

國立交通大學

資訊科學系

碩士論文



同步進行機器人定位與場景建置之研究

Simultaneous Map Building and Localization of A Robot

研究生：鄭兆良

指導教授：莊仁輝 教授

中華民國九十三年六月

同步進行機器人定位與場景建置之研究  
Simultaneous Map Building and Localization of A Robot

研究生：鄭兆良                      Student：Zhao-Liang Zheng

指導教授：莊仁輝                    Advisor：Jen-Hui Chuang

國立交通大學  
資訊科學系  
碩士論文

A Thesis

Submitted to Institute of Computer and Information Science

College of Electrical Engineering and Computer Science

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Computer and Information Science

June 2004

Hsinchu, Taiwan, Republic of China

中華民國九十三年六月

# 同步進行機器人定位與場景建置之研究

學生：鄭兆良

指導教授：莊仁輝 博士

國立交通大學

資訊科學系

## 摘要



由二維的影像重建出物體的三維模型，在電腦視覺的領域裡已被研究多年，這個技術若應用在機器人上，便可以幫助機器人進行場景的重建及定位。隨著數位相機的普及，可選擇的相機種類也越來越多。目前市面上的相機大致可粗分為兩種等級，一種為畫素高，功能齊全，強調能拍出生動影像的數位相機，價格較高。另一種則是應用在電腦視訊上，體積小，不需電池，強調即插即用，影像品質較不穩定的 webcam。若考慮家用機器人視覺系統的應用，高價位的相機雖然有畫素高、影像品質佳等優點，相對的，畫素高將造成影像處理的時間增加。另一方面，成本的考量也是個重點，若能以較便宜且方便的 webcam，達到令人滿意的效果，可大大提升家用機器人之實用性。本論文即以一套現有的電腦視覺軟體系統，針對一些較低階的 webcam，做三維重建與相機定位之精確度分析。藉由分析的結果，探討使用低階相機為機器人視覺，幫助機器人做定位和場景重建的可行性，並列出幾點可能影響精確度的因素，以供日後改進做參考。

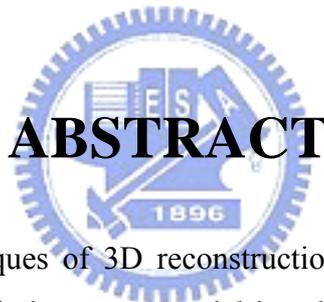
# Simultaneous Map Building and Localization of A Robot

Student : Zhao-Liang Zheng

Advisor : Dr. Jen-Hui Chuang

Institute of Computer and Information Science

National Chiao Tung University



## ABSTRACT

Computer vision techniques of 3D reconstruction from 2D images have been studied for decades. Such techniques are crucial in robot vision in that map building of environment and localization of a robot can be achieved. Recently, digital cameras with a great variety are available to the general public and can roughly be separated into two categories. High-end cameras can take images with high resolution and have a rich set of functions. The low-end ones are mostly small in size and used in applications such as net meeting, and are often called “webcam”. They don’t need batteries and usually support plug-and-play. However, the quality of image taken by webcams is not stable. Considering the application of computer vision in home robots, although high-end cameras have the advantage of better image quality, the high-resolution of images will result in higher processing time. On the other hand, if we can achieve satisfactory results by using low-cost and convenient webcams, the practicality of home robots will be improved greatly. In this paper, we analyze the results of 3D reconstruction and localization using webcams. Issues of using these low-end cameras, especially for data precision, as the “eye” of a robot are discussed.

## 致謝

感謝指導教授莊仁輝老師兩年來耐心並親切的予以指導，使我得以完成此論文，從中學習到從事研究的方法與應有的嚴謹態度。並於研究之外，不時提醒我們注意健康，多喝水，多運動，讓我健康地渡過充實愉快的碩士班生涯，謝謝老師。此外，由於三位口試委員，薛元澤教授、雷欽隆教授以及顏嗣鈞教授的意見及指教，讓本篇論文得以更加充實且完善，在此衷心感謝他們。另外，也感謝智慧型系統實驗室的各位同學們，有了他們的陪伴與互相扶持，使我不論在課業上或日常生活，都獲益良多。最後，感謝一直以來支持我的家人，以及所有幫助過我的人，謝謝大家。



# 目錄

摘要.....	I
<b>ABSTRACT</b> .....	<b>II</b>
致謝 III	
目錄 IV	
圖目錄 VI	
表目錄 VII	
第一章 簡介.....	1
第二章 相關研究及應用.....	2
2.1 多視角立體重建.....	2
2.2 單相機多視角重建.....	3
2.3 主動視覺重建.....	4
2.4 相機姿勢估計.....	5
第三章 投影幾何.....	7
3.1 齊次座標表示法.....	7
3.2 階層轉換.....	8
3.2.1 投影層.....	8
3.2.2 仿射層.....	9
3.2.3 公制層.....	10
3.2.4 歐幾里德層.....	11
3.2.5 各階層的特性整理.....	11
3.3 相機模型.....	13
3.4 極線幾何與基本矩陣.....	17
3.4.1 極線幾何.....	17
3.4.2 基本矩陣.....	18
3.4.3 必要矩陣.....	19
3.5 相機校正.....	20
第四章 三維重建.....	21
4.1 三維重建之步驟.....	21
4.1.1 估算基本矩陣.....	21
4.1.2 估算投影矩陣.....	22
4.1.3 歐氏三維重建.....	24
4.2 估測相機位置.....	25
第五章 三維重建之精確度分析.....	27
5.1 使用高階數位像機Agfa 1680 進行三維重建.....	27

5.2	使用Orite MY120 進行三維重建.....	37
5.3	使用Orite MC310 進行三維重建.....	41
5.4	使用Logitech QuickCam Express實驗.....	46
第六章	結論與未來工作.....	50
參考文獻	.....	51



# 圖目錄

圖 2.1	日常生活中，常見的由三邊所構成的角。.....	5
圖 3.1	不同幾何轉換下的立方體。.....	13
圖 3.2	針孔相機幾何 (a)針孔相機模型 (b)影像點的座標。.....	14
圖 3.3	極線幾何。.....	18
圖 4.1	由 $E$ 做重建後四種可能的幾何表示。.....	23
圖 4.2	相機座標與點座標的旋轉及平移關係。.....	26
圖 5.1	三維重建實驗流程圖。.....	28
圖 5.2	正交方塊圖板之 (a)側視影像 (b)俯視影像。.....	28
圖 5.3	實驗時固定相機，令兩方塊圖板平面相交形成的軸為旋轉軸，旋轉正交方塊圖板，取 6 張不同視角影像，相鄰影像視角間隔 $10^\circ$ 。.....	29
圖 5.4	(a)~(f)不同角度對同一校正板所拍攝的六張影像及其特徵點。.....	31
圖 5.5	(a)邊長標準差 (b)邊夾角平均誤差 (c)邊夾角標準差 之長條圖。.....	34
圖 5.6	(a)左平面誤差變異量 (b)右平面誤差變異量 (c)平面夾角誤差 之長條圖。.....	35
圖 5.7	(a)左距離誤差 (b)右距離誤差 (c)視角誤差 之長條圖。.....	36
圖 5.8	(a)~(f)不同角度對同一校正板所拍攝的六張影像及其特徵點。.....	38
圖 5.9	(a)~(f)由相機Orite MC310 所取得的立體校正板影像。.....	42
圖 5.10	(a)~(f)之影像，分別為圖 5.10 中(a)~(f)影像還原扭曲後的結果。.....	43
圖 5.11	(a)Logitech QuickCam Express (b)Orite MY120 所取之校正板影像。..	46
圖 5.12	此圖中之(a)、(b)影像，分別為圖 5.11 之(a)、(b)影像，利用Harris corner detection 的方法自動取角點，並人工刪除不必要點後的結果。.....	47

# 表目錄

表 3.1	不同幾何轉換之比較.....	12
表 3.2	不同投影轉換層級的投影矩陣.....	12
表 5.1	邊長與邊夾角重建結果分析.....	32
表 5.2	平面重建與相機定位分析.....	33
表 5.3	邊長與邊夾角重建結果分析.....	39
表 5.4	平面重建與相機定位分析.....	40
表 5.5	邊長與邊夾角重建結果分析.....	44
表 5.6	平面重建分析.....	45
表 5.7	邊長與邊夾角重建結果分析.....	48
表 5.8	平面重建與相機定位分析.....	49



# 第一章 簡介

由二維的影像得到物體三維的立體資訊，在電腦視覺中是一個廣泛被研究的題目，藉由這些立體資訊，許多實際的應用可以達成。電腦繪圖、虛擬實境等等，都會利用到影像重建的技術，比如一日常的物品，利用多個視角的照片，便可在電腦中建出它的三維模型，再輔以一些貼圖的技術，即可做出許多栩栩如生的模型，提供給動畫、電玩甚至電影來使用。若將這些重建出來的模型，搭配上一個較大型的三維場景，則可以建立出一個虛擬實境的環境。這些應用實現在機器人視覺上，便是機器人所在地之場景及物體重建。進一步的，如果於場景三維重建時，可以還原相機位置資訊，即機器人所在位置資訊，則有了這些資訊的幫助，將可達成機器人導航以及物體處理(例如抓取物體)的應用。

