

國立台灣師範大學
資訊工程研究所碩士論文

指導教授： 陳世旺 博士

由 RGB-D 影像資訊產生多視角立體電影用之
影像

**Multiview Stereo Images Generation from
RGB-D Images**

研究生： 劉貞佑 撰
中華民國 一〇三 年 六 月

摘要

由 RGB-D 影像資訊產生多視角立體電影用之影像

劉貞佑

目前 3D 顯示技術已是很成熟的技術，然而目前可用的 3D 內容來源有限，造成推廣難度較高，即使用戶有經濟能力購買高檔 3D 影音設備也受限於可用的數位內容不夠多，讓 3D 功能等同虛設。本研究提出的將 RGB+D 影像轉換成 3D 立體內容影像即為了舒緩部分立體內容不足的問題。

近年來有許多關於如何產生 3D 立體影像的研究，而從事這類研究，無可避免常會面臨到深度資訊估測以及產生新視角影像後修補的問題。深度估測的方式很多，有人工判斷、有根據深度線索判斷或是使用深度攝影機取得，人工判斷及根據深度線索判斷都相較耗時，尤其根據深度線索做判斷誤差的機率也較高。而深度攝影機則避免了深度資訊取得的困難，其誤差也較根據深度線索判斷來的低。然而使用深度攝影機取得的影像會有一個很大的問題，依照拍攝的情景不同，其影像可能會有不同程度的深度影像破洞，而深度資訊的完整及精準度直接影響了立體內容的好壞，因此需要對深度影像做合理的填補。而影像填補在過去亦有許多研究，主要考慮了顏色、紋理結構等因素，本研究使用兩種方式作處理，第一種是利用矩陣完整化技術的修補方式；第二種是使用以影像分割為基礎的修補方式，其中矩陣完整化是依據影像本身低 rank 的性質對影像做合理的填補，而基於影像分割的修補方式則是考慮顏色跟空間上位置的關係對影像作分割後再做填補。

實驗結果顯示我們的立體內容深度層次感相較於 2D 轉 3D 的立體內容明顯更佳，且相較於 2D 轉 3D 的技術，我們的研究使用的是深度攝影機所取得的深度資訊，因此深度資訊具有較低的誤差，產生出的立體影像能給予觀看者更好的立體感受。

關鍵字：2D 轉 3D、矩陣完整化、影像分割、3D、影像修補、深度影像繪圖法

Abstract

Multiview Stereo Images Generation from RGB-D Images

Chen-Yu Liu

Nowadays, 3D display technology has been well developed and gradually became a matured technology. However, limited 3D content resources obstruct this technology to be popularized to the market. Even if the customers can afford expensive media equipment, there is still lack of useable resources to function 3D display technology. This research provides the solution of converting RGB+D image to 3D image to partially improve the shortage of 3D resources.

In recent decades, many researches are already working on how to create 3D images, which always involved depth measurement and generating image with another perspective. Depth measurement can be done by implementing the solutions such as manual judgments, depth cues, or using depth cameras. The former two solutions are relatively time consuming than the latter one. Especially the depth cues usually cause inaccuracy. Moreover, using depth cameras simplifies the difficulties of getting the depth data and decreases the inaccuracy as well. But there is a problem when using the cameras to collect the depth data, the images may have holes occurs which depends on shooting scenarios. The depth data need to be repaired under a reasonable condition

because these two factors impact the 3D images' qualities. In the past, solution to image inpainting has been proposed from many researches. The main considerations are about the colors and the texture. This research implements two methods to process the missing value of depth images. One is based on images' low rank feature to use matrix completion technique; the other is based on image segmentation technique to do

the depth image repairing.

The results of experiment show that our 3D depth quality is obviously higher than the traditional 2D convert to 3D method. Furthermore, depth camera collects the depth data with higher accuracy so we can provide viewers a better experience in 3D display technology.

Key word: 2D to 3D、matrix completion、image segmentation、3D、image inpainting、DIBR



誌謝

在國立台灣師範大學的這兩年日子真的學到了很多、認識了很多很有才能的人也得到了很多人的幫助。首先，非常感謝我的指導教授陳世旺老師，在這兩年帶領我走進廣闊的影像處理領域，老師不只指導了我許多相關知識，更教會了我許多人生態度。這兩年來我開始頻繁接觸過往一直不擅長的數學運算，老師也總在我真的遇到困難的時候幫助我一起思考、跨越障礙。做研究的時候很容易使人灰心，因為有可能會一直嘗試一直失敗，但是老師總是鼓勵我們不要氣餒、要勇於嘗試和學習新事物，這兩年來學到的事物真的很寶貴。

感謝 IPCV 實驗室的所有夥伴彼此扶持努力，梁祐銘老師、王俊明學長常常回來跟跟我們團咪研究與討論，羅安鈞學長更是在這一年多來耐心地引領我步入他所擅長的 3D 領域，讓我在做研究的過程中心中一直比較踏實。許之凡學長及呂佳儒學姐也常常會細心聆聽我們研究上遇到的困難，給予我們很實際的建議與方向，教導我們他們在碩士所學到的寶貴經驗，讓我們少走了很多冤枉路。感謝俊宇、軒嘉兩年來互相扶持，有你們的日子充滿溫馨跟笑聲，讓這一路走來雖然疲憊卻依然滿載笑容。感謝淳雅學妹、昆賢學弟、冠宇學弟總是我們需要幫助的時候鼎力支持，在我們需要歡笑的時候帶給我們快樂的氣氛，你們真的好可愛好貼心，謝謝你們。

也謝謝 CVIU 實驗室的巧珊、KDD 實驗室的舜宸及懿萱，這兩年來與我分擔、分享生活中遇到的喜怒哀樂，有了你們讓我一路走來總是戰力十足。謝謝李皓這兩年來常常陪著我熬夜趕工，分擔我的心情與壓力，當我的後盾。最後，謝謝我最愛的爸爸、媽媽及哥哥，一直是我最大的心靈支柱，在我成長的歲月裡一直給予我很多很多的鼓勵與安慰，讓我可以樂觀的面對人生遇到的一切難題。

目錄

摘要	II
ABSTRACT	III
誌謝	V
目錄	VI
圖目錄	VIII
表目錄	X
第一章 緒論	1
第一節 研究動機	1
第二節 文獻探討	2
第三節 研究方法	9
第四節 論文架構	10
第二章 系統的架構與流程	11
第一節 系統架構	11
第二節 系統運作流程	12
第三章 深度影像的修補	17
第一節 矩陣完整化之修補技術	17
第二節 基於影像分割之修補技術	24
第四章 多視角影像的產生及處理	28

第一節 深度圖的前處理	28
第二節 3-D IMAGE WARPING	30
第三節 空洞的填補	33
第五章 實驗結果	35
第一節 實驗工具	35
第二節 深度圖修補技術的比較	36
第三節 3D 立體影像的結果比較	39
第六章 結論及未來方向	42
參考文獻	44



圖目錄

圖 1.1：雙眼視差。	3
圖 1.2：左右眼所見之影像。	3
圖 1.3：大腦接收的影像。	3
圖 1.4：立體顯示技術(STEREOSCOPIC DISPLAY TECHNOLOGY)的分類。	7
圖 1.5：DIBR 的系統流程。	9
圖 2.1：系統架構圖。	11
圖 2.2：系統流程圖。	12
圖 2.3：KINECT 感測器。	12
圖 2.4：(A) 彩色攝影機取得之彩色影像，(B) 深度攝影機(未經 REGISTRATION)所取得之深度影像， (C) 圖(A)彩色影像跟圖(B)深度影像的疊合影像。	13
圖 2.5：(A) 彩色攝影機取得之彩色影像，(B) 深度攝影機(經 REGISTRATION)所取得之深度影像， (C) 圖(A)彩色影像跟圖(B)深度影像的疊合影像。	14
圖 2.6：(A) 利用矩陣完整化的技術修補深度圖，(B)平滑化修補完的深度圖。	15
圖 2.7：(A) 利用基於影像分割的技術修補深度圖，(B)平滑化修補完的深度圖。	15
圖 2.8：MULTIVIEW 示意圖。	16
圖 2.9：MULTIVIEW 示意圖(放大)。	16
圖 3.1：(A)有缺值之不完整矩陣，(B)缺值填補後之矩陣。	17
圖 3.2：(A)一張 588×453 的影像，(B)R-CHANNEL 的奇異值分解，(C)G-CHANNEL 的奇異值分解， (D)B-CHANNEL 的奇異值分解。	18
圖 3.3：深度資訊做奇異值分解。	19
圖 3.5：減少 SUPERPIXEL 搜尋範圍。左圖為傳統K-MEANS 的演算法，右圖為 SLIC 的演算法，只需要在 一個限制的區域中做尋找。	25
圖 3.6：對彩色影像做 SLIC 分割後的結果。	26

圖 3.7：將 SLIC 分割後的結果對應到深度影像。.....	27
圖 3.8：左圖為原有缺值的深度圖，右圖為修補後的結果。.....	27
圖 4.1：DIBR 的系統流程。.....	28
圖 4.2：視差平面。.....	29
圖 4.3：8-BIT 的深度圖。.....	29
圖 4.4：經 BILATERAL FILTER SMOOTH 後的深度圖。.....	30
圖 4.5：產生虛擬立體影像的攝影機設置圖。.....	31
圖 4.6：(A)原本的彩圖，(B)深度圖未經平滑化處理 WARPING 後的結果，(C)深度圖經 BILATERAL FILTER SMOOTH 處理 WARPING 後的結果。.....	32
圖 4.7：填補原理。.....	33
圖 5.1：可接受的 2D 轉 3D 的輸入格式。.....	35
圖 5.2 深度影像修補的比較，(A)原始影像，(B)遮蓋掉一些資訊後的深度影像，(C) MATRIX COMPLETION VIA TNNR –METHOD A，(D) MATRIX COMPLETION VIA TNNR –METHOD B，(E)IMAGE SEGMENTATION，(F)[CAI08]SVT，(G)[BER01]，(H)[TEL04]。.....	37

表目錄

表 1.1 常見之深度線索特徵	4
表 1.2 立體影像顯示技術	6
表 1.3 CSV 與 2D+D 之比較	8
表 5.1 深度影像修補的評估	38
表 5.2 3D 立體影像是否俱有由近而遠的層次感	40
表 5.3 3D 看起來的舒適度	40
表 5.4 深度感正不正確	41
表 5.5 影像邊緣重疊	41

第一章緒論

第一節研究動機

這些年來顯示器設備持續不斷地精進改善，從早期的黑白電視到後來的彩色電視，再到高畫質(High Definition, HD)電視。尤其近年來，立體電視與電影顯示器蓬勃發展，例如：日本為了推廣 3D 顯示器，以顯示器廠商為主，聯合其他相關製造商、發展商等 70 多家公司成立了“3D 聯盟”，共同研發 3D 顯示技術的產品及應用；又韓國政府提出了“2010 年 3D 視覺”政策，計畫到了 2010 年，能實現大多數顯示產品和記錄設備與 3D 立體格式的轉換；台灣也創立了“台灣 3D 互動影像顯示產業協會”積極整合面板、系統、晶片等不同領域廠商一同建立 3D 軟體的產業鏈。然而隨著 3D 立體顯示器的發展，業者及消費者開始意識到市面上沒有夠多的 3D 立體內容可供使用者選擇，其中最主要的原因乃是目前立體影像擷取設備的發展遠遠落後於立體顯示器的進展。目前市面上雖然有非常平價的雙鏡頭手機立體內容擷取設備（例如：LG Optimus 3D、HTC EVO 3D），但所獲得的景深資訊，其精確度與解析度均不高，而兩者都要高的設備，其價格動則百萬，不只如此，其操作難度也高，這些都是造成立體內容取得不易的原因，而立體內容的缺乏成為進入 3D 多媒體時代的最大阻礙。

立體內容的來源主要分成三種：1.拍攝時直接使用多台彩色攝影機排列成陣列取得多對影像，因此不需透過深度值的估測即可生成立體影像 2.使用彩色攝影機搭配深度攝影機（例如：Kinect 感測器）做拍攝，再利用其取得的深度資訊及彩色影像來產生立體影像 3.將舊有的 2D 影像做深度估測，再結合原彩色影像來產生立體影像。尤其過去所收集的龐大 2D 數位內容（例如：Discovery Channel、National Geographic、Animal Planet 及 BBC 等等這些科學、歷史、國家地理的紀錄片），其取得並不容易且內容亦極為珍貴，因此有許多人開始轉向思考如何將其轉換成 3D 立體內容，如此至少可以部分解決立體內容缺乏的問題。[Feh04][賴 99]研究中指出，2D 至 3D 影像轉換的過程首先是由已知的 2D 影像推估景物的深度，再結合原 2D 影像建立起立體影像。而產生立體影像的關鍵乃在於建

立視差資訊（因兩眼所見之影像不同），視差資訊可由深度圖轉換求得。所以當給予一 2D 影像，又我們知道其相對應的深度影像後，即可在虛擬平面上根據深度資訊將給予的 2D 影像像素重新投影(re-projection)以產生出相鄰視角的影像，也就是產生兩眼中另一眼所見之影像。估測景物的深度資訊在過去已有許多人在做相關的研究，主要分成三類，分別為 1. 全自動：全部的 frames 都由程式自動做深度估測; 2. 半自動：部分 frames 的深度估測由人工作判斷，其他部分則由程式自動估測; 3. 全人工：全部的 frames 都由人工作深度估測的動作。其中深度估測的過程中越涉入人工的參與，其費用也就越貴，然而所得的立體內容也越佳。若一部影片的深度資訊全由人工標記，其所費不貲，因此很多人開始研究半自動和全自動的深度估測方式如：[Ide08] [Bat04] [Com97] [Ang10] [Che10] Depth from cues、[Li06] Structure from Motion(SfM)、[Har02] Machine Learning Algorithm(MLA) 等多種深度估測方式。

近年來因有許多深度感測器發展，大大簡化了深度資訊取得的困難，進而促進了產生 3D 內容的研究發展，如：2010 年微軟(Microsoft)開發的 Kinect 感測器，使用者可於拍攝時同時取得景物影像(RGB)及景深影像(Depth)，統稱為 RGB-D 資訊; 2011 年華碩(Asus)展出的 WAVE Xtion，可提供景深(Depth)的資訊。

第二節 文獻探討

一、深度線索

[劉 10][Rob88]提到人類一般是藉由多種深度線索(Depth cues)的組合來判斷物體在空間中的位置，因此 Depth from cues 是眾多深度估測方式中很常被提出來的方式。而深度線索主要分成雙眼線索(binocular cue，或稱生理視深線索)及單眼線索(monocular cue，或稱心理視深線索)。例如：雙眼視物時，可由兩眼彼此之間的協調作用來獲得深度資訊，如圖 1.1 所示，當零視差(zero parallax)時會覺得物體在圖中的 screen 上面，正視差(Positive parallax)會覺得物體在 screen 後面，負視差(negative parallax)則會覺得物體在螢幕前面，此稱為雙眼視差(binocular parallax)。

