BP neural network is applied to approximate the nonlinear functional relationship between the central values of chromaticity histogram and coefficients of basis function of illuminant sources. Finally, the spectral power distribution of the surrounding illuminant, $\hat{E}(\lambda)$, estimated by the BP network is substituted into the finite-dimensional linear model of surface reflectance to complete the color correction with the standard illuminant (as D65).

The rest of this chapter is organized as follows. Section 5.2 describes the finite-dimensional linear model of surface reflectance for correcting the image color pixel-wisely based on the estimation of scene illumination. The details of the proposed illumination estimation algorithm are presented in Section 5.3. The chromaticity histogram is used to extract the distribution of image colors and the

spectral power distribution of the surrounding illuminant is estimated using a BP network. In Section 5.4, two sets of experiments are performed to evaluate the proposed approach, and the results are compared to those of two popular existing color constancy schemes. Conclusions are summarized in the last section.

## 5.2. Finite-Dimensional Linear Model of Surface Reflectance

As shown in Fig. 1-1, the complete procedure of the color constancy process involves two steps. The first step estimates the spectral power distribution of the surrounding illumination of an image, and based on which the second step corrects the image color pixel-wisely to the colors under desired illuminant such as the standard illuminant D65. In this section, we shall introduce the method of the second step, i.e., color correction of an image from a known (estimated) illuminant to another desired (standard) illuminant. Most algorithms [65]-[74] have solved this problem based on *finite-dimensional linear models* of surface reflectance or called illuminant functions. The model condenses all spectral information into a few numbers by supposing that illumination and reflectance can each be approximated by weighted sums of basis functions. In this chapter, we also exploit this scheme to perform the color correction in the second stage of color constancy.

The linear models (see Eq. (5-1) below) were developed to explain the relationship of the illuminant and the object surface reflections, and they also played a significant role in the study of color recovery. Given a spectral power distribution of the illuminant $E(\lambda)$ and the spectral reflectance function of a surface $R(\lambda)$, we have the definition of the color signal as:

$$C(\lambda) = E(\lambda)R(\lambda). \tag{5-1}$$

The researches of Maloney and Wandell [65], Buchsbaum [66], Sallstrom [67], and

Brill [68] assume a linear model for the illuminant and the reflectance, and it can be written as a sum of weighted basis functions:

$$E(\lambda) = \sum_{i=1}^{m} e_i E_i(\lambda) \quad \text{and} \tag{5-2}$$

$$R(\lambda) = \sum_{j=1}^{n} r_j R_j(\lambda), \tag{5-3}$$

where $E_i(\lambda)$ and $R_j(\lambda)$ are the sets of basis functions, $e_i$ and $r_j$ are coefficients of the basis functions that uniquely determine the illuminant and reflectance, and $m$ and $n$ are the numbers of basis functions corresponding to $E_i(\lambda)$ and $R_j(\lambda)$, respectively. This is called the *finite-dimensional linear model*. Substituting Eqs. (5-2) and (5-3) into Eq. (5-1), we have

$$C(\lambda) = \left[\sum_{i=1}^{m} e_i E_i(\lambda)\right]\left[\sum_{j=1}^{n} r_j R_j(\lambda)\right]$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{n} e_i r_j E_i(\lambda) R_j(\lambda). \tag{5-4}$$

Assume that there are $p$ distinct classes of receptors in a visual system. The response of the class-$k$ receptor is

$$q_k = \int_{\lambda} C(\lambda) S_k(\lambda) d\lambda, \tag{5-5}$$

where $S_k(\lambda)$ is the corresponding spectral sensitivity function. Transforming the above equations into the discrete case, we have

$$q_k = \sum_{\lambda_n} C(\lambda_n) S_k(\lambda_n)$$

$$= \sum_{\lambda_n}\left[\sum_{i=1}^{m}\sum_{j=1}^{n} e_i r_j E_i(\lambda_n) R_j(\lambda_n)\right] S_k(\lambda_n)$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{n} e_i r_j a_{ijk}, \quad k=1, 2, \ldots, p, \tag{5-6}$$

where $\lambda_n$ is sample wavelengths and

$$a_{ijk} = \sum_{\lambda_n} E_i(\lambda_n) R_j(\lambda_n) S_k(\lambda_n) .$$

There are several restrictions in the finite-dimensional linear model. It assumes that the object surfaces are nonfluorescent and geometric properties of surfaces are embedded in surface spectral reflectance. Another necessary constraint is imposed to the dimensionality of surface reflections: *n* must be more than and equal to the number of classes of receptors; otherwise the solution is underdetermined [13].

If the finite-dimensional linear model is used, the spectral power distribution of the illuminant has to be estimated so that colors can be recovered. Rewriting Eq. (5-6) as

$$\begin{aligned} q_k &= \sum_{\lambda_n} E(\lambda_n) \sum_{j=1}^{n} r_j R_j(\lambda_n) S_k(\lambda_n) \\ &= \sum_{j=1}^{n} r_j \sum_{\lambda_n} E(\lambda_n) R_j(\lambda_n) S_k(\lambda_n) . \end{aligned} \tag{5-7}$$

Assume that there are three classes of receptors *X*, *Y*, and *Z*. Use three basis functions to describe surface reflectance, and rewrite Eq. (5-7) by using matrix form, we have

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{bmatrix} \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix} \quad \text{or}$$

$$\mathbf{q} = \mathbf{M}\mathbf{r} , \tag{5-8}$$

where $M_{kj} = \sum_{\lambda_n} E(\lambda_n) R_j(\lambda_n) S_k(\lambda_n)$, $j, k$=1, 2, 3.

Assume that the column vector of tristimulus values of a surface with the original illuminant is $\mathbf{q}_{ill}$ and the transformation matrix is $\mathbf{M}_{ill}$. From Eq. (5-8), we obtain

$$\mathbf{r} = \mathbf{M}_{ill}^{-1}\mathbf{q}_{ill} . \tag{5-9}$$

Therefore, the tristimulus values with the *CIE* Standard Illuminant D65 can be calculated by substituting Eq. (5-9) into Eq. (5-8), and then we have

$$\mathbf{q}_{D65} = \mathbf{M}_{D65}\mathbf{M}_{ill}^{-1}\mathbf{q}_{ill} . \tag{5-10}$$

The above derivation indicates that if we can estimate the spectral power distribution, $E(\lambda_n)$, of the surrounding illuminant for a given color image, we can obtain the corresponding color image of the same scene under the standard illuminant D65 according to Eq. (5-10). The estimation of $E(\lambda_n)$ from a color image directly is a hard task. As mentioned in the last section, the existing estimation approaches usually rely on impractical assumptions. For example, the two popular schemes, Max-RGB [13], [18]-[20] and Gray-World [14]-[17] methods, assume the white-point value of the illuminant to be the maximum value of the RGB channels or the average of all colors in an image, respectively. These assumptions do not usually hold practically, and thus make these estimation schemes susceptible to dominant color in an image. Another existing estimation scheme is based on the color distribution of the image on the chromaticity map. Although this scheme avoids the unrealistic assumptions, it takes into account the chromaticity only and ignores the luminance. Its assumption that all the colors on the chromaticity map have the same luminance usually results in obvious estimation errors. In the following section, we shall propose a new illumination estimation scheme based on chromaticity histogram. This scheme is reliable and efficient since it considers both factors of chromaticity and luminance, and eludes the affect of dominant colors in an image.

## 5.3. A Neural-Network-based Illumination Estimation Scheme

As we knew from Chapter 1, an object might appear differently in color when it was illuminated with different light sources. For example, when a white object is illuminated with the light source of a lower color temperature, the reflection becomes reddish. On the other hand, the light source of the high temperature causes bluish in color to the same white object. It will be very obvious by observing the phenomena of color unbalance of one image in the chromaticity histogram. Therefore, we propose to estimate the spectral power distribution of the surrounding illuminant by detecting the central value of the distribution of image colors in the chromaticity histogram.

Figure 5-1 is the block diagram of the proposed estimation algorithm. It consists of seven steps, which are divided into two parts. The first part constructs the chromaticity histogram of the input image, determines the center point of the chromaticity histogram to represent the major color distribution of the image on the chromaticity map, and also eludes the affect of dominant color. The second part uses a pre-trained neural network with back-propagation (BP) learning algorithm for illumination estimation. The BP network is trained to model the nonlinear functional relationship between the center value of the chromaticity histogram and the spectral power distribution of the surrounding illuminant. In the rest of this section, we shall explain the details of each step in the block diagram of Figure 5-1.