在考慮實際的應用時，許多因素會影響到重建與定位結果。有些因素與相機所攝得的影像品質有關，高階的數位像機較不會有影像品質不良的問題。本篇論文重點即在於使用不同的低階相機，分析不同的影像品質對於重建與定位精確度的影響。全文共分為六章，除了此章為簡介外，第二章介紹一些三維重建及相機姿勢估計的相關研究與應用。第三章介紹投影幾何、相機模型、極線幾何、基本矩陣以及必要矩陣。第四章介紹三維重建的步驟以及相機位置之估算。第五章使用不同相機進行三維重建的各種，並分析其精確度。第六章為結論並說明本研究接下來的一些方向。

## 第二章 相關研究及應用

由二維影像進行三維場景重建是一個已被研究多年的題目，而根據不同應用的需求，便會有不同的研究重點。比如機器人導航時並不需要詳細地重建出如相片般真實的 (photo-realistic) 場景，只需要重建出重要的環境結構，強調的是計算速度以及重建結果的精確度。若是要將一真實物體重建成電腦模型，在拍攝相片時相機的內外部參數可以加上一些已知的限制條件，比如假設 principal point 在影像中央，相機沒有 skew，或是相機的焦距  $f_x = f_y$ ，甚至已知的相機旋轉及平移，這些限制都可以用來簡化重建的過程。有些重建方法不需相機事先校正內部參數，並在拍攝影像的過程中，允許相機焦距的變化，此易適性將使得相機取像時，較不會受地勢限制，而能得到較好的待重建物體影像，於實際應用時獲得較佳的重建結果。除了物體及場景的重建外，相機姿勢的估計，在某些應用也是很有用的資訊，如機器人定位及導航。

在以下的幾個小節裡，我們將會介紹一些相關的三維重建及相機姿勢估計的實例，並且簡單介紹其特色，而透視理論及影像三維重建的詳細原理，將會在下一章做說明。

### 2.1 多視角立體重建

多視角立體重建是從多個不同視角的位置取得影像資訊，將這些影像兩兩比較，透過一些相似度比較 (correlation-based) 的方法，找出影像中相對應的特徵，通常為點，進行三維重建。這種重建的好處是利用多台相機同時拍攝影像，在一個特定的時刻就可以取得足夠的資訊來進行重建，如此一來便可針對運動中的物體進行追蹤與軌跡重建。

Narayanan、Rander 和 Kanade [1] 實作了一個系統，他們稱為 3D Dome。3D Dome 是一個由51台同步且完全校正好 (fully calibrated) 的相機所結合而成

的半球形環境，由於所有的相機皆是完全校正過的，於是每一台相機的投影矩陣便可以事先得到。

當有人在 3D Dome 中心進行活動時，所有的相機會紀錄下許多不同視角的影像資訊，透過一個多基線立體重建（multiple-baseline stereo reconstruction）[2]的過程，對於每一台相機可以得到一張密集的深度影像（dense depth graph），也就是影像中的像素灰階代表的是該點的深度資訊。接下來將原始影像上的材質資訊貼上去，便可得到一個簡單的重建模型，作者稱之為可見表面模型（visible surface model, VSM）。但是由於 VSM 是每一台相機所重建的結果，難免會因為重建物體本身的遮蔽關係導致有些地方會沒辦法重建出來，因此作者接下來又將許多的 VSM 結合為一，透過一些最佳化的選取及整合方式，將數個角度所重建出的場景結合成完整表面模型（complete surface model, CSM）。最後同樣將材質貼上 CSM，得到一個完整的重建結果。這個系統的特色是相機已經完全校正好了，因此在三維重建的過程中不需要太複雜的計算便可算出每一個點的三維資訊。但是由於這個系統使用了高達51台相機，因此系統在應用上的彈性將會受到限制，相機間的不同步技術也將會是高難度的挑戰。

## 2.2 單相機多視角重建

相對於2.1節所介紹的方法皆使用兩台以上的相機，本節所要介紹的方法是以單一的相機，由多個不同的視角來取得影像進行重建，這樣的重建有時候又稱為從運動中重建結構（structure from motion）。這麼做的好處是只需要利用一台相機便可以進行重建，所花費的成本較少，其缺點是在某一個時刻只能取得一張影像。

Pollefeys 和 Van Gool [3] 便實作了這樣的一個系統，這個系統的輸入是一連串同一個場景的影像，經過了初步的特徵點擷取步驟，對每張影像分別找出一些特徵點，接下來進行相似度比較找出其對應關係。由於數位的相機在成像時一

定會有些微的誤差，加上影像本身會有不穩定的雜訊，因此 Pollefeys 他們將所找出的特徵點，透過了 RANSAC (RANdom Sampling Consensus) 的方式，算出許多組可能的基本矩陣 (fundamental matrix)，再從這些基本矩陣中選出一個最穩定的。基本矩陣即代表了每兩張影像之中每個點的轉換關係，關於基本矩陣的定義與極線幾何將會在第三章中有詳細的介紹。

找出基本矩陣之後，即可進行投影重建 (projective reconstruction)。透過一些常見的相機內部參數假設，如 principle point 在影像中央、相機沒有 skew、相機的焦距  $f_x = f_y$ ，便可簡化相機的校正過程。經過校正以後，即得到完整的相機內部參數，於是可以重建出一個和真實場景只差一個比例常數的公制重建 (metric reconstruction) 結果。到了這個階段，每兩張影像的相機內外部參數以及投影矩陣都已經得到了，於是作者同樣地計算出密集的深度影像 (dense depth graph)，將影像中的每一個點都算出其深度資訊，最後貼上材質完成三維模型的建立。



## 2.3 主動視覺重建

在2.1節中的系統，相機內外部參數必需事先校正，並且，這些參數在擷取影像的過程中不能改變，即相機的位置、焦距必需保持不變，這限制了它在某些方面的應用。而2.2節中的系統，相機不需事先的校正，在拍攝影像的過程中，允許相機焦距 (即相機內部參數) 改變，但其理論是基於焦距變化時，principle point 不會改變的假設，實際應用時，這個假設可能不成立。

Y. F. Li 和 R. S. Lu [4] 所提出的系統，是一個不需事先相機校正的主動視覺三維重建系統，由一台相機和一投影機所組成。所謂的主動視覺，就是用相機擷取待重建物體影像時，同時使用投影機投射特定圖案 (如棋盤方格) 到待重建物體上。投影機的參數必需於拍攝影像時事先校正好，然而相機內、外部參數可以隨時改變，不需事先校正。這些特性使此系統於實際應用時，較容易適應其所

在的工作環境。

相較於2.1節所述之方法，此系統於重建時不需特徵點的擷取與對應，因此對於特徵點較少的自由形態表面（free-form surface），如人臉模型，也能完成令人滿意的重建結果。

## 2.4 相機姿勢估計

除了三維重建外，相機姿勢估計（estimation of camera pose）也是電腦視覺中一個研究項目，常見的應用有機器人、車輛導航、飛行器定位、物體追蹤等等。本論文所用的相機定位方法，為利用三維重建過程中所得的相機外部參數，進而求得相機與重建物體之相對位置。而一般的相機姿勢估計，除了能夠得知相機的位置（location）外，還包括了相機朝向（orientation）的資訊。

Fanhuai Shi, Xiaoyun Zhang, Yuncai Liu [5] 提出一個 correspondence-based 相機姿式估計方法。其相機焦距需事先校正完成，並且過程中保持不變。利用日常生活中，人造物體常見的特徵——由三個邊所構個的角，如圖2.1，假設此角的



圖2.1 日常生活中，常見的由三邊所構成的角。

三個角邊的夾角皆為已知，藉由二維影像與三維實際物體中角邊的對應關係，可求得相機座標系和角座標系間的旋轉矩陣，此即相機之朝向。再加上一個已知的

非角點（此點不為角點，且到角點的距離為已知），可求得相機到角點平移（translation）矩陣，也就是得到了相機之位置。此方法由於必需事先知道三個角邊的夾角與一個已知點，因此在實際應用上會有些限制。



## 第三章 投影幾何

本章將介紹一些與本論文所採用系統有關的觀念。依次為齊次座標表示法、階層轉換、相機模型、極線幾何與基本矩陣，最後說明本系統所使用的相機校正方法。由於此方面研究已有相當長遠的一段時間，在這裡提到的僅是與本系統較相關的部分，若想要深入瞭解本章沒有提到的部分，請參考電腦視覺相關文獻 [6][7][8]。

