

國立交通大學

電機與控制工程學系

碩士論文

應用於立體顯示器之二維轉立體靜態影像技術

A Novel 2D to 3D Image Conversion Technique
Applied on Stereo Display

研究生：林峻永

指導教授：林進燈 博士

周志成 博士

中華民國九十四年七月

應用於立體顯示器之二維轉立體靜態影像技術

A Novel 2D to 3D Image Conversion Technique
Applied on Stereo Display

研究生：林峻永
指導教授：林進燈

Student : Chun-Yeon Lin
Advisor : Dr. Chin-Teng Lin
Dr. Chi-Cheng Jou

國立交通大學

電機與控制工程學系



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 Master
in
Electrical and Control Engineering
July 2005
Hsinchu, Taiwan, Republic of China

中華民國九十四年七月

應用於立體顯示器之二維轉立體靜態影像技術

研究生：林峻永

指導教授：林進燈 博士

周志成 博士

國立交通大學電機與控制工程研究所

摘要

本論文提出一個將 2D 影像全自動轉換成 3D 立體影像的技術。即將一張 2D 影像轉換成可供左、右眼分別觀測的兩張影像，並應用於 3D 立體顯示器。此轉換技術可以分為兩部分，第一部分將 2D 影像編碼出具有深度資訊之深度圖，第二部分為將此深度圖解碼還原成模擬左、右雙眼觀測之立體影像。編碼部分又可以分為兩個步驟，第一步驟為基於 3D 立體影像特性的影像分割技術。利用 3D 立體影像的特性，經由去除反光效果的 SSR(Single Scale Retinex)處理。同時也利用色調、飽和度、亮度為特徵，及使用傳統分類法 FCM(Fuzzy C-means clustering method)，再經由 Size Filter 篩選區域大小。第二步驟為深度擷取，依據影像的深度線索及物件之間深度規則，推測出此物件所具有的單一或漸進深度。如此在 3D 顯示器能得到更真實且舒適之視覺效果。第二部分解碼亦分為兩個步驟。第一步驟為深度圖轉換成供左、右眼觀測之兩張影像，在此提出兩種方法，第一種方法為線性平移法，此為簡化之方法，能有效加快轉換的速度。第二種方法為模擬人眼視覺之平移法，此方法為依據人眼的特性為基礎所開發之技術。但由於影像經由平移運算後會遺失部分像素的資訊，故第二部分為”破洞補償”，依照破洞的特性，選擇平均或鏡射的方法填補破洞。經過編碼和解碼的過程，可以將單一 2D 影像自動產生可用於立體顯示器之立體影像。並建立計算深度正確性的方法，最後以人之實際觀感和科學化數據與 SANYO 公司提出的 CID(Computed Image Depth)演算法作為比較之依據。

A Novel 2D to 3D Image Conversion Technique Applied on Stereo Display

Student: Chun-Yeon Lin

Advisor: Dr. Chin-Teng Lin

Dr. Chi-Cheng Jou

Department of Electrical and Control Engineering
National Chiao-Tung University

Abstract

In this thesis, a novel automatic 2D to 3D image conversion technique is proposed. It means to convert a 2D image into the left and right eye images and apply on the 3D stereo monitor. We can divide the conversion technique into two parts. The first part is to encode the 2D image into the depth map which includes the depth information, and the second part is to decode the depth map into the stereo images which simulate the left and right eyes. The part of encoder can be divided into two steps. The first step is the image segmentation technique based on the property of 3D stereo images. According to the property of 3D stereo image, we use SSR (Single Scale Retinex) to decrease the light reflection effect. We also use hue, saturation, and intensity to be our features and chose the traditional clustering method FCM (Fuzzy C-means clustering). And we use the size filter to filter the regions with size .The second step is depth extraction. According to the depth cues and the depth rules between each object, we estimate the objects belong to the fixed or gradual depths. We can get more real and comfortable effects on the 3D monitor. The second part, decoder can be divided into two steps. The first step is to convert the depth map into the left and right eye images. We proposed two methods. The first method is linear shift algorithm, and this is the simplified method which can increase the speed of conversion effectively. The second method is binocular vision shift algorithm, and this method is based on the property of the human vision. After shifting, some information of pixels will be lost. So the second part is “interpolating holes method”. According to the property of holes, we choose the method of average or mirror to interpolate. After the process of encoder and decoder, we can automatically generate the stereo images displayed in 3D monitor from a single 2D image and establish the method of calculating the accuracy of the depths. Finally, we compare the results of our proposed method with SANYO’s method “CID” (Computed Image Depth) algorithm by the human perception and scientific data.

誌謝

在電控所的這兩年求學生涯中，首先感謝我的指導教授林進燈及周志成博士不僅在學業方面，給予悉心指導做研究的方法，更在為人處世及求學態度上給予啟蒙及規範，而使得本論文方能順利完成。

其次要感謝曾給我相當多協助的鶴章、文昌、群立、宇文、剛維及世安學長，以及陪伴我一同走過研究生活的宗恆、盈彰、弘昕、Linda、裕程及瑞正等同學，還有有德、庭瑋、家昇及峻谷等學弟的協助，也要感謝陳瑩在這一段時間給予我精神上的刺激與鼓勵。在離開交大之前，該感謝的人與事相當多，無法一一道盡，只有將感激收藏在心裡。

最後，謹以此論文獻給我最愛的家人，感謝父母以及妹妹、弟弟的支持，讓我能無後顧之憂地專心於課業上的研究，順利完成碩士學位。



Contents

摘要	i
Abstract	ii
誌謝	iii
Contents.....	iv
List of Tables	vii
List of Figures	viii
Chapter 1 Introduction	1
Chapter 2 Background Knowledge and Related Works	4
2.1 Depth Perception	4
2.1.1 The Formation of Binocular Images.....	5
2.1.2 The Cause of Generating Depth Perception	5
2.2 The Fundamental of 2D Image on 3D Display.....	9
2.3 The Proposed 2D to 3D Image System	12
2.4 SANYO 2D to 3D Conversion Adaptive Algorithm	13
2.4.1 MTD (Modified Time Difference Method)	14
2.4.2 CID (Computed Image Depth Method).....	15
Chapter 3 Depth Map Estimation from 2D Image	18
3.1 Image Segmentation	19
3.1.1 The Proposed 2D to 3D Image Segmentation Algorithm.....	20
3.1.2 SSR (Single Scale Retinex)	24
3.1.3 Fuzzy C-means Clustering.	26
3.1.4 Connected Component Searching	29
3.2 Depth Extraction.....	30
3.2.1 The Proposed Depth Extraction Algorithm	31
3.2.2 Depth Cues	33
3.2.3 Zero-Plane	34
3.2.4 Inference the Distance Relation of Each Object.....	35
3.2.5 The Index of Depth.....	36
Chapter 4 Binocular 3D Image Construction Based on Depth Map	38
4.1 Shift Algorithm.....	38

4.1.1 Camera Capturing Method	39
4.1.2 Linear Shift Algorithm	41
4.1.3 Binocular Vision Shift Algorithm.....	41
4.2 Interpolating Holes Algorithm.....	45
Chapter 5 Experimental Results	48
5.1 Experimental Results of Our Proposed Methods	48
5.1.1 The 2D to 3D Image Conversion Algorithm by Using Linear Shift Algorithm	48
5.1.2 The 2D to 3D Image Conversion Algorithm by Using Binocular Vision Shift Algorithm.....	50
5.2 Experimental Results of the SANYO Method	51
Chapter 6 Conclusions and Future Works.....	54
REFERENCES	55



List of Tables

Table 2-1 The proposed 2D to 3D image system 12



List of Figures

Fig. 2-1 The formation of image generating from binocular disparity.....	6
Fig. 2-2 The formation of generating depth perception.....	8
Fig. 2-3 The simulation of display of 3D space	10
Fig. 2-4 The stereo image of simulating 3D space by using stereoscope.....	10
Fig. 2-5 DTI 2015 XLS	13
Fig. 2-6 The principle of our 3D display	13
Fig. 2-7 The principle of MTD	15
Fig. 2-8 The principle of CID	16
Fig. 3-1 The flowchart of our 2D to 3D image conversion algorithm.....	19
Fig. 3-2 The flowchart of our proposed image segmentation algorithm.....	21
Fig. 3-3 The SSR effect on the image (a) The original image. (b) The image after SSR. (c) The pseudo-color images without SSR. (d) The pseudo-color images with SSR.....	25
Fig. 3-4 The SSR effect on the segmentation image. (a) The original image. (b) The image after SSR. (c) The image segmentation result without SSR. (d) The image segmentation result without SSR.....	26
Fig. 3-5 The connected component effect on the segmentation image.	30
Fig. 3-6 Inference the object is horizontal or vertical by the relation between two objects. (a) Two vertical objects. Determine depths by bottom lines. (b) A vertical object on a horizontal object. Determine depths by the boundaries and areas of regions.	32
Fig. 3-7 Flowchart of our proposed depth extraction approach.	33
Fig. 3-8 Zero-plane and bottom line.....	35
Fig. 3-9 The gradual effect on the depth map.....	35
Fig. 3-10 The example of depth index.....	36
Fig. 3-11 The example of correct order	37
Fig. 3-12 The example of wrong order.....	37
Fig. 4-1 Illustration of the camera shift capturing method.....	40
Fig. 4-2 Camera center capturing method.	40
Fig. 4-3 The results after linear shift algorithm.....	41
Fig. 4-4 Simulating the right eye image of binocular vision.....	42
Fig. 4-5 The projection point of relative regions.....	43
Fig. 4-6 The results after binocular vision algorithm.	44
Fig. 4-7 The flowchart of interpolating by horizontal smooth filter.....	46
Fig. 4-8 Interpolating left or right side pixels. (a) Depth map. (b) Interpolating	

holes.....	46
Fig. 4-9 Illustration of interpolating holes method.....	47
Fig. 4-10 The results of interpolating holes method.....	47
Fig. 5-1 The 2D to 3D image conversion process of an artificial image by linear shift algorithm.	49
Fig. 5-2 The 2D to 3D image conversion process of a natural image by linear shift algorithm.....	49
Fig. 5-3 The 2D to 3D image conversion process of an artificial image by binocular vision shift algorithm.....	50
Fig. 5-4 The 2D to 3D image conversion process of a natural image by binocular vision shift algorithm.....	51
Fig. 5-5 The 2D to 3D image conversion process of an artificial image by CID algorithm.....	52
Fig. 5-6 The 2D to 3D image conversion process of a natural image by CID algorithm.....	53



Chapter 1 Introduction

Recently three-dimensional (3D) images have become more popular in the amusement parks. It is difficult to get the 3D images in the real world, because a 3D camera is heavy to carry around and it is difficult to adjust the convergence, the sharpness and the zoom of the 3D camera. So, even if 3D-related technologies have been actively developed and many 3D displays are used in the professional field. 3D fields have not opened up yet. One reason is 3D image software had not been provided enough. We make efforts on the research of a single 2D image converting into stereo images. We hope that people can feel more real on the 3D display.

Some authors have proposed a real time method to convert ordinary 2D images into 3D [1], [2] on television. They use the “Computed Image Depth Method” (CID) [1] to evaluate the depth by a single image and/or the “Modified Time Difference” (MTD) [2] to convert the images, in different contest conditions, depending on the motion content. Other authors consider multiple camera sources to establish the “Structure from Motion (SFM)” defining a system with multiple viewpoints [3], [4]. There are some methods of shape recovery from shading (SFS) by using ICA-based reflectance model [5]. Some authors proposed depth estimation from image structure [6]. They proposed a source of information for absolute depth estimation based on the whole scene structure that does not rely on specific objects. Others authors proposed a technique used to generate stereo pair images starting from a single image source and its related depth map [7]. But some hypothesis must be done and a manual processing is needed in depth map extraction.

“Geometry recovering” (GR) [8] tries to identify basic contest structure such as polygons, vanishing point, and so on to identify the best 3D spider mesh to move the camera and the viewpoint around into the picture. This technique requires a big off-line

effort to recover and to reconstruct the missing data.

The real goal is the reconstruction of the binocular view of a source for a monocular sampled source. To do this, we first calculate the parallax [9]-[11] value of each object in the image. Then, we build the left and right eye images to the final user in a 3D perspective entertainment. And our 3D experimental equipment is a DTI 201XLS monitor which uses barriers to let the users see two different images on the left and right eyes.

We proposed a novel automatic 2D to 3D image conversion technique. It means to convert a 2D image into the left and right eye images and apply on the 3D stereo monitor. The conversion technique can be divided into two parts. The first part is to encode the 2D image into the depth map which includes depth information, and the second part is to decode the depth map into the stereo images which simulate the left and right eyes.

The part of encoder can be divided into two steps. The first step is the image segmentation technique based on the property of 3D stereo images. According to the property of 3D stereo image, we use SSR (Single Scale Retinex) to decrease the light reflection effect. We also use hue, saturation, and intensity to be our features. The traditional clustering method FCM (Fuzzy C-means clustering) is used. We set the threshold of the size of the image segmentation region. If the size of the region is less than the threshold, the region will be merged into the neighbor regions until all the sizes of the image segmentation regions are more than the threshold. The second step is depth extraction. According to the depth cues and the depth rules between each object, we estimate the objects belong to the fixed or gradual depths. So we can get more real effects.

The second part, decoder can be divided into two steps. The first step is to convert the depth map into the left and right eye images. We proposed two methods. The first method is linear shift algorithm, and this is the simplified method which can increase the speed of conversion effectively. The second method is binocular vision shift algorithm, and this method is based on the property of the human vision. After shifting, some points can't be

calculated. So the second step is “interpolating holes method”. After the process of encode and decode, we can automatically generate the stereo images used in 3D display from a single 2D image.

We also implement the SANYO’s method "CID" (Computed Image Depth Method) algorithm which is converting a single 2D image into 3D images. And we will introduce that algorithm in section 2.3. We also establish the index for calculating the accuracy of the depth extraction. We compare the results of our proposed method with CID algorithm by the human perception and scientific data.

The rest of the thesis is organized as follows. Chapter 2 describes the background knowledge and related works. We introduce the depth perception and the fundamental of 2D image on 3D display. Before we design the 2D/3D software system, we must know why people have depth perception and the fundamental of 2D image on 3D display. Thus, we can design a 2D/3D software system based on the human vision and applied on the 3D display. In the post of chapter 2, we describe our 2D/3D system and the 2D/3D conversion adaptive algorithm proposed from SANYO. Chapter 3 describes the encoder of our 2D/3D conversion adaptive algorithm which estimate depth map from a 2D image. Chapter 4 describes the decoder of our 2D/3D conversion adaptive algorithm which constructs binocular 3D image based on depth map. And the experimental results of our methods and CID algorithm are discussed in chapter 5. Finally, conclusions and the future work are summarized in chapter 6.

Chapter 2 Background Knowledge and Related Works

In order to construct a 2D/3D image conversion algorithm, we must know the background knowledge of stereo images. We introduce some information about depth perception and the fundamental of 2D image on 3D display. We also introduce the 2D/3D conversion adaptive algorithm proposed from SANYO. This algorithm is to convert a single 2D image to 3D effect image automatically. Other author proposed the design of the stereo image on 2D monitor [12]. They design the stereo image based on the optical properties. In this chapter, we will introduce the depth perception first. In order to discover the stereo images, we must define the depth perception. And in section 2, we introduce the fundamental of 2D images on 3D display. We introduce the principle of stereo display based on 2D image. There are four kinds of stereo display based on 2D image methods. They are polarization-division, time-division, wavelength-division, and spatial-division. In section 3, we introduce our 2D/3D system in hardware and the depth cues which we used in our system. In the last of the chapter, we introduce the 2D/3D conversion adaptive algorithm proposed from SANYO. We also implement CID algorithm proposed from SANYO. We will compare our results with the results from CID algorithm in the thesis.