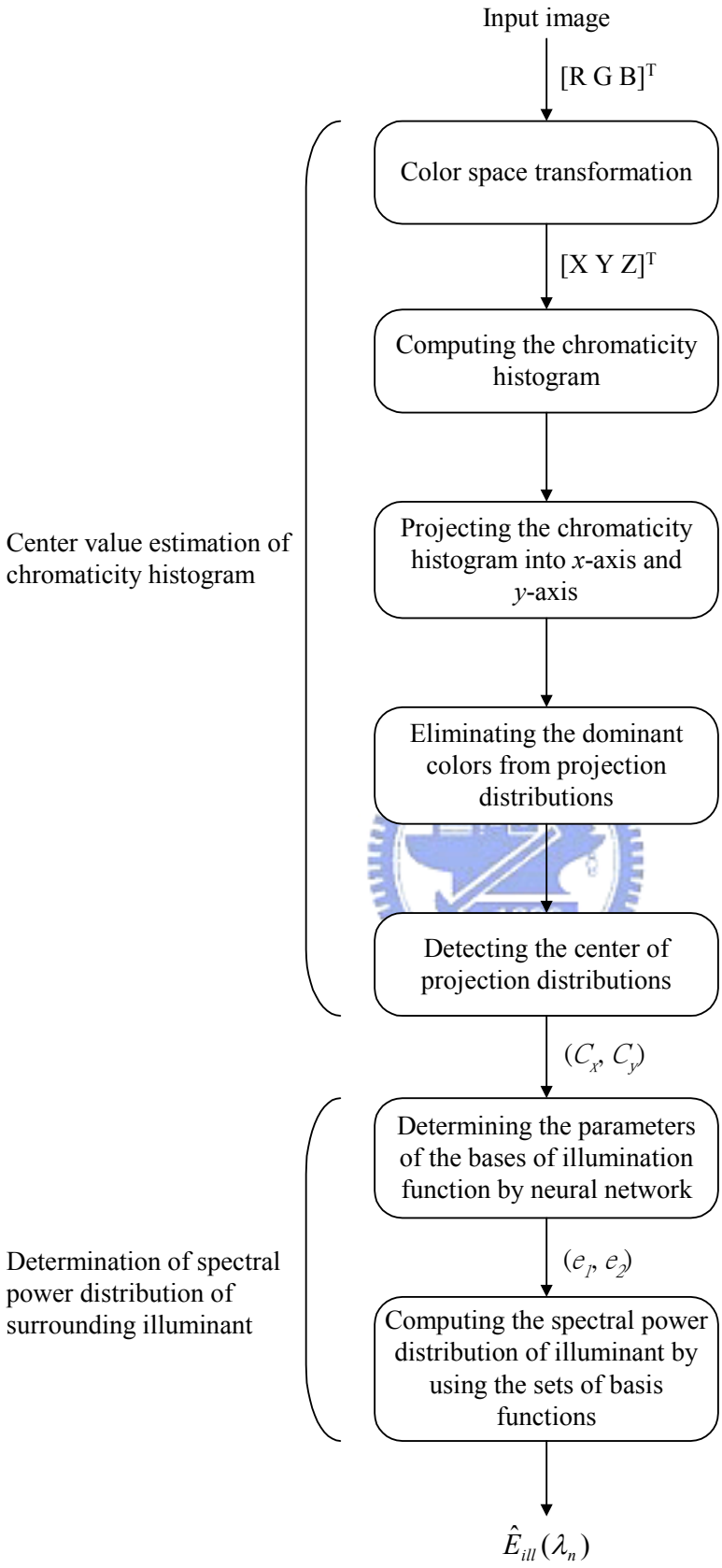



Figure 5-1. Diagram of the proposed illumination estimation scheme.

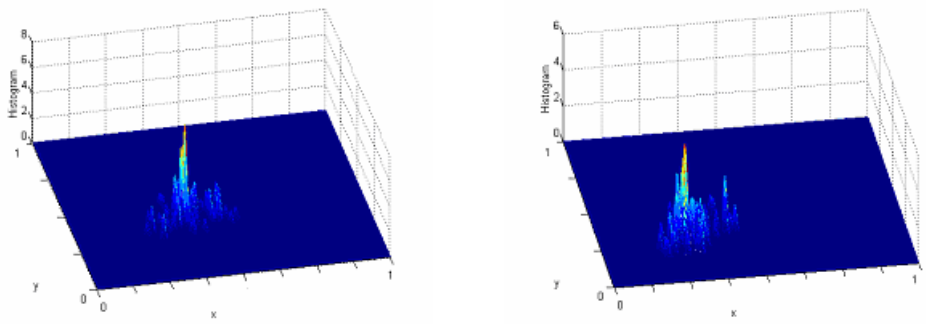
### 5.3.1. Center value estimation of the chromaticity histogram

*(a) Color space transformation and chromaticity histogram computation*

In our approach, we first convert and map the values of three color channels of an image to *CIE*'s *XYZ* color space and then get the ($x$, $y$) values of chromaticity through normalization according to the following equation:

$$x = \frac{X}{X+Y+Z} \quad \text{and} \quad y = \frac{Y}{X+Y+Z}, \tag{5-11}$$

where the values of $x$ and $y$ are in the range [0, 1]. A chromaticity histogram of a color image is a function of $x$ and $y$, denoted by $f(x, y)$. It is built by the accumulation of every pair of ($x$, $y$) values of chromaticity corresponding to all colors in the image. A plot of this function for all chromaticity provides a global description of the appearance of a color image. For example, Fig. 5-2 shows the chromaticity histograms of two images of the same scene with two different surrounding illuminants. The chromaticity histogram in Fig. 5-2(a) corresponds to the illuminant with a low color temperature, and that in Fig. 5-2(b) corresponds to the illuminant with a high color temperature. Thus the chromaticity histogram can reflect the information of white points of the surroundings faithfully. We shall next detect the central values of all color distributions of chromaticity histogram. The central values can represent how all the colors in an image are distributed in the chromaticity histogram.




(a) Low color temperature (b) High color temperature

Figure 5-2. Chromaticity histograms of two images of the same scene under two different illuminants with two different color temperatures.

*(b) Projecting the chromaticity histogram into x-axis and y-axis*

In order to detect the main color distribution region on the chromaticity histogram more efficiently and fast, we project the chromaticity histogram to *x*-axis and *y*-axis, respectively, to get two projection function curves. Its sample results are shown in Figs. 5-3(a) and (b), and can be expressed in the following mathematical forms:

$$g_{proj\_x}(x) = \frac{\max_{1 \le y \le D_y} f(x,y)}{\max(f(x,y))} \quad \text{and} \quad g_{proj\_y}(y) = \frac{\max_{1 \le x \le D_x} f(x,y)}{\max(f(x,y))}, \tag{5-12}$$

where $g_{proj\_x}(x)$ and $g_{proj\_y}(y)$ are the projection distribution functions in *x*-axis and *y*-axis of the chromaticity histogram whose values are between 0 and 1 through normalization by the maxima of projection functions, $D_x$ and $D_y$ are the dimensions of the chromaticity histogram in *x*-axis and *y*-axis, respectively, and $f(x, y)$ is the chromaticity histogram. By the projection operation, we can also decrease the calculation cost of chromaticity histogram from 2D to 1D spaces. This leads to a faster and more efficient detection scheme of the main color distribution of an image.

*(c) Eliminating the dominant colors from projected signals and detecting the center of the projection distributions*

For most cases, the assumption that there could be enough colors (full-color) in the image is reasonable. But in some situations, some colors would occupy most range of the image, which are called dominant colors. For examples, green and blue will be the dominant colors for the images gotten from a large grass field and from a whole blue sky or ocean, respectively. Obviously, if an image possesses a dominant color, we can easily locate the corresponding position of the dominant color in the chromaticity histogram, which appears to be an impulse. The same phenomena can be observed in the projection curves of chromaticity histogram. As Figs. 5-3(a) and (b) show, there

exists a dominant color at $x$=190 and $y$=100, where these two figures are relative to the projection of chromaticity histogram in $x$-axis and in $y$-axis, respectively.

The dominant color will bias the estimation of major colors in the image and causes obvious estimation error of surrounding illumination. Hence, we have to remove the effect of dominant colors from the projection curves of chromaticity histogram on $x$-axis and $y$-axis before we find the central values of color distribution. As shown in Figs. 5-3(a) and (b), the projection distributions of the dominant colors onto $x$-axis and $y$-axis both belong to the variations of high frequency. So we can pass the $g_{proj\_x}(x)$ and $g_{proj\_y}(y)$ functions in Eq. (5-12) through an ideal low pass filter to remove the effect of dominant colors. The low-pass filtered projection distributions of chromaticity histogram can more authentically represent the distribution of all colors of the image in the chromaticity histogram for central point location. The low-pass filtering process is demonstrated in Figs. 5-3(c)~(f). Figs. 5-3(c) and (d) are the results of DCT (discrete cosine transform) of Figs. 5-3(a) and (b), respectively. After we apply the ideal low-pass filtering on the curves in Figs. 5-3(c) and (d), and then take the inverse DCT (IDCT), we obtain the filtered projection curves of chromaticity histogram as shown in Figs. 5-3(e) and (f). Comparing Figs. 5-3(a), 5-3(b) and 5-3(e), 5-3(f), we observe that the effect of dominant colors have been removed effectively, and the central values can be estimated more correctly on the filtered curves.