### 3.1 齊次座標表示法

在代數學中，平面上的一條直線可以用一等式  $ax+by+c=0$  來表示，而不同的  $a$ 、 $b$ 、 $c$  參數便代表了不同的直線，因此我們可以向量的形式  $(a,b,c)^T$  來代表平面上的一條直線。如果一個點  $x=(x_0,y_0)^T$  在一條直線  $l=(a,b,c)^T$  上，則將點座標  $(x_0,y_0)$  代入直線方程式，得  $ax_0+by_0+c=0$ 。為了能使用向量內積來表達這個式子，我們可將  $x$  座標重新寫成  $x=(x_0,y_0,1)^T$  的形式，則：

$$ax_0+by_0+c=(x_0,y_0,1)(a,b,c)^T=x^Tl=0$$

其中  $(x_0,y_0,1)^T$  和  $(a,b,c)^T$  就是平面上點  $x$  和直線  $l$  的齊次座標表示法。

需要注意的是， $ax_0+by_0+c=0$  和  $kax_0+kby_0+kc=0$  代表的是同一條直線，因此在  $k \neq 0$  的情況下，由向量形式表示的  $(a,b,c)^T$  和  $(ka,kb,kc)^T$ ，代表同一條直線。同樣的情形， $ka(x_0)+kb(y_0)+kc=0$  和  $a(kx_0)+b(ky_0)+c(k)=0$  的意義是相同的，因此在  $k \neq 0$  的情況下， $(x_0,y_0,1)^T$  和  $(kx_0,ky_0,k)^T$  表示的是同一個點。也就是說，在  $x_3 \neq 0$  的情況下， $(x_1,x_2,x_3)^T$  可以對應到我們熟悉的歐幾里德平面上的  $(x_1/x_3,x_2/x_3)$  這一點。另外，在齊次座標表示法中， $(0,0,0)^T$  是沒有定義的。

## 3.2 階層轉換

在我們生活的世界中，所有我們看見的物體都可以用歐幾里德三度空間來表示，但是有時候我們不需要用到完整三度空間的歐幾里德座標轉換，而只需要針對投影幾何中比較簡單且限制較少的轉換來處理。我們可以把這些轉換看成是一個階層性的關係[9][10]。從最簡單的投影 (projective) 層，接下來是仿射 (affine) 層，公制 (metric) 層，最後是歐幾里德 (Euclidean) 層。這些階層性的觀念，彼此在幾何上有著一些轉換的關係，並且有一些不變性存在。和投影層有關的是投影轉換群，和仿射層有關的是仿射轉換群，和公制層有關的是相似群 (similarities)，最後和歐幾里德層有關的是歐幾里德轉換群。很重要的一點是，不同的轉換群之間有著父子群的關係，比方說相似群是仿射群的子群，而他們兩者都是投影群的子群。有關這些轉換群間的詳細介紹，請參考[7]。

在每一種轉換群中，有著一個很重要的特性，那就是不變性，是一種在某一個特定群中，不管經過什麼轉換，都依然保持不變的特性。不同的幾何轉換對應著不同的不變性。這些不變性通常和幾何實體 (entity) 有關，若能恢復並得知其值，便可將幾何轉換做些改變，通常是用來提昇成較高階層的轉換。例如，若我們能夠得知幾何實體的角度和相對長度的資訊，就可以將仿射結構提升到相似轉換。在本節中，我們將會個別對每一層做簡單的介紹，並且以圖例來表示在各個轉換層級下正方體可能的形變，並以表格表示各種轉換的不變性。

### 3.2.1 投影層

首先介紹投影層 (projective stratum)，它是限制最少的一層，因此含有最少的不變性，而投影轉換 (projective transformation) 也是最普通的線性轉換。

三維空間的點轉換若滿足  $\rho y = T_p x$ ，則可稱為投影轉換。其中， $\rho$  為一個常數， $x$ 、 $y$  分別為轉換前和轉換後空間點的齊次座標， $T_p$  為  $4 \times 4$  的可逆矩陣：

$$T_p = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix} \quad (3.1)$$

由於這個轉換矩陣可以有一個比例常數，因此共有15個自由度。

在投影轉換群裡，除了共線的不變性質外，交比（cross-ratio）也是不會隨著投影轉換而改變的。假設有四個共線的點  $M_1, M_2, M_3$  和  $M_4$ ，則可將這四個點表示成  $M_i = M + \lambda_i M'$ ，其中  $\lambda_i$  為純量， $M'$  為該直線的方向向量，則交比便可定義為：

$$R(M_1, M_2, M_3, M_4) = \frac{(\lambda_1 - \lambda_3)(\lambda_2 - \lambda_4)}{(\lambda_2 - \lambda_3)(\lambda_1 - \lambda_4)} \quad (3.2)$$

由式(3.2)可知，交比和  $M$  或  $M'$  的選擇無關，只與共線之四點所形成的各線段之相對長度有關。



### 3.2.2 仿射層

仿射層（affine stratum）和投影層的不同點是，它建立在一個特別的參考平面上，這個平面我們稱之為無窮遠處平面（the plane at infinity）。若我們於投影空間  $P^3$  中選一平面，在不失一般性的情況下，我們可令其齊次方程式為  $x_4 = 0$ ，則此平面便是無窮遠處的平面。

在  $X = [X \ Y \ Z]^T \rightarrow [X \ Y \ Z \ 1]^T$  的對應關係下，我們可以將投影空間看成是包含了仿射空間。對於每個不在無窮遠處平面  $x_4 = 0$  上的點，假設其空間座標為  $(x_1, x_2, x_3, x_4)$ ，我們可以得到  $X = x_1/x_4$ ， $Y = x_2/x_4$ ， $Z = x_3/x_4$ 。

仿射轉換可以表示成：

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} a_{14} \\ a_{24} \\ a_{34} \end{bmatrix} \quad \text{with } \det(a_{ij}) \neq 0, 1 \leq i, j \leq 3 \quad (3.3)$$

若寫成齊次座標系則可改寫為  $M' = T_A M$ ，其中

$$T_A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \quad (3.4)$$

因此仿射層含有12個自由度，而在仿射的情況之下，平行性 (parallism) 是不變的。

### 3.2.3 公制層

公制層 (metric stratum) 相較於仿射層，多了角度與相對長度的限制，而其與歐幾里德層 (Euclidean stratum) 的轉換，其間只差了一個常數倍的比例關係。當我們無法得知物體實際尺寸大小時，我們便只能從影像中得到具有最多限制的這種幾何結構。公制轉換可表示如下：

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \sigma \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} t_{14} \\ t_{24} \\ t_{34} \end{bmatrix} \quad (3.5)$$

其中， $[X \ Y \ Z]^T$  和  $[X' \ Y' \ Z']^T$  分別代表轉換前後的三維點座標， $R = [r_{ij}]$  是正交矩陣，它會滿足旋轉矩陣  $RR^T = I$  且  $\det(R) = 1$  的關係。將三維座標點以齊次座標表示，可再改寫成  $M' = T_M M$ ，其中

$$T_M = \begin{bmatrix} \sigma r_{11} & \sigma r_{12} & \sigma r_{13} & t_x \\ \sigma r_{21} & \sigma r_{22} & \sigma r_{23} & t_y \\ \sigma r_{31} & \sigma r_{32} & \sigma r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3.6)$$

因為正交矩陣只有3個自由度，再加上3個平移及1個常數，所以共有7個自由度。在相似轉換的情形下，有兩個重要的不變性就是相對長度及角度。若可以得到相對長度和角度的資訊，我們便可將仿射轉換提昇為相似轉換。

### 3.2.4 歐幾里德層

最後要介紹的是歐幾里德層 (Euclidean stratum)。它和公制層差不多，只不過把比例常數給固定下來，而使得相對長度成為絕對長度。歐幾里德轉換有6個自由度，包括了3個方位及3個平移。一個歐幾里德轉換可表示為：

$$T_E = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3.7)$$

其中  $r_{ij}$  是正交矩陣的元素。

### 3.2.5 各階層的特性整理

表3.1列出不同幾何轉換的特性，其中包含每個轉換的自由度、轉換矩陣及不變性。表3.2則列出了由三度空間投影到二度空間的投影矩陣[7]。另外，圖3.1則顯示出對於不同的幾何轉換下，立方體所可能呈現的樣子。由圖3.1可知，至少要到正規轉換以上的階層，才比較像立方體原來的樣子。

表3.1 不同幾何轉換之比較

幾何轉換	自由度	轉換矩陣	不變性
投影(projective)	15	$T_p = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix}$	交比 共線
仿射(affine)	12	$T_A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix}$	平行 共線、交比
公制(metric)	7	$T_M = \begin{bmatrix} \sigma_{r_{11}} & \sigma_{r_{12}} & \sigma_{r_{13}} & t_x \\ \sigma_{r_{21}} & \sigma_{r_{22}} & \sigma_{r_{23}} & t_y \\ \sigma_{r_{31}} & \sigma_{r_{32}} & \sigma_{r_{33}} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}$	角度、相對長度 平行 共線、交比
歐幾里德(Euclidean)	6	$T_E = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_X \\ r_{21} & r_{22} & r_{23} & t_Y \\ r_{31} & r_{32} & r_{33} & t_Z \\ 0 & 0 & 0 & 1 \end{bmatrix}$	角度、絕對長度 平行 共線、交比

表3.2 不同投影轉換層級的投影矩陣

投影層級	投影矩陣	說明
投影(projective)	P	P:從三度空間投影到影像平面
仿射(affine)	$[Pp]$	P:無限遠平面和影像平面之間的 homography p:世界座標系原點的投影
歐幾里德(Euclidean)	$A[Rt]$	A:相機內部參數 (R, t):在世界座標中的相機位置