2.1 Depth Perception

Why do people have stereo perception when they see the images? People have thought this problem long time ago. In order to understand the generation of the stereo image, we must define the meaning of depth perception first. In the year of 280, Euclid explained the fact that as to the same object, if our two eyes see the different images at the same time, we would have depth perception. In modern psychology, this is called binocular parallax and is

the important reason of causing depth perception. Besides, convergence and the ability of eye accommodation also play important roles on depth perception.

2.1.1 The Formation of Binocular Images

The scientists considered that left eye and right eye see the different images is the most important reason of causing depth perception. The image on the retinex of a single eye is a 2D image. In the basis, the image on the retinex of a single eye doesn't generate depth perception. It needs another image generating from the other eye, and let the retinex receives the different 2D image. By transmission of the optic nerves to the brain and the decision made by brain of two images, there is the depth perception of the distance relation between near or far of the objects.



2.1.2 The Cause of Generating Depth Perception

People know the existence and shape of objects in 3D space by the reflection of light on the retinex. The fundamental faculty of retinex is like the principle of camera capturing. There are many optical nerve cells distributed on the retinex. So we can know the positions of the observed objects. However, it is not enough to know the depth perception of the observed objects only by retinex. In the viewpoints of and Physiology and Psychology, there are some causes of the depth perception.

(a) Physiological factors:

(1) Binocular disparity:

The distance between left and right eye is about 6cm. By the difference of formation of images from two eyes, people can perceive and determine the depth of the observed objects. In Fig 2.1, the left and right eyes see the pyramid, and get

different image. It is binocular disparity. By the formation of images of the binocular disparity, left and right eye get the part of real line of the image shape. And the part of dot line of the image shape is the composition stereo shape of the pyramid.

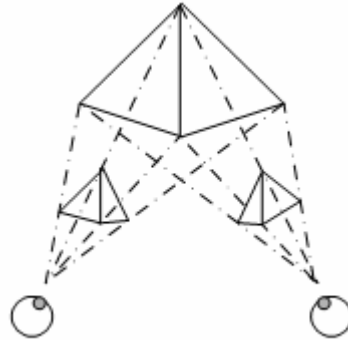


Fig. 2-1 The formation of image generating from binocular disparity

(2) Convergence:

It means the angle of the lines of left and right eye vision. The angle changes according to the distance between the observer and objects. If the distance is nearer, convergence is larger. On the contrary, if the distance is farther, convergence is smaller. The angle of two eyes needs to modify according to the distances of the objects. Thus, people can feel the depth of the objects. In near distance, the change of convergence has significant contributions to the depth perception, especially when it cooperates with eye accommodation. But when the distance is over 10m, because of the tiny change of convergence, people can't perceive the depth of objects.

(3) Accommodation:

The lens of eye is relative to the focus lens of camera. It can raise lens to let the image project on the surface of retina. According to the degree of raise, people can perceive the depth of objects. Generally, when the object is over 2 m, it is difficult to actually perceive the depth.

(b) Psychology factors:

(1) When the observer relatively moves to the objects, the binocular disparity is generated by the motion of near and far scene. We said that it is the disparity of a single eye motion. The disparity of a single eye motion may provide the perception of the depth of objects.

(2) By telling from the difference of the size of the formation of images on retina, even if they are the same objects, people also can distinguish the near distance from far distance.

(3) From the arrangement of objects in 3D space, we can compose perception of depth and stereo.

(4) When people see the parallel lines of a distant place, people also can perceive the depth by their eyes.

(5) Even if people see the uniform objects of a distant place, for example the terraced field and the gradient surface, people also can perceive the depth by the different density.

(6) When people see the same contrast ratio object of a far distance, people also can perceive the depth and distance of objects by the decrease of the contrast ratio.

(7) By the shading of objects generating from the light, people also can feel the stereo shape of objects.

(c) The formation of generating depth perception:

Basically, when people see the scene, they approximately can divide the scene into three levels (far, middle, near). Thus, our eyes will adjust to comfortable angles to observe the objects. When the two eyes see the different positions and angles, binocular disparity will exist naturally. And when the eyes see the far, middle, and near scene, the change of lines of visions can be classified as follow:

(1) Uncrossed-Parallax.

(2) Zero-Parallax.

(3) Crossed-Parallax.

In order to simulate the situation of seeing the scene, we will let the screen be a reference surface and individually present to left eye and right eye. In Fig. 2-2, we will explain the change of focusing visual.

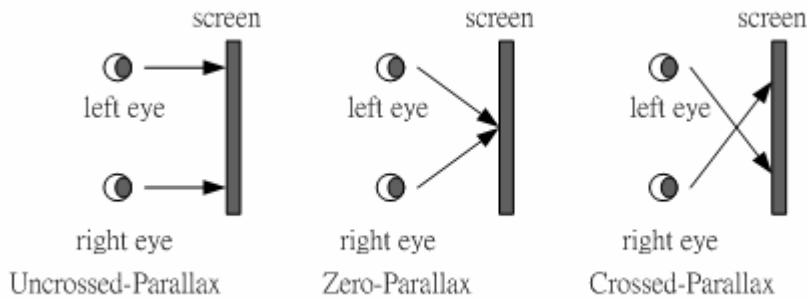


Fig. 2-2 The formation of generating depth perception

(1) Uncrossed-Parallax:

Two eyes don't focus in front of the screen (two eyes see the distant scene parallel), and the image will be shown in back of the screen.

(2) Zero-Parallax:

Two eyes focus on the screen, and the image will be shown on the screen.

(3) Crossed-Parallax:

Two eyes focus in front of the screen, and the image will be shown in front of the screen.

The three methods will generate different stereo effect. And when we know the physical characters of the formation of stereo images and the binocular disparity, it is helpful for 2D to 3D image conversion algorithm to simulate binocular vision.

(d) Monoscopic Depth cues:

There are some monoscopic depth cues:

(1) Interposition:

An object that occludes another is closer.

(2) Shading:

The shadow includes shape information.

(3) Size:

Usually, the larger object is closer.

(4) Linear perspective

Parallel lines converge at a single point.

(5) Surface texture gradient:

Closer objects are more detail.

(6) Atmospheric effects:

Further away objects are blurrier.

(7) Brightness

Further away objects are dimmer.



2.2 The Fundamental of 2D Image on 3D Display

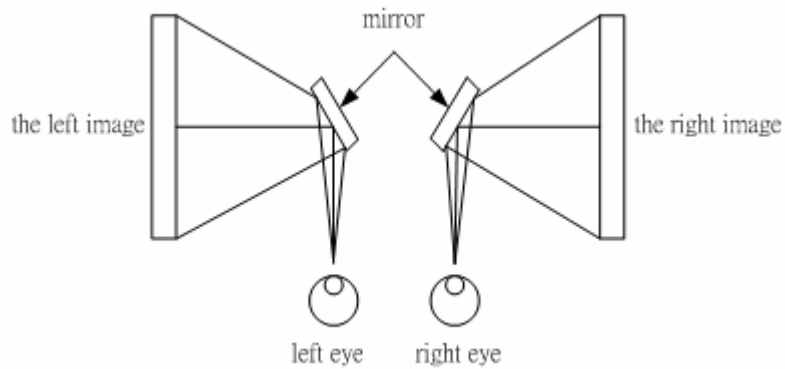
In the last section, it mainly describes the reasons why people have depth perception in 3D space. And we apply these reasons of generating depth perception on the 2D image on 3D display. And the methods of 3D display are in the following:

(1) The reasons of observers in physiology and psychology aspects. Even if it is a 2D image, it can be described as the perception of far or near and belongs to simulate the display of 3D. Left and right eye see the same image of the display at the same time. Fig. 2-3 is an example of simulating 3D image. It gains the density of the parallel lines to simulate that people see the far distant scene as the density increases. So it will generate the depth perception. In general, 3D computer graphic is based on this conception.

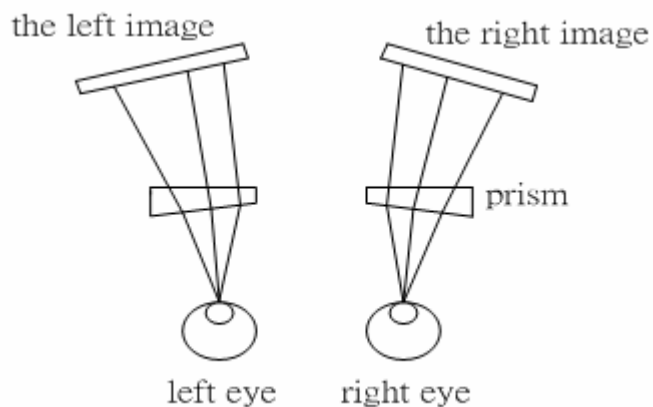


Fig. 2-3 The simulation of display of 3D space

(2) Because of the difference of images received by left and right eye seeing the same object, observers will feel the relation of far or near. It belongs to simulate the display of 3D, too. The basic method is to let the left and right eye see the different images on the same position and generate the stereo perception. In Fig.2-4, it makes the left and right eye see the left and right images individually and use the mirrors and prisms to refract the light. In general, the device is named Wheatstone & Brewster's Stereoscope.



(a) Wheatstone's stereoscope



(b) Brewster's stereoscope

Fig. 2-4 The stereo image of simulating 3D space by using stereoscope

This method simulates the situation of eyes see the scene in 3D space and use two 2D images to achieve the 3D effect. But, the depth perception proposed from the method has limits.

(3) The third method is like the method proposed from (2). However, observers can't get the real depth perception from a single position, and they must change their observed position. The images which the observers see must be changed according to the positions of observers. Thus, people can get real feeling of 3D space and this is real 3D display. In order to let the left and right eye individually see the relative images, there are four techniques:

(1) Polarization division:

Before the images input into the monitor, they pass through the polarized light plate, then make the images into two different polarized images. One's is polarized direction is horizontal, and the other polarized direction is vertical. The users wear polarized glasses which are polarized orthogonally. That is to say the left glass is horizontal polarized prism which can only pass through the horizontal polarized image. The right glass is vertical polarized prism which can only pass through the vertical polarized image.

(2) Time division:

Let the left and right eye image display at the different time. Assume if we want to display the left eye image, the window of left eye will open and the window of right eye will close. And at next time, we display the right eye image. And the window of right eye will open and the window of left eye will shot down. Thus, the left eye sees the left eye image only, and the right eye sees the right eye image only. If the transmission matches up with the stereo image, the two lenses will open and close 60 times every second. And by the principle of vision residual, our brain will combine the left eye image with right eye image and compose a stereo image.

(3) Wavelength division:

Let the images of left and right eye undergo biased red and green treatment. And the

observer wears red glass on the left eye and green glass on the right eye. Then the observer will see the red image on the left eye and the green images on the right eye. It uses the different colors on the left and right images to construct 3D effect image.

(4) Spatial division:

Let the images of left and right eye individually shown on the display. The left eye sees the left eye image on the display, and the right eye sees the right eye image on the display. For example: the monitor with barriers will let the left eye and right eye see the different images at the same time.

2.3 The Proposed 2D to 3D Image System

In the last section, we introduce the depth cues in physiology and psychology. And we also introduce the fundamental 2D image on 3D display, and there are four kinds of 3D displays. Table 1 is the information of our 2D to 3D image system. We use binocular disparity in physiology and contrast, size, and arrangement in psychology to be our depth cues in our system. And the display we use in our 2D to 3D image system is DTI 2015 XLS (Fig.2-5), which uses barriers to let the observer see the different left and right eye images. Fig. 2-6 is the principle of our 3D display. When we move, our eyes will see the different images in left and right eyes.

Table 2-1 The proposed 2D to 3D image system

Depth Cues		Display
Physiology	Psychology	DTI 2015 XLS (Spatial Division)
Binocular Disparity	Contrast	
	Size	
	Arrangement	



Fig. 2-5 DTI 2015 XLS

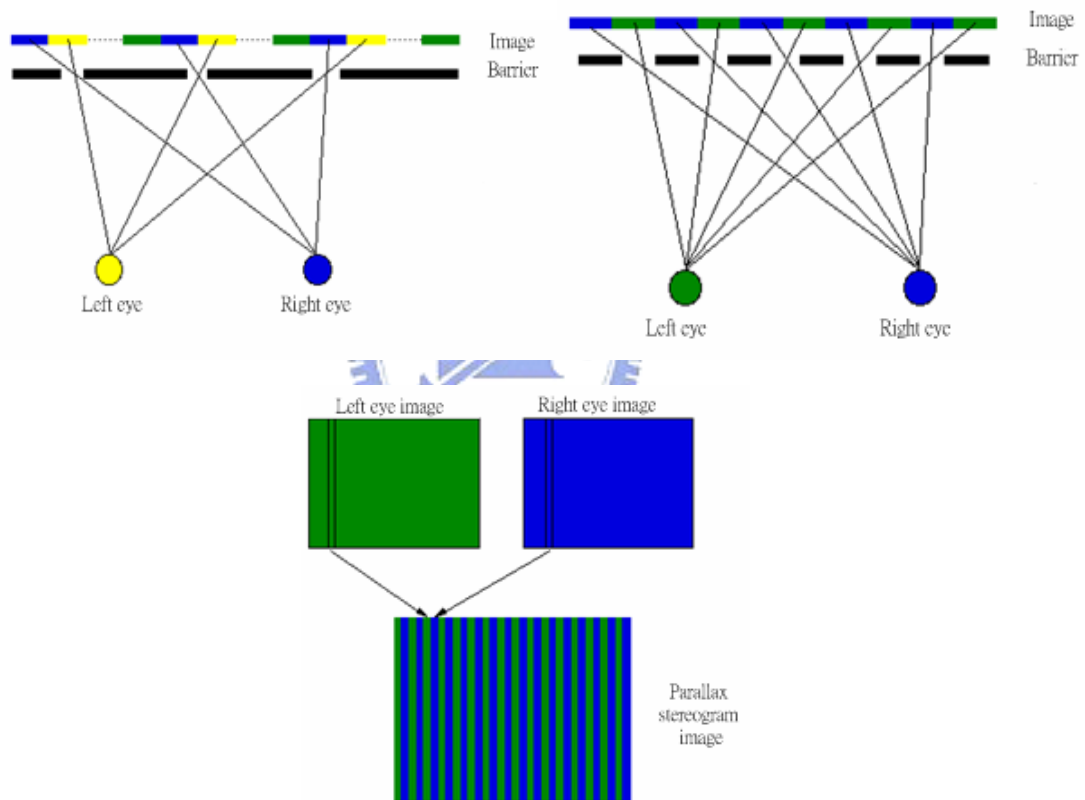


Fig. 2-6 The principle of our 3D display

2.4 SANYO 2D to 3D Conversion Adaptive Algorithm [1], [2]

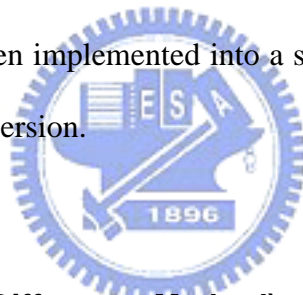
The 2D-to-3D image conversion technique using the “Modified Time Difference method” (MTD) had been developed in 1995 [2]. The MTD allows to converting from 2D

images into 3D images by selecting images that would be a stereo-pair according to the detected motions of objects in the input sequential images.

The 2D images that have the objects with the simple horizontal motion can be converted into 3D images by the MTD, but it is not good for converting from the still images or the images that have the objects with the complicated motions. So the new technique converting from these 2D images into 3D images is required.

The “Computed Image Depth Method” (CID) has been developed to solve this subject. The CID allows to converting from all kinds of 2D images into 3D images. Especially the CID is suitable for converting from still images. The 3D images are generated by computing the depth of each separated area of the input 2D images with their contrast, sharpness and chrominance with the CID.

These techniques have been implemented into a single-chip LSI for the automatic and real-time 2D-to-3D image conversion.