The center ($C_x$, $C_y$) of the distribution of chromaticity histogram of the surrounding illuminant can be computed by the following formulas on the filtered projection distributions:

$$C_x = \arg\max_{1 \le x \le D_x} \left[ g_{proj\_x}(x) * h(x) \right] \text{ and}$$

$$C_y = \arg\max_{1 \le y \le D_y} \left[ g_{proj\_y}(y) * h(y) \right], \tag{5-13}$$

where * is the convolution operator and $h(\cdot)$ is the inverse DCT function of an ideal low pass filter $H(u)$ with cut-off frequency $u_c$ as described below:

$$H(u) = \begin{cases} 1, & |u| \le u_c \\ 0, & \text{otherwise} \end{cases}. \tag{5-14}$$

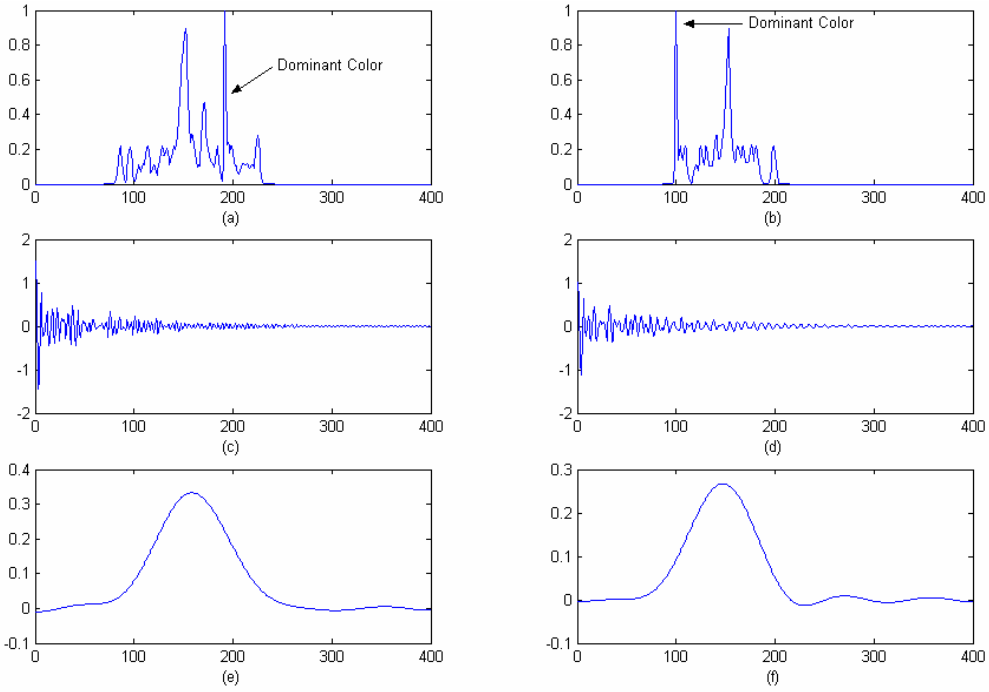



Figure 5-3. (a) and (b) are the projections of chromaticity histogram onto *x*-axis and *y*-axis, respectively, where the impulses correspond to the dominant colors in the image; (c) and (d) are the results of DCT of (a) and (b), respectively; (e) and (f) are the results of IDCT of ideal low-passed filtered (c) and (d), respectively, with cut-off frequency being 30.

### 5.3.2. Determination of spectral power distribution of illuminant

From the previous discussions, any light source $E(\lambda)$ can be expressed as the discrete form $E(\lambda_j)$ in a whole visible spectrum, where $j$=1, 2, …, $P$, and $P$ is the number of samples in the spectrum. Since $P$ is usually a big number, it is impossible that the spectral power distribution of illuminant can be fully reconstructed directly based on the detected center values of chromaticity histogram. From the finite-dimensional linear model described in Section 5.2, we know the spectral power distribution in the discrete time can be represented as the summation of weighted basis functions:

$$E(\lambda_j) = \sum_{i=1}^{m} e_i E_i(\lambda_j), \quad j=1, 2, \ldots, P, \tag{5-15}$$

where $E_i(\lambda_j)$ is the set of basis functions, $e_i$ is the coefficient of the basis functions that uniquely determine the illuminant, and $m$ is the number of the sets of basis functions corresponding to $E_i(\lambda_j)$. When the sets of basis functions $E_i(\lambda_j)$ are already known, it only remains to decide the coefficients of the basis functions in the estimation of the spectral power distribution of the illuminant. In that way we can obtain a complete distribution of light sources from the detection of the center values of chromaticity histogram.

The basis functions in Eq. (5-15) can be computed by the principal component analysis (PCA) technique. Assume that we have $N$ known sets of light sources of spectral power distributions. We can then arrange it as a $P \times N$ matrix:

$$\mathbf{E} = \begin{bmatrix} E^{(1)}(\lambda_1) & E^{(2)}(\lambda_1) & \cdots & E^{(N)}(\lambda_1) \\ E^{(1)}(\lambda_2) & E^{(2)}(\lambda_2) & \cdots & E^{(N)}(\lambda_2) \\ \vdots & \vdots & \ddots & \vdots \\ E^{(1)}(\lambda_P) & E^{(2)}(\lambda_P) & \cdots & E^{(N)}(\lambda_P) \end{bmatrix}. \tag{5-16}$$

Each column represents an independent light source, and it includes a vector of $P$

sampling values. All information in the matrix **E** can be denoted as the covariance matrix **S** defined by

$$\mathbf{S} = \frac{1}{N}\sum_{j=1}^{N}(\mathbf{E}^{(j)} - \boldsymbol{\mu})(\mathbf{E}^{(j)} - \boldsymbol{\mu})^{T}, \tag{5-17}$$

where $\boldsymbol{\mu}$ is a mean vector of all illuminants in **E**, and $N$ is the number of light sources included in the set. From the theory of PCA, the sets of basis functions $E_i(\lambda_j)$ can be obtained by the calculations of eigenvectors of the covariance matrix **S**.

In this chapter, we mainly take the standard *D* light source in different color temperatures according to the administration of *CIE* for acquiring the information required in the calculations of the basis functions. We take one light source every 200°K in the range of 4000°K~6000°K, every 400°K in the range of 6400°K~8000°K, every 500°K in the range of 8500°K~10000°K, every 1000°K in the range of 11000°K~15000°K, and 17000°K, 20000°K and 25000°K (as shown in the first column of Table 5-2). Hence we have 28 light sources totally. We can then use the method mentioned above to calculate the corresponding covariance matrix and its eigenvectors. The two eigenvectors corresponding to the two largest eigenvalues and the mean vector of all illuminants are taken to form the sets of basis functions shown in Table 5-1. Next, we convert all the standard *D* light sources in the matrix **E** to the corresponding coefficients of the basis functions through linear transformation, and the results are given in Table 5-2. The first column represents the light sources of different color temperatures. The second to fifth columns list the chromaticity values $(x_D, y_D)$ and $(u'_D, v'_D)$ relative to different color temperatures. The sixth and seventh columns in Table 5-2 list the corresponding coefficients of the basis functions in different color temperatures.

With the data available in Table 5-2, we can use them to train a neural network

with back-propagation (BP) learning algorithm for the coefficient estimation of basis functions of the illuminants with various color temperatures. The BP network structure is a fully-connected 2-10-2 network with ten hidden nodes. The BP network describes the nonlinear relationship between the center values of chromaticity histogram of an image and the coefficients of basis functions of the illuminant. The two input variables of the BP network is the center ($C_x$, $C_y$) of chromaticity histogram of the surrounding illuminant detected in Eq. (5-13). The two output variables of the BP network are the coefficients $e_1$ and $e_2$ of the two basis functions in Eq. (5-15). Hence, once we obtain the center values of the chromaticity histogram as described in Section 5.3.1, we can feed them into the pre-trained BP network and obtain the two coefficients of basis functions as the network outputs. Substituting these coefficients of basis functions of the illuminant into the following equation,

$$\hat{E}_{ill}(\lambda_j) = e_1 E_1(\lambda_j) + e_2 E_2(\lambda_j) + \mu(\lambda_j), \quad j=1, 2, \ldots, P, \tag{5-18}$$

we can obtain the estimated spectral power distribution of the surrounding illuminant $\hat{E}_{ill}(\lambda_j)$. Finally, the colors of the color-biased images can be corrected according to the illumination estimation $\hat{E}_{ill}(\lambda_j)$ by using Eqs. (5-8) and (5-10) given in Section 5.2.