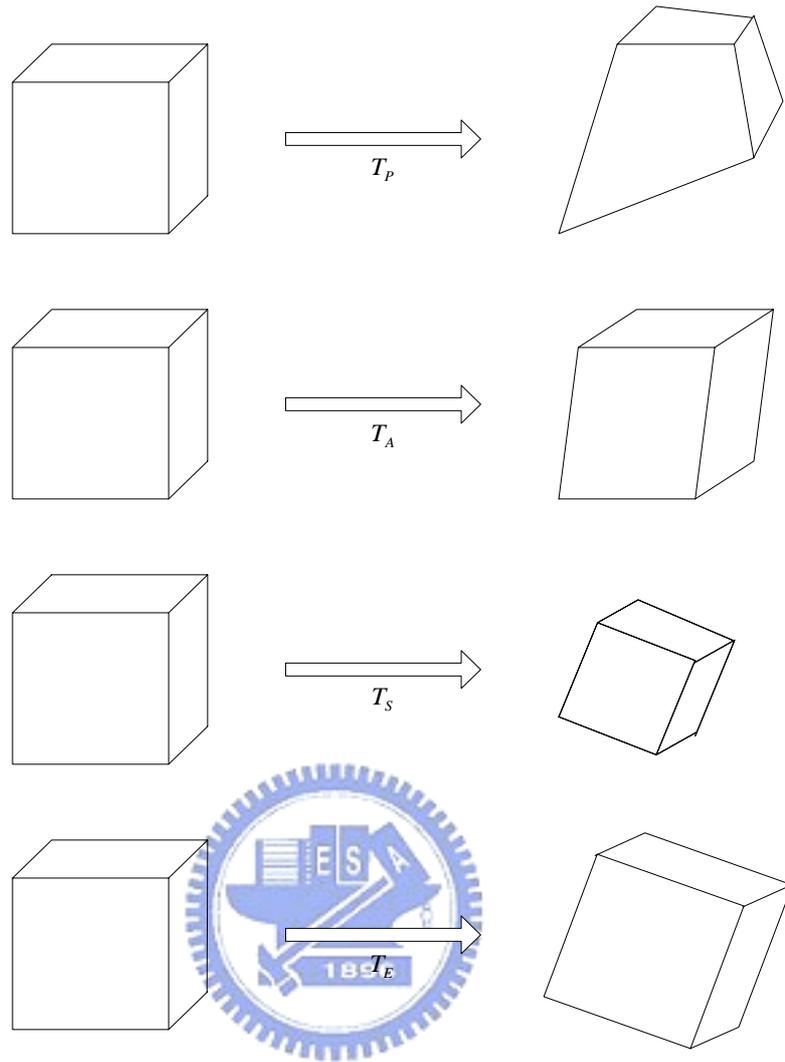


圖3.1 不同幾何轉換下的立方體。

### 3.3 相機模型

在這一節中，我們將會透過數學算式的漸進推導，介紹本系統所使用的一般化相機模型[6]。考慮用中心投影的方法，將一個空間中的點投影到一個平面，我們讓投影中心位在歐幾里德座標系的原點，並且讓成像的平面是  $Z = f$ ，稱為影像平面。考慮一個針孔相機的模型如圖3.2(a)，圖中  $C$  是相機光學中心、 $P$  是 principal point、 $f$  是焦距 (focal length)、 $P$  和  $C$  所形成的直線稱為 principal axis、通過  $P$  和  $Z$  軸垂直的平面就是影像平面。空間中的點  $X = (X, Y, Z)^T$  在影像平面上的對應點，即  $X$  和相機中心  $C$  所形成的直線和影像平面的交點  $x$ 。

用相似三角的幾何關係，參考圖3.2(b)，我們可以知道  $x = (fX/Z, fY/Z, f)^T$ 。  
 如果不考慮最後一個座標，針孔相機可以把歐幾里德三維空間的點對應到歐幾里德二維空間的點：

$$(X, Y, Z)^T \mapsto (fX/Z, fY/Z)^T \quad (3.8)$$

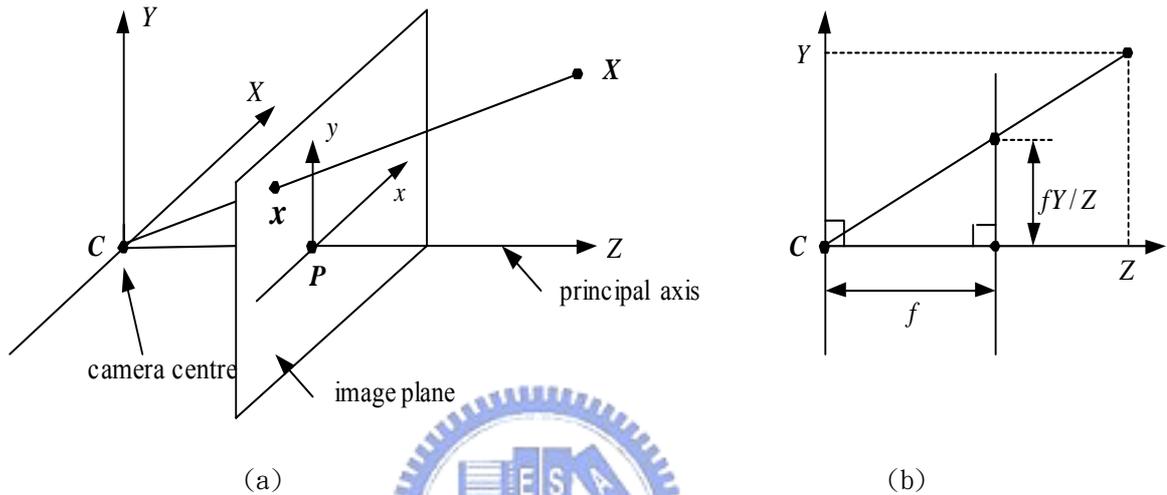


圖3.2 針孔相機幾何 (a)針孔相機模型 (b)影像點的座標。

如果將空間中的點和影像上的點用齊次座標來表示，式(3.8)可以寫成以下的線性方程式：

$$x = \begin{bmatrix} fX \\ fY \\ Z \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = P X \quad (3.9)$$

其中， $X$  表示三維空間中的點的齊次座標， $x$  表示影像平面上的點的齊次座標， $P$  是一個  $3 \times 4$  的矩陣，稱為投影矩陣。

在式(3.9)中，我們假設影像平面的原點剛好就是 principal point，但是實際上可能不是如此。若  $(p_x, p_y)^T$  是影像上 principal point 的座標，則其歐幾里德對應關係為：

$$(X, Y, Z)^T \mapsto (fX/Z + p_x, fY/Z + p_y)^T \quad (3.10)$$

上述的對應我們可以用齊次座標表示如下：

$$x = \begin{bmatrix} fX + Zp_x \\ fY + Zp_y \\ Z \end{bmatrix} = \begin{bmatrix} f & 0 & p_x & 0 \\ 0 & f & p_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = P X \quad (3.11)$$

若將相機內部參數定義為：

$$A = \begin{bmatrix} f & 0 & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{bmatrix} \quad (3.12)$$

那麼，式子(3.11)變可以寫成：

$$x = A[I_3 | 0]X_{cam} \quad (3.13)$$

在式(3.13)中，我們把  $(X, Y, Z, 1)^T$  寫成  $X_{cam}$ ，主要是要強調這個座標是屬於相機座標系統。然而，在一般情況下，三維空間中的點是以世界座標系統表示，在世界座標系統和相機座標系統之間存在著旋轉和平移的關係。

假設  $\tilde{X}$  代表三維空間中一個點的歐幾里德座標， $\tilde{X}_{cam}$  代表在相機座標系統中同一個點的座標， $\tilde{C}$  代表相機中心在世界座標系統的位置， $R$  是一個  $3 \times 3$  的旋轉矩陣，代表著相機座標系統的方向。那麼， $\tilde{X}$  和  $\tilde{X}_{cam}$  之間的關係可以表示成  $\tilde{X}_{cam} = R(\tilde{X} - \tilde{C})$ ，這個方程式可以齊次座標表示如下：

$$\tilde{X}_{cam} = \begin{bmatrix} R & -R\tilde{C} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} R & -R\tilde{C} \\ 0 & 1 \end{bmatrix} \tilde{X} \quad (3.14)$$

將式(3.14)代入式(3.13)可以得到：

$$x = AR[I | -\tilde{C}]X \quad (3.15)$$

其中  $X$  已經是相對於實際的世界座標系統，而式(3.15)已經將理想的針孔相機模型推進為較具一般性的針孔相機模型了。

我們可以看出在一般性針孔相機模型中，相機投影矩陣  $P = AR[I | -\tilde{C}]$  有9個自由度： $A$  有  $f$ 、 $p_x$ 、 $p_y$  共3個自由度， $R$  有3個自由度， $\tilde{C}$  有3個自由度。 $A$  之中的參數稱為相機內部參數， $R$  和  $\tilde{C}$  分別與相機在世界座標系統中的方向和位置有關，稱為外部參數。通常我們用  $t$  來表示  $-R\tilde{C}$ ，所以投影矩陣可以改寫成  $P = A[R | t]$ ，而式(3.15)的針孔模型可以替換如下：

$$x = A[R | t]X \quad (3.16)$$

上面所述的針孔模型是假設影像平面在  $x$  軸和  $y$  軸方向的比例是相等的。但是如果是用數位的 CCD 相機，它的影像平面上  $x$  軸和  $y$  軸方向的比例就有可能不是相等，也就是說經由 CCD 相機所得到的影像的每一個像素(pixel)可能不是正方形。我們假設在影像座標中  $x$  軸和  $y$  軸方向的單位長度內分別有  $m_x$  和  $m_y$  個像素。那麼，在式(3.12)的左邊乘上一個對角矩陣  $diag(m_x, m_y, 1)$ ，就可以把世界座標的點轉換到影像座標的點。因此，CCD 相機的內部校正矩陣就會變成：

$$A = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3.17)$$

其中， $f_x = fm_x$ 、 $f_y = fm_y$ 、 $u_0 = p_x m_x$ 、 $v_0 = p_y m_y$ 。 $f_x$  和  $f_y$  代表相機在  $x$  軸和  $y$  軸方向上的焦距，單位為像素。 $(u_0, v_0)$  代表影像上用像素為單位的 principal point。因此，一個相機投影矩陣有10個自由度。