2.4.1 MTD (Modified Time Difference Method)

Fig. 2-7 shows how the MTD converts from images into 3D images.

In the sequences of these 2D images, a bird is flying from the left to the right in front of the mountains. In this scene, if the fourth field of these images is given to the left eye and the second field is given to the right eye, the motion of the bird is perceived as the binocular parallax, whereby the bird can be seen as if it were popping out of the mountains. This technique is based on the principle well known as the Pulfrich Effect.

In order to adapt this technique to the automatic and real-time 2D-to-3D image conversion, they had developed the MTD. In the case such as Fig. 7, the images include the objects moving toward the horizontal direction. The result of analyzing the motion vectors that are detected from the input 2D images should decide the time difference between the

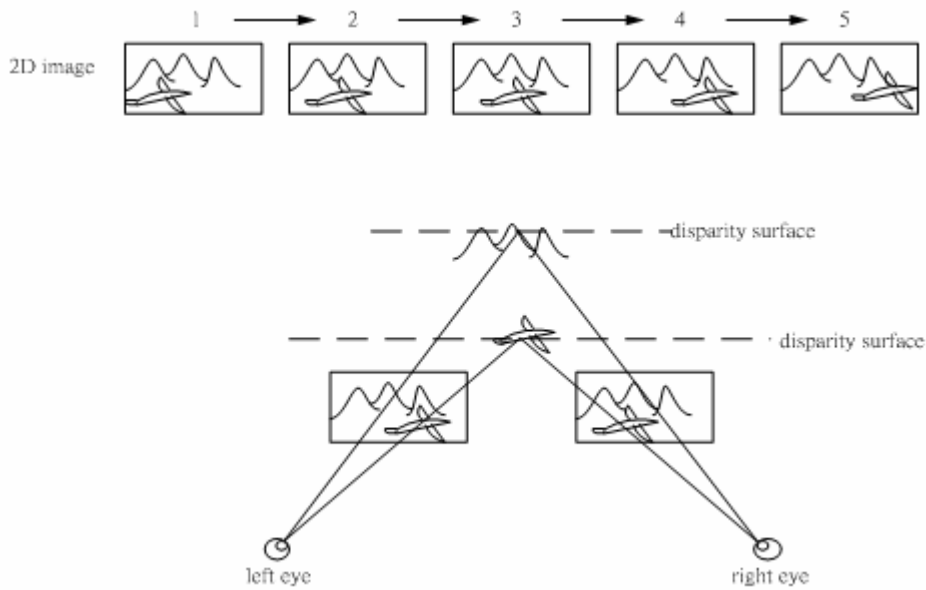


Fig. 2-7 The principle of MTD

left and the right eye images, and either eye that the delayed image is given to. So, the MTD is suitable for converting from the images with the simple horizontal-moving object such that the object spins or the object moves to the horizontal direction against the background. But the MTD does not work well for the images that have the objects with the complicated motions or no motion.

2.4.2 CID (Computed Image Depth Method)

The “Computed Image Depth method” (CID) is proposed for converting from still 2D picture, we generally recognize the far-and-near positional relationship between the objects in the picture by some information in it. These information are how the objects are covered each other, how sharp the images are, how much contrast the images have, what the images are, what the objects are, and so on. These information are supposed to be useful for the 2D-to-3D image conversion. So we use the sharpness and the contrast of the input images for computing the far-and-near positional relationship of the objects in the CID.

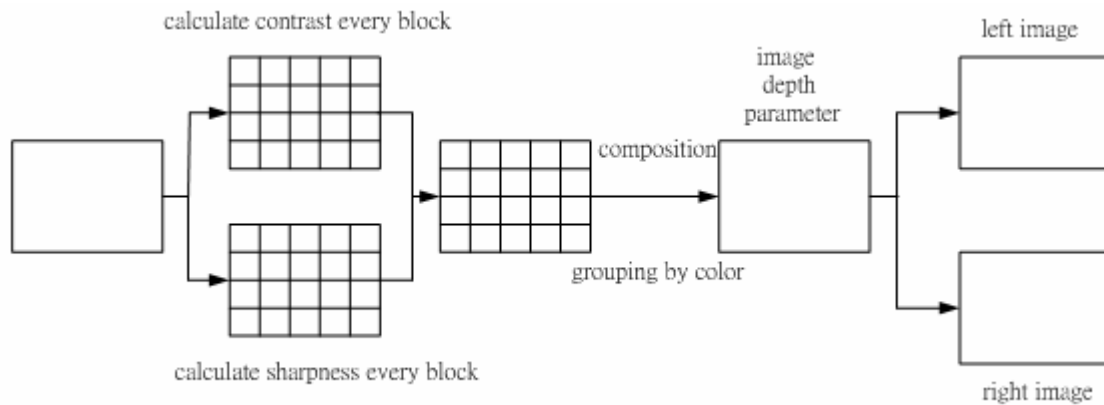


Fig. 2-8 The principle of CID

The CID consists of the following two processes. One is the image depth computation process that computes the image depth parameters with the contrast, the sharpness and the chrominance of the input images. The other is the 3D image generation process that generates the 3D images according to the image depth parameters. Fig.2-8 shows the basic principle of the CID.

At first, each sharpness, contrast and chrominance values of the separated areas in the input images is detected respectively. The sharpness means the high frequency element of the luminance signal of the input images. The contrast means the middle frequency element of the luminance signal. The chrominance means the hue and the tint of the color signal of the input images.

Furthermore, the adjacent areas that have close color are grouped according to the chrominance values. The image depth computation works to be based on the grouped areas.

The image depth computation process uses the contrast values and the sharpness values. Generally in the photographs and the TV images, the near-positioned objects have higher sharpness and higher contrast than the far-positioned objects and the background image. Therefore, these contrast and sharpness values are inversely proportional to the distance from the camera to the objects. If only these values are used for the image depth computation, it often occurs that the center of the images become nearer than both sides, top

and bottom of the images. This cause is that the focused object is generally positioned at the center of the images, and the ground or the floor is positioned the bottom of the images that has flat surface generally and few contrast and sharpness values are taken from the bottom areas. So, it is adopted to compensate these values by the image's composition. The composition has the tendency that the center or the bottom side of the images is nearer than the upper side in the general images. So, each image depth parameter is decided by the average of each area's sharpness and contrast value that is weighted by the image's composition. This compensation would be better way to get good 3D effect, but it should be changed according to the applications.

Secondly, the 3D image generation process generates the left and the right eye images according to the image depth parameter of each grouped area, If the parameter of an area indicates near, the left image are made by shifting the input images to the right, and the right images are made by shifting to the left. If the parameter of an area indicates farm both images are made by shifting to each opposite direction. The horizontal shift value of each separated area is proportional to the 3D effect. Furthermore, when the image depth parameters are changed quickly or frequently, the converted images become hard to be watching. Therefore, each shift value is adjusted to decrease the quick changes of the image depth parameters between the adjacent areas. As a result of these processes, the 3D images that are easy to watch can be generated.

The CID is especially suitable for converting from still images, because it does not need any motions of the objects in the images. Of course, the CID can be also adapted for the images with moving objects.

Chapter 3 Depth Map Estimation from 2D Image

We develop a new architecture for converting a single 2D image into the 3D effect images. This architecture consists of the following two processes. The first part is to encode the image into the depth map, and the second part is to decode the depth map into the left and right eye images. The flowchart is shown in Fig. 3-1. The first part includes two steps. The first step is image segmentation, and the second step is depth extraction. After image segmentation, we can get more complete objects. At the depth extraction stage, we can estimate the distances between each object according to the depth cues. We also set some rules to infer that the object is horizontal or vertical distribution. Thus, we can feel more correct and comfortable on the 3D monitor. After the first part, we can get the 3D coordinates values of any objects in the scene. Thus, we have the image depth parameter. We can encode the 2D image with depth information into the depth map. In the first section of this chapter, we introduce our image segmentation method based on 3D effect images. And we also introduce the image process method which we used in our image segmentation algorithm. And in the second section, we proposed our novel depth extraction method. We show how we extract depths in a single 2D image.

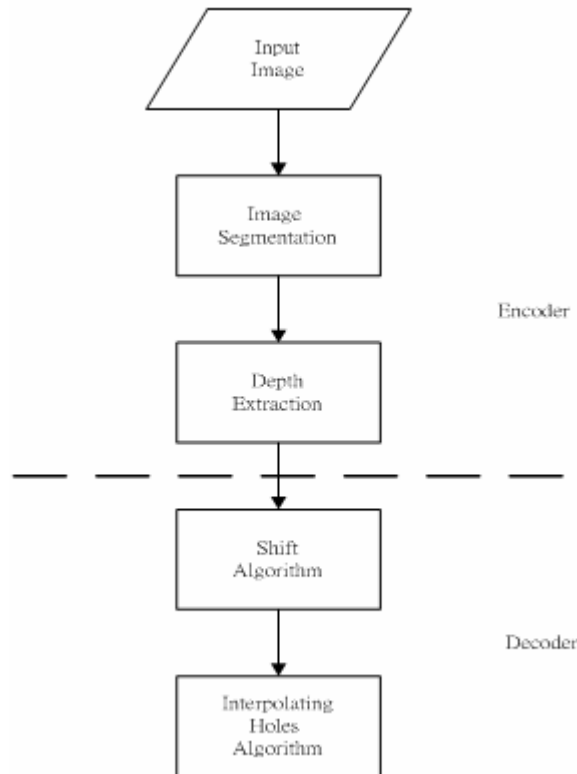
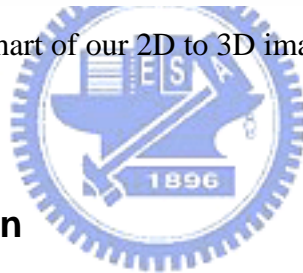


Fig. 3-1 The flowchart of our 2D to 3D image conversion algorithm



3.1 Image Segmentation

Image segmentation is a process that partitions an image into the different objects composing it. Generally, color image segmentation approaches can be divided into the following categories: statistical approaches, edge detection, region splitting and merging approaches, methods based on physical reflectance models, methods based on human color perception, and the approaches using fuzzy set theory [8], [13]. And some authors used other classified methods to do the image segmentation. They used neural fuzzy network to classify the image [14] and an on-line ICA-mixture-model-based self-constructing fuzzy neural networks to segment the image [15]. Other authors also gain the texture information which derived from wavelet into the image segmentation [16].

Histogram thresholding is one of the widely used techniques for monochrome image

segmentation [17]. As for color images, the situation is different due to the multifeatures [18]. Other authors proposed a novel approach to color image segmentation [19], extend the general concept of the histogram, and define a homogeneity histogram. Histogram analysis is applied to both the homogeneity and color feature domains. To calculate the homogeneity feature, both local and global information is taken into account. And they employ a hierarchical histogram analysis method based on homogeneity and color features. We also implemented the color image segmentation proposed from [19]. It has good effect in color image segmentation. But it is not suitable for the 3D effect images, because it lacks the spatial information. So we must use some methods to gain the spatial information. We construct an image segmentation algorithm which is suitable for the 3D effect images.

3.1.1 The Proposed 2D to 3D Image Segmentation Algorithm

We proposed an image segmentation which is suitable for the 3D effect images. The flowchart of our proposed image segmentation algorithm is show in Fig. 3-2. This image segmentation method can be divided into three parts. The first part is preceding process. Due to the image is obtained by CDD camera, it will be affected with the environment. The light reflection is cause of the fluorescent lamps and sunlight outdoors. The light reflection has influence on our segmentation results. It will result in error segmentation results. And we can use SSR (Single Scale Retinex) [20] to reduce this problem. SSR imitates human retina characteristic to avoid environment lighting effect. The SSR method can separate light source and object reflectance from the image. We will introduce SSR in the next subsection and compare the image segmentation method using SSR with the image segmentation method without using

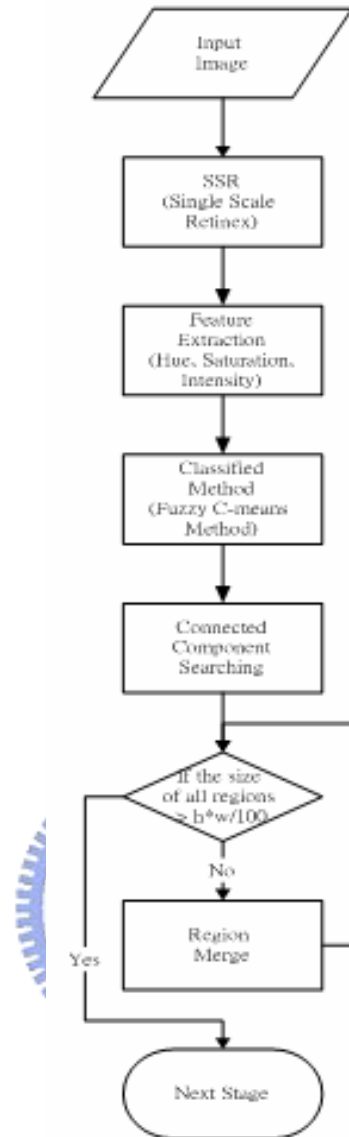


Fig. 3-2 The flowchart of our proposed image segmentation algorithm

SSR. We will show some examples to explain that SSR has the effect to decrease the light reflection. It has a good effect on our image segmentation because we use intensity to be one of our features. After SSR processing, the new image will keep each object specular rate and present clear and undistortion details.

And the second part is feature extraction and classified method. There are two important ideas for color segmentation [21], which are uniform chromaticity scale (UCS) and human perception. The UCS color space means that the color difference of human

perception can be directly expressed by an Euclidean distance in the color space [18]. (L^* , a^* , b^*) and (L^* , u^* , v^*) color spaces are approximately UCS. On the contrary, (R , G , B) and (X , Y , Z) color spaces are non-UCS. Achieving an adequate segmentation result depends on segmentation techniques by detecting similarity among the attributes of image pixels. The UCS is a mathematical system to match sensitivity of human eyes with computer processing. Color has psychological attributes; the perceptual color space is usually described by hue, saturation and intensity (H , S , I). The followings are the equations of converting colors from RGB to HSI.

$$H = \begin{cases} \theta & \text{if } B \leq G \\ 360 - \theta & \text{if } B > G \end{cases} \quad (3-1)$$

with

$$\theta = \cos^{-1} \left\{ \frac{\frac{1}{2}[(R-G) + (R-B)]}{[(R-G)^2 + (R-B)(G-B)]^{1/2}} \right\} \quad (3-2)$$

$$S = 1 - \frac{3}{(R+G+B)} [\min(R, G, B)] \quad (3-3)$$

$$I = \frac{1}{3}(R+G+B) \quad (3-4)$$

The intensity is a measure of total reflectance in the visible region of spectrum, and it is an achromatic component of color. The hue is the attribute of color perception denoted by red, yellow, green, blue, and so on. Saturation is used to describe how pure a color is or how much white is added to a pure color. Using perceptual color space for color image segmentation has two advantages: (1) specifying and controlling color is more suitable for intuition of human than using the primary color RGB; (2) it can control intensity and chromatic components more easily and independently. Other authors proposed circular histogram thresholding for color image segmentation [23]. They also concluded that using 1-D clustering (Hue) may be not good as that of 3-D clustering (Hue, Saturation, Intensity). We have done some experiments with above color systems. We found that in most images

using hue, saturation, and intensity to be the features is better than using (L, a, b) space. And using more features (hue, saturation, intensity) is better than only one feature (hue). So we choose hue, saturation, and intensity to be our features. And classified method, we use a traditional clustering method (the FCM algorithm) in image features (H, S, I). With FCM method, the pixels belonging to a valid class are clustered. A cluster is determined if the maximum value of the membership function is below the threshold T. FCM is a supervised classified method, so we need to decide the number of clusters. We use CE (classification entropy) to measure for the fuzziness of the cluster partition. We use 2-6 to be the number of the clusters in FCM and calculate CE. When the CE is minimum, it regards as a good partition. We decide the number of clusters which has minimum value of CE. In the experimental experience, there is good image segmentation result when the ranges of the clusters number between 2 and 6.