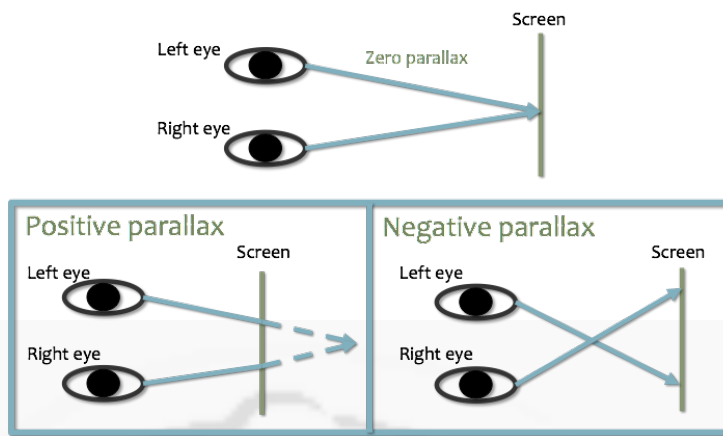


圖 1.1：雙眼視差。

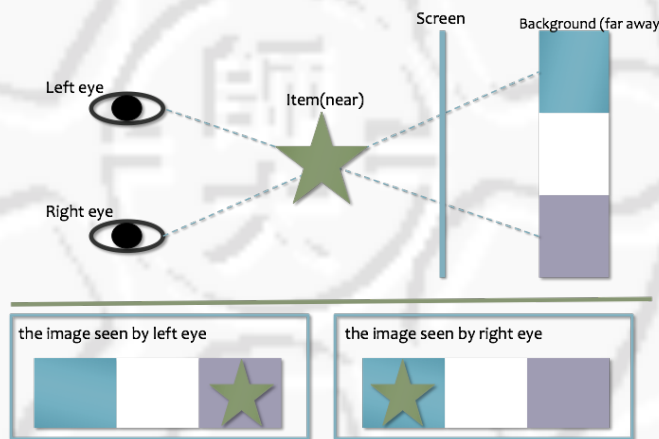


圖 1.2：左右眼所見之影像。



圖 1.3：大腦接收的影像。

雙眼在獲得視差值後即能藉此判斷深度資訊，這是因為左右兩眼間隔約為 6.5cm，使得雙眼所見的影像存在著位置與角度的差別，其中越近的物體會位移的越多，越遠的物體會位移的越少，如圖 1.2 所示。大腦會將兩眼所得的影像融合解釋出 3D 資訊，如圖 1.3 所示，此一解釋深度的過程也就是所謂的深度知覺(depth perception); 而以單眼視物時，通常可以利用物體本身的特徵得到深度知覺的線索，稱為單眼線索。過去的研究提出了許多單眼線索(depth cue) 可用，例如：當物體移動量越大的時候，會感覺物體的深度越深，稱之移動視差，亦稱運動視差; 又因空氣中微小粒子的影響使得遠處的物體看起來較模糊，稱之大氣透視; 而影像中的主軸線會交會在無窮遠處或地水線上的一點（消失點）(vanishing point)，稱之線性透視; 相同紋理的物體隨著距離改變,物體的紋路會呈現一種漸層的效果，當距離較遠時，紋理密度會感覺較為密集，稱之紋理梯度; 物體看起來較小時，會相對感覺深度較深，此深度線索受平常生活經驗認知影響很大，稱之網膜影像大小; 被遮蔽的物體，會感覺較遠，稱之遮蔽; 亮度較暗的物體，會感覺深度較深，稱之相對亮度; 高度較高的物體，會感覺深度越深。表 1.1 整理出單眼推導深度資訊時常用之線索，這對於由單張影像推估景深資訊相當重要。

表 1.1 常見之深度線索特徵

深度線索	線索之特徵
移動視差(Motion Parallax)	移動量越大的物體感覺深度越深
大氣透視(Atmospheric Perspective)	越模糊的物體感覺深度越深
線性透射(Linear Perspective)	消失點的位置感覺深度最深
紋理梯度(Texture Gradient)	紋理密度越大感覺深度越深
網膜影像大小 (Retina Size、Familiar Size)	物體越小感覺深度越深

遮蔽(Interposition、Occlusion)	被遮蔽物體感覺深度越深
相對亮度(Relative Brightness)	亮度越低的物體感覺深度越深
高度(Distance to Horizon)	物體最低點位置越高感覺深度越深

二、影像修補

影像修補(Image Inpainting)是指重建影像中丢失或損壞的部分。深度影像若缺洞不大時,可使用內插(Interpolation.)來作簡易的缺洞處理或是使用 interpolation 來解決,但是當缺洞很大的時候以上這些方法就會變得不可行,因為處理過程會使得深度影像失去很多資訊(例如:邊緣資訊),導致合成彩色影像產生誤差,因此也有許多人針對深度圖提出相關的填補方法,如 M. Camplani and L. Salgado [Cam12] 提出的 joint-bilateral filtering framework,考慮了時間和空間的資訊去對深度圖做修補; Xu *et al.* [Xu12] 提出的方式則是利用偵測移動物體,以沒有移動的物體像素為中心對空洞作 4-neighbor-pixels 的內插填補,缺點是對於處理像是鏡子、玻璃等等較高反射區域所造成的破洞此方法會較不穩定。而對於彩色影像的修補至今也有很多的研究,其中主要分為下面三類: Structural inpainting、Textural inpainting 或是[Cri04]同時結合了 Structural and Textural 兩者的影像修補方式。後來主要以第三類,同時考慮了結構與紋理的修補方式為主。

表 1.2 立體影像顯示技術

類別	原理	特色
<p>紅綠（又稱紅藍）眼鏡 (Anaglyph Glasses)</p> 	<p>將給左右眼的影像用互補色穿插顯示於同一畫面中。利用鏡片上顏色過濾使左右眼看到不同影像而形成立體視覺。</p>	<p>缺點是部分顏色被有色鏡片所濾除，影像色彩較差</p>
<p>快門眼鏡 (Shutter Glasses)</p> 	<p>透過眼鏡上的電子快門開關使交錯式的影片分別播放給左右眼觀看，又因為視覺暫留的原理，感覺起來是立體感的連續影片。</p>	<p>缺點是價格較貴</p>
<p>偏光眼鏡 (Polarization Glasses)</p> 	<p>分別投射出垂直及水平偏振的影像，而眼鏡上左右眼也分別使用垂直及水平的偏光片，如此左右兩眼可各自看到不同的影像，產生立體感。</p>	<p>可供多數人欣賞、體積輕巧、色彩較正確，現今已成為戴眼鏡式立體顯示的主流技術。</p>
<p>明暗眼鏡 (Pulfrich Glasses)</p> 	<p>明暗眼鏡是由一片深色鏡片及一片透明鏡片所組成，利用大腦對於明亮與陰暗反應速度的不同產生立體感覺。</p>	<p>用於立體電視節目、錄影帶及多媒體電腦影片。</p>
<p>頭盔式顯示器 (Head Mounted Display)</p> 	<p>直接在眼鏡上分別播放左右兩眼個別的畫面</p>	<p>缺點是造價不便宜、重量不輕且只能一人觀賞</p>

三、3D 立體顯示技術及其資料格式



圖 1.4：立體顯示技術(Stereoscopic Display Technology)的分類。

[劉 10][林 12]指出目前大部份的立體顯示器皆是利用雙眼視差來產生大腦中的立體感，將不同視角的影像利用光學設計分別投射到左右眼，再經過大腦融合即可建立出立體影像。圖 1.4 是立體顯示技術的分類，主要分成戴眼鏡式 (stereoscopic display)與裸眼式技術(auto-stereoscopic display)，目前的 stereoscopic display 都是 2-view，auto-stereoscopic display 則可實現 multi-view。目前家庭用的 3D 立體電視(3DTV)都是採用 stereoscopic display，其原理及特色整理如表 1.2，基本上都是利用眼鏡的設計以達到同一時間給予兩眼不同影像的目的，進而使得大腦建立出立體感。

而立體視訊的資料格式主要分為兩種：1. 傳統立體視訊(Conventional Stereo Video, CSV)，2. 2D 彩色影像搭配其對應的深度影像(2D+Depth)。其比較如表 1.3 所示。[賴 99]指出 CSV 在壓縮處理上比較簡單，但是資料量較大，而 2D+Depth 需注意 2D 攝影機及深度攝影機之間的校正，確保兩拍攝到的影像有相對應。

表 1.3 CSV 與 2D+D 之比較

格式	影像資訊	資料量	攝影機類型	成本
CSV	左右兩眼取得之影像	大	陣列式彩色攝影機	高
2D+Depth	彩色影像及其對應之深度影像	小	彩色攝影機配上深度攝影機	低

四、Depth Image Based Rendering (DIBR)

Fehn [Feh04] 在 2004 年首先提出了深度影像繪圖法 (Depth Image based Rendering, DIBR) (或稱虛擬視角映射技術) 的原理，使用一張原始視角的彩色影像搭配其相關的深度資訊，即可產生出多個不同視角的彩色影像以符合立體顯示器所需求的格式，其產生出的影像統稱為虛擬影像。

虛擬影像影像會因景物自身或互相遮蔽(occlusion)，而於影像中存在有一些空洞，所以需要對影像從事空洞修補的工作。後來許多研究皆基於 C. Fehn 提出的 DIBR 原理做立體內容的研究，如：L. Zhang and W.J. Tam [Zha05] 提出基於深度影像產生 3D 電視用之立體影像、S. Zinger[Zin10] 提出關於 Free-view point 深度影像繪圖法的研究、Lai *et al.* [Lai13] 提出一個有效多層次深度感的 3D 轉 3D 演算法等。

其中 Zhang *et al.* 更提出了一個簡單 DIBR 系統去實作 Fehn 所提出來的 DIBR 原理，其流程如圖 1.5 所示，包括了三個主要步驟：(1)深度圖的前處理 (preprocessing of depth map)，(2)3-D image warping 以及(3)空洞的填補(hole filling)，後來的研究幾乎都使用此三步驟做 DIBR 的處理。



圖 1.5：DIBR 的系統流程。

第三節 研究方法

本研究利用 Kinect 感測器來取得彩色(RGB)影像及深度資訊(Depth)影像，統稱為 RGB-D 影像，再由其產生出多視角的立體影像。如圖 1.6 所示，Kinect 感測器發射出的紅外線透過鏡頭前的光柵(diffuser)將雷射光均勻分布投射在測量空間中形成「散斑」，散斑具有高度隨機性，空間中任兩處的散斑都會視為不同的圖案，再經由紅外線攝影機記錄空間中的散斑(laser speckle)，然後交由硬體晶片計算出深度資訊。而 Kinect 所取得的深度資訊可能會有所遺失造成缺洞，因此需要再做修補，利用修補完的深度資訊結合 Kinect 的 2D 彩色影像，藉由演算法估測出視差資訊，最後生成多視角的影像，也就是所謂的 Depth Image Based Rendering (DIBR)。Kinect 取得之深度資訊簡單也準確，所建立出來的 3D 立體感也較貼近人類對 3D 空間的感知。

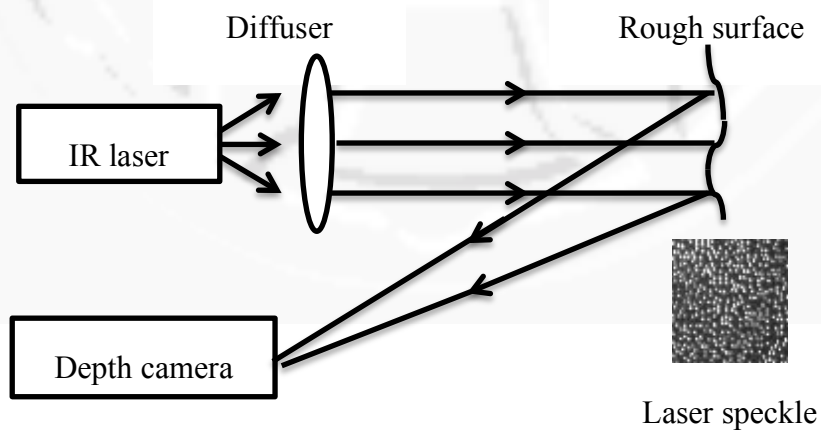


圖 1.6：Kinect 深度資訊的取得。

第四節 論文架構

本論文將於第二章介紹系統的設計架構及系統的運作流程；第三章詳細說明深度影像的處理；第四章說明多視角影像的產生及處理。第五章為實驗結果討論；第六章討論未來可能的研究方向與改進空間。



第二章系統的架構與流程

本章將對所提之系統架構（2.1 節）及其運作流程（2.2 節）詳細說明。

第一節系統架構

所提之系統架構主要包含三個部分，分別為 Kinect 感測器、3D LED 液晶電視以及電腦主機。Kinect 感測器負責提供同步的景物影像(RGB)及景深資訊(D)至電腦主機。利用取得之 RGB-D 影像在電腦做處理後產生立體影像，最後顯示在 3D 立體液晶電視上，整體的系統架構如圖 2.1 所示。Kinect 感測器的視野角度介於水平 57 度、垂直 43 度，有效的深度追蹤距離為 1.2m~3.6m，其他關於 Kinect 感測器的細節會在後面實驗章節詳細介紹。