通常，更具一般性的相機內部參數矩陣還有一個參數 skew ( $\gamma$ )，這個參數在大部分的相機中的值都是0，只有在某些狀況下不為0，將它加入內部參數矩陣中，可得到式(3.18)：

$$A = \begin{bmatrix} f_x & \gamma & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3.18)$$

那麼，一個一般性的相機投影矩陣就有11個自由度， $A$ 便成為一個完整的相機內部參數矩陣。

假設二維的點座標以  $m = [u \ v]^T$  來表示，而一個三維的點座標以  $M = [X \ Y \ Z]^T$  來表示。我們令他們的齊次座標表示法分別為  $\tilde{m} = [u \ v \ 1]^T$  和  $\tilde{M} = [X \ Y \ Z \ 1]^T$ ，一部相機的則用針孔相機的模型來表示，那麼，一個三維的點  $M$  和該點投影在影像上的二維點座標  $m$  就有以下的關係存在：

$$s\tilde{m} = \begin{bmatrix} f_x & \gamma & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} [R | t] \tilde{M} = A [R | t] \tilde{M} \quad (3.19)$$

其中  $s$  是一個比例常數， $[R | t]$  是由世界座標系統轉換到相機座標系統的旋轉矩陣和平移矩陣， $A$  則是相機的內部參數矩陣。

## 3.4 極線幾何與基本矩陣

### 3.4.1 極線幾何

在同一個三維場景，所取得兩張不同視角的影像之間，有一個很重要的限制關係，稱為極線限制（epipolar constraint）。我們以圖3.3來說明這個限制。假設  $M$  是三維空間中的一個點，則它在第一張影像所成像的影像點是  $m$ ，在第二張影像的影像點是  $m'$ 。 $C$  和  $C'$  分別是第一個和第二個相機的光學中心。如圖所示，三維空間的點  $M$  和影像點  $m$ 、 $m'$  以及相機的光學中心是共平面的，我們將這個平面稱為極線平面（epipolar plane） $\pi$ 。 $C$  和  $C'$  的連線分別與第一和第二個影像平面相交於  $e$  和  $e'$ ，我們稱這兩個點為極點（epipole）。

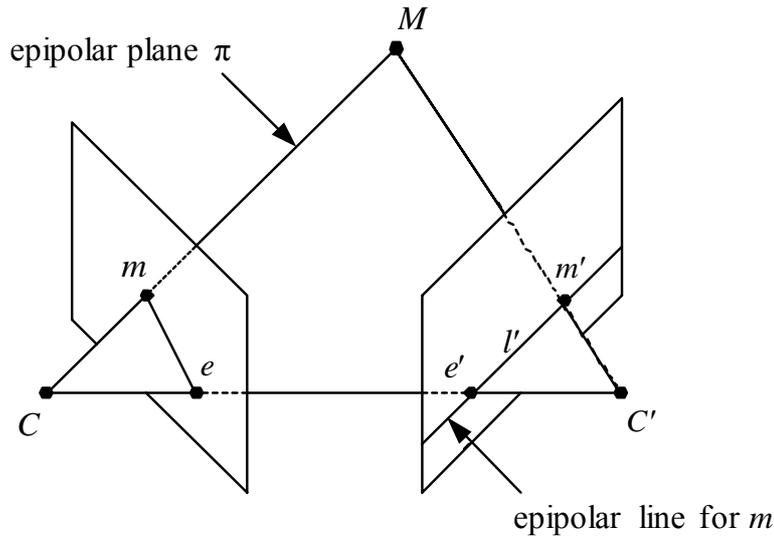


圖3.3 極線幾何。

根據極線幾何，可以發現一個特性，第一張影像上的任何一點和第一張影像的相機光學中心所形成的射線，投影在第二個影像平面上，將成像為一條線(極點除外)，這種線就稱做極線 (epipolar line)。如圖3.3中的線  $l'$ ，就是  $m$  在第二張影像上的極線。不同的三維空間點對於  $C$  和  $C'$  會產生許多不同的極線平面，而這些極線平面將會交於一條線，也就是相機光學中心  $C$  和  $C'$  的連線，稱為基線 (baseline)。從影像平面來看，也就表示所有的極線都會通過極點。

極線限制對於影像特徵點的對應有著極大的輔助功能，可以將搜尋的範圍有效地從整張影像減少到極線附近。另外極線皆會交於極點這個性質，也可以用來當作一個誤差評量的標準。

### 3.4.2 基本矩陣

基本矩陣 (fundamental matrix) 是用來描述在極線幾何中，任一點與其極線之間的對應關係。由圖3.3中我們可以看出，在第一張影像中的任一點  $m$ ，在第二張影像會有一條對應的極線  $l'$ ，在第二張影像中和  $m$  點相對應的  $m'$  點一定會位在這條極線  $l'$  上。 $m$  點和極線  $l'$  透過基本矩陣  $F$ ，以齊次座標表示

會有一個線性的關係：

$$l' = F m \quad (3.20)$$

因為  $m$  在第二張影像上的對應點  $m'$  會位在  $l'$  這條線上，所以  $m'$  的座標和  $l'$  的向量內積為零：

$$m'^T l' = m'^T F m = 0 \quad (3.21)$$

我們可以將兩張影像的角色互換，則基本矩陣會轉變為原來矩陣的轉置矩陣，也就是可以將上式轉換為：

$$m^T F^T m' = 0 \quad (3.22)$$

當第一張影像中的點是極點  $e$  的時候是一個特例，它所對應的極線退化成第二張影像的極點  $e'$ 。所以：

$$l' = F e = 0 \quad (3.23)$$

除此之外，基本矩陣可寫成一個和相機內、外部參數有關的形式：

$$F \cong A'^T [t]_{\times} R A^{-1} = [A' t]_{\times} A' R A^{-1} \quad (3.24)$$

其中， $[t]_{\times}$  為一個內含平移向量  $t = [t_1 \ t_2 \ t_3]^T$  的斜對稱矩陣，也就是：

$$[t]_{\times} = \begin{bmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{bmatrix} \quad (3.25)$$

### 3.4.3 必要矩陣

必要矩陣 (essential matrix) 為正規化影像座標的特別化基本矩陣，也可以看成是內含了相機內部參數的基本矩陣。

已知相機投影矩陣為  $P = A[R | t]$ ，且  $m = PM$ ， $m$  為影像上的點， $M$  為真實的三維空間點。若相機內部參數矩陣  $A$  已知，則我們可以在  $m$  左邊乘上  $A$  的反矩陣得到  $\tilde{m} = A^{-1}m$ ，則  $\tilde{m} = [R | t]M$ ， $\tilde{m}$  即為以正規化座標表示之影像點，可視為點  $M$  使用了具有單位校正矩陣  $I$  相機投影在影像上的二維點，投

影矩陣  $A^{-1}P = [R | t]$  稱為正規化投影矩陣。

必要矩陣的定義如下：

$$\tilde{m}'^T E \tilde{m} = 0 \quad (3.26)$$

$\tilde{m}$  和  $\tilde{m}'$  分別代表相對於點  $m$  和  $m'$  之正規化影像座標。將  $\tilde{m}$  和  $\tilde{m}'$  分別用  $m$  和  $m'$  取代，則式(3.26)可改寫為  $m'^T A'^{-T} E A^{-1} m = 0$ 。與式(3.21)做比較，則可知必要矩陣與基本矩陣間的關係為：

$$E = A'^T F A \quad (3.27)$$

將上式代入(3.24)，得到：

$$E = [t]_x R \quad (3.28)$$

到這裡我們可以發現，當我們需要重建出相機的投影矩陣時，只要能算出基本矩陣，便能還原出旋轉矩陣  $R$  和平移向量  $t$ ，詳細的重建過程我們將會在下一章中做詳細的介紹。



### 3.5 相機校正

由兩張影像重建三維場景時，若完全不知道場景的結構以及相機的資訊，則所有的資訊都將隱含在基本矩陣中。然而，若我們限制在拍攝影像的過程中，不改變相機本身內部設定，如焦距，則相機內部參數並不會改變，因此我們便可以透過事先校正 (calibration) 的方式先算出這些參數，降低重建時的複雜度。

從式(3.18)中我們可以知到，一般相機的內部參數有  $f_x$ 、 $f_y$ 、 $\gamma$ 、 $u_0$  和  $v_0$  這五個， $\gamma$ 、 $u_0$  和  $v_0$  這三個參數較為固定，而且可以有一些假設值比如  $\gamma = 0$ ， $u_0$  和  $v_0$  分別為影像中心等，但是相機的焦距  $f_x$ 、 $f_y$  卻對於三維重建有很大的影響，因此相機校正也是近年來電腦視覺的一個重要的課題。