And other authors proposed to use Self-organizing feature map (SOFM) to select which features are important in an image [24]. We also try to use SOFM to select our import features. But we want to simplify our algorithm, we choose the traditional clustering method, FCM. And we will introduce FCM later.

After FCM method, the original image will be divided into many regions. Each region has the unique label number. But it will generate under-segmentation problem. Namely, different objects are associated to the same region. This circumstance represents the lack of spatial information. Hence, we solve this problem by connected component searching method. We will introduce connected component searching method later. And we set a threshold of the size of the regions. Because of the 3D effect images, we need few clustering number in order to get more complete objects on the 3D monitor. If the size of the region is less than the threshold, the region will be merged into the neighbor region which the size is smallest. And we merge little regions iteratively until all the sizes of the regions are not less than the threshold.

3.1.2 SSR (Single Scale Retinex)

Due to the image is obtained by the CCD camera, it will be affected with the light reflection. For example, the objects in color images can exhibit variations in color saturation with little or no correspondence in luminance variation. In order to reduce this problem, we adopt the technology of imitating human retina characteristic to avoid environment lighting effect. There have been suggested over the years on many variants of the retinex. The last version that Land proposed is now referred to as the Single Scale Retinex (SSR) [10], The Single Scale Retinex for a point (x, y) in an image is defined as being:

$$R_i(x, y) = \log I_i(x, y) - \log [F(x, y) \otimes I_i(x, y)] \quad (3-5)$$

where $R_i(x, y)$ is the retinex output and $I_i(x, y)$ is the image distribution in the i^{th} spectral band. In the above equation the symbol \otimes represents the convolution operator and $F(x, y)$ is the Gaussian surround function: $F(x, y) = Ke^{-\frac{(x^2+y^2)}{c}}$, and c is the Gaussian surround constant - analogous to the σ generally used to represent standard deviation. The Gaussian surround constant c is what is referred to as the scale of the retinex. The SSR method can separate light source and object specular rate from image. After SSR processing, the new image will keep each object specular rate and present clear and undistortion details. Fig. 3-3 (a) and (b) are the original image and the processed image by single scale retinex method. Fig. 3-3 (c) and (d) show the pseudo-color images without SSR and with SSR respectively. We can find there are many highlight places in the image without SSR processing, and the highlight places are concentrated in an area in the image with SSR processing.

We did the indoor experiment. We take a picture of the cup on the table. Fig. 3-4 (a) is the original image in intensity channel. We can see there is light reflection on the table obviously. And the light reflection will produce the wrong image segmentation. Fig. 3-4 (c)

is the image segmentation result without SSR. We can see that the table is divided into more than one part because of the reflection. And we can see that Fig. 3-4 (d) is the image segmentation result after SSR. The table after image segmentation is complete, because the SSR decreases the light reflection.

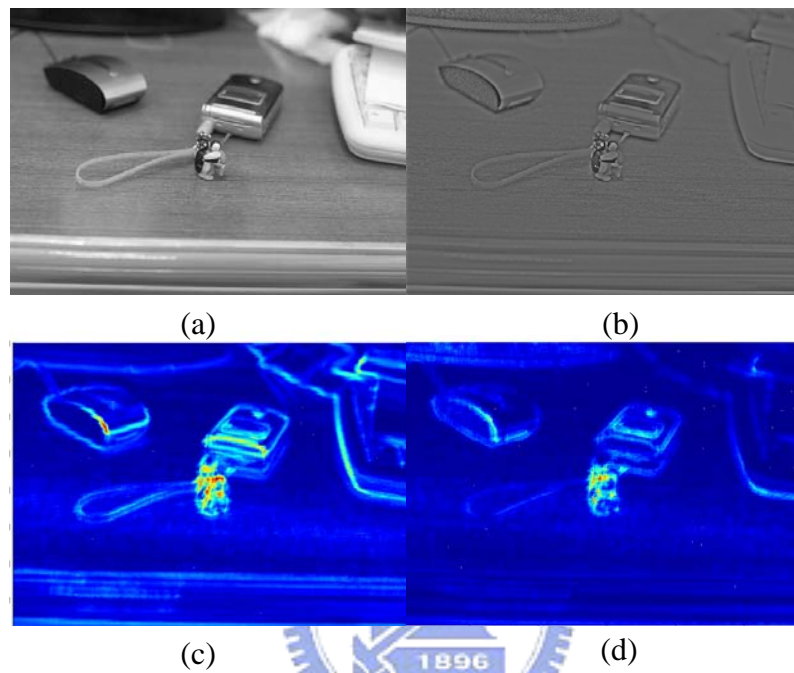


Fig. 3-3 The SSR effect on the image (a) The original image. (b) The image after SSR. (c) The pseudo-color images without SSR. (d) The pseudo-color images with SSR.

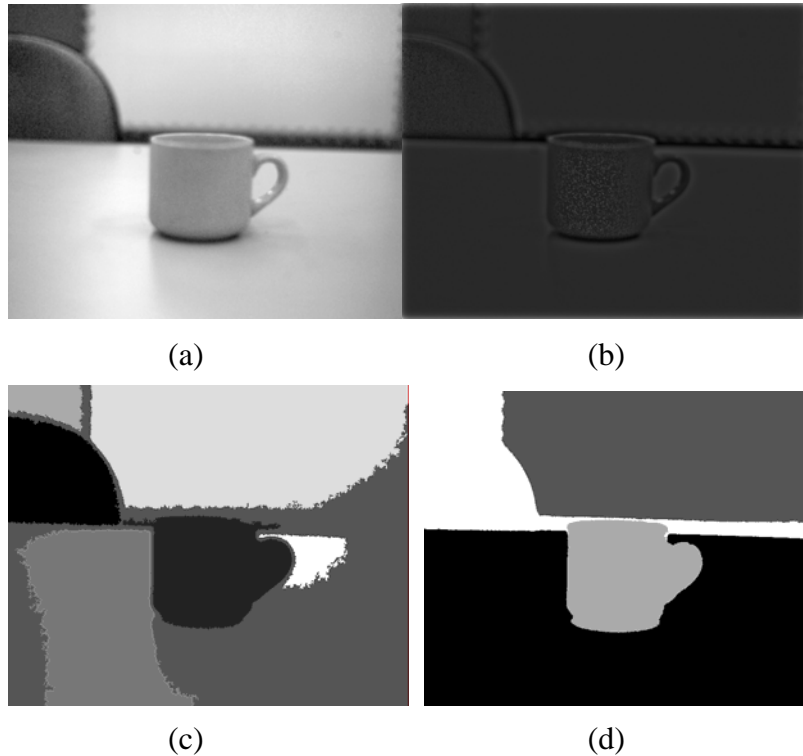
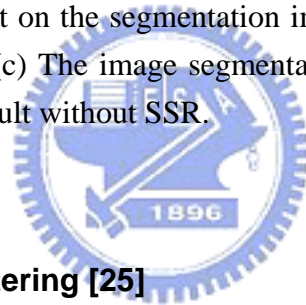


Fig. 3-4 The SSR effect on the segmentation image. (a) The original image. (b) The image after SSR. (c) The image segmentation result without SSR. (d) The image segmentation result with SSR.



3.1.3 Fuzzy C-means Clustering [25]

Clustering essentially deals with the task of splitting a set of patterns into a number of more-or-less homogeneous classes (clusters) with respect to a suitable similarity measure such that the patterns belonging to any one of the clusters are similar and the patterns of different clusters are as dissimilar as possible. The similar measure used has an important effect on the clustering results since it indicates which mathematical properties of the data set, for example, distance, connectivity, and intensity, should be used and in what way they should be used in order to identify the clusters. In nonfuzzy “hard” clustering, the boundary of different clusters is crisp, such that one pattern is assigned to exactly one cluster. On the contrary, fuzzy clustering provides partitioning results with additional information supplied by the cluster membership values indicating different degrees of belongingness.

Consider a finite set of elements $X=\{x_1, x_2, \dots, x_n\}$ as being elements of the p -dimensional Euclidean space R^P , that is, $x_j \in R^P, j=1, 2, \dots, n$. The problem is to perform a partition of this collection of elements into c fuzzy sets with respect to a given criterion, where c is a given number of clusters. The criterion is usually to optimize an *object function* that acts as a performance index of clustering. The end result of fuzzy clustering can be expressed by a *partition matrix* U such that

$$U=[u_{ij}]_{i=1, \dots, c, j=1, \dots, n}, \quad (3-6)$$

where u_{ij} is a numerical value in $[0, 1]$ and expresses the degree to which the element x_j belongs to the i th cluster. However, there are two additional constraints on the value of u_{ij} . First, a total membership of the element $x_j \in X$ in all classes is equal to 1.0; that is,

$$\sum_{i=1}^c u_{ij} = 1 \quad \text{for all } j=1, 2, \dots, n. \quad (3-7)$$

Second, every constructed cluster is nonempty and different from the entire set; that is,

$$0 < \sum_{j=1}^n u_{ij} < n \quad \text{for all } i=1, 2, \dots, c. \quad (3-8)$$

A general form of the objective function is

$$J(u_{ij}, v_k) = \sum_{i=1}^c \sum_{j=1}^n \sum_{k=1}^c g[w(x_j), u_{ij}] d(x_j, v_k), \quad (3-9)$$

where $w(x_j)$ is the a priori weight for each x_j , $g[w(x_j), v_{ij}]$ influences the degree of fuzziness of the partition matrix, and $d(x_j, v_k)$ is the degree of dissimilarity between the data x_j and the supplemental element v_k , which can be considered the central vector of the k th cluster. The degree of dissimilarity is defined as a measure that satisfies two axioms:

$$(i) \quad d(x_j, v_k) \geq 0, \quad (3-10)$$

$$(ii) \quad d(x_j, v_k) = d(v_k, x_j),$$

and thus it is a concept weaker than distance measures.

With the above background, fuzzy clustering can be precisely formulated as an optimization problem:

$$\text{Minimize } J(u_{ij}, v_k), i, k= 1,2,\dots, c; j=1,2,\dots, n \quad (3-11)$$

Subject to Eqs. (3-7) and (3-8).

One of the widely used clustering methods based on Eq. (3-7) is the fuzzy c-means (FCM) algorithm developed by Bezdek [1981]. This objective function of the FCM algorithm takes the form of

$$J(u_{ij}, v_i) = \sum_{i=1}^c \sum_{j=1}^n (u_{ij})^m \|x_j - v_i\|^2, \quad m > 1, \quad (3-12)$$

where m is called the exponential weight which influences the degree of fuzziness of the membership (partition) matrix. To solve this minimization problem, we first differentiate the objective function in Eq. 3-12 with respect to v_i (for fixed u_{ij} , $i = 1, \dots, c$, $j = 1, \dots, n$) and to u_{ij} (for fixed v_i , $i=1, \dots, c$) and apply the conditions of Eq. (3-7), obtaining

$$v_i = \frac{1}{\sum_{j=1}^n (u_{ij})^m} \sum_{j=1}^n (u_{ij})^m x_j, \quad i = 1, 2, \dots, c, \quad (3-13)$$

$$u_{ij} = \frac{(1/\|x_j - v_i\|^2)^{1/(m-1)}}{\sum_{k=1}^c (1/\|x_j - v_k\|^2)^{1/(m-1)}}, \quad i = 1, 2, \dots, c; j = 1, 2, \dots, n. \quad (3-14)$$

The system described by Eqs. (3-13) and (3-14) cannot be solved analytically. However, the FCM algorithm provides an iterative approach to approximating the minimum of the objective function starting from a given position. Besides there are some validity indices such as partition coefficient (PC) and classification entropy (CE) use the information of fuzzy membership grades to evaluate clustering result.

$$\text{PC}(c) = \frac{1}{N} \sum_{i=1}^c \sum_{j=1}^N (u_{ij})^2, \quad \text{CE}(c) = -\frac{1}{N} \sum_{i=1}^c \sum_{j=1}^N u_{ij} \log(u_{ij})$$

u_{ij} ($i=1,2,\dots,c; j=1,2,\dots, N$) is the membership of data point j in cluster i .

FCM algorithm is summarized in the following:

FCM (Fuzzy c-Means Algorithm):

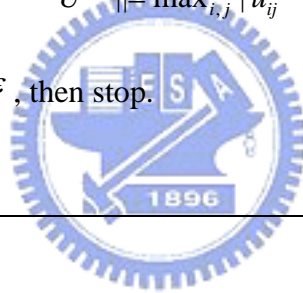
Step 1: Select a number of clusters $c(2 \leq c \leq n)$ and exponential weight $m(1 < m < \infty)$. Choose an initial partition matrix $U^{(0)}$ and a termination criterion ε . Set the iteration index l to 0.

Step 2: Calculate the fuzzy cluster centers $\{v_i^{(l)} \mid i = 1, 2, \dots, c\}$ by using $U^{(l)}$ and Eq. 3-9.

Step 3: Calculate the new partition matrix $U^{(l+1)}$ by using $\{v_i^{(l)} \mid i = 1, 2, \dots, c\}$ and Eq. 3-10.

Step 4: Calculate $\Delta = \|U^{(l+1)} - U^{(l)}\| = \max_{i,j} |u_{ij}^{(l+1)} - u_{ij}^{(l)}|$. If $\Delta > \varepsilon$, then set $l=l+1$ and go to step 2. If $\Delta \leq \varepsilon$, then stop.

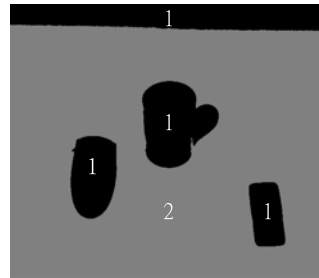
END FCM



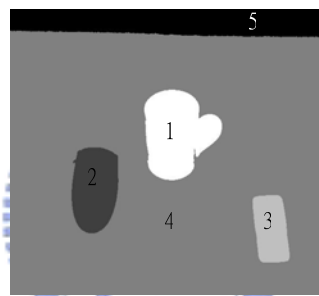
3.1.4 Connected Component Searching

After image segmentation, the original image will be divided into a lot of regions. Each region has the individual label number. But it will generate under-segmentation problem. Namely, different objects are associated to the same region. Fig. 3-5 (a) shows the under-segmentation result. There are many black areas, which have different spatial positions, in labeled image. This circumstance represents the lack of spatial information. Hence, we use connected component searching method to solve this problem. Connected components labeling (CCL) of an image is a fundamental step in the segmentation process and consists in identifying and labeling the separate different regions of interest of the image. This operation has applications in the fields of image understanding, geometric

modeling and computer vision and so on. It scans an image and groups its pixel into components based on pixel connectivity. After the connected component searching, the objects in different spatial position are separated. Fig. 3-5 (b) shows the result. The processing will result in the generation of many new labels.



(a)



(b)

Fig. 3-5 The connected component effect on the segmentation image.

(a) The under-segmentation image.

(b) The result after connected component research.

3.2 Depth Extraction

After image segmentation stage, we would assign the depths to the objects. Traditionally, the ability to perceive depth from two-dimensional images has been accomplished by binocular method [26], [27]. Binocular depth cues, such as disparity, have been used. More recently, researchers have developed monocular approach [28], [29]. Monocular depth cues, such as blur, have been used to perceive depth. In these approaches, point matching was not required; however, multiple images were required for depth

perception. Researchers have begun looking at integration of binocular and monocular cues. Several researchers have studied the result of combining cues to perceive depth from 2D images [30], [31]. Others have studied how these different cues interact in creating a depth effect [32]-[34]. In these studies, binocular cues have been considered large contributors to depth perception in 2D images.