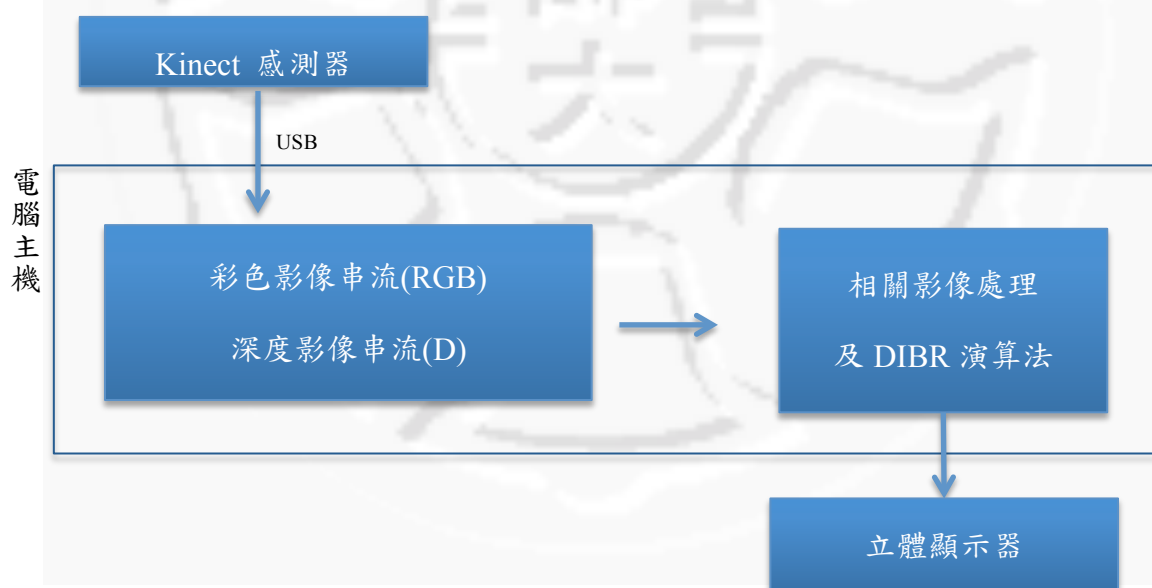


圖 2.1：系統架構圖。

第二節 系統運作流程

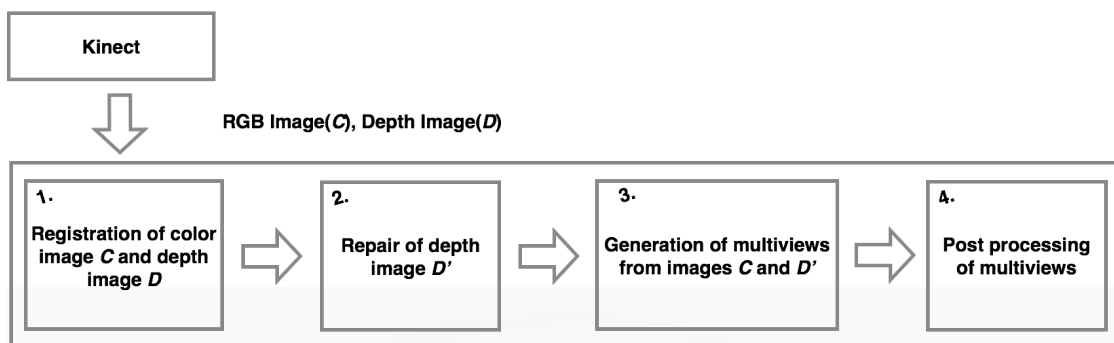


圖 2.2：系統流程圖。

圖 2.2 為所提系統的流程圖，Kinect 感測器取得 RGB-D 影像後會經過四個主要的處理步驟，分別是彩色(RGB)影像跟深度(depth)影像的 registration(registration of color and range images)、深度圖的修補(repairing of range image)、多視角影像的產生(generation of multiviews)以及多視角影像的處理(post processing of multiviews)，以下說明上述各個步驟。



圖 2.3: Kinect 感測器。

由於Kinect感測器上的彩色攝影機(RGB camera)及深度攝影機(depth camera)位於感測器上不同的位置，如圖 2.3 所示，且彩色及深度兩影像產生之方式也不同，所以取得的彩色影像及深度影像像素與像素之間會無法一一對應，如圖 2.4 所示，圖(a)為彩色攝影機取得之彩色影像，圖(b)為深度攝影機(未經registration)所取得之深度影像，圖(c)為圖(a)彩色影像跟圖(b)深度影像的疊合影像，可以明顯地從疊合影像圖2.4(c)中發現物件邊緣都沒有貼齊，這會導致在產生多視角影像時

的結果錯誤，因此須先對兩張影像做registration的動作，為使得兩者畫面能有較好的貼合，如圖 2.5 所示。

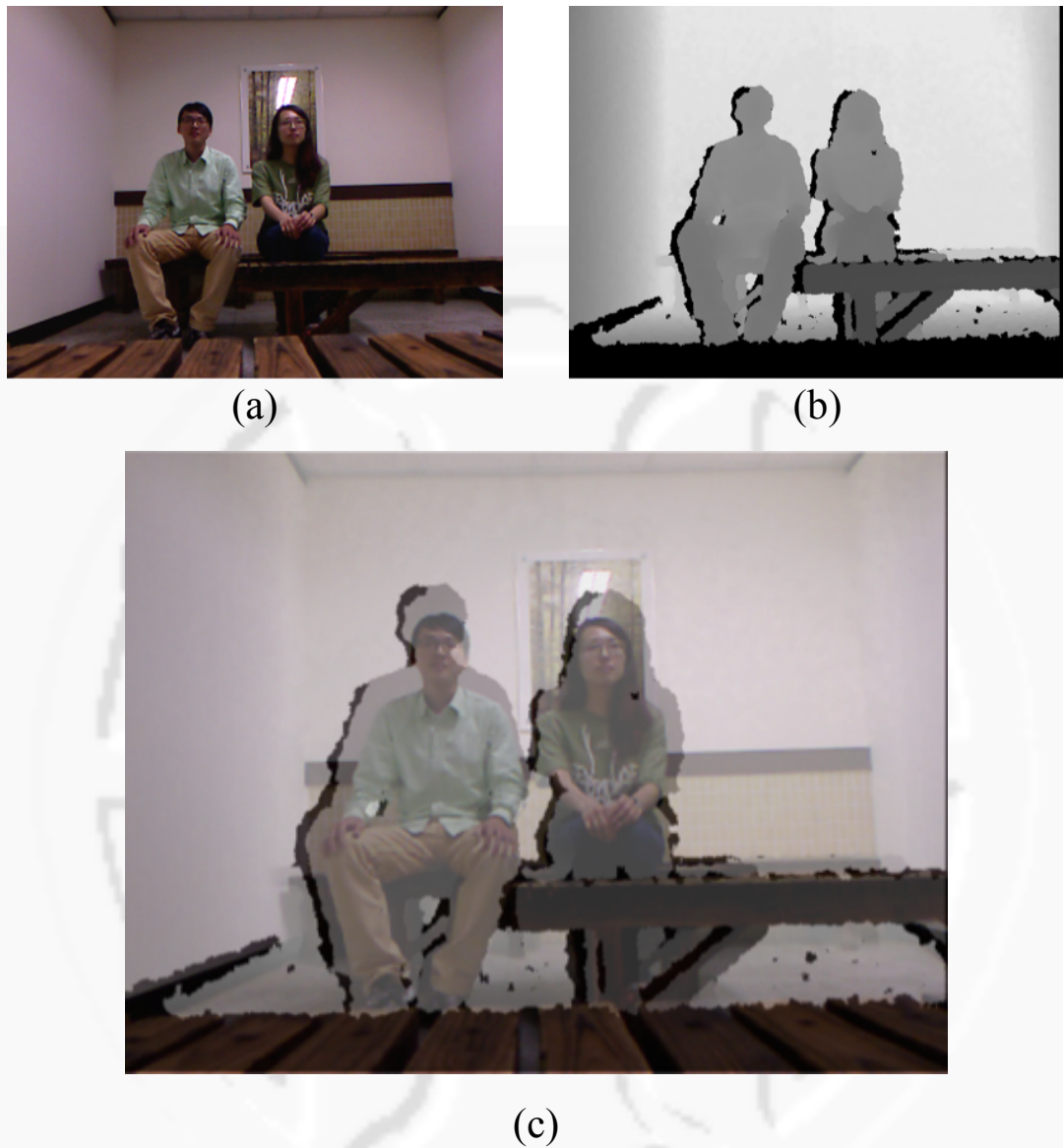


圖 2.4 : (a) 彩色攝影機取得之彩色影像，(b) 深度攝影機（未經 registration）所取得之深度影像，(c) 圖(a)彩色影像跟圖(b)深度影像的疊合影像。

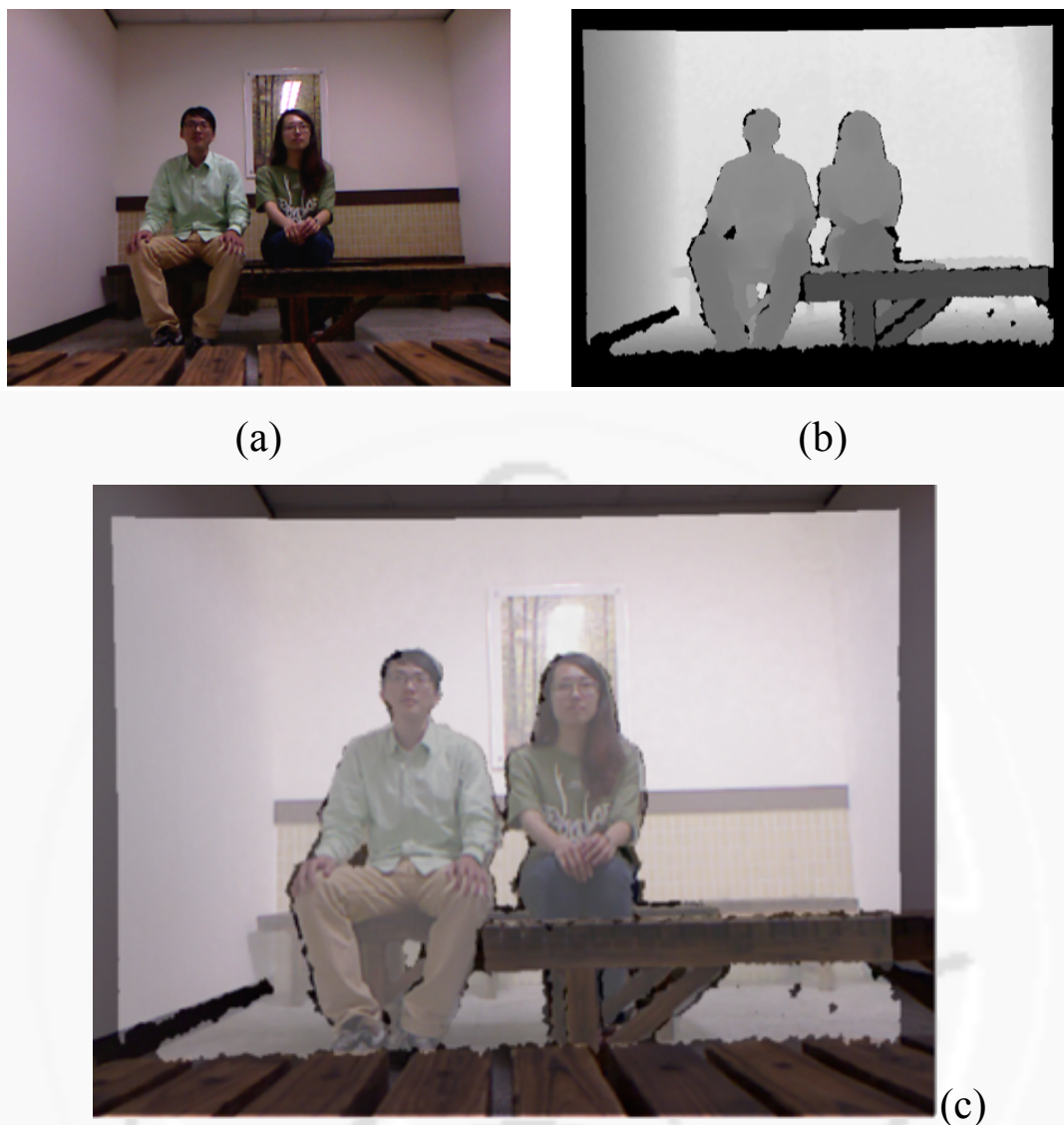


圖 2.5 : (a) 彩色攝影機取得之彩色影像，(b) 深度攝影機（經 registration）所取得之深度影像，(c) 圖(a)彩色影像跟圖(b)深度影像的疊合影像。

而 Kinect 感測器所取得的深度資訊，有時會因為某些狀況（如：玻璃與鏡子過高的反射、物件邊緣反射情況不佳、黑色或是影子導致的較低反射等等情況）使得紅外線攝影機無法準確判定該位置的距離。針對這些狀況，Kinect 感測器所提供的深度資訊圖對於完全沒有深度資訊的位置會給予數值 0 表示為未知狀況，如圖 2.4(b)、2.5(b)所示。而有時也會因為雜訊（如：些微反射）造成有些位置的值與周遭極不協調，因此我們設一個閾值，將那些位置上的值皆設為 0，之後再來修補數值為 0 的破洞。

這裡的深度圖修補我們使用了兩種方法作處理，第一種是利用矩陣完整化 (matrix completion) 的技術，第二種是使用以影像分割為基礎的修補技術。修補完

的深度影像會再經過 bilateral filter 平滑化處理，可以在保留邊緣資訊的情況下，減低深度劇烈變化的情況、可有效減少將來產生立體影像內部的破洞。圖 2.6 及 2.7 顯示上述兩種方法應用在修補圖 2.4(b)中不完整的深度圖及其平滑化後的深度圖。

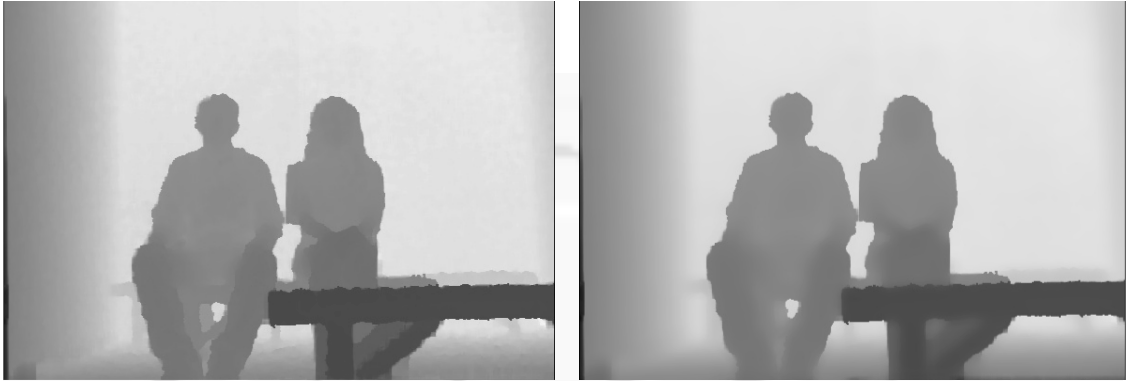


圖 2.6 : (a) 利用矩陣完整化的技術修補深度圖，(b)平滑化修補完的深度圖。



圖 2.7 : (a) 利用基於影像分割的技術修補深度圖，(b)平滑化修補完的深度圖。

修補完深度圖後再利用基於深度影像繪圖法(Depth Image Based Rendering)之立體影像合成技術，將原彩色影像的像素重新投影到虛擬影像平面上，就可以製作出多視角(multi-view)的影像，如圖 2.8 所示，箭頭下面中間為原始拍攝影像，越往左邊的影像代表觀測者越往左邊移動所看到的影像，反之亦然。