Table 5-1. The sets of basis functions of the standard *D* illuminant.

| Wave-length (nm) | Mean | First | Second |
|---|---|---|---|
| 380 | 63.40 | 38.50 | 3.00 |
| 390 | 65.80 | 35.00 | 1.20 |
| 400 | 94.80 | 43.40 | -1.10 |
| 410 | 104.80 | 46.30 | -0.50 |
| 420 | 105.90 | 43.90 | -0.70 |
| 430 | 96.80 | 37.10 | -1.20 |
| 440 | 113.90 | 36.70 | -2.60 |
| 450 | 125.60 | 35.90 | -2.90 |
| 460 | 125.50 | 32.60 | -2.80 |
| 470 | 121.30 | 27.90 | -2.60 |
| 480 | 121.30 | 24.30 | -2.60 |
| 490 | 113.50 | 20.10 | -1.80 |
| 500 | 113.10 | 16.20 | -1.50 |
| 510 | 110.80 | 13.20 | -1.30 |
| 520 | 106.50 | 8.60 | -1.20 |
| 530 | 108.80 | 6.10 | -1.00 |
| 540 | 105.30 | 4.20 | -0.50 |
| 550 | 104.40 | 1.90 | -0.30 |
| 560 | 100.00 | 0.00 | 0.00 |
| 570 | 96.00 | -1.60 | 0.20 |
| 580 | 95.10 | -3.50 | 0.50 |
| 590 | 89.10 | -3.50 | 2.10 |
| 600 | 90.50 | -5.80 | 3.20 |
| 610 | 90.30 | -7.20 | 4.10 |
| 620 | 88.40 | -8.60 | 4.70 |
| 630 | 84.00 | -9.50 | 5.10 |
| 640 | 85.10 | -10.90 | 6.70 |
| 650 | 81.90 | -10.70 | 7.30 |
| 660 | 82.60 | -12.00 | 8.60 |
| 670 | 84.90 | -14.00 | 9.80 |
| 680 | 81.30 | -13.60 | 10.20 |
| 690 | 71.90 | -12.00 | 8.30 |
| 700 | 74.30 | -13.30 | 9.60 |
| 710 | 76.40 | -12.90 | 8.50 |
| 720 | 63.30 | -10.60 | 7.00 |
| 730 | 71.70 | -11.60 | 7.60 |
| 740 | 77.00 | -12.20 | 8.00 |
| 750 | 65.20 | -10.20 | 6.70 |
| 760 | 47.70 | -7.80 | 5.20 |
| 770 | 68.60 | -11.20 | 7.40 |
| 780 | 65.00 | -10.40 | 6.80 |

Table 5-2. Chromaticity co-ordinates $x_D$, $y_D$, chromaticity co-ordinates $u'_D$, $v'_D$, and coefficients $e_1$, $e_2$ of basis functions for the illuminants with different color temperatures.

| Color temp. (°K) | $x_D$ | $y_D$ | $u'_D$ | $v'_D$ | $e_1$ | $e_2$ |
|---|---|---|---|---|---|---|
| 4000 | 0.3823 | 0.3838 | 0.2236 | 0.5049 | -1.5046 | 2.8265 |
| 4200 | 0.3737 | 0.3786 | 0.2200 | 0.5014 | -1.4223 | 2.1271 |
| 4400 | 0.3658 | 0.3734 | 0.2168 | 0.4979 | -1.3329 | 1.5505 |
| 4600 | 0.3585 | 0.3684 | 0.2139 | 0.4946 | -1.2384 | 1.0759 |
| 4800 | 0.3519 | 0.3634 | 0.2114 | 0.4913 | -1.1403 | 0.6860 |
| 5000 | 0.3457 | 0.3587 | 0.2091 | 0.4882 | -1.0401 | 0.3667 |
| 5200 | 0.3401 | 0.3541 | 0.2071 | 0.4851 | -0.9387 | 0.1061 |
| 5400 | 0.3349 | 0.3497 | 0.2053 | 0.4822 | -0.8371 | -0.1055 |
| 5600 | 0.3302 | 0.3455 | 0.2036 | 0.4795 | -0.7358 | -0.2759 |
| 5800 | 0.3258 | 0.3416 | 0.2021 | 0.4768 | -0.6354 | -0.4119 |
| 6000 | 0.3217 | 0.3378 | 0.2007 | 0.4743 | -0.5363 | -0.5190 |
| 6400 | 0.3144 | 0.3308 | 0.1983 | 0.4695 | -0.3434 | -0.6639 |
| 6800 | 0.3082 | 0.3245 | 0.1963 | 0.4652 | -0.1586 | -0.7391 |
| 7200 | 0.3027 | 0.3189 | 0.1946 | 0.4613 | 0.0175 | -0.7646 |
| 7600 | 0.2980 | 0.3138 | 0.1932 | 0.4578 | 0.1844 | -0.7548 |
| 8000 | 0.2938 | 0.3092 | 0.1919 | 0.4545 | 0.3419 | -0.7198 |
| 8500 | 0.2892 | 0.3041 | 0.1906 | 0.4508 | 0.5262 | -0.6518 |
| 9000 | 0.2853 | 0.2996 | 0.1894 | 0.4475 | 0.6971 | -0.5665 |
| 9500 | 0.2818 | 0.2956 | 0.1884 | 0.4446 | 0.8556 | -0.4706 |
| 10000 | 0.2788 | 0.2920 | 0.1876 | 0.4419 | 1.0027 | -0.3689 |
| 11000 | 0.2737 | 0.2858 | 0.1861 | 0.4373 | 1.2663 | -0.1599 |
| 12000 | 0.2697 | 0.2808 | 0.1850 | 0.4335 | 1.4946 | 0.0450 |
| 13000 | 0.2664 | 0.2767 | 0.1841 | 0.4303 | 1.6934 | 0.2390 |
| 14000 | 0.2637 | 0.2732 | 0.1834 | 0.4275 | 1.8675 | 0.4194 |
| 15000 | 0.2614 | 0.2702 | 0.1828 | 0.4252 | 2.0209 | 0.5857 |
| 17000 | 0.2578 | 0.2655 | 0.1818 | 0.4214 | 2.2778 | 0.8780 |
| 20000 | 0.2539 | 0.2603 | 0.1809 | 0.4172 | 2.5714 | 1.2308 |
| 25000 | 0.2499 | 0.2548 | 0.1798 | 0.4126 | 2.9069 | 1.6551 |

## 5.4. Experiments and Results

In this section, two experiments are performed to assess the performance of the proposed approach. In the first experiment, we test the algorithm on synthetically generated images for the estimation of surrounding illumination of images of the same scene. In the second experiment, we test the algorithm on a number of real images captured from a still digital camera and show the color recovery results under the desired illuminant by using the finite-dimensional linear model of surface reflectance. All the experimental results are compared to those of the Max-RGB and Gray-World algorithms tested on the same images. These two algorithms are chosen for comparisons because they are easy to implement and are widely used to perform the function of auto-white-balance in digital cameras currently.

### 5.4.1. Illumination estimation of synthetic images

In this experiment, we used a simple feedforward neural network with two input and two output nodes to approximate the nonlinear mapping between the center of chromaticity histogram and the coefficients of basis functions of illuminants. The weights of this neural network were adjusted by using back-propagation (BP) learning rule. To train the neural network, we need a set of images under different illuminants whose spectral power distributions are known in advance. However, such spectral information is not easy to measure in natural surrounding environments. To solve this problem, we use synthetic images with known illumination distributions in this experiment to produce the required training samples for the BP network. We also produce a set of testing images using the same procedure for testing the performance of the trained network.

To synthesize the images of a scene under different illuminants, we need to have the spectral reflectance $R(\lambda)$ of different colors. In this experiment, we use the

colors defined in the color chart, AgfaIT8.7/2 [75], to form the colors in our synthetic images. Although AgfaIT8.7/2 consists of only 288 colors, they are enough for representing the colors of normal pictures. With the spectral reflectance $R(\lambda)$ of all 288 colors and the spectral power distribution of the standard D illuminants for different color temperatures, we can then apply the linear combination of *CIE*'s color matching functions to synthesize the images of a scene under different illuminants. As listed in Table 5-2 and explained in Section 5.3.2, we select 28 standard D illuminants with 28 different color temperatures falling in the range of 4000°K~25000°K as the light sources for image synthesis. On the other hand, to make the synthetic images having the scenes close to the real captured ones, we randomly chose 50 colors from the AgfaIT8.7/2 color chart to form the colors in the images for the training and testing of BP network. We then apply the algorithm proposed in Section 5.3.1 to find the center values of chromaticity histogram, ($C_x$, $C_y$), on the synthetic images. The ($C_x$, $C_y$) values of the synthetic images for each of the 28 color temperatures are collected and used as the inputs of the training network, and the coefficients of the basis functions of illuminants, ($e_1$, $e_2$), corresponding to each of the 28 color temperatures are used as the desired outputs of the network. The learning constant in the BP learning rule is set as 0.8, and the convergence criteria in the form of output RMS error defined below is set as 0.005.

To evaluate the learning accuracy of the BP network, we define the network output error, *Er*, as the RMS error on the *u*-*v* chromaticity space:

$$Er = \sqrt{(u'_D - \hat{u}_D)^2 + (v'_D - \hat{v}_D)^2}\ , \tag{5-19}$$

where $(u'_D, v'_D)$ and $(\hat{u}_D, \hat{v}_D)$ are the chromaticity co-ordinates of the real (synthetic) and estimated illuminants, respectively. Since the *u*-*v* color space is a uniform distributed color space, we use it as the basis for error calculation in order to

keep the same error scale. On the other hand, because the two output values of the BP network represent the coefficients of the two basis functions of illumination, we have to convert them into the values of *x*-*y* chromaticity for computing the RMS errors in Eq. (5-19). This conversion can be achieved by the following equation from the formula of standard D illuminants:

$$\hat{x}_D = \frac{-0.7467e_1 + 0.8496e_2 + 40.8192}{15.3775e_1 + 0.2149e_2 + 132.6327} \quad \text{and}$$

$$\hat{y}_D = \frac{0.7655e_1 + 0.3036e_2 + 42.5475}{15.3775e_1 + 0.2149e_2 + 132.6327}, \tag{5-20}$$

where $e_1$ and $e_2$ are parameters whose values are related to the chromaticity co-ordinates $(\hat{x}_D, \hat{y}_D)$. The values of $(\hat{x}_D, \hat{y}_D)$, $e_1$, and $e_2$ correlate color temperatures in the range $4000°K$ to $25000°K$. We can then use the *CIE*'s formula to transform $(\hat{x}_D, \hat{y}_D)$ into $(\hat{u}_D, \hat{v}_D)$, which can be used directly to calculate the RMS errors in Eq. (5-19).