We proposed a novel depth extraction algorithm which can extract depth from a single 2D image. Thus, we can encode a 2D image with depth information into the depth map. In the first sub-section, we will propose our depth extraction algorithm. And in other sub-sections, we will explain our method which used in our depth extraction algorithm.

3.2.1 The Proposed Depth Extraction Algorithm

According to human vision characteristic, people can perceive the depth relations between every object because of the focused position. There exists a proper balance surface. When we use left and right eye to observe the balance plane, it is on the same position on the left and right eye images. The balance plane is entitled “zero-plane”. We propose a method to find the zero-plane. Subsequently, the image will employ the mask operation method to calculate variance each pixel. We discovered that the focused area has higher variance than other areas. According to this characteristic, we can determine zero-plane which has maximum variance in the picture.

After getting zero-plane, we will give depth values to each object. The depth value is calculated according to the distances between the bottom of the object and zero-plane. All of the areas without being focused have blur phenomenon. Due to the arrangement of objects, we have to assume that objects under zero-plane have lower depth values. On the contrary, the objects above zero-plane have higher depth values.

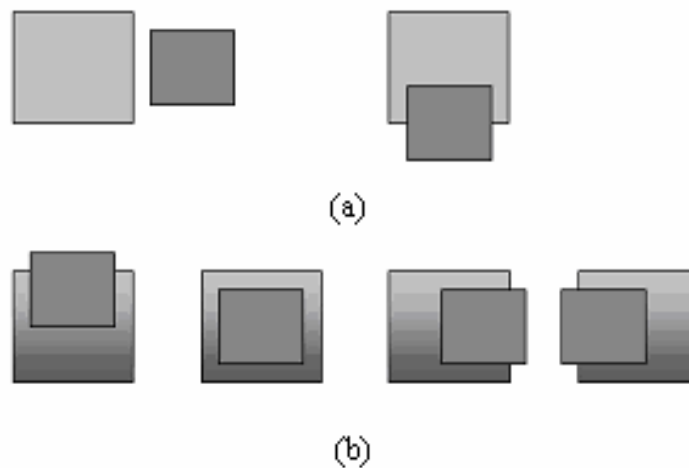


Fig. 3-6 Inference the object is horizontal or vertical by the relation between two objects. (a) Two vertical objects. Determine depths by bottom lines. (b) A vertical object on a horizontal object. Determine depths by the boundaries and areas of regions.

On the other side, we estimate that the object is horizontal or vertical distribution. Because we give the horizontal distribution object a gradual depth and the vertical distribution object a fixed depth, the observers will feel more correct 3D effect on 3D display. In order to estimate the relation, we set some logic rules. Fig. 3-6 illustrates some rules. Fig. 3-6a gives two examples of two vertical objects. We can assign depths to the two vertical objects by calculating their bottom lines. Fig. 3-6b gives the examples of that one object is horizontal distribution and the other is vertical distribution. We can assign depths to the vertical distribution object and the horizontal distribution object by the boundaries and the areas of regions. We set the depth relationship between two objects, but the segmentation image would include more than two objects. We can use the rules of depth relation one by one until the estimations of all pairs have finished. When we get the depths of objects and infer that the objects are vertical or horizontal distributions, we can arrange the lists of depths and get the depth map. Thus, we can get the 3D coordinates values of any objects in the scene. The flowchart of depth extraction is shown in Fig. 3-7.

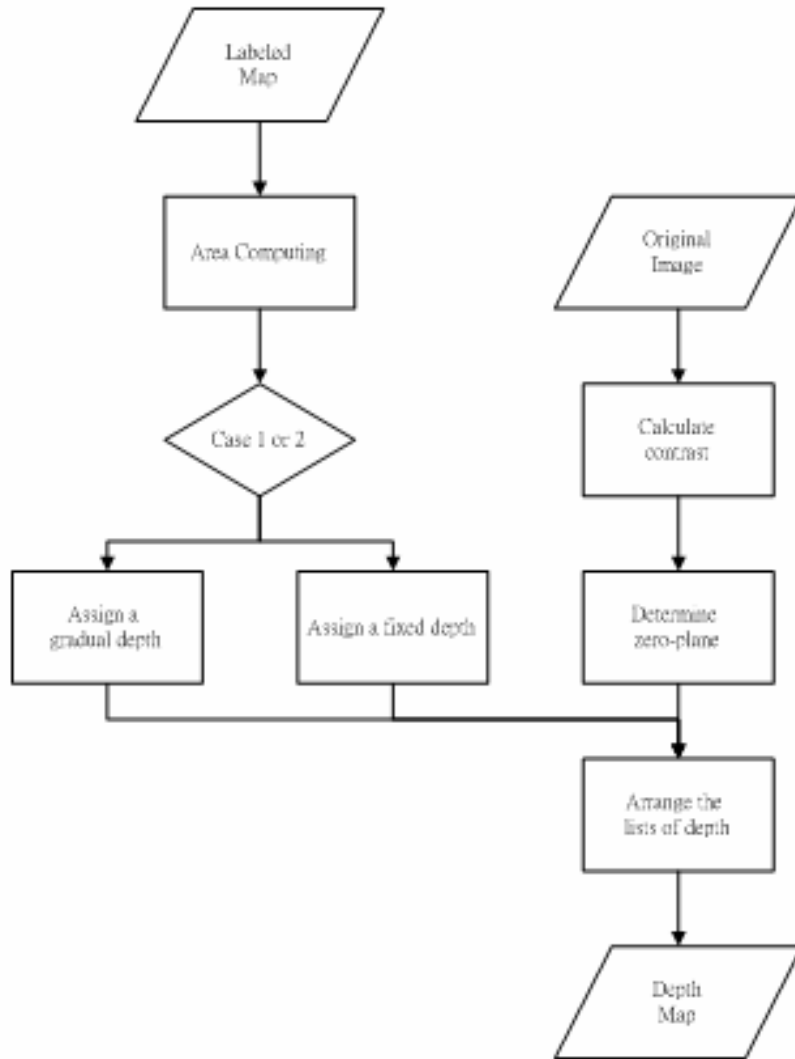


Fig. 3-7 Flowchart of our proposed depth extraction approach.

3.2.2 Depth Cues

In section 2.1, we proposed some depth cues of a 2D image. They are useful in depth extraction from a single 2D image. We use some of them in depth extraction. We use contrast, size, and arrangement in psychology. And in physiology, we simulate binocular disparity to generate the left and right eye images. We will introduce in the next chapter.

Generally in the photographs, the near-positioned objects have higher contrast than the far-positioned objects and the background image. Therefore, these contrast values are

inversely proportional to the distance from the camera to the objects. So we choose “contrast” to be one of our depth cues. But only contrast is not enough. Because it often occurs that the center of the images become nearer than both sides, top and bottom of the images. This cause is that the focused object is generally positioned at the center of the images, and the ground or the floor is positioned the bottom of the images that has flat surface generally and few contrast and sharpness values are taken from the bottom. So we add the factor “arrangement” to depth extraction. We consider the position of the object in the vertical direction of the image. If the object is on the top of the image, we give it a farer depth map. On the contrary, if the object is on the bottom of the image, we give it a nearer depth map. And the factor “size” is also used in depth extraction. We inference the object is vertical or horizontal object by using the factors “arrangement” and “size”. We will introduce the inference the distance relation of every object at section 3.2.4.



3.2.3 Zero-Plane

We employ the mask operation method to calculate variance each pixel. We determine zero-plane which has the maximum variance in the picture. It simulates that two eyes focus on the plane, and the property of the zero-plane is that the pixels on the plane don't shift on the left and right eye images. After image segmentation, we get the regions of objects. We named the horizontal line which is on the bottom of the object bottom line. We assign the depth to the object by calculating the distance between zero-plane and bottom line. We illustrate an example in Fig. 3-8. We can see the highest variance is in the center of the image. So we determine zero-plane on the penguin. And we calculate the depth of mouse by the distance between the zero-plane and bottom line of mouse.

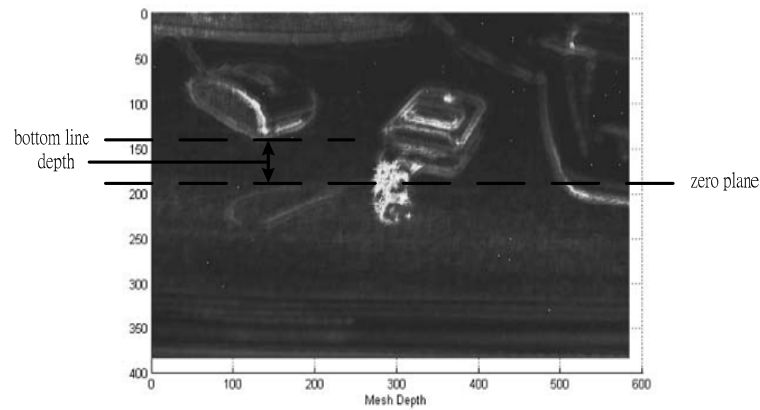


Fig. 3-8 Zero-plane and bottom line.

3.2.4 Inference the Distance Relation of Each Object

We set some rules to infer the relation between every object. We estimate that the object is horizontal or vertical by the rules shown in Fig. 3-7. We calculate the left, right, up, and down boundaries of objects. And we can infer that the object is vertical or horizontal distribution. We assign depths to the two vertical distribution objects by bottom lines. And we assign the depths to the vertical and horizontal distribution objects by the boundaries and areas of regions. If there are more than two objects, we can use the depth relations rules one by one until the estimations of all pairs have finished. We can see Fig. 3-9 are the examples of fixed and gradual depths. On the 3D monitor, we can see the object with a fixed depth stands on the object with gradual depths.

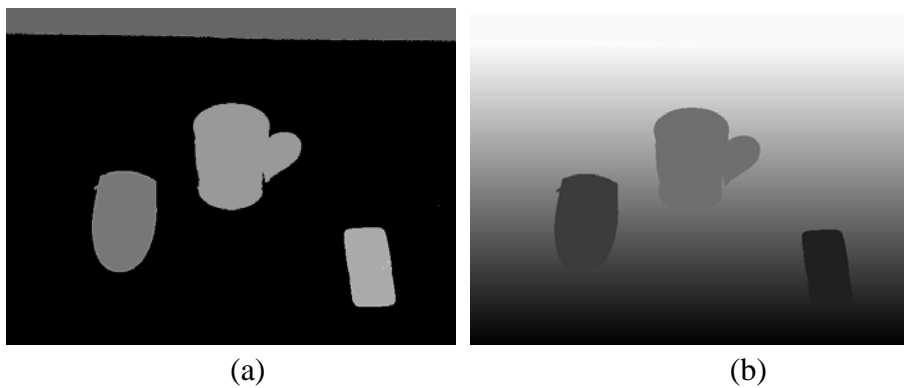


Fig. 3-9 The gradual effect on the depth map.

(a) The example of fixed depth. (b) The example of gradual depths.

3.2.5 The Index of Depth

We proposed a method to set an index to determine the accuracy of depth extraction. The principle of our method is to detect the front and rear relation of one object and the other objects individually. And calculate the error rate. We illustrate an example in Fig. 3-10. We can see there are five objects in this image. And we labeled them in label 1-5. And the order of actual depths is 5, 4, 1, 3, and 2. Assume the order of depths we computed is 5, 4, 1, 2, and 3. Fig. 3-11 illustrates how we compute the depth accuracy of object 1. We compute if other objects are far or near relative to the object 1. And we compare the relation of actual depths with computed depths. If it is the same, the depth estimation of the object is correct. On the contrary, if it is different, the depth estimation of the object is wrong. In Fig. 3-11, we can see that the objects 2, 3, 4, and 5 are rear of the object 1 both in actual depths and computed depths. So the depth estimation of object 1 is correct. In Fig. 3-12, we estimate the depth of object 4. We can see that the object 5 is front of the object 4 in actual depths. But it is rear of the object 4 in computed depths. So the depth estimation of object 4 is wrong. After we computed the depth accuracy of the five objects, we can find that object 1, object 2, and object 3 are correct. And the object 4 and object 5 are wrong. The depth accuracy of the image is 60 percent.

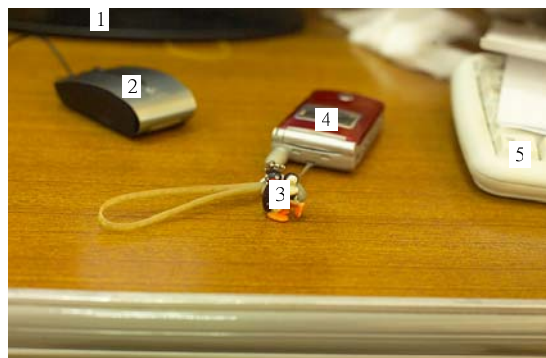


Fig. 3-10 The example of depth index.

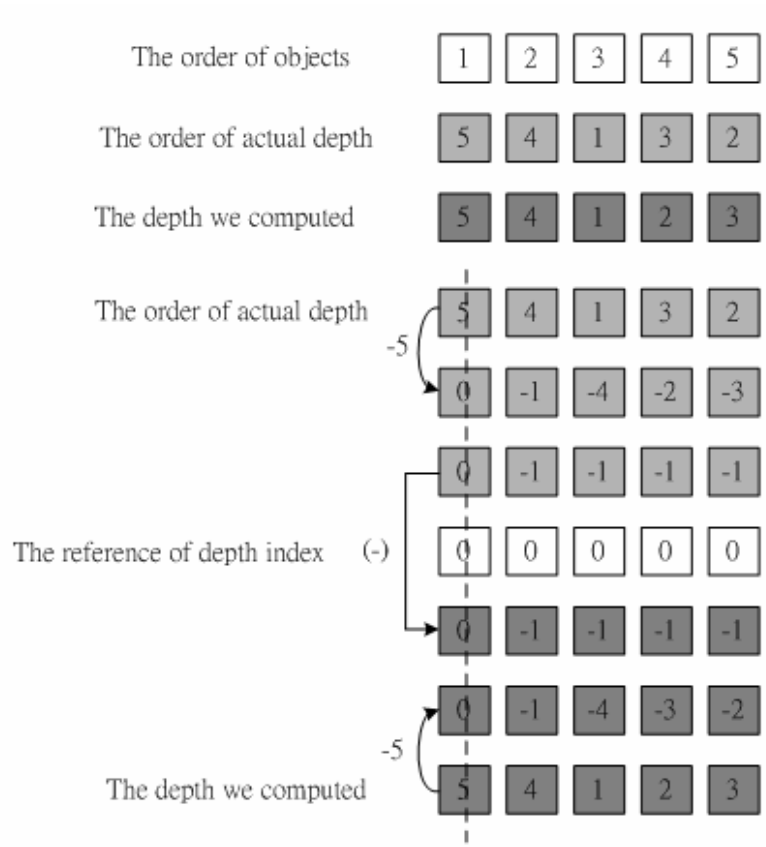


Fig. 3-11 The example of correct order

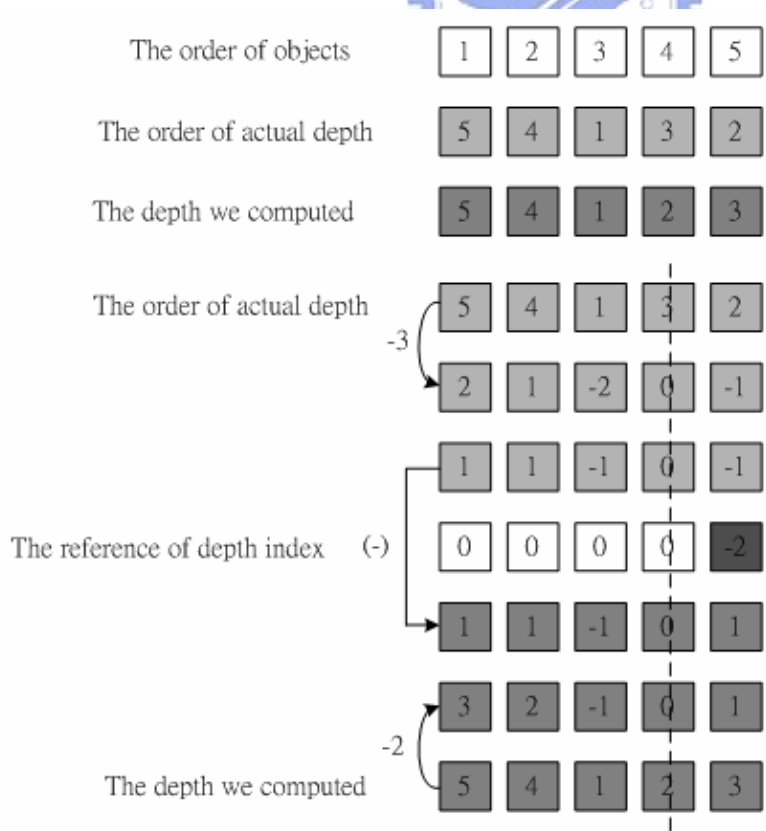
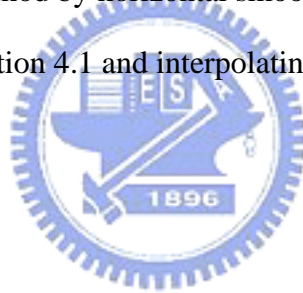


Fig. 3-12 The example of wrong order

Chapter 4 Binocular 3D Image Construction Based on Depth Map

After we have the depth image parameters, we can decompress the depth image parameters to the left and right eye images. We proposed two kinds of shift algorithm. They are inferred from two kinds of camera capturing methods which generate stereo images. And the shift algorithm is to generate two images from one image, so there are some holes in the left and right eye images. We need to interpolate these holes to let the observers feel more comfortable. We test some interpolating holes algorithm to interpolate holes. We found that the interpolating method by horizontal smooth filter has better effect. We will introduce shift algorithm in section 4.1 and interpolating holes algorithm in section 4.2.