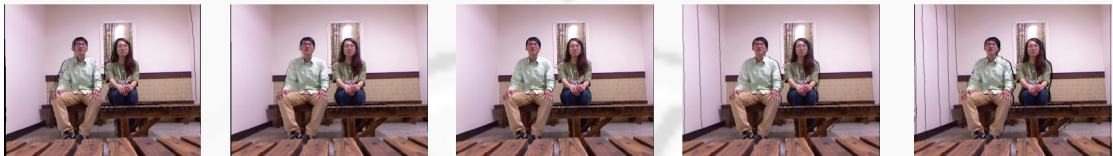


圖 2.8：multiview 示意圖。

最後，經過深度影像繪圖法合成的多視野影像可能會因為遮蔽(occlusion)效應產生許多破洞，如圖2.8所示，但圖太小，看不清楚，另外加上圖2.9放大其中兩張作說明，新合成影像上黑色的區域（牆壁上黑色線條及人物邊緣等等），這些洞主要是因為不同視角物體所延伸出來的部分或是新視角的背景所造成的，尤其當視角差異大時，洞的問題會越大，因此最後需再對合成出的影像做破洞處理，讓立體影像的品質更好。



圖 2.9：multiview 示意圖（放大）。

第三章深度影像的修補

由於 Kinect 感測器常因某些狀況（如：紅外線雷射光遇粗糙、黑色表面反射後，只有少部分的反射光被紅外線攝影機捕捉到；或紅外線雷射光遇鏡子、玻璃全反射被攝影機捕捉到過高的反射光；或紅外線雷射光遇物件邊緣導致反射情況不佳及其他原因造成的雜訊等），使其取得的深度資訊既不完美也不完整。從 Kinect 感測器取得之深度圖中，沒有深度資訊位置點上的值都被設為 0 以使用來區別，然而還有少數雜訊（如：些微反射）參雜其中，因此我們設定一個閾值，閾值以下全部設定為零，其統稱為缺值(missing value)點，再對深度圖做修補。

深度圖修補的目的就是希望對這些缺值點填入適當的值，能與其鄰近點的值合諧一致，本研究採用兩種技術作深度影像修補的處理，其分別為 1. 利用 Hu *et al.* [Hu13]所提的矩陣完整化(matrix completion)技術來對深度圖做修補（3.1 節）；2. 利用 Achanta *et al.* [Ach12] 先對整張彩色影像做超像素的分割，再對每個超像素區塊中對應的深度資訊作處理（3.2 節），以下我們對此兩種技術分別闡述。

第一節矩陣完整化之修補技術

以圖 3.1 為例，我們將有缺值之深度圖視為一不完整之矩陣，如圖 3.1(a)所示，其中缺值的部分(有?標記)就是我們所要填補的地方，圖 3.1(b)為以 Hu *et al.* 技術填補後之矩陣。以下我們先介紹 Hu *et al.*的矩陣完整化技術，之後再探討其填補結果的合理性。

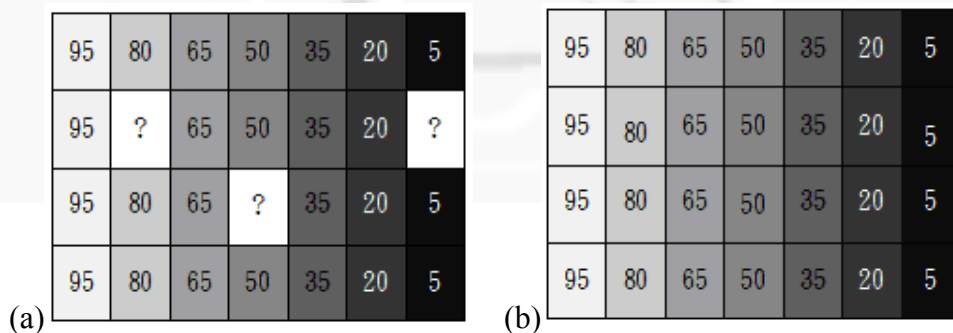
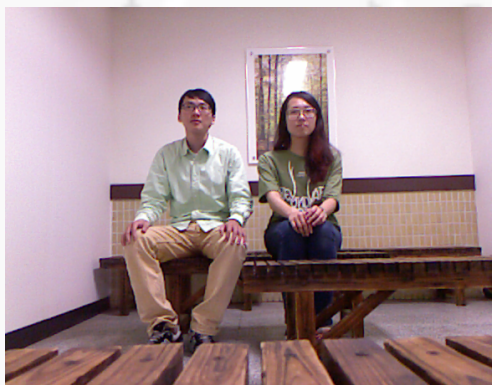


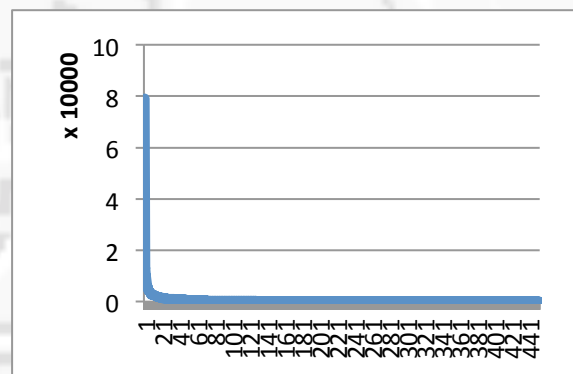
圖 3.1：(a)有缺值之不完整矩陣，(b)缺值填補後之矩陣。

矩陣的完整化在影像處理與電腦視覺領域中已有被用來從事 image inpainting [Kom06] [Ras05]以及 denoising [Ji10]等的工作。然而矩陣的完整化是一個 ill-posed

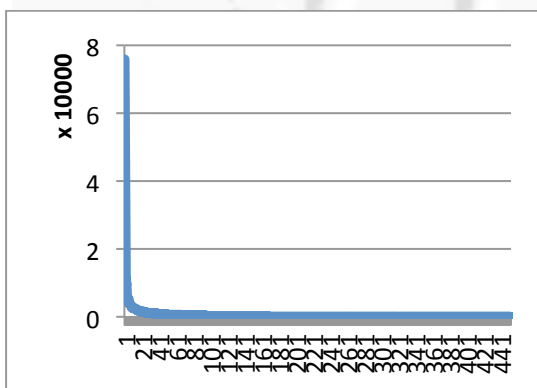
的問題；簡單地說，就是它會有許多可能的解，這是因為有許多填補缺值的方式，我們需要額外的條件來限制解的個數。Lee and Verleysen [Lee10]在其論文中指出，在高維空間中的真實數據，通常都只存在於非常低秩(low rank)的流形結構(manifold)中。以影像來說，若視一影像為一矩陣，則一張影像是一非常高維的數據。根據 Lee and Verleysen 的說法，因影像是真實數據，其矩陣屬低階矩陣結構。再以圖 3.2(a)中的彩色影像(588×453)為例，彩色影像可視為由三個矩陣所組成，分別對應影像的三個色頻(R, G, B channels)。我們將三個矩陣分別做奇異值分解(Singular Value Decomposition, SVD)顯示在圖 3.2(b), (c), (d)中，由這些結果，我們可以看見，無論那一個色頻，除了前 20 個較大的奇異值外，其餘都非常小接近 0。從資訊的觀點，影像內容只掌控在極少數較大的奇異值；又從數理的觀點，一矩陣的階數即為其非 0 奇異值的個數，影像矩陣中大部份的奇異值接近 0，也就是說影像矩陣實際上是一種非常低階的矩陣。



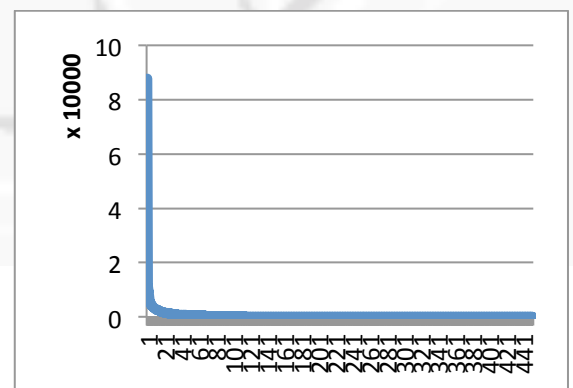
(a)an image example



(b)B-channel 的奇異值分解



(c)G-channel 的奇異值分解



(d)R-channel 的奇異值分解

圖 3.2：(a)一張 588×453 的影像，(b)R-channel 的奇異值分解，(c)G-channel 的奇異值分解，(d)B-channel 的奇異值分解。

上述的討論主要是針對 intensity 影像，然而此處我們所考慮的景深影像屬於一種 range 影像，它們是否也具有與 intensity 影像同樣的性質，亦即景深影像矩陣是否也是一種非常低秩的矩陣。圖 3.3 顯示兩景深影像及它們的奇異值分析結果，我們可以看見，景深影像同樣受控於少數較大的奇異值，所以景深影像矩陣也是一種非常低秩的矩陣。Candès and Recht [Can08]於是將矩陣的完整化以數學方式描述如下

$$\begin{aligned} \min_X \quad & \text{rank}(X) \\ \text{s.t.} \quad & X_{ij} = M_{ij}, (i,j) \in \Omega, \end{aligned} \quad (1)$$

其中 $X \in \mathbb{R}^{m \times n}$ 代表期待填補完成的結果矩陣， $M \in \mathbb{R}^{m \times n}$ 是有缺值的原始矩陣， (i,j) 指的是第 i 列第 j 行的值，而則 Ω 是指已知值的矩陣元素位置之集合。

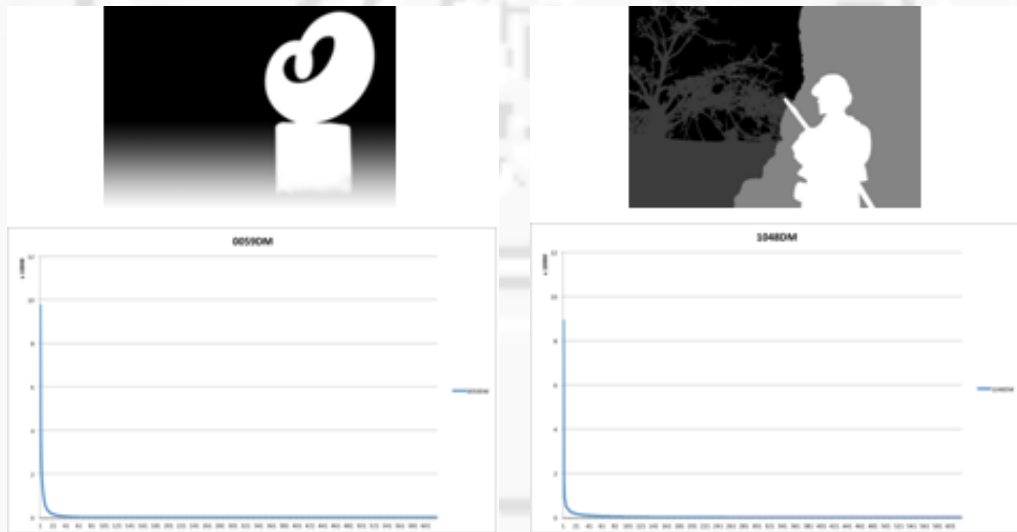


圖 3.3：深度資訊做奇異值分解。

上述極小化的問題(1)是一種 NP-hard 的問題，主要是因為 $\text{rank}()$ 函數是一非凸(non-convex)且不連續的函數。凸函數是規劃凸程序(convex programming)必要的條件，而凸程序可以在多項式時間內求得問題解。Fazel [Faz02]以及 Recht et al. [Rec10]的研究指出，nuclear norm (NN)，亦即矩陣的奇異值總和，是最接近 $\text{rank}()$ 函數之較低界(lower bound)的凸函數，於是(1)式中的極小化的問題被近似改寫如下[Cai10] [Toh10] [Wri09]：

$$\min_x \{\|X\|_*\} \text{ s.t. } X_{ij} = M_{ij}, (i, j) \in \Omega \quad (2)$$

其中 $\|X\|_* = \sum_{i=1}^{\min(m,n)} \sigma_i(X)$ 為 $m \times n$ 矩陣 X 的 nuclear norm， $\sigma_i(X)$ 為 X 第 i 大的奇異值。然而 Hu *et al.* [Hu13] 認為一矩陣的秩是對應到較大的 r 個非零奇異值，所以應該極小化的是 $\min\{m, n\} - r$ 個較小奇異值和，亦即 $\|X\|_r = \sum_{i=r+1}^{\min(m,n)} \sigma_i(X)$ ，要比 nuclear norm 更貼近矩陣的秩。Hu *et al.* 稱 $\|X\|_r$ 為 truncated nuclear norm (TNN)，而將(2)式改成

$$\min_x \{\|X\|_r\} \text{ s.t. } P_\Omega(X) = P_\Omega(M), \quad (3)$$

其中 $(P_\Omega(X))_{ij} = \begin{cases} X_{ij} & \text{if } (i, j) \in \Omega \\ 0 & \text{otherwise} \end{cases}$ 。只是上述的 $\|X\|_r$ 函數並非凸函數，Hu *et al.*

於是提出了一兩段式反覆架構來解(3)式的極小化問題。

假設矩陣 X 的奇異值分解為 $X = U\Sigma V^T$ ，其中 $U = (u_1, \dots, u_m) \in R^{m \times m}$ ， $\Sigma \in R^{m \times n}$ ，以及 $V = (v_1, \dots, v_n) \in R^{n \times n}$ 。令 $A = (u_1, \dots, u_r)^T$ 和 $B = (v_1, \dots, v_r)^T$ ，所以 $A \in R^{r \times m}$ ， $B \in R^{r \times n}$ ，而且 $AA^T = I^{r \times r}$ ， $BB^T = I^{r \times r}$ 都是 unitary 矩陣 ($A^{-1} = A^T, B^{-1} = B^T$)，滿足 $\max_{AA^T=I, BB^T=I} \text{Tr}(AXB^T) = \sum_i^r \sigma_i(X)$ ，因此改寫 $\|X\|_r = \|X\|_* - \max_{AA^T=I, BB^T=I} \text{Tr}(AXB^T)$ ，其中 $\|X\|_*$ 和 $-\text{Tr}(AXB^T)$ 均為凸函數。(3)式於是改寫成

$$\min_x \left\{ \|X\|_* - \max_{AA^T=I, BB^T=I} \text{Tr}(AXB^T) \right\} \text{ s.t. } P_\Omega(X) = P_\Omega(M) \quad (4)$$

在 Hu *et al.* 的兩段式反覆架構中，首先令 $X_1 = M_\Omega$ ；而在每一次(l)的反覆時，先固定當時的矩陣 X_l ，再計算其對應的 unitary 矩陣 A_l 及 B_l ，然後固定 A_l 和 B_l 由下式來更新 X_{l+1}

$$X_{l+1} = \arg \min_x \{ \|X\|_* - \text{Tr}(A_l X B_l^T) \} \text{ s.t. } P_\Omega(X) = P_\Omega(M) \quad (5)$$

Algorithm 1 總結了上述兩段式反覆過程，其中步驟 2 求 X_{l+1} (亦即解(5)式)，Hu *et al.* 提了兩種方法，分別闡述於后。

首先將(5)式改寫成 $\min_x \{ \|X\|_* - \text{Tr}(A_l W B_l^T) \} \text{ s.t. } X = W, P_\Omega(W) = P_\Omega(M)$ 。令 $L(X, Y, W, \beta) = \|X\|_* - \text{Tr}(A_l W B_l^T) + \frac{\beta}{2} \|X - W\|_F^2 + \text{Tr}(Y^T (X - W))$ ，稱為 augmented Lagrange function。藉著極小化函數 L 分別反覆更新其中變數如下：令

$X_1 = M_\Omega$, $W_1 = X_1$ 以及 $Y_1 = X_1$, 接下來計算 $X_{k+1} = \mathcal{D}_{1/\beta} \left(W_k - \frac{1}{\beta} Y_k \right)$, 其中 $\mathcal{D}_\alpha(X) = \arg \left(\min_x \left\{ \frac{1}{2} \|X - Y\|_\alpha^2 + \alpha \|X\|_* \right\} \right)$; 再計算 $W_{k+1} = X_{k+1} + \frac{1}{\beta} (A_l^T B_l + Y_k)$, 然後固定矩陣中的觀測值後 , 計算 $W_{k+1} = P_{\Omega^c}(W_{k+1}) + P_\Omega(M)$; 最後計算 $Y_{k+1} = Y_k + \beta(X_{k+1} - W_{k+1})$; 如此反覆計算 , 直到矩陣 X 收斂。以上過程綜結於下面的 Algorithm 2 ADMM 中。

Algorithm 1

Input: Incomplete matrix M_Ω , where Ω is the position of the observed entries, tolerance ϵ_0 .