在本系統中所使用的相機，是一個事先校正好的相機。這個校正的方法是由 Zhang [11]所提出的，利用平面的 homography 性質來計算相機的內部參數，並且在計算出初始參數之後再進行一些微調，以達到最佳化。

## 第四章 三維重建

有了前一章的理論基礎，本章將會進一步介紹利用不同視角影像完成三維重建的步驟，除了重建出三維場景之外，也會討論如何利用必要矩陣找出相機的旋轉矩陣及平移向量，進而還原出相機的三維位置資訊。

### 4.1 三維重建之步驟

在具備了投影幾何的基本觀念之後，本節將介紹由兩張影像重建場景的方法，其步驟包括：

1. 由對應點計算基本矩陣，見4.1.1 節。
2. 由基本矩陣計算投影矩陣，見4.1.2 節。
3. 對每組對應點，計算其三維空間之點座標，見4.1.3 節。

以下就針對各個步驟進行詳細的說明。



#### 4.1.1 估算基本矩陣

重建的第一個步驟，是要估算出兩張影像間的基本矩陣。由式(3.22)中我們可以知道，在兩張影像上任意的對應點  $m$  和  $m'$  必定滿足  $m^T F^T m' = 0$  這個條件。假設兩張影像中的特徵點  $m_i$  和其對應點  $m_i'$  已知，則根據式(3.22)，每一組對應點可以產生一個線性方程式，由於基本矩陣  $F$  扣除了一個縮放的比例常數之後，會剩下八個未知數，因此只需要八組對應點，即可以線性方法解出一個允許改變大小比例的基本矩陣，這個方法稱為八點演算法 (8-point algorithm) [12]。

然而考慮到數位相機在成像時可能的取樣誤差，加上影像本身會有不穩定的雜訊，甚至在特徵點的擷取與對應上發生錯誤，使得只利用八個點來計算基本矩陣，並不保證能得到正確的結果，因此有很多方法也被提出來改進基本矩陣。比

如使用超過八組對應點計算出多組基本矩陣，再透過最小平方 (least-square) 法來求出最佳的解，以增加基本矩陣的可靠度。有關基本矩陣的最佳化方法比較，可以參考[13][14][15]。

## 4.1.2 估算投影矩陣

利用前述之方法得到基本矩陣之後，第二個步驟我們要算出兩台相機各自的投影矩陣。由於本論文中，相機是已經校正完畢的，因此可以看成兩台相機的內部參數矩陣  $A$  和  $A'$  皆已知，加上在第一步中算出的基本矩陣  $F$ ，透過式 (3.27)，我們便可以算出必要矩陣  $E$ 。必要矩陣有以下的兩個性質：

1.  $3 \times 3$  矩陣  $E = [t]_x R$  為一必要矩陣若且唯若  $E$  的兩個奇異值相等，第三個為零[16]。因為必要矩陣  $E$  可分解為  $[t]_x R$ ，其中  $[t]_x$  為斜對稱矩陣，考慮

$$W = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{且} \quad Z = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (4.1)$$

則  $3 \times 3$  斜對稱矩陣  $[t]_x$  可分解為  $[t]_x = kUZU^T$ ， $U$  為正交矩陣，而  $Z = \text{diag}(1,1,0)W$ ，此即具有兩個相等奇異值之  $E$  的奇異值分解(SVD)。

2. 假設  $E$  的 SVD 為  $E = U \text{diag}(1,1,0)V^T$ ，則使用上述的表示法， $E = [t]_x R$  有以下兩種可能的分解方式：

$$[t]_x = UZU^T, \quad R = UWV^T \quad \text{或} \quad R = UW^T V^T \quad (4.2)$$

上式中的  $R$  即為相機的旋轉矩陣，而平移向量  $t$  則包含在  $[t]_x$  裡。由於  $[t]_x$  的富氏長度為2，表示若  $[t]_x$  包含比例，則  $\|t\|=1$ 。因為  $[t]_x t = 0$ ，可知  $t = U(0,0,1)^T = u_3$ ，也就是  $U$  的最後一行。

由以上兩個性質，我們可以看出可能的  $R$  和  $t$ ，由於  $R$  有兩種可能性，

而  $t$  因為沒有辦法決定其正負號，所以也會有兩種可能性，因此第二部相機相對於第一部相機的外部參數，便會有四種可能的選擇。也就是說，給一個必要矩陣  $E = U \text{diag}(1,1,0)^T$ ，第一部相機的投影矩陣為  $P = [I | 0]$ ，則第二部相機的投影矩陣  $P'$  會有四種可能的選擇，分別為：

$$P' = [UWV^T | u_3] \text{ 或 } [UWV^T | -u_3] \text{ 或 } [UW^T V^T | u_3] \text{ 或 } [UW^T V^T | -u_3] \quad (4.3)$$

圖4.1說明由  $E$  做重建後四種可能的幾何表示， $A$  代表第一部相機中心， $B$  和  $B'$  代表第二部相機可能的相機中心， $M_1 \sim M_4$  即為重建出之三維點。由於重建後的點只有可能會在兩部相機的前方，因此只要測試一個點是否落在相機前方，就可以從四種可能的相機位置中決定第二部相機的位置。

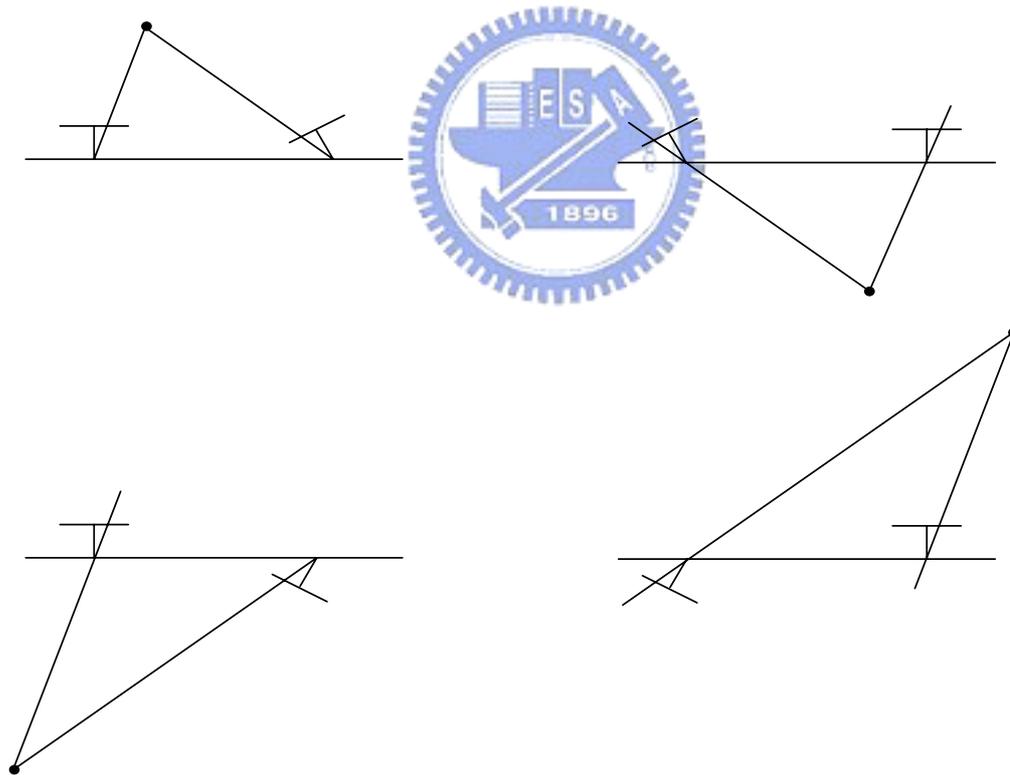


圖4.1 由  $E$  做重建後四種可能的幾何表示。

### 4.1.3 歐氏三維重建

在得到兩部相機的投影矩陣  $P$  和  $P'$  之後，接下來便可進行點的歐式重建 (Euclidean reconstruction)。首先將空間中的三維點  $M$  以及其在影像中對應的特徵點  $m_1$ 、 $m_2$  寫成以下的關係式：

$$m_i = s_i P_i M, \quad i \in \{1, 2\} \quad (4.4)$$

其中  $s_i$  為一個代表大小的比例常數 (scale)。若我們將式(4.4)裡之影像點改寫成齊次座標系，並將投影矩陣以列向量表示，可得到：

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} p_1^T \\ p_2^T \\ p_3^T \end{bmatrix}_{3 \times 4} M_{4 \times 1} \quad (4.5)$$

由式(4.5)可推導出：

$$\begin{bmatrix} up_3^T - wp_1^T \\ vp_3^T - wp_2^T \end{bmatrix}_{2 \times 4} M_{4 \times 1} = 0 \quad (4.6)$$

令  $B_i = \begin{bmatrix} up_3^T - wp_1^T \\ vp_3^T - wp_2^T \end{bmatrix}_{2 \times 4}$ ，則對於每一台相機，皆有  $B_i M = 0$  的關係存在。