4.1 Shift Algorithm

After obtaining the depth map, the left and right eye images can be obtained by stereovision characteristic. Namely, obtaining the left and right eye images is based on the image depth and the hardware parameters of 3D display. If we do not consider the hardware parameters of 3D display, the binocular vision image is relative to the image depth.

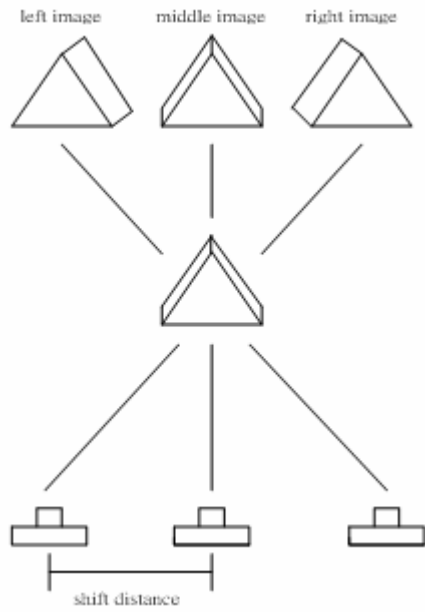
We proposed two kinds of shift algorithms. They are inferred from two kinds of camera capturing methods which generate stereo images. One is camera shift capturing method and the other is camera center capturing method. Camera shift capturing method is to shift a camera and get the left and right eye images. And camera center capturing method is to let the focus centers of two images on the same position. Using camera center capturing method will decrease parallax and let observers feel more comfortable. We infer

linear shift algorithm from camera shift capturing method and binocular vision shift algorithm from camera center capturing method.

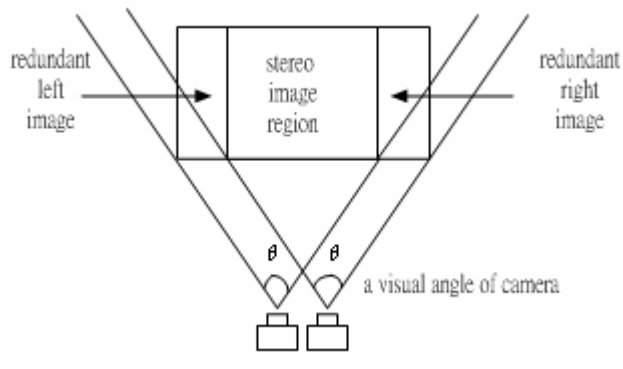
4.1.1 Camera Capturing Method

Basically, a camera is like the structure of eyes. When we observe the objects, the scene passes through the eye ball (lenses) and form an image on the retina (a negative of a photo). When we see the stereo objects in 3D space, the left and right eye must transmit images to the brain individually. In other words, the brain will compose a stereo image by binocular parallax and retinal disparity. According to the principle, we shift the camera to simulate our eyes. Fig. 4-1(a) illustrates how to shift the camera to capture stereo images. And Fig. 4-1(b) is the camera capturing shift method.

When the eyes see the scene, the vision angles of two eyes will focus on the same position of the object. And this will generate less binocular parallax. Based on the principle, we can let the camera to simulate our eyes to capture the scene. We will let the focus center of left and right images on the same position of the objects. By simulating the binocular vision which focuses on the same position, we can let the parallax of left and right remained images decrease at the least degree. And we can feel more comfortable. Fig. 4-2 is camera center capturing method.



(a)



(b)

Fig. 4-1 Illustration of the camera shift capturing method.

(a) Shifting the camera to capture the stereo image.

(b) Camera shift capturing method.

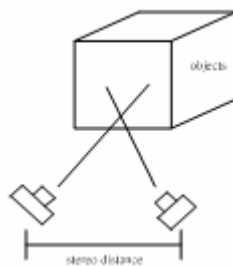


Fig. 4-2 Camera center capturing method.

4.1.2 Linear Shift Algorithm

When people see a scene, the nearer objects will shift more and the farer objects will shift less on the two eyes. Based on the thesis, we can simplify the shift by linear method. On the right eye image, if the depth of pixel is higher than the zero-plane, the pixel will rightwards shift. On the contrary, if the depth of pixel is less than the zero- plane, the pixel will leftwards shift. And the number of shift pixels is direct proportion to the distance between the depth value of pixel and zero-plane. On the contrary, we can get the left eye image in the opposite direction. Fig. 4-3 is the example of the result after linear shift algorithm.

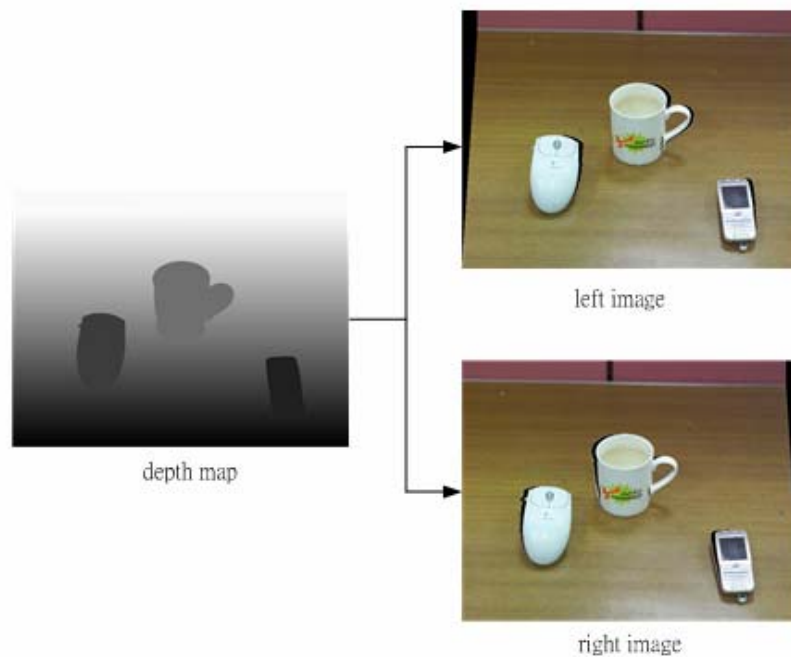


Fig. 4-3 The results after linear shift algorithm.

4.1.3 Binocular Vision Shift Algorithm

We propose a method that simulates the human vision. The scene projects to the left and right eye images with an angle rotation. The diagram of simulating the right eye image

is shown in Fig. 4-4. We assume that two eyes focus on the point whose vertical coordinate is half of the maximum depth of the scene and horizontal coordinate is half of the width. We can get the projection depth (D') on the following formulas:

$$\theta_3 = \tan^{-1} \frac{\frac{b}{2}}{D + \frac{d_m}{2}} \quad (4-1), \quad \theta_4 = \tan^{-1} \frac{\frac{w}{2} + \frac{b}{2}}{D + \frac{d_m}{2}} \quad (4-2), \quad \theta_5 = \tan^{-1} \frac{\frac{w}{2} - \frac{b}{2}}{D + \frac{d_m}{2}} \quad (4-3),$$

$$\theta_1 = \theta_4 - \theta_3 \quad (4-4), \quad \theta_2 = \theta_1 + \theta_3 \quad (4-5), \quad D' = \frac{w}{\tan \theta_1 + \tan \theta_2} \quad (4-6)$$

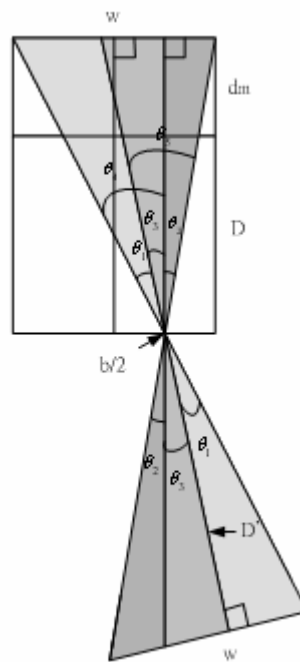


Fig. 4-4 Simulating the right eye image of binocular vision. w : image width, d_m : scene depth, D : the depth between right eye and scene, b : the distance between two eyes, D' : the projection depth.

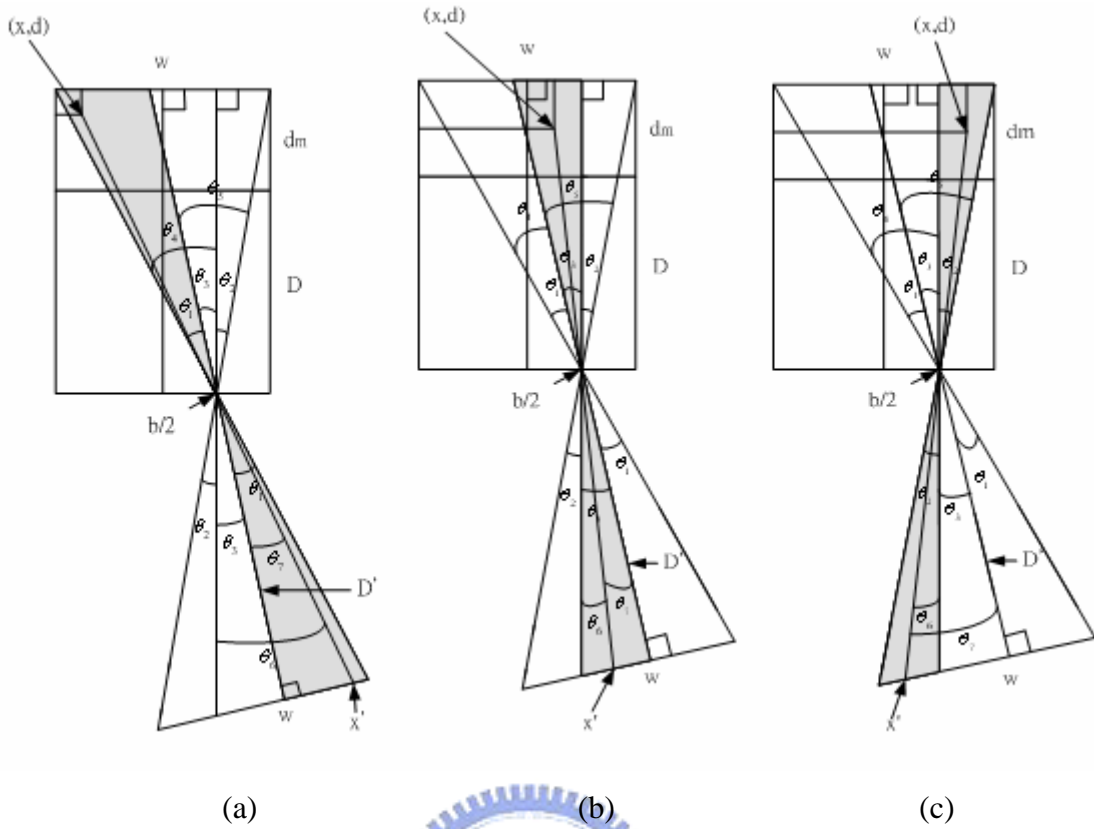


Fig. 4-5 The projection point of relative regions.

(a) Region 1, (2) Region 2, (3) Region 3.

After getting the projection depth (D'), we can infer the relative point (x, d) individually in region 1 (Fig. 4-5a), region 2(Fig. 4-5b), and region 3(Fig. 4-5c).

And followings are the results:

Region 1:

$$\theta_1 = \tan^{-1} \frac{\frac{b}{2} + \frac{w}{2} - x}{D+d} \quad (4-7)$$

$$\theta_7 = \theta_1 - \theta_3 \quad (4-8)$$

$$x' = D'(\tan \theta_3 + \tan \theta_7) \quad (4-9)$$

Region 2:

$$\theta_6 = \tan^{-1} \frac{\frac{b}{2} + \frac{w}{2} - x}{D+d} \quad (4-10)$$

$$\theta_7 = \theta_3 - \theta_4 \quad (4-11)$$

$$x' = D(\tan\theta_3 - \tan\theta_7) \quad (4-12)$$

Region 3:

$$\theta_6 = \tan^{-1} \frac{x - \frac{b}{2} - \frac{w}{2}}{D+d} \quad (4-13)$$

$$\theta_7 = \theta_3 + \theta_6 \quad (4-14)$$

$$x' = D(\tan\theta_3 - \tan\theta_7) \quad (4-15)$$

And the pixel which is lower in depth map has high priority than the pixel which is higher in the depth map, because the near objects are in front of the far objects in the 3D view. When we get the projection image, we mirror the image in the horizontal direction. On the contrary, we can get the left eye image in the opposite direction. And the projection image has rotation effect and less parallax. We can see the results in Fig. 4-6.

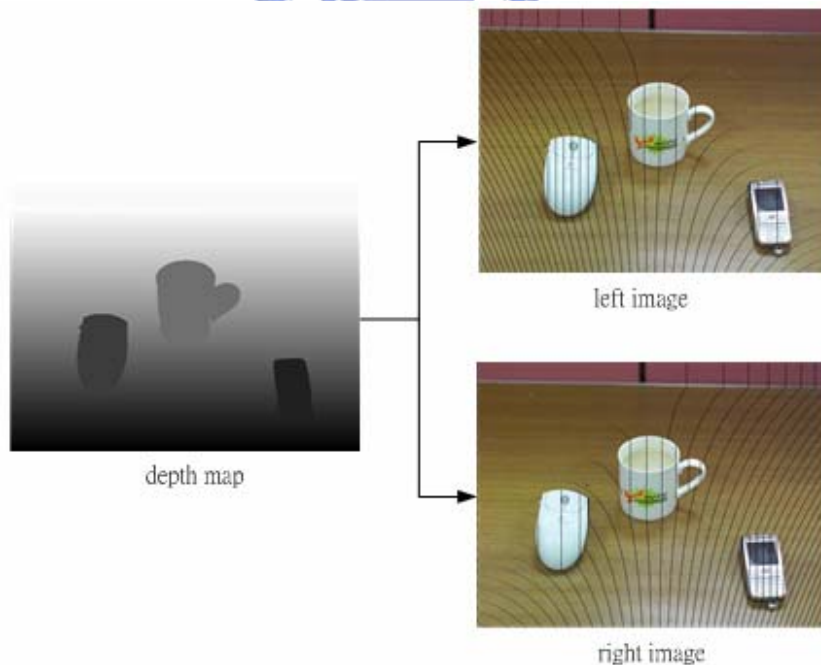


Fig. 4-6 The results after binocular vision algorithm.