Table 5-3 lists the estimation RMS errors of the proposed algorithm under different color temperatures in the second column. The estimation RMS errors of the Max-RGB and Gray-World algorithms on the same synthetic images are also listed in the third and fourth columns of Table 5-3 for comparison. Table 5-3 clearly shows the superiority of the proposed approach over the compared ones; the proposed algorithm always produces much smaller RMS errors than the other two. Although the Max-RGB and Gray-World algorithm produces smaller RMS errors in a few synthetic images, the RMS errors produced by the proposed algorithm on these images are also quite small. These exceptional cases are possible and reasonable, since we choose different combinations of colors randomly from the AgfaIT8.7/2 color chart to form the synthetic images, and the compared algorithms might have better performance on

the chosen colors. Overall, the proposed algorithm has better and stable estimation accuracy than the compared counterparts.

Table 5-3. The average RMS errors of illumination estimation of the proposed and compared algorithms on training synthetic images for an ideal camera (number of colors is 50).

| Color temp. (°K) | Error (RMS) | | |
|---|---|---|---|
| | Our Approach | Gray-World | Max-RGB |
| 4000 | 0.0002 | 0.0069 | 0.0035 |
| 4200 | 0.0024 | 0.0044 | 0.0138 |
| 4400 | 0.0011 | 0.0099 | 0.0068 |
| 4600 | 0.0014 | 0.0057 | 0.0032 |
| 4800 | 0.0020 | 0.0052 | 0.0046 |
| 5000 | 0.0023 | 0.0072 | 0.0016 |
| 5200 | 0.0025 | 0.0071 | 0.0015 |
| 5400 | 0.0012 | 0.0036 | 0.0014 |
| 5600 | 0.0010 | 0.0062 | 0.0048 |
| 5800 | 0.0020 | 0.0065 | 0.0033 |
| 6000 | 0.0030 | 0.0070 | 0.0050 |
| 6400 | 0.0023 | 0.0083 | 0.0032 |
| 6800 | 0.0022 | 0.0061 | 0.0062 |
| 7200 | 0.0013 | 0.0059 | 0.0067 |
| 7600 | 0.0010 | 0.0053 | 0.0021 |
| 8000 | 0.0025 | 0.0064 | 0.0066 |
| 8500 | 0.0048 | 0.0083 | 0.0033 |
| 9000 | 0.0011 | 0.0058 | 0.0037 |
| 9500 | 0.0030 | 0.0065 | 0.0034 |
| 10000 | 0.0014 | 0.0054 | 0.0036 |
| 11000 | 0.0026 | 0.0050 | 0.0153 |
| 12000 | 0.0025 | 0.0043 | 0.0058 |
| 13000 | 0.0018 | 0.0076 | 0.0020 |
| 14000 | 0.0024 | 0.0062 | 0.0131 |
| 15000 | 0.0032 | 0.0066 | 0.0020 |
| 17000 | 0.0019 | 0.0057 | 0.0049 |
| 20000 | 0.0022 | 0.0065 | 0.0035 |
| 25000 | 0.0036 | 0.0058 | 0.0017 |
| Avg. Error | **0.0021** | **0.0063** | **0.0049** |

We use the same synthesis procedure as mentioned in the above to produce another set of synthetic images for testing of the trained BP network. These testing sets of images are synthesized according to the illuminants different from those used in producing the training set of synthetic images. A total of 40 colors are randomly chosen from the 288 colors in the AgfaIT8.7/2 color chart to form a testing image. The proposed and compared approaches are then used to estimate the illumination of each testing image. The estimation RMS errors are listed in Table 5-4, where each value in the table is the average RMS error over 20 synthetic images of a specific illuminant. The results still indicate the superiority of the proposed scheme over the other two compared algorithms even in the cases of unlearned surrounding illuminants.

Finally, we want to find out how the number of colors in an image affects the estimation accuracy of the surrounding illumination. Under a specific illuminant, we randomly choose different numbers of colors from the AgfaIT8.7/2 color chart to form the synthetic images, starting from the number of five and increasing by five colors each time until 35 colors in total. Hence, we have seven synthetic images for a specific illuminant, with the numbers of colors in the seven images being 5, 10, 15, …, 50, respectively, all chosen randomly from the AgfaIT8.7/2 color chart. For a specific illuminant and a specific number of colors, a total of 25 images are synthesized. Again, we use the proposed and compared algorithms to estimate the illumination of these synthetic images. The estimation RMS errors are shown in Fig. 5-4, where each value in the figure is the average RMS error over 175 (25x7) synthetic images of a specific illuminant. The results indicate that the estimation error of each algorithm increases as the number of colors in an image decreases. This is reasonable since higher number of colors can provide more information of spectral power distribution of the surrounding

illuminant. However, in any case, the proposed algorithm still possesses the best performance.

Table 5-4. The average RMS errors of illumination estimation of the proposed and compared algorithms on testing synthetic images for an ideal camera (number of colors is 40).

| Color temp. (°K) | Error (RMS) | | |
|---|---|---|---|
| | Our Approach | Gray-World | Max-RGB |
| 4300 | 0.0025 | 0.0072 | 0.0053 |
| 5300 | 0.0022 | 0.0065 | 0.0043 |
| 5800 | 0.0040 | 0.0086 | 0.0103 |
| 6600 | 0.0039 | 0.0067 | 0.0038 |
| 7400 | 0.0053 | 0.0061 | 0.0109 |
| 8200 | 0.0025 | 0.0070 | 0.0048 |
| 10500 | 0.0029 | 0.0067 | 0.0035 |
| Avg. Error | **0.0033** | **0.0070** | **0.0061** |

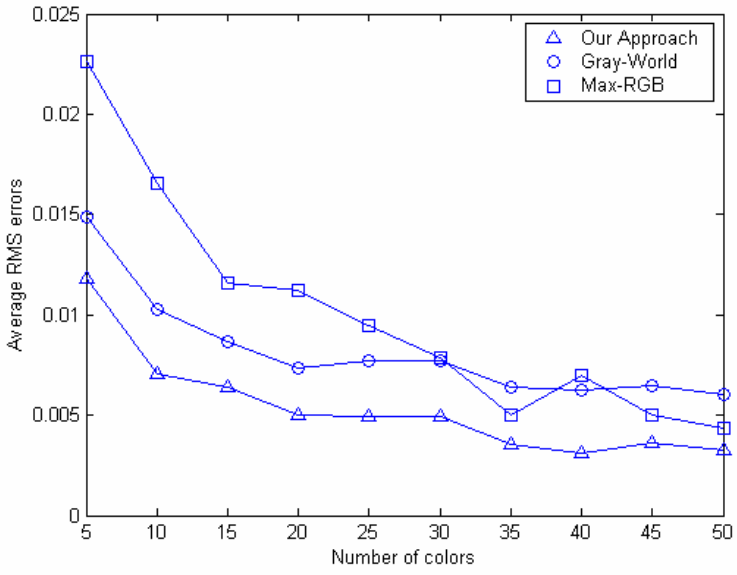



Figure 5-4. The average RMS errors of illumination estimation of the proposed and compared algorithms on testing synthetic images with respect to different number of colors in an image.

### 5.4.2. Illumination estimation of real images for color recovery

In the second experiment, we test the performance of the new algorithm on real images under natural illumination. These images are captured from Olympus DC460 still digital camera under different color temperatures. For proper control of different color temperatures, these images are captured inside a Color-Viewing (CV) box (made by GAIN Associates Inc., No. D1729). In the CV box, there are four different light sources, D、F、A and UV, producing four different color temperatures (D=6500°K, F=4200°K, A=2850°K). When capturing the images inside the CV box, we have to turn off the auto-white-balancing function in the digital camera such that the proposed approach can be fully exploited. To apply the proposed illumination estimation scheme to the captured images, we have to find the matrix that transforms the camera *R*, *G*, and *B* signals into *CIE* 1931 *X*, *Y*, and *Z* values at first. The $3\times3$ matrix is constructed so that the final *CIE*'s *XYZ* mean square errors between the reflection print and the camera image are minimized under the constraint that the neutral colors will be kept neutral after the color transformation. Let us take the Olympus DC-460 still camera as an example. Its chromaticity transformation between the phosphor primaries and the *CIE*'s *XYZ* primaries are listed in Table 5-5. Using the relations developed for camera calibration [76], [77], we have

$$V=\begin{bmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{bmatrix}=\begin{bmatrix} 0.6237 & 0.2640 & 0.2033 \\ 0.3313 & 0.6044 & 0.0892 \\ 0.0450 & 0.1316 & 0.7075 \end{bmatrix}, \tag{5-21}$$

$$U=\begin{bmatrix} r_1 & r_2 & r_3 \\ g_1 & g_2 & g_3 \\ b_1 & b_2 & b_3 \end{bmatrix}=\begin{bmatrix} 1.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 1.0 \end{bmatrix}, \tag{5-22}$$