每一個三維點  $M$ ，若有一組影像對應點，就可以獲得兩個如式(4.6)的關係式，將兩式合而為一，便為一滿足  $BX=0$  的  $4 \times 4$  矩陣  $B$ 。欲求得三維點座標  $M$ ，理論上可利用特徵值分解之對應特徵值 (eigenvalue) 為零的特徵向量 (eigenvector)，可是當有雜訊時，則需由對應於最小特徵值的特徵向量加以求得，求得的方法通常是利用矩陣的 SVD 分解，來得到三維點  $M$ 。

經過了這些步驟，便可重建出特徵點的三維座標，到了這個階段，我們得到的是一個與歐式重建具有縮放比例關係的公制重建 (metric reconstruction)。接著，利用場景中任何一個已知的尺寸，便可以將這個比例固定下來，得到真正的歐式重建，也就是與實物等尺寸的三維重建。

## 4.2 估測相機位置

在我們的方法中，於歐式重建的過程裡，我們可以從基本矩陣中直接抽取出兩台相機間的旋轉矩陣及平移向量，因此我們可以很直接地估算出每台相機相對於重建場景的位置。多了這項資訊，在應用上，對於機器人或自動車導航與定位，將有很大的幫助。

我們在式(4.3)中提到，旋轉矩陣及平移向量各有兩種可能，而選擇的方法為判斷重建結果是否在兩台相機的前方。但是實際上在計算的時候，由於使用了SVD分解法，因此所分解出來的旋轉矩陣和平移向量，對於正負符號的決定會有著不確定性，導致相機位置無法求出，也無法檢查場景是否在兩台相機前方。

為了解決這個問題，我們可以利用一些額外的限制，來找出正確的正負號。對於旋轉矩陣，我們可以利用旋轉矩陣一個重要的特性，也就是行列式（determinant）等於1，來決定出它的正負號。而平移向量則可從已知的相機位



移方向決定，比如事先知道相機是向右移動，則可知  $x$  方向的平移為正。確定了旋轉矩陣及平移向量的正負號之後，我們也就可以檢查四種可能的三維重建結果，哪一個是在兩台相機的前方了。

找出相機座標系間的旋轉及平移之後，也就代表著每一組三維重建的兩台相機相對位置已經找到，只要利用這些資訊，便可以重建出所有相機的位置。令  $C_1$  為第一台相機的相機中心， $R_1$  及  $t_1$  為第一台相機座標系與世界座標系間的旋轉和平移； $C_2$  為第二台相機的相機中心， $R_2$  及  $t_2$  為第二台相機座標系與世界座標系間的旋轉和平移，則我們可以求得兩台相機之間，座標系的相對旋轉及平移  $R_{12}$  及  $t_{12}$  分別為：

$$R_{12} = R_2 R_1^T \quad (4.7)$$

$$t_{12} = t_2 - R_2 R_1^T t_1 \quad (4.8)$$

若以平面的幾何上來看，如圖4.2所示，第二台相機的座標系即是第一台相機的座標系逆時針旋轉了  $R_{12}$  之後，再向右平移  $t_{12}$ 。

對於所重建出的三維點，若它們的座標系不同，我們也可以利用這些已知的相機座標系關係，將重建點轉換為以同一個座標系為基礎。如圖4.2， $M_1$  代表重建物體於  $C_1$  座標系下的座標向量，若我們要計算同一個物體相對於  $C_2$  座標系下的座標向量  $M_2$ ，相當於算出  $M_1$  經過與相機座標系相反的旋轉及平移之後，所得到相對於  $C_1$  座標系的點座標，也就是圖中的  $M_1'$ 。而由  $M_1$  到  $M_1'$  的這個過程，剛好是和相機座標系的旋轉和平移相反，如圖4.2 所示， $M_1$  先向左平移了  $-t_{12}$  之後，再順時針旋轉  $R_{12}^T$ ，得到  $M_1'$ 。因此，我們可以將這樣的轉換寫成以下的關係式：

$$M_2 = M_1' = R_{12}^T (M_1 - t_{12}) \quad (4.9)$$

其中， $M_1$  為原物體相對於  $C_1$  座標系下的座標向量， $M_2$  為原物體相對於  $C_2$  座標系下的座標向量，而  $M_1'$  則為原物體經過與相機座標系相反的旋轉及平移之後，所得到相對於  $C_1$  座標系下的座標向量。反之，若我們已知  $M_2$  而欲求  $M_1$ ，則可透過相反的旋轉及平移過程，關係式如下：

$$M_1 = R_{12} M_2 + t_{12} \quad (4.10)$$

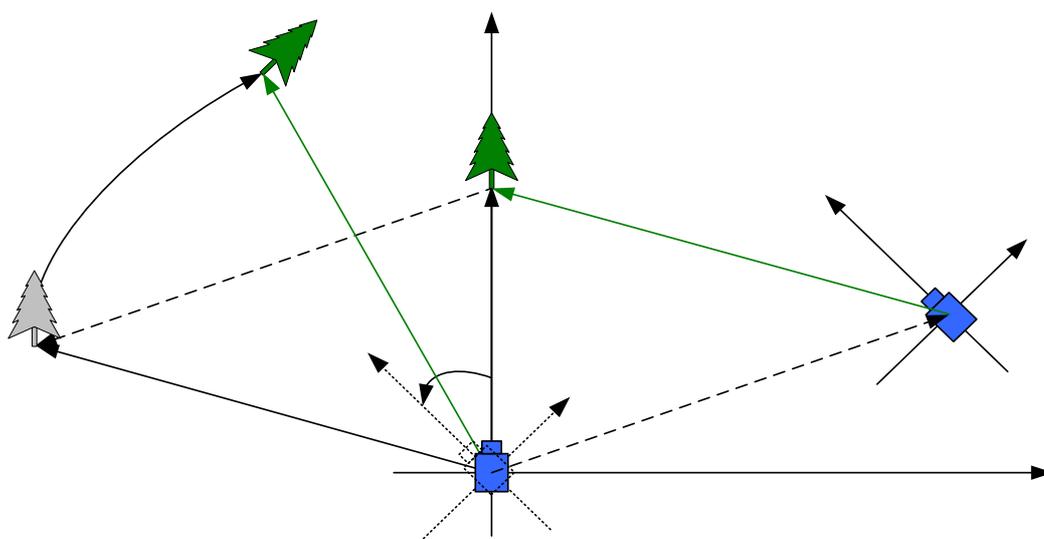


圖4.2 相機座標與點座標的旋轉及平移關係。

## 第五章 三維重建之精確度分析

由前面的理論基礎，我們已發展一套初步的視覺系統[17]，幫助機器人做場景的重建和定位，本章就是利用這一套已發展的系統，使用不同的相機，進行機器人定位與場景建置之研究。

目前市面上數位相機已很普遍，大致可粗分為兩種等級。一種為畫素高，功能齊全，強調能拍出生動影像的數位相機，價格較高。另一種則是應用在電腦視訊上，體積小，不需電池，強調即插即用，影像的品質較不穩定的 webcam。若考慮家用機器人視覺系統的應用，高價位的相機雖然有畫素高、影像品質佳等優點，相對的，畫素高將造成影像處理的時間增加。另一方面，成本的考量也是個重點，若能以較便宜且方便的 webcam，達到令人滿意的效果，可大大提升家用機器人之實用性。因此，本章節著重於平價 webcam 之探討，並以高階相機的實驗結果與之比較。

以下我們將對各種不同的相機進行三維重建實驗，再藉由實驗數據之分析，來探討相機品質對重建結果的影響。我們一開始會使用較高階的數位相機 Agfa 1680 所取得的  $640 \times 480$  的影像來作實驗，並且對於實驗的環境及過程做初步的介紹。之後我們將使用三種較低階的 webcam，取  $320 \times 240$  的影像進行實驗，包括：Orite MY120、Orite MC310 以及 Logitech QuickCam Express。

### 5.1 使用高階數位像機 Agfa 1680 進行三維重建

為了說明實驗流程，並且呈現完整的實驗數據，以便稍後與較低階的相機做一比較，我們首先使用高階數位像機 Agfa 1680 來進行三維重建實驗，實驗流程圖如圖 5.1。相機校正、特徵點的選取、特徵點排序、三維重建等步驟，均是沿用[17]之程式以進行實驗。其中，經過相機校正，事先取得的相機內部參數為：

$$\begin{bmatrix} 727.7667 & -0.3100 & 305.4135 \\ 0 & 728.1041 & 242.7015 \\ 0 & 0 & 1 \end{bmatrix}$$