4.2 Interpolating Holes Algorithm

After getting the left and right eye images from the depth map, there are some holes in the left and right eye images. Due to we generate two images from one image. We tried some kinds of interpolating holes algorithm. We finally found that interpolating by horizontal smooth filter has better effect in 3D effect images.

We proposed a horizontal smooth filter. The flowchart is shown in Fig. 4-7. If the size of hole is one pixel, we use the average of the left side and right side pixel to interpolate it. And if the size of hole is more than one pixel, we will mirror the hole with the side which is lower in depth map. The reason that we use mirror is that we can feel continuous in the edge of the holes. And the reason that we use the side which is lower in depth map is that we use bottom lines to give objects depths. It illustrates in Fig. 4-8. We can see that in Fig.4-8(a) the right side of the holes is lower in the depth maps. So we interpolate the holes with the pixels of the right side (Fig. 4-8(b)).

Fig. 4-9 illustrates the methods of mirror, shift, and corner. The black blocks are the holes. When we use the methods of mirror and shift, we mirror and shift the near pixels. And when we use the method of corner, we interpolate the hole with the average of the left and up pixels from the left-up corner to the right-down corner of the image.

Fig. 4-10 shows the results of these interpolating methods. We can see that the pixels on the edge are continuous when we use the mirror method. It is discontinuous when we use the shift method. When we use the corner method, there are some holes remaining on the left-up corner of the image. So, we choose the method of mirror.

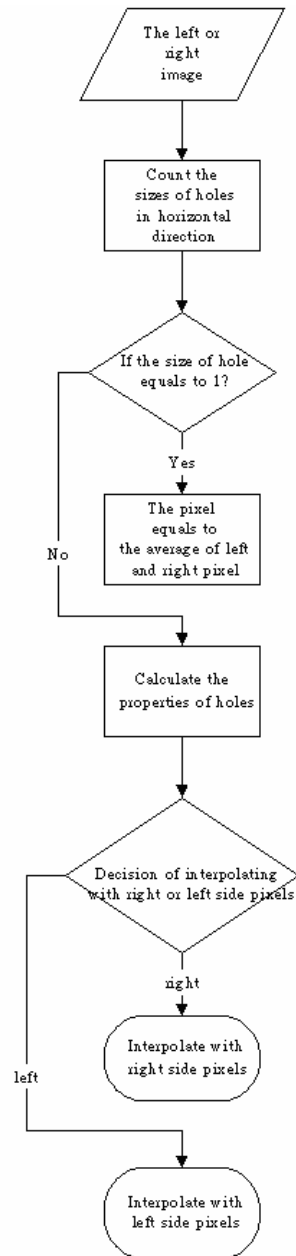


Fig. 4-7 The flowchart of interpolating by horizontal smooth filter.



Fig. 4-8 Interpolating left or right side pixels. (a) Depth map. (b) Interpolating holes.

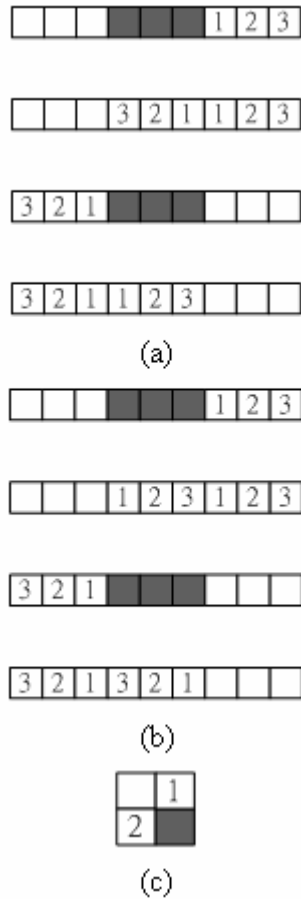


Fig. 4-9 Illustration of interpolating holes method.
 (a) The method of mirror. (b) The method of shift. (c) The method of corner.

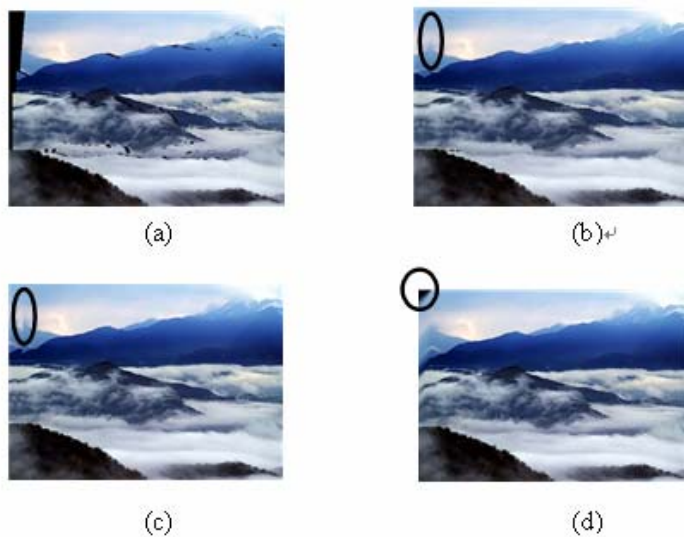


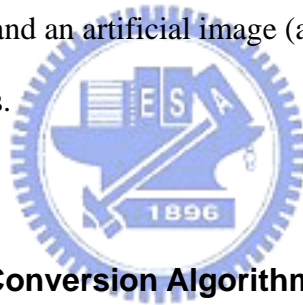
Fig. 4-10 The results of interpolating holes method.
 (a) The image with holes. (b) Interpolating by mirror.
 (c) Interpolating by shift. (d) Interpolating by corner

Chapter 5 Experimental Results

In the chapter, we demonstrate our experimental results. We show the experimental results of our proposed methods. We show the experimental results of 2D to 3D image conversion algorithm by linear shift algorithm and binocular vision shift algorithm. We also show the experimental results of the SANYO method.

5.1 Experimental Results of Our Proposed Methods

In this section, we demonstrate our experimental results of our propose method. We show a natural image (clouds) and an artificial image (a table) according to the order of the process of our propose methods.



5.1.1 The 2D to 3D Image Conversion Algorithm by Using Linear Shift

Algorithm

In this section, we demonstrate our 2D to 3D image conversion algorithm by using linear shift algorithm. Fig. 5-1a and Fig. 5-2a are the original image. Fig. 5-1b and Fig. 5-2b are the images after segmentation without spatial information. Fig. 5-1c and Fig. 5-2c are the segmentation images after size filter. Fig. 5-1d and Fig. 5-2d are the depth maps. We can see that in Fig. 5-1d the depth of table is gradual and depths of the objects on the table are fixed. In Fig. 5-2d the depths of clouds are gradual and mountains are fixed. So on the 3D monitor, we can see the mountains stand on the clouds and the objects stand on the table. Figs. 5-1e~5-1h and Figs. 5-2e~5-2h are the left and right eye images of the linear shift

algorithm and interpolating holes algorithm. We can see the images have intense shift.

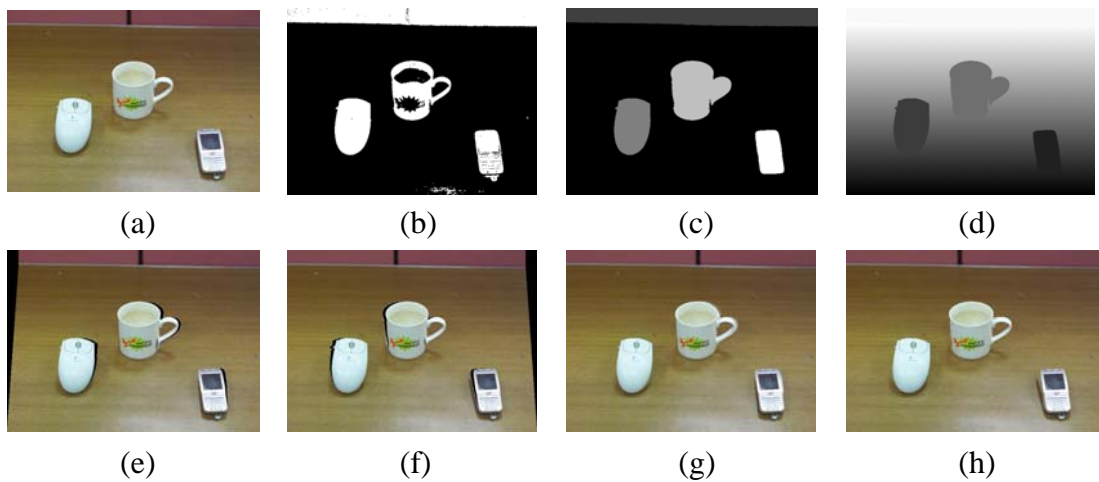


Fig. 5-1 The 2D to 3D image conversion process of an artificial image by linear shift algorithm. (a) The original image. (b) The image after segmentation. (c) The image after size filter. (d) The depth map. (e) The left eye image generating from linear shift algorithm. (f) The right eye image generating from linear shifter algorithm. (g) The final left eye image by linear shift algorithm. (h) The final right eye image by linear shift algorithm.

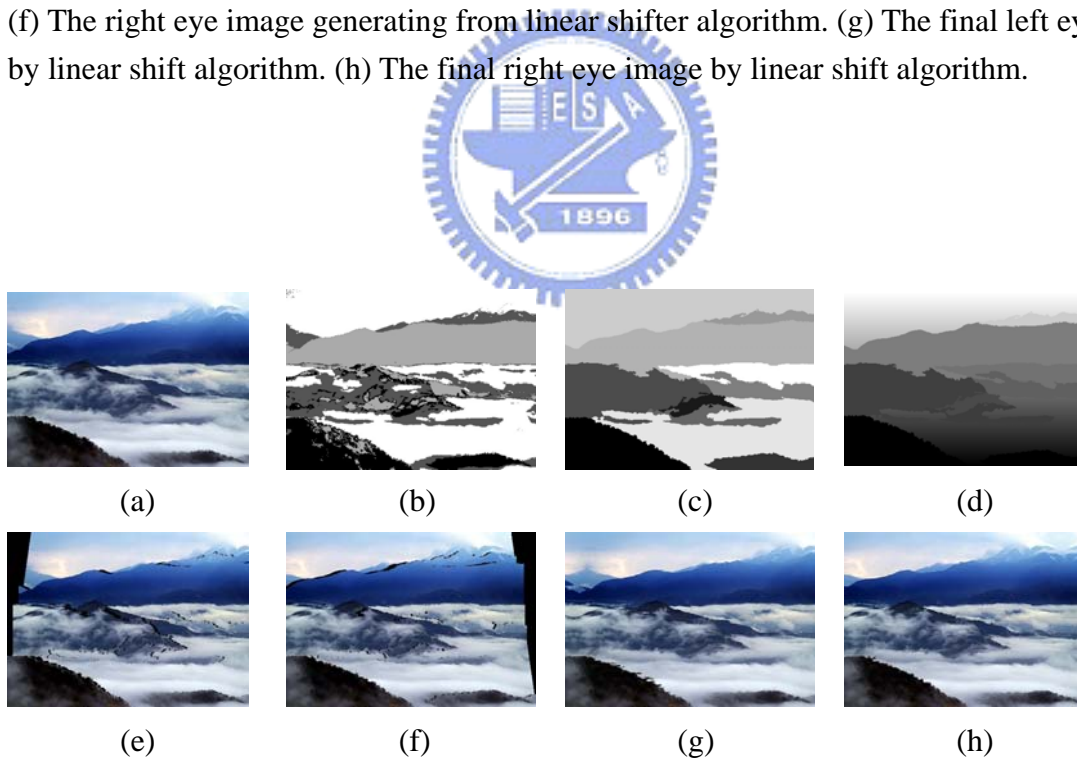


Fig. 5-2 The 2D to 3D image conversion process of a natural image by linear shift algorithm. (a) The original image. (b) The image after segmentation. (c) The image after size filter. (d) The depth map. (e) The left eye image generating from linear shift algorithm. (f) The right eye image generating from linear shifter algorithm. (g) The final left eye image by using linear shift algorithm. (h) The final right eye image by using linear shift algorithm.

5.1.2 The 2D to 3D Image Conversion Algorithm by Using Binocular Vision

Shift Algorithm

In this section, we demonstrate our 2D to 3D image conversion algorithm by using binocular vision shift algorithm. Fig. 5-3a and Fig. 5-4a are the original image. Fig. 5-3b and Fig. 5-4b are the images after segmentation without spatial information. Fig. 5-3c and Fig. 5-4c are the segmentation images with size filter. Fig. 5-3d and Fig. 5-4d are the depth maps. Figs. 5-1e~5-1h and Figs. 5-2e~5-2h are the left and right eye images of the binocular vision shift algorithm and interpolating holes algorithm. In Figs. 5-3e~5-3f and Figs. 5-4e~5-4f, we can see the images have rotation effects in horizontal objects. It simulates binocular vision that two eyes have vision angles. We can see the image has slow shift. And it has less parallax than using linear shift algorithm. With less parallax, we would feel more comfortable on the 3D monitor.

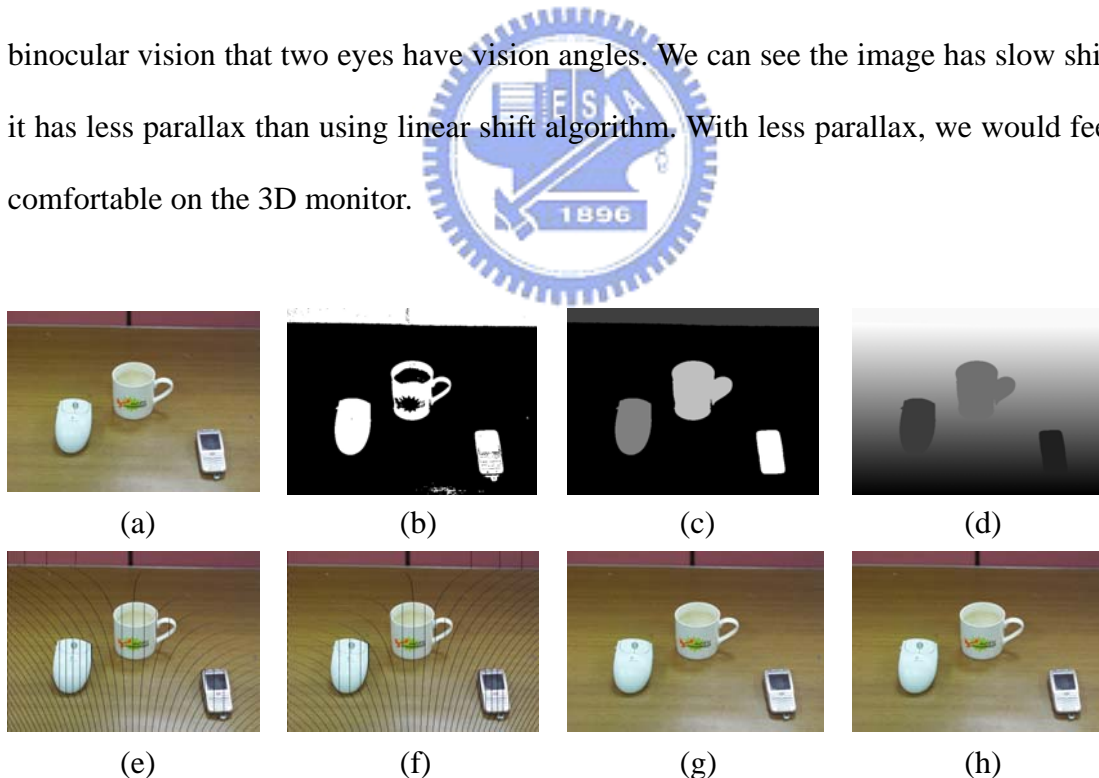


Fig. 5-3 The 2D to 3D image conversion process of an artificial image by binocular vision shift algorithm. (a) The original image. (b) The image after segmentation. (c) The image after size filter. (d) The depth map. (e) The left eye image generating from binocular vision shift algorithm. (f) The right eye image generating from binocular vision shift algorithm. (g) The final left eye image by binocular vision shift algorithm. (h) The final right eye image by binocular vision shift algorithm.