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}=U^{-1}\begin{bmatrix} r_4 \\ g_4 \\ b_4 \end{bmatrix}=\begin{bmatrix} 1.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 1.0 \end{bmatrix}^{-1}\begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}=\begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}, \tag{5-23}$$

$$\begin{bmatrix}\beta_1\\ \beta_2\\ \beta_3\end{bmatrix} = V^{-1}\begin{bmatrix}x_4\\ y_4\\ z_4\end{bmatrix} = \begin{bmatrix}0.6237 & 0.2640 & 0.2033\\ 0.3313 & 0.6044 & 0.0892\\ 0.0450 & 0.1316 & 0.7075\end{bmatrix}^{-1}\begin{bmatrix}0.3501\\ 0.3693\\ 0.2806\end{bmatrix} = \begin{bmatrix}0.2904\\ 0.4073\\ 0.3024\end{bmatrix}, \quad (5\text{-}24)$$

and

$$D = \begin{bmatrix}\beta_1/\alpha_1 & 0 & 0.\\ 0 & \beta_2/\alpha_2 & 0\\ 0 & 0 & \beta_3/\alpha_3\end{bmatrix}. \quad (5\text{-}25)$$

The transformation matrix $A = cVDU^{-1}$ is determined up to a constant factor $c$. The convention is to choose the constant $c$ so that $y$ is equal to 1 when $r$, $g$, and $b$ are set to 1. By doing so, we obtain the transformation matrix $A$ as

$$A = \begin{bmatrix}0.5807 & 0.2705 & 0.1651\\ 0.3084 & 0.6192 & 0.0724\\ 0.0419 & 0.1348 & 0.5744\end{bmatrix}. \quad (5\text{-}26)$$

Therefore, we have

$$\begin{bmatrix}x\\ y\\ z\end{bmatrix} = A\begin{bmatrix}r\\ g\\ b\end{bmatrix} \text{ or } \begin{bmatrix}r\\ g\\ b\end{bmatrix} = A^{-1}\begin{bmatrix}x\\ y\\ z\end{bmatrix}. \quad (5\text{-}27)$$

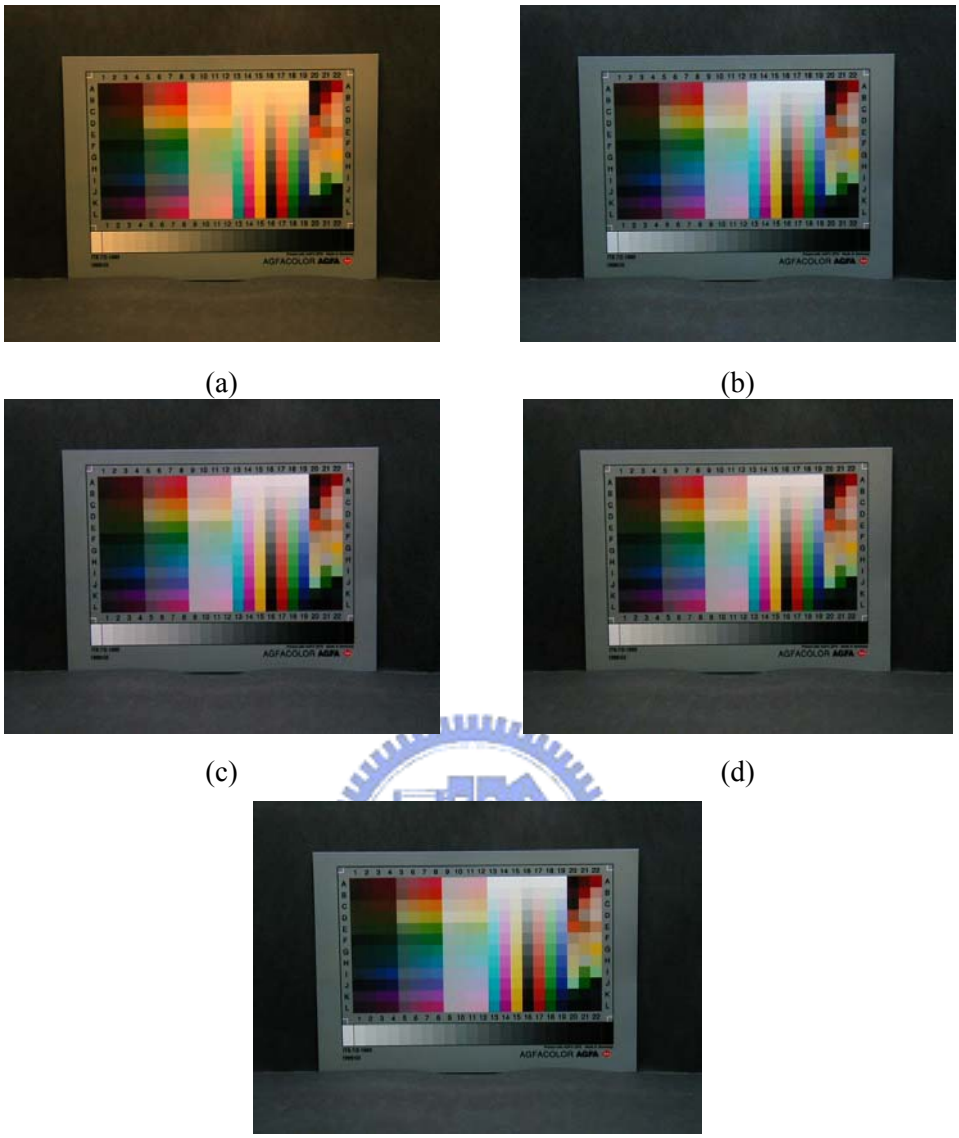
Table 5-5. Camera primaries (r, g, and b are normalized by the maximum stimulus).

| Stimulus | *r* | *g* | *b* | *x* | *y* | *z* |
|---|---|---|---|---|---|---|
| Red phosphor | 1.0 | 0.0 | 0.0 | 0.6237 | 0.3313 | 0.0450 |
| Green phosphor | 0.0 | 1.0 | 0.0 | 0.2640 | 0.6044 | 0.1316 |
| Blue phosphor | 0.0 | 0.0 | 1.0 | 0.2033 | 0.0892 | 0.7075 |
| White | 1/3 | 1/3 | 1/3 | 0.3501 | 0.3693 | 0.2806 |

Table 5-6. The average RMS errors of AgfaIT8.7/2 photographic print for the image rerendered based on different illuminant estimation algorithms.

| Average RMS Error | Our Scheme | Gray-World | Max-RGB |
|---|---|---|---|
| Error (RMS) | 5.53 | 8.39 | 10.38 |

Next, we can obtain the raw data of *CIE*'s *XYZ* of images by using the above transformation matrix *A*. The chromaticity histogram is computed at the same time. Finally, we use our approach to estimate the surrounding illumination. By substituting the estimation result into Eq. (5-10), we can obtain the color-corrected images for the desired surrounding illuminant. Figs. 5-5~5-7 show the exemplar images in this experiment. For each scene, we show four images: (a) the raw image captured by the camera with different illuminants, (b) the rendered image using the illuminant estimated by the proposed new algorithm, (c) the rendered image using the illuminant estimated by the Gray-World algorithm, and (d) the rendered image using the illuminant estimated by the Max-RGB algorithm. Besides, in Fig. 5-5, we add a raw image captured by the camera with D65 illuminants. Fig. 5-5(a) shows the AgfaIT8.7/2 color chart captured under the F illumination, and Figs. 5-5(b)~(d) are the three images rendered for illuminant D65 by the proposed, Gray-World, and Max-RGB algorithms, respectively. It is obvious that the rendered image based on the surrounding illuminant estimation of the proposed algorithm is very close to the image captured under the D65 illumination directly. In contrast, the performance of the Gray-World and Max-RGB algorithms is worse. The quantitative performance measurement is listed in Table 5-6. Table 5-6 shows the average RMS errors of colors in all the charts between the images captured under the D65 illumination and those rendered by different illumination estimation schemes. Fig. 5-6 shows the results on the image of the Macbeth ColorChecker captured under the D50 and D75 illuminants, respectively. Fig. 5-7 is the photo scene of the campus of our university. There is a dominant color caused by the grassplot on the photo. The results show that the proposed algorithm is less sensitive to the dominant color than the other two compared algorithms.