Initialize: $X_1 = P_\Omega(M)$

repeat

STEP 1. Given X_l , $[U_l, \Sigma_l, V_l] = \text{svd}(X_l)$,
where $U_l = (u_1, \dots, u_m) \in \mathbb{R}^{m \times m}$,
 $V_l = (v_1, \dots, v_n) \in \mathbb{R}^{n \times n}$.
Compute $A_l = (u_1, \dots, u_m)^T$,
 $B_l = (v_1, \dots, v_n)^T$.

STEP 2. Solve $X_{l+1} = \arg \left(\min_x \|X\|_* - \text{Tr}(A_l X B_l^T) \right)$
s. t. $P_\Omega(X) = P_\Omega(M)$

until $\|X_{l+1} - X_l\|_F \leq \epsilon_0$

Return X_{l+1}

Algorithm 2 ADMM

Input: A_l, B_l, M_Ω and tolerance ϵ .

Initialize: $X_1 = M_\Omega, W_1 = X_1, Y_1 = X_1$, and $\beta = 1$

repeat

STEP 1. $X_{k+1} = \mathcal{D}_{1/\beta} \left(W_k - \frac{1}{\beta} Y_k \right)$

STEP 2. $W_{k+1} = X_{k+1} + \frac{1}{\beta} (A_l^T B_l + Y_k)$
Fix values at observed entries
 $W_{k+1} = P_{\Omega^c}(W_{k+1}) + P_\Omega(M)$.

STEP 3. $Y_{k+1} = Y_k + \beta(X_{k+1} - W_{k+1})$

until $\|X_{k+1} - X_k\|_F \leq \epsilon$

由於 ADMM 在處理矩陣中的數據帶有雜訊時，效果不是很好，Hu *et al.* 再將 (5) 式改寫成

$$\min_X \{ \|X\|_* - \text{Tr}(A_l X B_l^T) \} + \frac{\lambda}{2} \|P_\Omega(X) - P_\Omega(M)\|_F^2 \quad (6)$$

於是矩陣 X 反覆更新如下

$$\begin{aligned} X_{k+1} &= \arg \left(\min_X \left\{ \|X\|_* + \frac{1}{2t_k} \|X - (Y_k - t_k \nabla f(Y_k))\|_F^2 \right\} \right) \\ &= \mathcal{D}_{t_k}(Y_k - t_k \nabla f(Y_k)) = \mathcal{D}_{t_k} \left(Y_k - t_k \left(A_l^T B_l - \lambda (P_\Omega(Y_k) - P_\Omega(M)) \right) \right), \end{aligned}$$

其中 $t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}$, $Y_{k+1} = X_{k+1} + \frac{t_k - 1}{t_{k+1}} (X_{k+1} - X_k)$ 。以上過程總結於下面的 Algorithm 3 中。

Algorithm 3 APGL

Input: A_l, B_l, M_Ω and tolerance ϵ .

Initialize: $t_1 = 1, X_1 = M_\Omega, Y_1 = X_1$

repeat

STEP 1. Update X_{k+1} as

$$X_{k+1} = \mathcal{D}_{t_k} \left(Y_k + t_k \left(A_l^T B_l - \lambda (P_\Omega(Y_k) - P_\Omega(M)) \right) \right)$$

STEP 2.

$$t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}$$

STEP 3.

$$Y_{k+1} = X_k + \frac{t_k - 1}{t_{k+1}} (X_k - X_{k-1})$$

until

$$\|X_{k+1} - X_k\|_F \leq \epsilon$$



圖 3.4：左圖為原有缺值的深度圖，右圖為使用 APGL 修補後的結果。

前述的兩個方法都是對(5)式中的兩個限制條件，分別去處理，Hu *et al.*建議將它們合併一起考慮，如此可以大大增加程序運作的速度。

$$X_{K+1} = \mathcal{D}_{\frac{1}{\beta}}(W_k - \frac{1}{\beta}(Y_k)_{11})$$

$$W_{k+1} = \frac{1}{2\beta} P_{\Omega}[\beta(M - X_{k+1}) - (A_l^T B_l + (Y_k)_{11} + (Y_k)_{22})] + X_{k+1} + \frac{1}{\beta} (A_l^T B_l + (Y_k)_{11})$$

$$Y_{K+1} = Y_k + \beta(X_{k+1} + W_{k+1})$$

其中參數 β 隨反覆次數改變如下： $\beta_{k+1} = \min(\beta_{max}, \rho\beta_k)$ ， β_{max} 為 β 的上界，

$$\rho = \begin{cases} \rho_0 & \text{if } \frac{\beta_k \max\{\|X_{k+1} - X_k\|_F, \|W_{k+1} - W_k\|_F\}}{\|C\|_F} \\ 1 & \text{otherwise} \end{cases}$$

以上計算過程綜結於下面的 Algorithm 4 ADMMAP 中， $\mathcal{A}(X) = \begin{pmatrix} X & 0 \\ 0 & 0 \end{pmatrix}$,

$\mathcal{B}(W) = \begin{pmatrix} -W & 0 \\ 0 & P_{\Omega}(W) \end{pmatrix}$, $\mathcal{C} = \begin{pmatrix} 0 & 0 \\ 0 & P_{\Omega}(M) \end{pmatrix}$ ，其中 \mathcal{A} 和 \mathcal{B} 為線性算子： $\mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{2m \times 2n}$ 。

Algorithm 4 ADMMAP

Input: A_l, B_l, M_{Ω} and tolerance ϵ .

Initialize: $X_1 = M_{\Omega}, W_1 = X_1, k = 10^{-3}, \rho_0, Y_1 = X_1$ and β_0

repeat

STEP 1. Set Y_k and W_k fixed

$$X_{k+1} = \mathcal{D}_{\frac{1}{\beta_k}}\left(W_k - \frac{1}{\beta_k}(Y_k)_{11}\right).$$

STEP 2. Set Y_k and X_{k+1} fixed

$$W_{k+1} = \frac{1}{2\beta_k} P_{\Omega}[\beta_k(M - X_{k+1}) - (A_l^T B_l + (Y_k)_{11} + (Y_k)_{22})] + X_{k+1} + \frac{1}{\beta_k} ((A_l^T B_l + (Y_k)_{11})).$$

STEP 3. Set X_{k+1} and W_{k+1} fixed

$$Y_{k+1} = Y_k + \beta_k(\mathcal{A}(X_{k+1}) + \mathcal{B}(W_{k+1}) - \mathcal{C})$$

STEP 4. Update β_k by $\beta_{k+1} = \min(\beta_{max}, \rho\beta_k)$

until $\|X_{k+1} - X_k\|_F \leq \epsilon$

第二節基於影像分割之修補技術

在第二種深度影像的修補時，我們使用了 Achanta *et al.* [Ach12] 所提出來的 Simple Linear Iterative Clustering(SLIC)方法。首先將彩色影像分割成多個超像素(superpixel)區塊，利用彩色影像分割出來的 superpixel 對應回深度影像的位置，再針對每一深度影像中的 superpixel 中有缺值的部分做修補。影像分割的研究在過往已非常廣泛，但是大部份的計算量都相當大、相當耗時，然而在實際應用時 superpixel 需要滿足运算速度快且簡單使用的條件，因此 SLIC 提出了一種低計算量的方式來做分割影像。

一開始先將 k 個 cluster 中心點散佈在彩色影像上，中心點與中心點間間隔 S 個 pixels，然後再將中心點移到 3×3 臨近點中最 lowest gradient 的位置，此一處理是為了避免 superpixel 的中心點位於 edge 上，接下來一步步的更新 cluster 中心點。SLIC 利用顏色相似度和距離鄰近度來生成像素的聚類，其定義 clustering 的生成是依據一 5-D 的特徵向量， $[l, a, b, x, y]^T$ ，其中 $[l, a, b]^T$ 指的是 CIE Lab 顏色空間中的坐標， l 表示亮度， a 和 b 表示顏色對立維度， $[x, y]^T$ 是指像素的位置。由於顏色與空間位置有不同的單位，其距離代表不同差異度，作者提出了一種新的距離測量方法 D' 同時考慮顏色和像素位置

$$d_c = \sqrt{(l_j - l_i)^2 + (a_j - a_i)^2 + (b_j - b_i)^2},$$
$$d_s = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2},$$
$$D' = \sqrt{\left(\frac{d_c}{N_c}\right)^2 + \left(\frac{d_s}{N_s}\right)^2}.$$

其中， N_c 為顏色距離的最大值， $N_s = S = \sqrt{N/k}$ ， k 為 superpixels 的數目，而 S 的限制也減少了 superpixel 的尋找範圍，SLIC 只計算 $2S \times 2S$ 區域中 cluster 中心點到各個 pixel 間的距離，不像傳統的 k -means 需要在整張影像上尋找，如圖 3.5 [Ach12] 所示。傳統 k -means 的複雜度為 $O(kNI)$ ， I 為遞迴的次數， N 為影像中的像素數目，而 SLIC 的複雜度則只需要 $O(N)$ 。

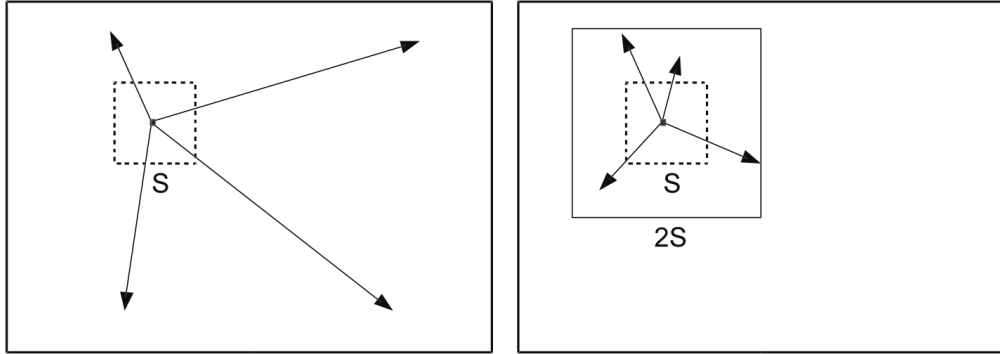


圖 3.5: 減少 superpixel 搜尋範圍。左圖為傳統 k -means 的演算法，右圖為 SLIC 的演算法，只需要在一個限制的區域中做尋找。

然而顏色距離的最大值 N_c 並不直觀，因為 cluster 與 cluster 或是影像與影像之間顏色可能變化的非常劇烈，但是此問題藉由固定 N_c 為一常數 m 就可以被避免，因此上式的 D' 變成

$$D' = \sqrt{\left(\frac{d_c}{m}\right)^2 + \left(\frac{d_s}{S}\right)^2}.$$

而在實際應用上整個距離測量可以寫成

$$D' = \sqrt{(d_c)^2 + \left(\frac{d_s}{S}\right)^2} m^2.$$

藉由這樣定義 D' ，使得 m 得以掌控顏色相似度及空間距離的相對權重。當 m 很大的時候，空間距離變得相對重要，superpixel 的 clustering 結果會變得比較緊湊（即面積周長比的值較低），而當 m 很小的時候，superpixel 的 clustering 結果會比較緊貼影像中的邊緣(edge)，因此大小跟形狀較不規則。其中當使用 CIELAB 顏色空間時， m 可以被定義在 [1,40] 的區間。此距離測

量方式也可以使用在灰階影像，只要將 $d_c = \sqrt{(l_j - l_i)^2}$ ，而當要拓展到處理 3D supervoxel 的問題時，則可把深度(z)的維度資訊考慮進空間距離

$$d_s = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2 + (z_j - z_i)^2}$$

整個 SLIC 演算法如下所示：

```

/* Initialization */
Initialize cluster centers  $C_k = [l_k, a_k, b_k, x_k, y_k]^T$  by sampling
pixels at regular grid steps  $S$ .
Move cluster centers to the lowest gradient position in a  $3 \times 3$ 
neighborhood.
Set label  $l(i) = -1$  for each pixel  $i$ .
Set distance  $d(i) = \infty$  for each pixel  $i$ .