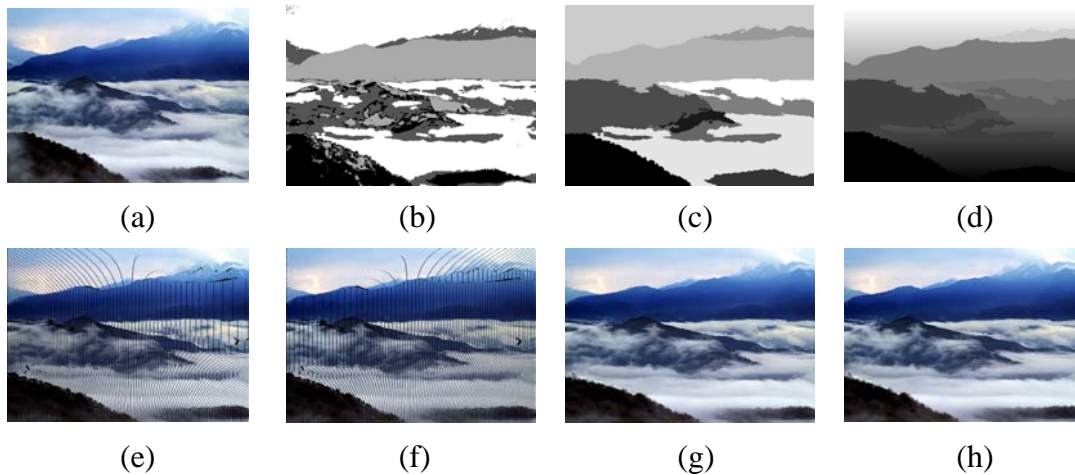


Fig. 5-4 The 2D to 3D image conversion process of a natural image by binocular vision shift algorithm. (a) The original image. (b) The image after segmentation. (c) The image after size filter. (d) The depth map. (e) The left eye image generating from binocular vision shift algorithm. (f) The right eye image generating from binocular vision shift algorithm. (g) The final left eye image by binocular vision shift algorithm. (h) The final right eye image by binocular vision shift algorithm.

5.2 Experimental Results of the SANYO Method

We also implement the “Computed Image Depth Method” (CID) algorithm [2] to compare with our proposed methods. The CID consists of the following two processes. One is the image depth computation process that computes the image depth parameters with the contrast, the sharpness and the chrominance of the input images. Another is the 3D image generation process that generates the 3D images according to the image depth parameters. Thus, we first calculate the sharpness and contrast value every block. And classify every block into three depth degree. Then group regions by color, and do some compensation with the positions of objects. When we get image depth parameter, we can generate the left and right eye images. If the parameter of an area indicates near, the left images are made by shifting the input images to the right, and the right images are made by shifting to the left. If the parameter of an area indicates far, both images are made by shifting to each opposite direction. At last, we use the smooth filter to interpolate the holes.

The results are show in Fig. 5-5 and Fig. 5-6. Because CID algorithm extracts the depth value every block and doesn't segment the image, we get broken objects. So on the 3D monitor, we feel uncomfortable because of the broken blocks. Figs. 5-5e~5-5f and Figs. 5-6e~5-6f are the left and right eye images after shift. Figs. 5-5g~5-5h and Figs. 5-6g~5-6h are the left and right final result images after the smooth filter.

We also use the depth index to evaluate Fig.5-5(a) by our method with CID algorithm. The depth accuracy is 100% of our proposed method and 33.33% of CID algorithm.

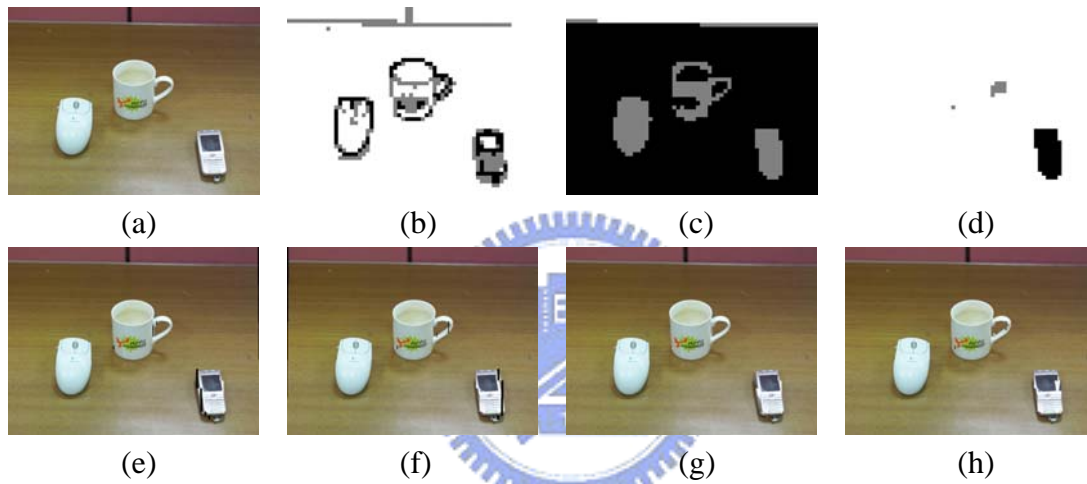


Fig. 5-5 The 2D to 3D image conversion process of an artificial image by CID algorithm. (a) the original image, (b) depth extraction from every block, (c) color grouping, (d) the depth map, (e) the left eye image from shift, (f) the right eye image from shift, (g) the final left eye image, (h) the final right eye image.

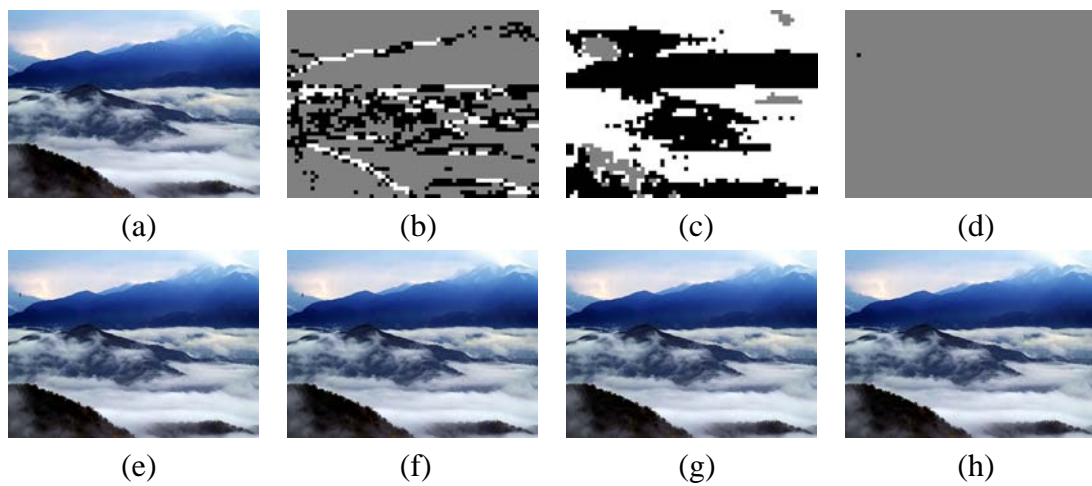


Fig. 5-6 The 2D to 3D image conversion process of a natural image by CID algorithm. (a) the original image, (b) depth extraction from every block, (c) color grouping, (d) the depth map, (e) the left eye image from shift, (f) the right eye image from shift, (g) the final left eye image, (h) the final right eye image.



Chapter 6 Conclusions and Future Works

We proposed a novel 2D to 3D images conversion algorithm which automatically converts a single 2D image into the 3D effect images.

Compared our proposed algorithm with CID algorithm, we can find that our proposed algorithm has these advantages:

(1) This algorithm which combines image segmentation with depth extraction, so we can see the objects more complete on the 3D display.

(2) In the part of image segmentation, we proposed an image segmentation method which contains SSR (Single Scale Retinex) and the size filter. This image segmentation method is suitable for the 3D effect images.

(3) We proposed a novel method of depth extraction which could infer the object is horizontal or vertical, so we can feel more real stereo effect than CID algorithm on the 3D display.

(4) In order to reconstruct the left and right eye images from the depth map, we simulate the binocular vision to get the left and right eye images from the depth image. Thus, we feel more comfortable about the 3D effect images on 3D display than CID algorithm.

In the future work, we could apply our algorithm on the 3D effect video. We can separate our algorithm into two parts. One part is image segmentation and the other is the part after image segmentation. We can combine some video segmentation methods with the part after image segmentation of our algorithm. Thus, we can generate the 3D effect video. We think that we can add more depth cues to depth extraction. Thus, we can get more correct 3D effect images.

REFERENCES

- [1] H. Murata, Y. Mori, S. Yamashita, A. Maenaka, S. Okada, K. Pyamada, and S. Kishimoto, "A real- Time Image Conversion Technique Using Computed Image Depth," *SID SYM*, Vol. 29, pp. 919-922.
- [2] T. Okino, H. Murata, K. Taima, T. Iinuma, and K. Oketani, "New television with 2D to 3D image conversion technologies," *Proc. SPIE Stereoscopic Displays and Virtual Reality Systems III*, Vol. 2653, pp. 96-103, 1996.
- [3] T. S. Jebara, and A. Pentland, "Parametrized structure from motion for 3D adaptive feedback tracking of faces," *Proceedings IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 144-150, 1997.
- [4] E. Steinbach, A. Hanhalic, and B. Girod, "3D Motion and Scene Structure Estimation with Motion Dependent Distortion of Measurement Windows," *Proceedings IEEE International Conference on Computer Vision*, Vol. 1, pp. 434-422, 1999.
- [5] C. T. Lin, and W. C. Cheng, "A Novel Post-Nonlinear ICA-Based Reflectance Model for 3D Surface Reconstruction," *IEEE International Symposium on Circuits and Systems*, May, 2005.
- [6] A. Torralba and A. Oliva, "Depth Estimation from Image Structure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 9 September 2002.
- [7] S. Curti, D. Sirtori, and F. Vella "3D Effect Generation from Monocular View," *IEEE Proceedings of the First International Symposium on 3D Data Processing Visualization and Transmission (3 DPVT'02)*, 2002.
- [8] H. D. Cheng, X. H. Jiang, Y. Sun, and J. L. Wang, "Color image segmentation: Advances and prospects", *Pattern Recognition*, 34(12), 2001.
- [9] I. P. Howard, J. R. Brian, and A. P. Howard, *Binocular Vision and Stereopsis*, Oxford University Press, Chapter 7, 1995.
- [10] R. Klette, K. Schluns, and A. Koschan, *Computer Vision*, Springer, pp. 29 - 33, 1998.
- [11] E. Trucco, and A. Verri , *Introductory Techniques for 3D Computer Vision*, Prentice Hall, pp. 140-143, 1998.
- [12] R. Z. Liu, "The design of stereo image on 2D monitor", The Master Thesis, 2001.
- [13] N. Pal and S. Pal, "A review on image segmentation techniques," *Pattern Recognition.*, vol. 26, no. 9, pp. 1277-1294, 1993.
- [14] C. T. Lin, Y. C. Lee, and H. C. Pu, "Satellite sensor image classification using cascaded architecture of neural fuzzy network," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 38, No. 2, pp.1033-1043, March 2000.

- [15] C. T. Lin, W. C. Cheng, and S. F. Liang, "An on-line ICA-mixture-model-based self-constructing fuzzy neural network," *IEEE Transactions on Circuits and Systems I*, vol. 52, no. 1, pp. 207-221, 2005.
- [16] V. Mezaris, I. Kompatsiaris, and M. G. Strintzis, "Still image segmentation tools for object-based multimedia applications," *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 18, No. 4, PP. 701-725, June 2004.
- [17] E. Littmann and H. Ritter, "Adaptive color segmentation—a comparison of neural and statistical methods," *IEEE Trans. Neural Networks*, vol. 8, Jan. 1997.
- [18] T. Uchiyama and M. A. Arbib, "Color image segmentation using competitive learning," *IEEE Trans. Pattern Anal. Machine Intelligence*, vol. 16, pp. 1197-1206, Dec. 1994.
- [19] H. D. Cheng, and Y. Sun, "A hierarchical approach to color image segmentation using homogeneity", *IEEE Trans. on Image Processing*, Vol. 9, No. 12, Dec 2000, p2071-2082.
- [20] D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *IEEE Trans. Image Processing*, vol. 6, pp. 451-462, Mar. 1997.
- [21] D. C. Tseng and C. H. Chang, "Color segmentation using perceptual attributes," in *Proc. of 11th International Conf. on Pattern Recognition*, vol.C, pp.228-231, 1992.
- [22] J. Tajima, "Uniform color scale applications to computer graphics," *Computer Vision, Graphics, and Image Processing*, vol.21, pp.305-325, 1983.
- [23] D. C. Tseng, Y. F. Li, and C. T. Tung, "Circular histogram thresholding for color image segmentation", *Proceedings of the Third International Conference on Document Analysis and Recognition*, Vol. 2, p. 673, August 14-155, 1995.
- [24] N. Li and Y. F. Li, "Feature encoding for unsupervised segmentation of color images," *IEEE Trans. On System Man and Cybernetics- Part B: Cybernetics*, Vol. 33, No. 3, June 2003
- [25] C. T. Lin and G. Lee, *Neural Fuzzy Systems: A Neuro-Fuzzy Synergism to Intelligent Systems*, Prentice Hall, 1996.
- [26] D. Marr and T. Poggio, "Cooperative computation of stereo disparity," *Science*, 194:283-287, 1976.
- [27] S. Barnard and M. Fischler, "Computational stereo," *Computing Surveys*, 14(4):553-572, December 1982.
- [28] A. Pentland, "A new sense for depth of field," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 9(4):523-531, July 1981.
- [29] G. Surya and M. Subbarao, "Depth from defocus by changing camera aperture: a spatial domain approach," *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, pp. 61-6, 1993.
- [30] N. Gershon, "Visualizing 3D PET images," *IEEE Computer Graphics and Applications*,

11(3):11-13, 1991.

- [31] S. Marapane and M. Trivedi, "An active vision system for depth extraction using multi-primitive hierarchical stereo analysis and multiple depth cues," *SPIE*, 1956:250-262, 1993.
- [32] R. Surdick, E. Davis, R. King and G. Corso, "Relevant cues for the visual perception of depth: is where you see it where it is?" *Proc. 38th mtg. Human Factors and Ergonomics Society*, pp. 1305-1309, 1994.
- [33] R. Srinivasan, M. Shridhar, and M. Ahmadi, "Computing surface information from monocular and binocular cues for vision applications," *Proc. 27th Conf. Decision and Control*, pp. 1085-1089, 1988.
- [34] S. Das and N. Ahuja, "Dynamic integration of visual cues for position estimation," *SPIE*, 1382:341-352, 1990.