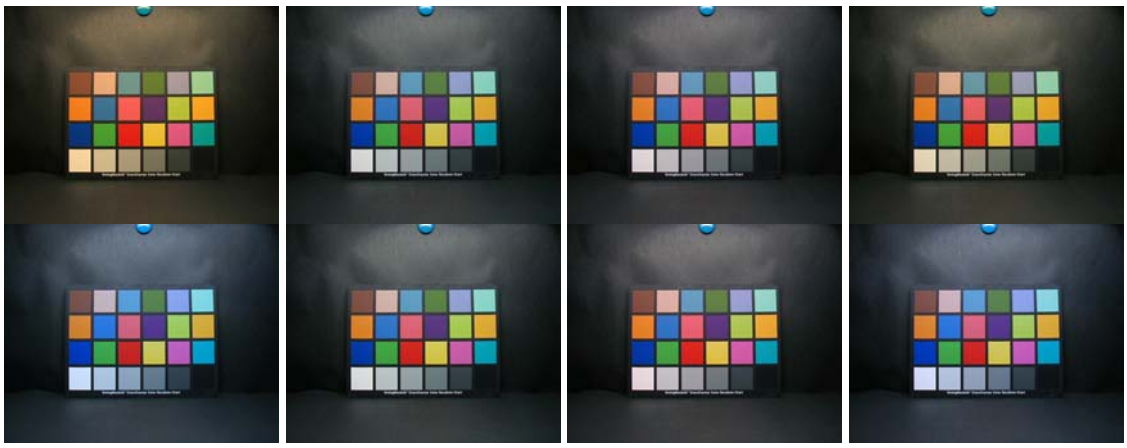


(a) (b)

(c) (d)

(e)

Figure 5-5. (a) Raw camera image captured under the F illumination. (b)-(d) Color-corrected images for the D65 illumination by the proposed algorithm, the Gray-World algorithm, and the Max-RGB algorithm, respectively. (e) Raw camera image captured under the D65 illumination.



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

Figure 5-6. (a) Raw camera image captured under D50 (first row) and D75 (second row) illuminations. (b)-(d) Color-corrected images for the D65 illumination by the proposed algorithm, the Gray-World algorithm, and the Max-RGB algorithm, respectively.

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

Figure 5-7. (a) Raw camera image captured under daylight and evening illuminations with a dominant color of grassplot. (b)-(d) Color-corrected images for the D65.illumination by the proposed algorithm, the Gray-World algorithm, and the Max-RGB algorithm, respectively.

## 5.5. Concluding Remarks

In this chapter, a new approach to surrounding illumination estimation of an image was proposed. The proposed algorithm estimated the illuminant based on chromaticity histogram of the image. And a neural network with back-propagation (BP) learning algorithm was used to estimate the spectral power distribution of the illuminant according to the center values of the chromaticity histogram. The proposed algorithm also eliminated the interference of dominant color to illumination estimation through low-pass filtering of the chromaticity histogram. The illumination estimation based on the chromaticity histogram can avoid unrealistic assumptions on the color images and provide high efficient and robust estimation. On the other hand, the use of the BP network provides good interpolation over a small number of different illuminants and gives highly estimation accuracy. Two experiments were performed to evaluate the performance of the proposed algorithm. In the first experiment, the proposed algorithm was used to estimate the illumination of synthetic images, and the estimation RMS errors were calculated. In the second experiments, the proposed algorithm was used to estimate the spectral power distributions of the illuminants, and then the colors of the image were corrected based on the finite-dimensional linear model of surface reflectance. Two popular existing illumination estimation algorithms, the Gray-World and Max-RGB algorithms, were also applied to the same images in these two experiments. Performance comparisons have demonstrated the superiority of the proposed algorithm both in estimation accuracy and robustness for color constancy.

# 6. Conclusion and Perspectives

In this thesis, we proposed two new reflectance models for 3D surface reconstruction. First, a novel 3D image reconstruction model was proposed. This method considers the components of both diffusion and specular reflection in the reflectance model. We used two neural networks with symmetric structure to estimate these two reflection models separately and combined them with an adaptive ratio for each point on the object surface. The proposed network estimates the point-wise adaptive combination ratio of the diffusion and specular intensities such that the different reflecting properties of each point on the object surface can help to achieve better performance of surface reconstruction. The proposed symmetric neural network structure with adaptive learning procedure does not need any special parameter setting and the smoothing conditions. It is also easier to achieve the convergence condition and to make the system stable. The critical parameters, such as the light source and the viewing direction and so on, are also obtained from the learning process of the neural network. The obtained normal vectors of the surface can then be applied to 3D surface reconstruction by enforcing integrability approach.

Secondly, we further proposed another new nonlinear reflection model consisting of the diffusion and specular components. We do not need to separate the two components in the proposed nonlinear reflection model. Using the unsupervised non-linear ICA network for solving photometric stereo problems does not need any desired outputs and the smoothing conditions. It is easier to achieve the convergence

condition and make the system stable.

For 3D surface reconstruction, several conclusions are listed below. (a) When we estimate the surface shape, the success of the reflectance model depends on two major components, including the diffusion and specular components. (b) In our methods, we do not know the locations of light sources for solving the photometric stereo problems. (c) The proposed symmetric neural network structure and the unsupervised post-nonlinear ICA network do not need any special parameter setting and the smoothing conditions.

On the other hand, we proposed a new approach in the surrounding illumination estimation of an image for color reconstruction. The proposed algorithm estimated the illuminant based on chromaticity histogram of the image. And a neural network back-propagation (BP) learning algorithm is used to estimate the spectral power distribution of the illuminant according to the center values of the chromaticity histogram. The proposed algorithm also eliminated the interference of the dominant colors and illumination estimation through low-pass filtering of the chromaticity histogram. The illumination estimation based on the chromaticity histogram can avoid unrealistic assumptions on the color images and provide the highly efficient and robust estimation. Compared with the methods based on neural networks which we proposed before, the size of their architectures is huge. Those methods need many connecting parameters and are not easy to implement on hardware. The size of our architecture is much smaller; hence our methods can estimate the relative parameters of outside sources effectively and rapidly. Thus, we could reconstruct the color of 3D objects by using these parameters. The comparisons in performance have demonstrated the superiority of the proposed algorithm both in estimation accuracy and robustness for color constancy.

# Bibliography


[1] K. P. Horn, “Shape from Shading: A Smooth Opaque Object from One View,” Ph. D. dissertation, Massachusetts Inst. Technol., 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 and Pattern Recognition*, pp. 377-384, 1994.

[5] M. J. Brooks and K. P. Horn, “Shape and source from shading,” in *Proc. International Joint Conference on Artificial Intelligence*, pp. 932-936, 1985.

[6] Q. Zheng and R. Chellappa, “Estimation of illuminant direction, albedo, and shape from shading,” *IEEE Trans. Pattern Anal. And Machine Intelligence.* 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. And Machine Intelligence.* Vol. 15, No. 8, pp. 815-522, 1993.

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

[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 Second International Conf. on Computer Vision*, pp. 404-413, 1988.

[11] R. J. Woodham, “Photometric Method for Determining Surface Orientation from Multiple Images,” *Journal of Optical Engineering*, Vol. 19, No. 1, 1980.

[12] K. Barnard, V. Cardei, and B. Funt, “A comparison of computational color

constancy algorithms - Part I: Methodology and experiments with synthesized data", IEEE Trans. On Image Processing, Vol. 11, No. 9, pp. 972-984, Sept. 2002.

[13] F. H. Cheng, W. H. Hsu, and T. W. Chen, "Recovering colors in an Image with chromatic illuminant," *IEEE Trans. on Image Processing*, vol. 7, No. 11, pp. 1524-1533, Nov. 1998.

[14] R. Gershon, A. D. Jepson, and J. K. Tsotsos, "From [R, G, B] to surface reflectance: Computing color constant descriptors in images," *Perception*, pp. 755-758, 1988.

[15] H. Helson, "Fundamental problems in color vision. I. The principle governing changes in hue, saturation, and lightness of nonselective samples in chromatic illumination," *J. Exper. Psychol.*, vol. 23, pp. 439-476, 1938.

[16] E. H. Land, "Recent advances in retinex theory," *Vision Research*, vol. 26, pp. 7-21, 1986.

[17] M. D. Zmura and P. Lennie, "Mechanisms of color constancy," *J. Opt. Soc. Amer. A*, vol. 3, pp. 1662-1672, 1986.

[18] E. H. Land and J. J. McCann, "Lightness and retinex theory," *J. Opt. Soc. Amer. A* , vol. 61, pp. 1-11, 1971.

[19] E. H. Land, "The retinex theory of color vision," *Scientific American*, pp. 108-129, 1977.

[20] J. J. McCann, S. P. McKee, and T. H. Taylor, "Quantitative studies in retinex theory," *Vision Research*, vol. 16, pp. 445-458, 1976.

[21] S. Usui, S. Nakauchi, and Y. Miyamoto, "A decorrelating neural network for color constancy," *Proceedings of the Annual International Conference of the IEEE, Engineering in Medicine and Biology Society*, vol. 14, pp. 1030-1031, 1992.

[22] B. V. Funt, V. C. Cardei, and J. J. Barnard, "Method of estimating chromaticity of illumination using neural networks," *United States Patent*, Patent Number: 5907629, May 1999.

[23] G. D. Finlayson, P. M. Hubel, and Hordley, "Color by correlation," *Proceedings of IS&T/SID Fifth Color Imaging Conference*: *Color Science, systems, and Applications*, pp. 6-11, Scottsdale, AZ, 1997.

[24] P. M. Hubel, J. Holm, and G. D. Finlayson, "Illuminant estimation and color correction," *Colour Imaging in Multimedia*, Lindsay MacDonald.,(in press)

Wiley, 1999.