```

repeat

```

/* Assignment */
for each cluster center  $C_k$  do
  for each pixel  $I$  in a  $2S \times 2S$  region around  $C_k$  do
    Compute the distance  $D$  between  $C_k$  and  $i$ .
    if  $D < d(i)$  then
      set  $d(i) = D$ 
      set  $l(i) = k$ 
    end if
  end for
end for
/* Update */
Compute new cluster centers.
Compute residua error  $E$ .
until  $E \leq \text{threshold}$ 

```

最後，就像其他的一些superpixel演算法，SLIC也沒有明確的執行連通性，在clustering的過程結束後，不屬於任一個cluster的一些孤立像素可能依然存在，為了解決這個問題，最後使用了connected components演算法將孤立的pixel分配到最靠近的cluster。

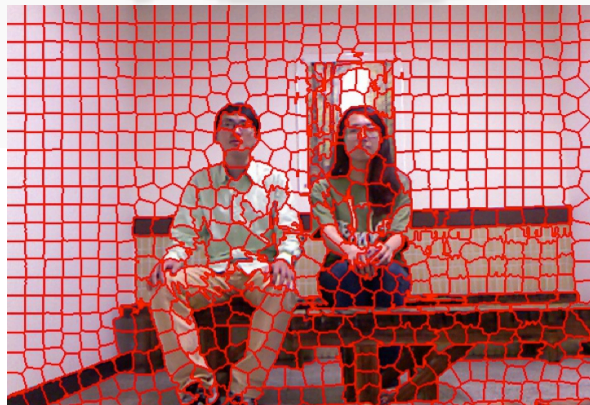


圖 3.6：對彩色影像做 SLIC 分割後的結果。

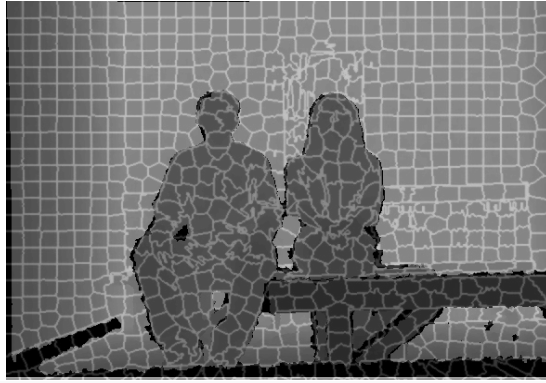


圖 3.7：將 SLIC 分割後的結果對應到深度影像。

將彩色影像使用SLIC分割成800塊 superpixels 後如圖3.6所示，大致上 superpixels 都可以很好的貼近 edge，而且形狀大小都還算規則。再把 superpixels 的分佈對應回深度影像，如圖3.7所示。對於一顏色及空間距離都相近的superpixels集合，我們假設該集合內的 pixels 皆為同一物體的一部份，其深度值應該相近，因此，若深度影像上的superpixel中某pixel有缺值，我們使用該superpixel中有值的平均深度值做填補。然而有時候會發生一superpixel中可能大部份都是缺值（例如鏡子反射，大部份都是缺值，只有一小部份因為些微反射而有值），為了避免少數深度資訊掌控整個superpixel，我們設了一個threshold t ，若該superpixel有值的數目沒有超過 t 則不使用該superpixel的平均值來填補，而是找尋相鄰且相近的superpixel與原本的合併後再判斷是否有達到threshold，圖3.8顯示一利用此方法修補之深度圖；其中左邊影像為有缺值的深度圖，右邊為修補後的結果。



圖 3.8：左圖為原有缺值的深度圖，右圖為修補後的結果。

第四章多視角影像的產生及處理

當要在一立體顯示器上觀看一景緻的立體內容時，我們需要該景緻左右眼的影像，供大腦解釋以產生立體感。然而一對左右眼影像所產生的立體感，通常只適合在一定範圍的視角內觀看，其他的視角則立體感效果不佳。而為解決此種缺點，就是準備更多視角用的影像對，本研究採用 Zhang *et al.* [Zha05]所提的 DIBR 技術，其流程如圖 4.1 所示，其中包括了三個主要步驟：(1)深度圖的前處理 (preprocessing of depth map)(4.1 節)，(2)3-D image warping(4.2 節)以及(3)空洞的填補(hole filling) (4.3 節)，分別在下面各節詳細討論。其概念為從一對彩色及深度影像，產生出多個不同視角用的影像對，統稱為虛擬影像，而此種影像會因景物自身或景物間互相遮蔽(occlusion)造成在影像中產生一些破洞，所以需要對這些影像從事空洞修補的工作。

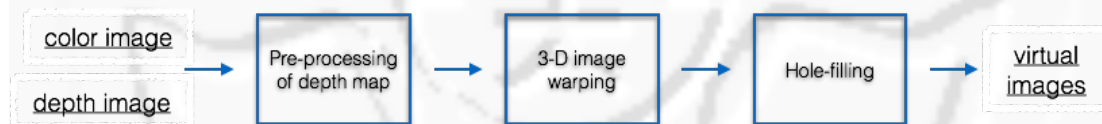


圖 4.1：DIBR 的系統流程。

第一節深度圖的前處理

深度圖的前處理主要包含了兩個部分：(a)設定深度圖中零視差(zero parallax setting, ZPS)的位置，以及(b)對深度圖做平滑化(smoothing)。零視差指的是兩眼所見之視差為零的位置，如圖 4.2 所示，若零視差所在的平面設為圖中螢光幕的位置，則正視差(positive parallax)會讓人感覺物體向螢幕內凹入(recessed)，而負視差(negative parallax)則讓人感覺物體向螢幕外凸出(prominent)。因此 ZPS 的設定會影響畫面的凸凹感，稱 ZPS 的景深為收斂距離(convergence distance, Z_c)。Zhang *et al.*建議取 $Z_c = \frac{Z_{near} - Z_{far}}{2}$ ，其中 Z_{near} 為深度圖中最近的截面， Z_{far} 為深度圖中

最遠的截面。在一個 8-bit 的灰階深度圖中，0 跟 255 分別代表的是 Z_{near} 跟 Z_{far} 的值，如圖 4.3 所示，所有深度資訊也被正規化(normalized)到 0 到 255 之間。

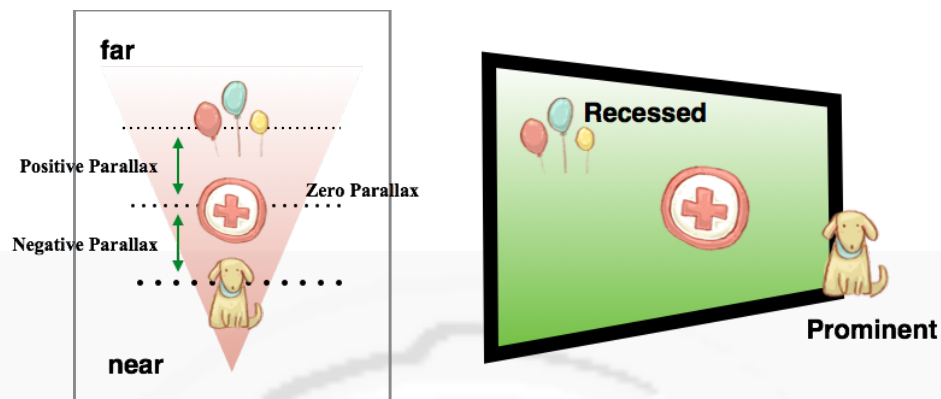


圖 4.2：視差平面。



圖 4.3：8-bit 的深度圖。

在深度圖的平滑化時，主要是為了讓深度變化較劇烈的地方趨於平緩。我們使用雙邊濾波器(bilateral filter, BF) [Tom98]，此種濾波器在平滑過程中，除了考慮像素間的距離外，還考慮了像素間的光度及彩度差異，使得 Bilateral Filter 能夠有效的將影像上的雜訊去除，同時保存影像上的邊緣資訊。BF 的定義為

$$I_p = \frac{1}{w_p} \sum_{q \in S} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|I_p - I_q|) I_q, \text{ 其中}$$

$w_p = \sum_{q \in S} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|I_p - I_q|)$ ， p 為目標像素， q 為目標像素周圍的一個像素， I_p 為目標像素的顏色， I_q 為目標像素周圍一個像素的顏色， S 為目標像素周圍的像素群， G_{σ_s} 是以像素的距離來決定權重高低的 Gaussian filter 式子， G_{σ_c} 是以像素色差來決定權重高低的 Gaussian filter 式子，Gaussian filter 定義如下所示

$$G(u, v) = \frac{1}{2\pi\sigma^2} e^{-(u^2 + v^2)/(2\sigma^2)}$$

其中 σ 是指常態分佈的標準偏差。圖 4.3，矩陣完整化修補完的深度圖為例，經 Bilateral filter smooth 後的結果如圖 4.4 所示，右下角灰色的深度很明顯被平滑化了，但是整張圖的邊緣資訊有被很好地保留。



圖 4.4：經 Bilateral filter smooth 後的深度圖。

第二節 3-D image warping

DIBR 的技術核心就是 3-D image warping，利用深度資訊將原本參考用的彩色影像投影到三維歐式空間上，然後再將三維空間的點投影到虛擬攝影機的成像平面上。為了簡化，我們只考慮平行擺設的攝影機，因此三維空間中垂直坐標(y) 投影到任一個平行攝影機上並不會有所改變，我們只考慮坐標點的水平移動。如圖 4.5 所示， Z 為像素 P 點的深度值， f 為攝影機焦距(focal length)， t_x 為兩虛擬攝影機的基線(baseline)長度， X_c 為中間影像的像素坐標， X_l 和 X_r 則分別為欲合成的像素在左、右眼影像中的位置， C_c 為原始影像拍攝的攝影機光學中心(optical center)， C_l 和 C_r 為左、右眼虛擬攝影機的光學中心。根據上述設置，一個深度為 Z 的點 P 被投影到三個攝影機的成像平面上，分別為 (x_l, y) ， (x_c, y) 及 (x_r, y) 。而從圖 4.5 的幾何關係就可以得到 image warping 的公式 $x_l = x_c + \frac{t_x f}{2 Z}$ ， $x_r = x_c - \frac{t_x f}{2 Z}$ 。用原來的彩色影像根據其相對應的深度值就可以在虛擬影像平面上將影像像素重新投影(re-projecting)產生出其相鄰視角的 virtual images (坐標點右移形成左影像，坐標點左移形成右影像)。

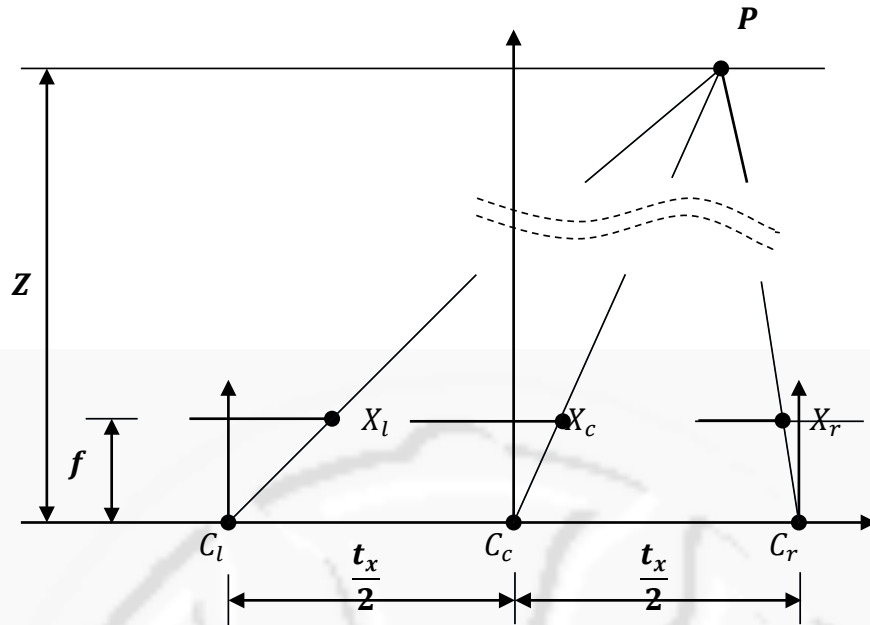


圖 4.5：產生虛擬立體影像的攝影機設置圖。

DIBR 最主要的困難在於 image warping 後產生之 virtual images，會因為深度圖不連續或是因為遮蔽效應（視角的不同而產生出影像的缺洞(holes),這些影像的缺洞可能是因為物體在不同視角所延伸出來的部份或是新視角的背景所造成）的關係出現一些 newly exposed 的區域或是洞。如圖 4.6(a)(b)(c)所示，(a)為原本的彩圖，(b)為深度圖未經平滑處理 warping 後的結果，其中(b)圖中黑色的點狀區域即為 newly exposed 的部分，(c)則為深度圖經 Bilateral filter smooth 處理 warping 後的結果，可以看出使用預先經過平滑化處理的深度圖所產生出的 virtual image 上導致的破洞較少。



(a)



(b)



(c)

圖 4.6：(a)原本的彩圖，(b)深度圖未經平滑化處理 warping 後的結果，(c)深度圖經 Bilateral filter smooth 處理 warping 後的結果。

第三節 空洞的填補

新產生的視角影像由於 image warping 的處理會產生一些 newly exposed 的洞，而這些洞的產生因素主要分為兩種，第一種為同一物件中有較大深度值上的差距，造成相鄰的兩個 pixel 在 image warping 時水平移動的幅度的不同，進而導致原本兩相臨的 pixels 在 image warping 後產生中間缺值的區域；第二種破洞的產生是因為視角改變時前方物體或是後方背景在不同視角所新延伸出來的部份是我們所未俱有的資訊，如：原本看到罐頭的正面，欲產生側面的視角，我們並不知道它側面的模樣，也不知道原本被罐頭遮住的背景影像資訊。而利用經 bilateral filter smooth 過後的深度圖產生之 virtual images 在視角改變有限的情況下，則有效的減少了深度資訊劇烈改變的可能性。

接下來，我們使用了 Telea [Tel04]提出的利用 Fast Marching Method(FMM) 來做新視角彩色影像的填補。其算法的概念思想是，先處理待修復區域邊緣上的像素點，然後逐步層層向內推進，直到修復完所有的像素點。而一像素填補的原理如下圖 4.7 所示， Ω : 為影像中欲修復的區域， $\delta\Omega$: 為欲修復的區域的邊界， p : 為 $\delta\Omega$ 邊界上的任何一點，在點 p 周圍已知區域中，選擇一個以 ε 為距離的區域 $B(\varepsilon)$ ，然後利用 $B(\varepsilon)$ 內部的 pixels 來計算 p 點可能的值。

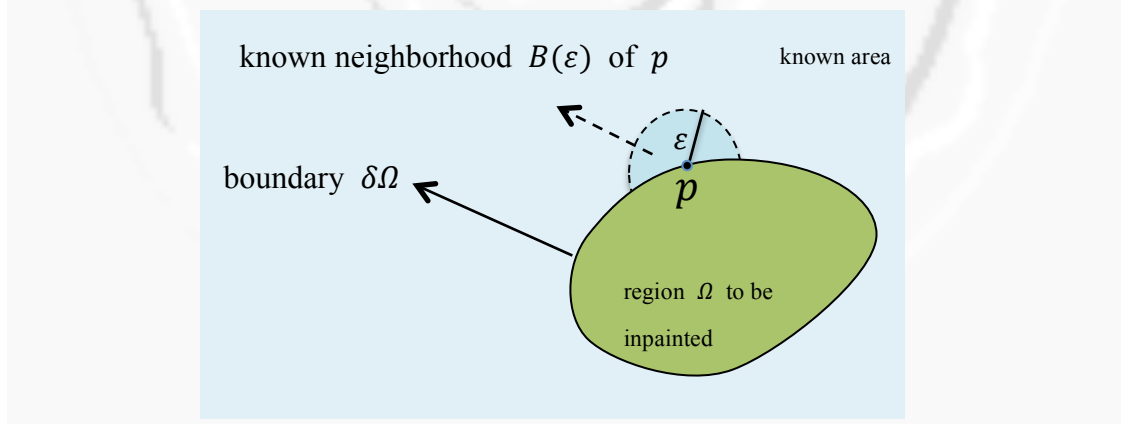


圖 4.7：填補原理。

點 q 為 $B(\varepsilon)$ 區域中的一點，已知點 q 的像素值 $I(q)$ 及其梯度 $\nabla I(q)$ ，若只考慮點 q 的資訊我們可以藉由下式得到點 p 可能的值：

$$I_q(p) = I(q) + \nabla I(q)(p - q)$$

而藉由考慮整個 $B(\varepsilon)$ 區域的 pixels 資訊，上式可以轉換成下式：

$$I(p) = \frac{\sum_{q \in B_\varepsilon(p)} w(p, q) [I_q(p) = I(q) + \nabla I(q)(p - q)]}{\sum_{q \in B_\varepsilon(p)} w(p, q)}$$

$w(p, q)$ 代表的是權重函數，用來調節 $B(\varepsilon)$ 區域中個像素的權重，其定義如下：

$$w(p, q) = dir(p, q) \cdot dst(p, q) \cdot lev(p, q)$$

其中 directional component: $dir(p, q)$ 確保越靠近法線方向 $N = \nabla T$ (藉由 FMM 計算出)，其權重越重; geometric distance component: $dst(p, q)$ 確保距離點 p 越近的像素點權重越重; level set distance component: $lev(p, q)$ 確保通過像素點 p 輪廓的像素點權重越重，如下所示：

$$dir(p, q) = \frac{p - q}{\|p - q\|} \cdot N(p)$$

$$dst(p, q) = \frac{d_0^2}{\|p - q\|^2}$$

$$lev(p, q) = \frac{T_0}{1 + |T(p) - T(q)|}$$

$N(p)$ 代表 p 點的法線方向，通常 T_0 及 d_0 設為像素間的距離，亦即:1。不停的迭代式子 $I(p) = \frac{\sum_{q \in B_\varepsilon(p)} w(p, q) [I_q(p) = I(q) + \nabla I(q)(p - q)]}{\sum_{q \in B_\varepsilon(p)} w(p, q)}$ 後，就可以逐步收縮待修復區域的邊界，直到整個區域都被修復為止。

第五章實驗結果

在本章我們介紹了本系統的實驗工具(5.1 節)、深度圖修補技術的比較(5.2 節)及 3D 立體影像的結果比較(5.3 節)。

第一節實驗工具

在本實驗中我們主要使用到兩個實驗工具，先是使用 Microsoft 設計的深度感測器 Kinect 做 RGB-D 影像的截取，在電腦裡處理完影像後再利用 LG 的 3D 立體顯示電視做輸出。

Kinect 可同時擷取三種資訊，分別是彩色影像、3D 深度影像、以及聲音訊號，其身上有 3 顆鏡頭，中間的鏡頭是一般常見的 RGB 彩色攝影機，左右兩邊鏡頭則分別為紅外線發射器和紅外線 CMOS 攝影機所構成的 3D 深度感應器。Kinect 感測深度的技術是利用發射出的紅外線透過鏡頭前的光柵(diffuser)將雷射光均勻分布投射在測量空間中形成「散斑」，散斑具有高度隨機性，空間中任兩處的散斑都會視為不同的圖案，再經由紅外線攝影機記錄空間中的散斑(laser speckle)，然後交由硬體晶片計算出深度資訊。

LG 的 3D 立體顯示電視使用的是搭配 3D 眼鏡偏光式顯示技術，播放影片時，同時將左右圖像顯示于螢幕，此時搭配上偏光鏡 (Polarizing glasses)，透過偏光原理我們的左右眼會分別得到一影像，再經由大腦解析後就會形成了 3D 立體影像。其可接受的 2D 轉 3D 的輸入格式如圖 5.1 所示，分別為 2D 影片、左右、上下、棋盤式、欄交錯、列交錯五種格式。

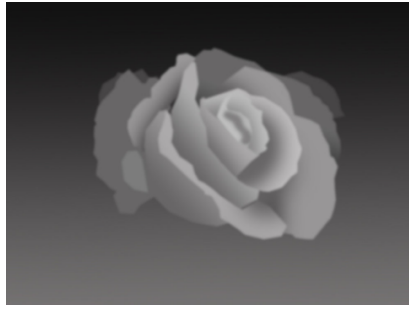


圖 5.1：可接受的 2D 轉 3D 的輸入格式。

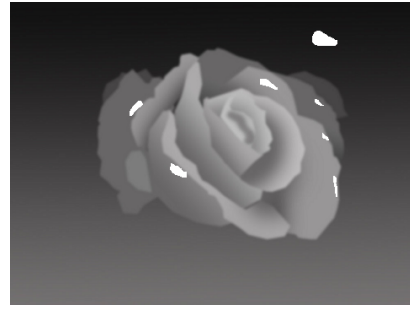
第二節 深度圖修補技術的比較

本節我們對五種深度影像修補技術的好壞做評估，分別是第三章第一節提到之 Matrix Completion via Truncated Nuclear Norm Regularization 中的兩個演算法，分別為 Method A 及 Method B、第三章第二節提到的基於影像分割的修補技術、Cai *et al.* [Cai08] 提出的 SVT、[Tel04]及[Ber01]五種影像修補技術。而我們使用了以下的評估方式來比較深度影像經修補後的結果好壞：1. Mean squared error(MSE) = $\frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \|I(i, j) - K(i, j)\|^2$: MSE 能反映一組數據間的離散程度，數字越小代表離散程度越小，在這裡的意義就是圖片越相似；2. Peak signal to noise ratio(PSNR) = $10 \cdot \log_{10} \left(\frac{MAX_I^2}{MSE} \right) = 20 \cdot \log_{10} \left(\frac{MAX_I}{\sqrt{MSE}} \right)$: MSE 的延伸，單位為 db，數字越大表示圖片越相似，然而 PSNR 偶爾會出現分數與人類觀感成反比的情形，因為人們對於數據上的誤差不一定很敏；3. Structure similarity(SSIM) = $[l(x, y)]^\alpha [c(x, y)]^\beta [s(x, y)]^\gamma$ ，其中 $\alpha > 0$ 、 $\beta > 0$ 、 $\gamma > 0$ 為調整 $l(x, y)$ 、 $c(x, y)$ 、 $s(x, y)$ 相對重要性的參數， $l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}$ ， $c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}$ ， $s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}$ ， μ 為平均值， σ 為標準差， C_1 、 C_2 、 C_3 皆為常數，用以維持 $l(x, y)$ 、 $c(x, y)$ 、 $s(x, y)$ 的穩定:SSIM 比較了兩圖片之間的對比度(c)、光線(l)和結構(s)後，估算出兩影像相似的程度，用百分比表示，越接近 100% 表示越相似，此衡量方式解決了 PSNR 中分數與人類觀感成反比的情況。

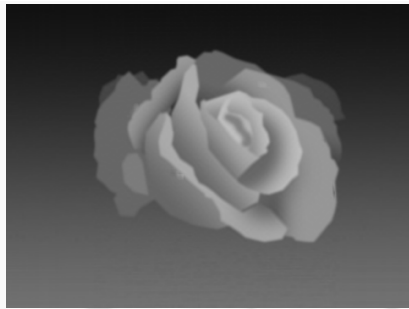
為了做影像修補好壞的評估我們需要有 Ground Truth，以圖 5.2 為例，我們先將圖 5.2(a)一張完美的深度影像設為 Ground Truth，模擬深度影像可能會有缺陷處遮蓋掉一部分的資訊造成影像有缺值如圖 5.2(b)所示，接著使用上述的五種影像修補技術對其作修補，其修補的結果如圖 5.2(c)~(h)所示，再對其結果使用上述三種的評估方式來衡量結果好壞，如表 5.1 所示即是對圖 5.2 的修補的結果做評估。



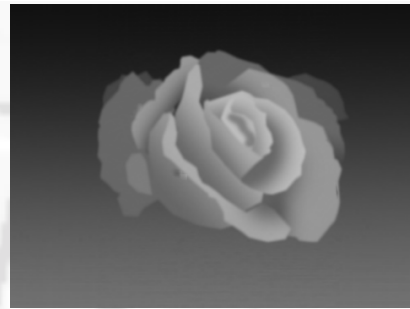
(a)原始影像



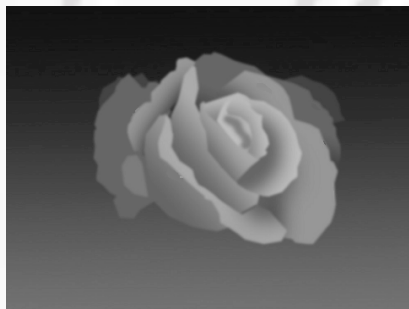
(b)遮蓋掉一些資訊後的深度影像



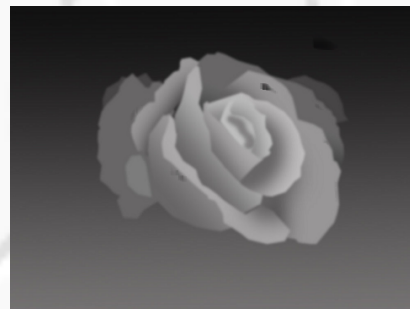
(c)Matrix Completion via TNNR –
Method A



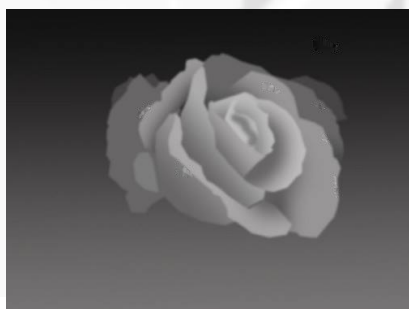
(d)Matrix Completion via TNNR –
Method B



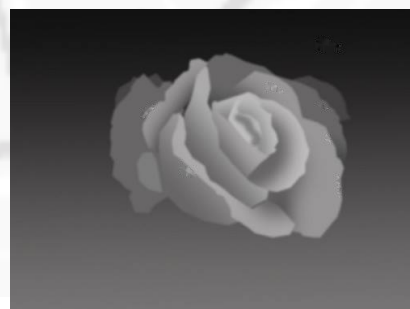
(e)Image Segmentation



(f)[Cai08]SVT



(g)[Ber01]



(h)[Tel04]

圖 5.2 深度影像修補的比較，(a)原始影像，(b)遮蓋掉一些資訊後的深度影像，(c) Matrix Completion via TNNR –Method A，(d) Matrix Completion via TNNR –Method B，(e)Image Segmentation，(f)[Cai08]SVT，(g)[Ber01]，(h)[Tel04]。

	MSE	PSNR	SSIM index	Execution time
Image Segmentation (Ours proposed method)	2.40302	44.3232db	99.02%	Below 1min
Matrix Completion via TNNR – ADMM	1.47114	46.4543db	99.38%	21min
Matrix Completion via TNNR – APGL	1.69269	45.845db	99.41%	17min
SVT	2.83728	43.6018db	98.23%	27min
[Tel04]	2.70559	43.8082db	98.51%	Below 1min
[Ber01]	2.74679	43.7426db	98.51%	Below 1min

表 5.1 深度影像修補的評估

做完影像修補測試後發現，Matrix completion 適合處理缺值是稀疏散佈的情況，對於大量缺值散佈的情況仍可以處理的很好，但是若為一大區塊的缺值就會比較不理想；而影像分割由於是根據顏色跟位置做分割，因此適合處理彩色影像俱有明顯差異的情況，若顏色相近就不理想，另外 SVT 的處理深度影像的結果明顯較弱於 TNNR 兩方法，尤其在缺值比例越大的情況下修補得結果也明顯較差；[Tel04]及[Ber01]兩方法則是使用較簡單的方式去考慮填補，因此雖然非常快速，但結果較不理想。其中比較特別的是在這種深度影像比較規律的情況下，Matrix Completion via TNNR - ADMM 的填補結果與 APGL 難分軒輊，但是考慮到 Kinect 取得的深度影像是比較不規律且較一般影像多雜訊，而 APGL 在處理有雜訊的情況時可以有較佳的表現，因此選擇了 APGL 修補的深度影像來產生立體影像。另外考慮到將來 Real-time 的可行性，希望可以控制影像修補的執行時間並且確保修補結果俱有一定的修補品質，選擇了 SSIM 數值較高的 Image Segmentation 修補之深度影像來產生另外一組立體影像。

以下是我們使用 Matrix Completion via TNNR –APGL、Based on Image Segmentation Repairing、[Tel04]及[Ber01]四種深度影像修補方式的對於 Kinect 深度影像的結果比較。

第三節 3D 立體影像的結果比較

參考深度影像修補的結果評估，我們在產生立體影像的部分採用了 Matrix Completion via TNNR - Method B 及 基於 Image Segmentation 技術來修補的兩種深度影像，並且與市面上販售的 3D 電視提供之 2D 轉 3D 立體影像做比較。關於 3D 立體影像的評估，由於還沒有較完善且客觀的評估方式，我們使用問卷評測的方式做比較。根據產生立體影像時造成觀察不佳的主要因素設計問題，藉以判斷結果好壞，問題總共為四題分為兩種：一種為分為五等第的問題；一種為是非題，樣本數總共為 50 個人。

問卷題目內容：

1. 3D 立體影像是否俱有由近而遠的層次感（分為五等：分別表示幾乎沒有層次感、很少層次感、一般層次感、很有層次感、非常有層次感）3D 立體感主要是來自於對畫面中的物體感覺到不同的相對距離，因此當深度影像層次比較多的時候可以感覺 3D 影像中俱有由近而遠的差別，進而使立體觀感更豐富，結果如表 5.2 所示，垂直坐標為人數，水平坐標分別為三種立體影像；

2. 3D 看起來的舒適程度（分為五等：分別表示非常不舒適、很不舒適、一般舒適、很舒適、非常舒適）建立出的立體影像在強調立體感之餘，應該考慮觀看者的舒適度，結果如表 5.3 所示，垂直坐標為人數，水平坐標分別為三種立體影像；