[25] P. M. Hubel and G. D. Finlayson, "White point determination using correlation matrix memory," *United States Patent*, Patent Number: 6038339, Mar. 2000.

[26] K. M. Lee and C. C. J. Kuo, "Shape Reconstruction from Photometric Stereo," *Journal of Optical Society of America: A*, Vol. 10, No. 5, 1993.

[27] C. Cho and H. Minanitani, "A New Photometric Method Using 3 Point Light Sources," *IEICE Trans. Inf. & Syst*. V.E76-D, No. 8, pp. 898-904, August 1993.

[28] Y. Iwahori, R. Woodham, and A. Bagheri, "Principal Components Analysis and Neural Network Implementation of Photometric Stereo," *Proceedings of the Workshop on Physics-Based Modeling in Computer Vision*, pp. 117-125, June 1995.

[29] G. Kay and T. Caelli, "Estimating the Parameters of an Illumination Model Using Photometric Stereo," *Graphical Models and Image Processing*, Vol. 57, No. 5, pp. 365-388, 1995.

[30] G. McGunnigle, "The Classification of Textured Surfaces under Varying Illuminant Direction," Ph.D. Thesis, Department of Computing and Electrical Engineering, Heriot-Watt University, Edinburgh, 1998.

[31] S. K. Nayar, K. Ikeuchi, and T. Kanade, "Determining Shape and Reflectance of Hybrid Surfaces by Photometric Sampling," *IEEE Transactions on Robotics and Automation*, Vol. 6, No. 4, pp. 418-431, 1990.

[32] E. Angelopoulou and J. P. Williams, "Photometric Surface Analysis in a Triluminal Environment," *Proceedings of IEEE International Conference on Computer Vision*, 1999.

[33] P. N. Belhumeur, D. J. Kriegman, and A. L. Yuille, "The Bas-Belief Ambiguity," *CVPR*, pp. 1060-1066, 1997.

[34] 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. on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 6, pp. 643-660, June 2001.

[35] S. Georghiades, "Incorporating the Torrance and Sparrow Model of Reflectance in Uncalibrated Photometric Stereo," *Proceedings of the Ninth IEEE International Conference on Computer Vision*, 2003.

[36] K. Ikeuchi, "Determining Surface Orientations of Specular Surfaces by Using

the Photometric Stereo Method," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 3, pp. 661-669, 1981.

[37] E. N. Coleman and R. Jain, "Obtaining 3Dimensional Shape of Textured and Specular Surfaces Using Four-Source Photometry," *Processing of Computer Vision, Graphics, and Image*, Vol. 18, pp. 309-328, 1982.

[38] H. D. Tagare and R. J. P. deFigueiredo, "Simulataeous Estimation of Shape and Reflectance Maps from Photometric Stereo," *IEEE International Conference on Computer Vision*, pp. 340-343, 1990.

[39] F. Solomon and K. Ikeuchi, "Extracting the Shape and Roughness of Specular Lobe Objects Using Four Light Photometric Stereo," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 466-471, 1992.

[40] H. Rushmeier, G. Taubin, and A. Gueziec, "Applying Shape from Lighting Variation to Bump Map Capture," *In Eurographics Rendering Techniques'97*, pp. 35-44, St. Etienne, France, June 1997.

[41] O. Drbohlav and A. Leonardis, "Detecting Shadows and Specularities by Moving Light," *Proceeding of Computer Vision Winter Workshop*, Ljubljana, Slovenia, pp. 60-74, 1998.

[42] O. Drbohlav and R. Sara, "Specularities Reduce Ambiguity of Uncalibrated Photometric Stereo," *Proceedings of the 7th European Conference on Computer Vision*, Copenhagen, Denmark, 2002.

[43] K. Hayakawa, "Photometric Stereo under a Light Source with Arbitrary Motion," *Journal of the Optical Society of America: A*, Vol. 11, No. 11, 1994.

[44] H. Wang, Y. Wang, and H. Wei, "Face Representation and Reconstruction under Different Illumination Conditions," *Proceedings of the 7th International Conference on Information Visualization*, 2003.

[45] G. Healey and T. O. Binford, "Local shape from specularity," *Computer Vision, Graphics, Image Processing*, Vol. 42, pp. 62-86, 1988.

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

[47] B. T. Phong, "Illumination for computer generated pictures," *Communications of the ACM*, Vol.18, No.6, pp. 311-317, Juni 1975.

[48] G. Q. Wei and G. Hirzinger, "Learning shape from shading by a multilayer network," *IEEE Trans. on Neural Networks*, Vol. 17, pp. 985-995, 1996.

[49] S.Y. Cho and T. W. S. Chow, "Shape recovery from shading by a new

neural-based reflectance model," *IEEE Trans. on Neural Networks*, Vol. 10, pp.1536-1541, 1999.

[50] S. Y. Cho and T. W. S. Chow, "Neural computation approach for developing a 3D shape reconstruction model," *IEEE Trans. on Neural Networks*, Vol. 12, No. 5, September 2001.

[51] R. T. Frankot and R. Chellappa, "A method for enforcing integrability in shape from shading algorithms", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 10, No. 4, pp. 439-451, July 1988.

[52] P. S. Tsai and M. shah, "Shape from shading with variable albedo," *Optical Engineering*, Vol. 37, No. 4, April 1998.

[53] C. H. Lee and A. Rosenfeld, "Albedo estimation for scene segmentation," *Pattern Recognition Lett.* 1. 155-160, 1983.

[54] 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.

[55] Online available = http://cvc.yale.edu/projects/yalefacesB/yalefacesB.html.

[56] 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. on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 6, pp. 643-660, June 2001.

[57] S. Y. Cho and T. W. S. Chow, "Learning parametric specular reflectance model by radial basis function network," *IEEE Trans. on Neural Networks*, Vol. 11, No. 6, pp. 1498-1503, November 2000.

[58] University of Notre Dame Biometrics Database Distribution, Online available = http://www.nd.edu/~cvrl/UNDBiometricsDatabase.html.

[59] A. Taleb and C. Jutten, "Nonlinear Source Separation: the Pose-Nonlinear Mixtures," *in Proc. ESANN*, Bruges, Belgium, pp. 279-284, Apr. 1997.

[60] A. Taleb and C. Jutten, "Source Separation in Post-Nonlinear Mixtures," *IEEE Trans. on Signal Processing*, Vol. 47, No. 10, October 1999.

[61] J. Bell and T. J. Sejnowski, "An Information-Maximization Approach to Blind Separation and Blind Deconvolution," *Neural Computation*, Vol. 7, pp. 1129-1159, 1995.

[62] R. T. Frankot and R. Chellappa, "A Method for Enforcing Integrability in Shape From Shading Algorithms," *IEEE Trans. on Pattern Analysis and*

*Machine Intelligence*, Vol. 10, No. 4, pp. 439-451, July 1988.

[63] S. I. Amari, "Natural Gradient Works Efficiently in Learning," *Neural Computation*, Vol. 10, pp. 251-276, 1998.

[64] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent component analysis*, John Wiley & Sons, Inc., 2001.

[65] L. T. Maloney and B. A. Wandell, "Color constancy: A method for recovering surface spectral reflectance," *J. Opt. Soc. Amer. A*, vol. 3, pp. 29-33, 1986.

[66] G. Buchsbaum, "A spatial processor model for object color perception," *J. Franklin Inst*., vol. 310, pp. 1-26, 1980.

[67] P. Sallstrom, "Color and physics: Some remarks concerning the physical aspects of human color vision," *Perception*, pp.73-09, Stockholm Inst. Phys., Univ. Stockholm, Sweden, 1973.

[68] M. H. Brill, "A device performing illuminant-invariant assessment of chromatic relations," *J. Theor. Biol.*, vol. 71, pp. 473-478, 1978.

[69] B. A. Wandell, "Color constancy and Natural Image," *Phycica. Scripta.,* vol. 39, pp. 187-192, 1989.

[70] J. Cohen, "Dependency of the spectral reflectance curves of munsell color chip," *Psychon. Scripta.*, vol 1, pp. 369-370, Aug. 1964.

[71] D. B. Judd, D. L. MacAdam, and G. Wyszecki, "Spectral distribution of typical daylight as a function of correlated color temperature," *J. Opt. Soc. Amer*., vol. 54, pp. 1031-1040, Aug. 1964.

[72] B. A. Wandell, "Color rendering of color camera data," *Color Research and Application*, vol. 11, pp. 30-33, 1986.

[73] L. T. Maloney, "*A Computational Approach to Color Constancy*," Ph.D. dissertation, Appl. Psychol. Lab., Stanford Univ., Stanford, CA., 1985.

[74] J. Ho, B. V. Funt, and M. S. Drew, "Separating a color signal into illuminant and surface reflectance components: Theory and Application," *IEEE Trans. on Pattern Recognition and Machine Intelligence*, vol. 12, pp. 966-977, 1990.

[75] AgfaIT8.7/2 Color chart, http://www.agfa.com.

[76] R. W. G. Hunt, "Measuring colour," Ellis Horwood series in applied science and industrial technology, Halsted Press, New York, NY, 1989.

[77] H. R. Kang, "Color technology for electronic imaging devices," SPIE- The international Society for Optical Engineering, Washington, 1997.