3. 深度感正不正確?(是/否) 是否有該凸部分的凹進去了，或是該凹的部分凸出來了，深度判斷若錯誤時，會導致觀看時有不好的感受，結果如表 5.4 所示；

4. 有感覺到鬼影嗎，也就是影像邊緣重疊的感覺(是/否):若產生立體影像用的新視角影像沒處理好便會導致此問題，結果如表 5.5 所示。

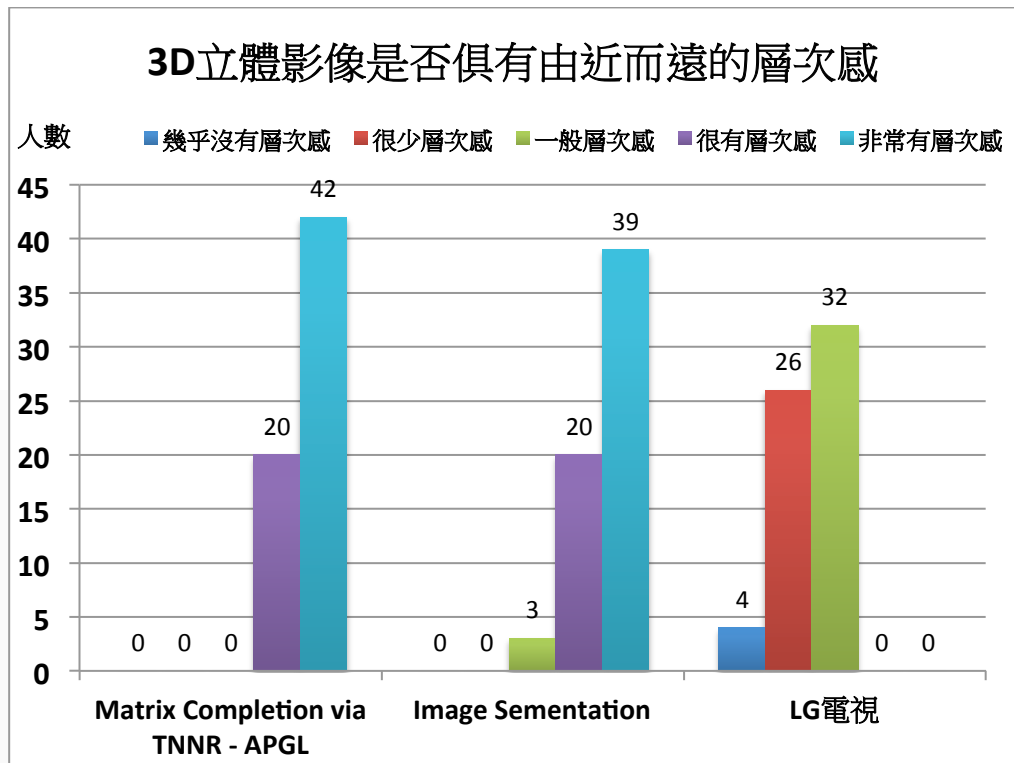


表 5.2 3D 立體影像是否俱有由近而遠的層次感

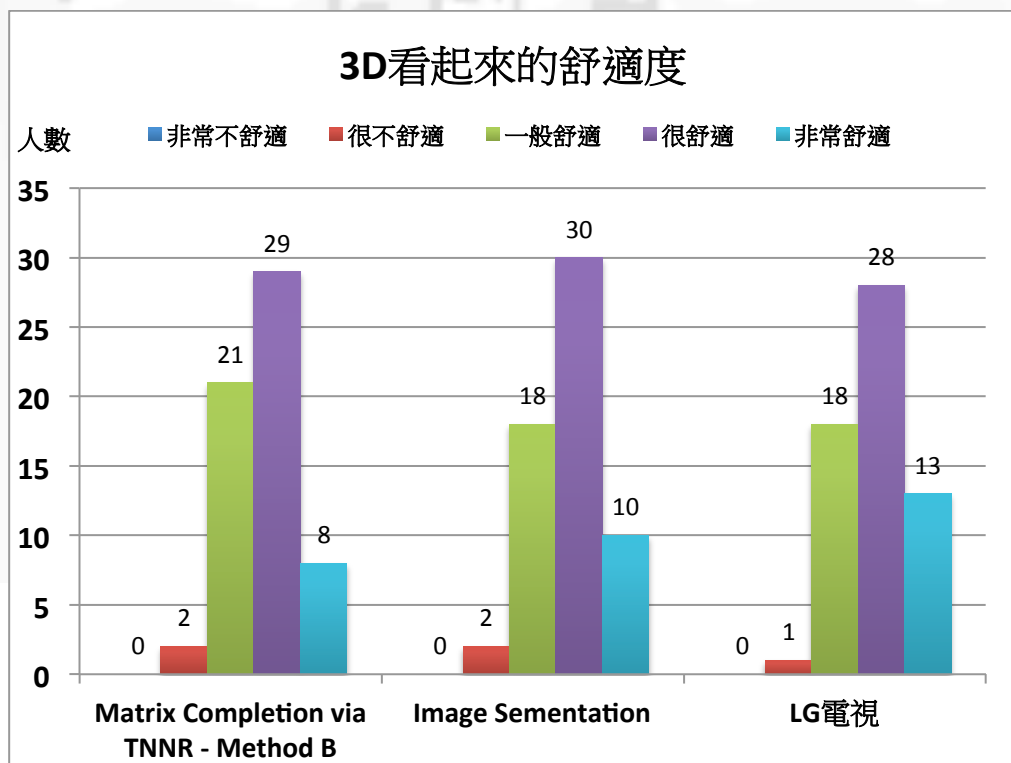


表 5.3 3D 看起來的舒適度

	Matrix Completion- APGL	Image Segmentation	LG 電視
深度感正確	51 人	46 人	8 人
深度感不正確	11 人	16 人	54 人

表 5.4 深度感正不正確

	Matrix Completion- APGL	Image Segmentation	LG 電視
影像邊緣有重疊的情況	7 人	25 人	22 人
影像邊緣沒有重疊情況	55 人	37 人	40 人

表 5.5 影像邊緣重疊

結果顯示使用 Kinect 取得之深度影像所生成的立體影像俱有較佳的深度層次感，也較為符合人眼在觀看一景致時，會感覺到較豐富的距離差異。而當立體感受越明顯時，也較容易造成觀看者的不舒適感，而這種不舒適感的層度也因人而有很大的差異，有些人可能光看到 3D 立體影像就開始覺得身體不適、頭暈。為了讓 3D 立體影像更貼近大眾，未來應該更深入研究導致觀看者不舒適的可能因素，提供更多的立體影像可調整性來適應不同的觀賞者。而大部份的人覺得利用 Matrix Completion- APGL 修補後所提供的深度資訊較為正確，且也不覺得影像邊緣有重疊的情況，因此此種立體影像頗受受訪者青睞。

第六章結論及未來方向

3D 立體顯示器的發展提供了三維立體空間的深度感受，使人們更接近真實自然的視覺感受。目前 3D 顯示技術已是很成熟的技術，除了搭配眼鏡的 3D 影像技術外，也有不需輔助配件的裸眼 3D 技術，然而目前可用的 3D 內容來源有限，造成推廣難度較高，也因此造成相關配件、產品單價過高。即使用戶有經濟能力購買高檔 3D 影音設備也受限于可用的數位內容不夠多，讓 3D 功能等同虛設。本研究提出的將 RGB+D 影像轉換成 3D 立體內容影像即為了舒緩部分立體內容不足的問題。

近年來有許多關於如何產生 3D 立體影像的研究，而從事這類研究，無可避免的常會面臨到深度資訊估測以及產生新視角影像後修補的問題。深度估測的方式很多，有人工判斷、有根據深度線索判斷或是使用深度攝影機取得，人工判斷及根據深度線索判斷都相較耗時，尤其根據深度線索做判斷誤差的機率也較高。而深度攝影機則避免了深度資訊取得的困難，其誤差也較根據深度線索判斷來的低。然而使用深度攝影機取得的影像會有一個很大的問題，依照拍攝的情景不同，其影像可能會有不同程度的深度影像破洞，而深度資訊的完整及精準度直接影響了立體內容的好壞，因此需要對深度影像做合理的填補。而影像填補在過去亦有許多研究，主要考慮了顏色、紋理結構等因素，本研究使用兩種方式作處理，第一種是利用矩陣完整化技術的修補方式；第二種是使用以影像分割為基礎的修補方式，其中矩陣完整化是依據影像本身低 rank 的性質對影像做合理的填補，而基於影像分割的修補方式則是考慮顏色跟空間上位置的關係對影像作分割後再做填補，矩陣完整化的修補結果較後者的修補結果佳，但是修補效率則是後者佳於前者，因此若考慮將此系統設計為即時系統可考慮採用後者的修補方式，另外亦可結合兩者填補方式使結果更好。

實驗結果顯示我們的立體內容深度層次感相較於 2D 轉 3D 的立體內容明顯更

佳，且相較于 2D 轉 3D 的技術，我們的研究使用的是深度攝影機所取得的深度資訊，因此深度資訊具有較低的誤差，產生出的立體影像能給予觀看者更好的立體感受。然而我們使用的實驗器具 Kinect 採用的深度感測技術是利用紅外線做測量，因此若於日光充足的情況下會受到干擾。

由於目前我們修補時只考慮當下彩色及深度影像的資訊，未來針對影片作處理的時候我們希望能夠將原本使用的 Matrix completion 改為使用 Tensor completion，把時間前後的影像資訊也當作一個特徵，增加修補影像可以參考的資訊，使其修補結果更趨穩定。3D 已經成為一種潮流，例如遊戲機 3DS、3D 電影及 3D 列印，而 Kinect 則提供了一種相對簡便取得深度資訊的方式，只要將其深度影像做適當的修補之後，便可以使用在更多應用上，例如：3DS 列印，若使用 Kinect 所取得之深度資訊便得以建立出拍攝物的立體模型。

參考文獻

- [Ide08] I. Ideses, L. Yaroslavsky, and B. Fishbain, “Depth Map Manipulation for 3D Visualization,” *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video*, pp.337-340, 2008.
- [Bat04] S. Battiato, S. Curti, M. La Cascia, E. Scordato, and M. Tortora, “Depth Map Generation By Image Classification,” *Proc. of SPIE IS&T/SPIE's 16th Annual Symposium on Electronic Imaging*, pp. 95-104, 2004.
- [Com97] D. Comaniciu, and P. Meer, “Robust Analysis of Feature Spaces: Color Image Segmentation,” *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 750-755, June 1997.
- [Ang10] L. J. Angot, W. J. Huang and K. C. Liu, “A 2D to 3D video and image conversion technique based on a bilateral filter,” *Proc. of SPIE-IS&T Electronic Imaging*, Vol. 7526, 2010.
- [Che10] C. C. Cheng, C. T. Li, and L. G. Chen, “A 2D-to-3D conversion system using edge information,” *Int'l Conf. on Consumer Electronics (ICCE)*, pp.377-378, 2010.
- [Li06] P. Li, and R. K. Gunnewiek, “On Creating Depth Maps from Monoscopic Video using Structure from Motion,” *Proc. of IEEE Workshop on Content Generation and Coding for 3D-television*, pp.508-515, 2006.
- [Har02] P. Harman, J. Flack, S. Fox and M. Dowley, “Rapid 2D to 3D Conversion,” *Proc. SPIE*, Vol. 4660, pp.78-86, 2002.
- [Rob88] L. S. Robert, (1994), “Cognition and the Visual Arts,” *Cambridge, MA, The MIT Pres.*
- [Cam12] M. Camplani and L. Salgado, “Efficient spatio-temporal hole filling strategy for Kinect depth maps,” *Proc. SPIE*, Vol. 8290, 2012.
- [Xu12] K. Xu, J. Zhou and Z. Wang, “A method of hole-filling for the depth map generated by Kinect with moving objects detection,” *IEEE on Broadband Multimedia Systems and Broadcasting (BMSB)*, pp.1-5, 2012.
- [Cri04] A. Criminisi, P. Perez, and K. Toyama, “Region filling and object removal by exemplar-based image inpainting,” *IEEE Transactions on Image Processing*, pp. 1200 - 1212, 2004 .

- [Feh04] C. Fehn, “Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV,” *Proc. SPIE, Stereoscopic Displays and Virtual Reality Systems*, 2004.
- [Zin10] S. Zinger, L. Do and P. H. N. de With, “Free-viewpoint depth image based rendering,” *Journal of Visual Communication and Image Representation*, Vol. 21, pp.533-541, 2010
- [Lai13] Y. K. Lai, Y. F. Lai, and Y. C. Chen “An Effective Hybrid Depth-Generation Algorithm For 2d-To-3d Conversion In 3d Displays,” *Journal of Display Technology*, vol.9, pp. 154-161, 2013.
- [Ach12] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua and S. Süsstrunk, S. “SLIC Superpixels Compared to State-of-the-Art Superpixel Methods,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 34, Issue 11, pp.2274-2282, 2012.
- [Kom06] N. Komodakis and G. Tziritas, “Image Completion Using Global Optimization,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2006.
- [Ras05] C. Rasmussen and T. Korah, “Spatiotemporal Inpainting for Recovering Texture Maps of Partially Occluded Building Facades,” *Proc. IEEE Int’l Conf. Image Processing*, 2005.
- [Ji10] H. Ji, C. Liu, Z. Shen, and Y. Xu, “Robust Video Denoising Using Low Rank Matrix Completion,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2010.
- [Lee10] J. A. Lee and M. Verleysen “Unsupervised dimensionality reduction: Overview and recent advances,” *The 2010 International Joint Conference on Neural Networks (IJCNN)*, pp. 1-8, 2010.
- [Can08] E. J. Candes and B. Recht, “Exact low-rank matrix completion via convex optimization,” *Annual Allerton Conference on Communication, Control, and Computing*, pp. 806-812, 2008
- [Zha05] L. Zhang and W. J. Tam, “Stereoscopic Image Generation Based on Depth Images for 3DTV,” *IEEE Trans. Broadcast*, vol. 51, pp. 191–199, 2005.
- [Tom98] C. Tomasi and R. Manduchi, “Bilateral Filtering for Gray and Color Images,” *Proc. IEEE Int’l Conf. on Computer Vision*, vol. 51, pp. 191–199, 2005
- [Cai10] J. F. Cai, E. J. Cande`s, and Z. Shen, “(ASi),” *SIAM J. Optimization*, vol. 20,

pp. 1956-1982, 2010.

[Hu13] Y. Hu, D. Zhang, J. Ye, X. Li, and X. He, “Fast and Accurate Matrix Completion via Truncated Nuclear Norm Regularization,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 35, No. 9, Sept. 2013.

[Rec10] B. Recht, M. Fazel, and P.A. Parrilo, “Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization,” *SIAM Rev.*, Vol. 52, No. 3, pp. 471-501, 2010.

[Toh10] K. C. Toh and S. Yun, “An Accelerated Proximal Gradient Algorithm for Nuclear Norm Regularized Least Squares Problems,” *Pacific J. Optimization*, pp. 615-640, 2010.

[Wri09] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, “Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Matrices via Convex Optimization,” *Proc. Advances in Neural Information Processing Systems*, 2009.

[Tel04] A. Telea, “An Image Inpainting Technique Based on the Fast Marching Method,” *Graphics, GPU, & Game Tools*, 2004.

[Ber01] B. Bertalmio, A. L. Bertozzi, and G. Sapiro, “Navier-stokes, fluid dynamics, and image and video inpainting,” *Computer Vision and Pattern Recognition*, Vol. 1, pp.355-362, 2001

[劉 10] 劉楷哲、吳其霖、黃偉豪、陳信榮、李錕、羅豐祥， “基於3D顯示器格式之即時3D內容合成技術” ， *影像與識別期刊(IPPR)* 2010年， Vol. 16， No. 2。

[林 12] 林晃巖， “試看紅塵的深度 —— 漫談 3D 立體顯示器的發展”
<http://www.ee.ntu.edu.tw/hischool/doc/2012.04.pdf>。

[賴 99] 賴文能，陳韋志， “淺談 2D 至 3D 視訊轉換技術” ， *影像與識別期刊(IPPR)* 2010年， Vol. 16， No. 2。