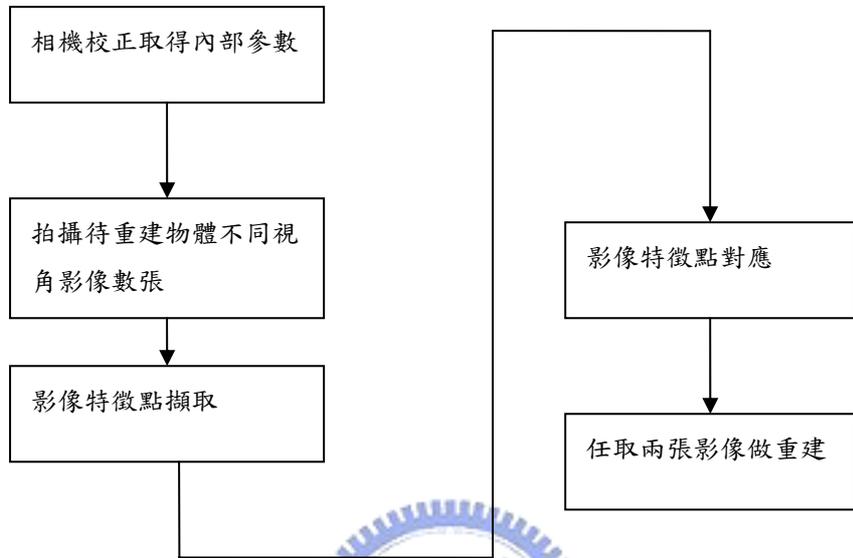
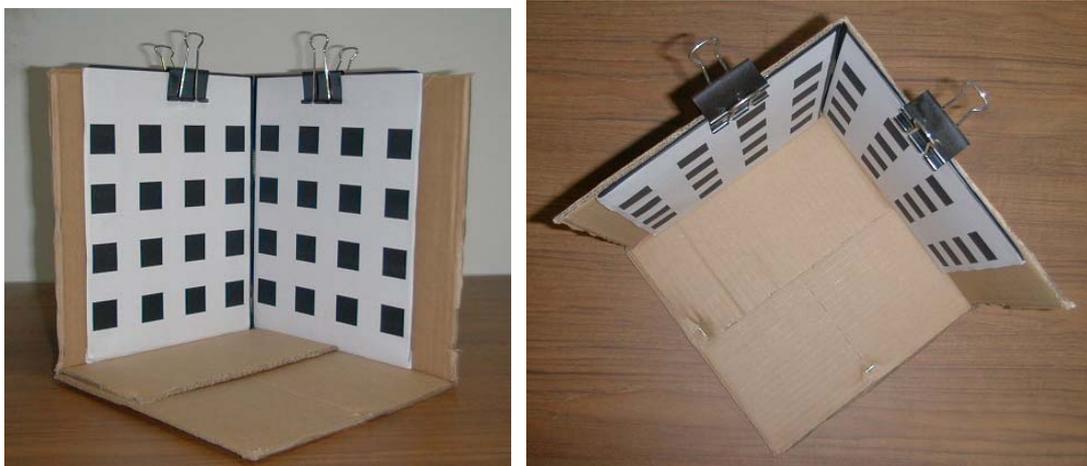


圖5.1 三維重建實驗流程圖。

圖5.2中，為本次三維重建實驗中的待重建物體，是由兩個夾角約為90度的平面所構成的正交方塊圖板，每個平面上有16個方塊。實驗之最終目的為針對這些方塊圖案平面進行三維重建。



(a)

(b)

圖5.2 正交方塊圖板之 (a)側視影像 (b)俯視影像。

為了取得不同視角的影像，相機可以繞著兩方塊圖板平面相交形成的軸來取像。然而為了減少相機移動時所造成的誤差，並且可以比較精確的控制相機相對於正交方塊圖板的位置及視角差，本實驗之器材配置如圖5.3，而實際上是固定相機，並以兩方塊圖板平面相交形成的軸為旋轉軸，旋轉正交方塊圖板，取6張不同視角影像，相鄰影像視角間隔 $10^\circ$ 。因此，相機到旋轉軸的距離理論上是不變的，在實驗中，此距離為71.4公分。

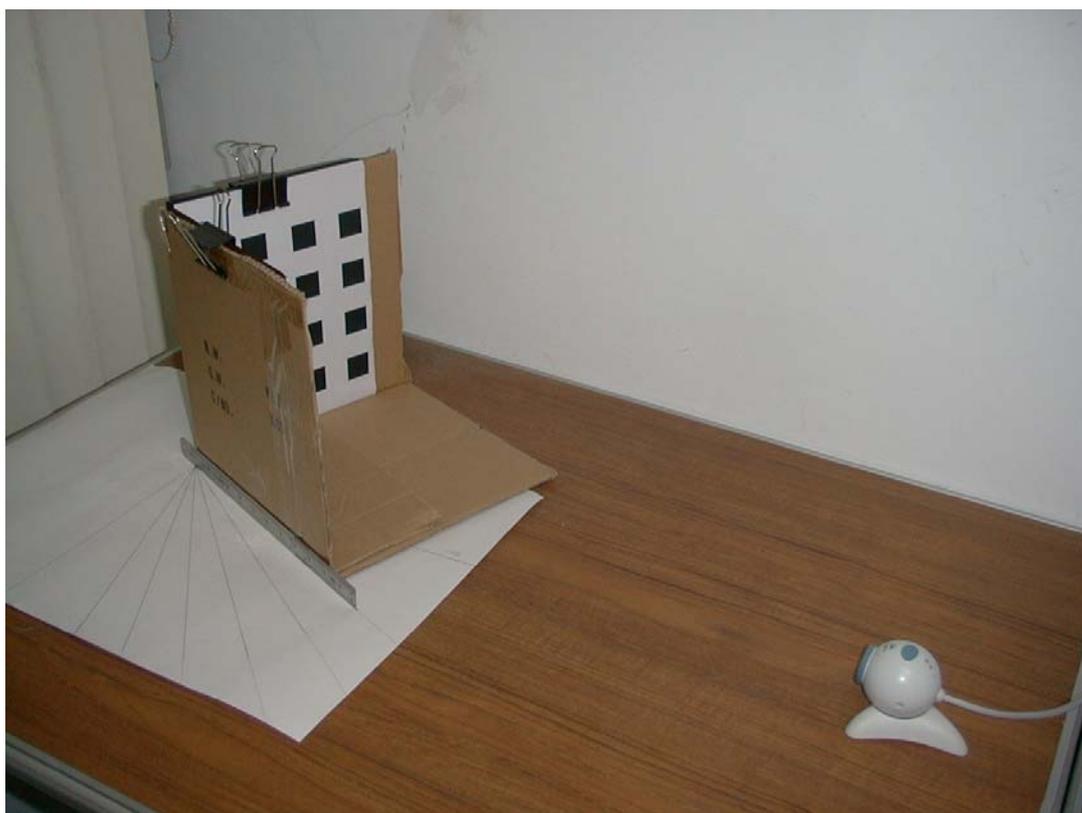


圖5.3 實驗時固定相機，令兩方塊圖板平面相交形成的軸為旋轉軸，旋轉正交方塊圖板，取6張不同視角影像，相鄰影像視角間隔 $10^\circ$ 。

由相機 Agfa 1680取得的影像，解析度為  $640 \times 480$  個像素。根據圖5.1所示之流程圖，在取得不同視角景物影像後，本實驗之主要步驟包括：

1. 六張影像經過 Harris 角點偵測，並人工刪除不必要的點後，在正交方塊圖板左右平面各可得到64個特徵點。關於Harris 角點偵測的詳細作法請參考[18]。結果如圖5.4所示。
2. 從六張影像中任取兩張為一組，以人工方式完成128個特徵點對應。

3. 任取兩張影像做三維重建，共15組重建結果。

上述步驟完成後，本節中對於不同相機所作重建精確度分析包括：

**1. 方塊圖案中，各方塊邊長之變異程度（表5.1右上部份）：**

以重建後每一個方塊的平均邊長為單位長度，算出128個邊長的標準差（standard deviation）。

**2. 方塊圖案裡，各方塊中鄰邊夾角精確度（表5.1左下部份）：**

以邊夾角的平均值與標準差來分析。

**3. 方塊圖案所構成之平面精確度（表5.2右上部份）：**

將重建結果調為實際尺寸，即一校正方格邊長調為2.75公分。左右兩個平面各64個點，利用主軸分析（principal components analysis）找出短軸，用它來代表平面的法向量，再利用兩個平面的法向量來計算二者之夾角。另外，以主軸分析所得到最小的eigenvalue，來代表平面上的點距離該平面的平均誤差。

**4. 相機定位精確度（表5.2右下部份）：**

由還原的相機位置，算出左右影像相機到正交方塊圖板中心軸的距離與視角誤差。

我們將表5.1及表5.2所得資料，依不同視角差，繪製成長條圖，結果如圖5.5、圖5.6、圖5.7。其中不同顏色的長條，代表以某張影像為左邊的影像，取視角差分別為 $10^\circ$ 、 $20^\circ$ 、 $30^\circ$ 、 $40^\circ$ 、 $50^\circ$ 的重建結果。例如圖中藍色的長條，就是以(a)影像為左影像，取某一視角間隔做重建的結果。

由這些分析資料中可看出，重建方格之形狀、兩面夾角，均可獲得頗為正確之資訊。在大部分的情況之下，當重建影像組的間隔愈大，也就是視角差愈大時，所重建出平面上的特徵點距離平面的誤差較小，而由較為相鄰的影像組，所重建出的誤差會較大。另外，校正板夾角都接近90度。

在校正板與攝影機之相對位置方面，仍然存在若干誤差。此誤差除了與前述章節所提及之深度不確定性有關外，實驗時所測距離，為相機鏡頭到中心軸之長

度，與真正相機中心的位置並不相同，這也是造成誤差的原因。整體來說，誤差在6公分以內。至於左右影像相機的視角差則相當精確，誤差皆在 $1^\circ$ 以內。

以後的實驗裡，我們將針對較低階的相機做測試，除了影像解析度降為 $320 \times 240$ ，其他方面的問題也會影響到重建結果，將在下面各節依不同相機詳述。

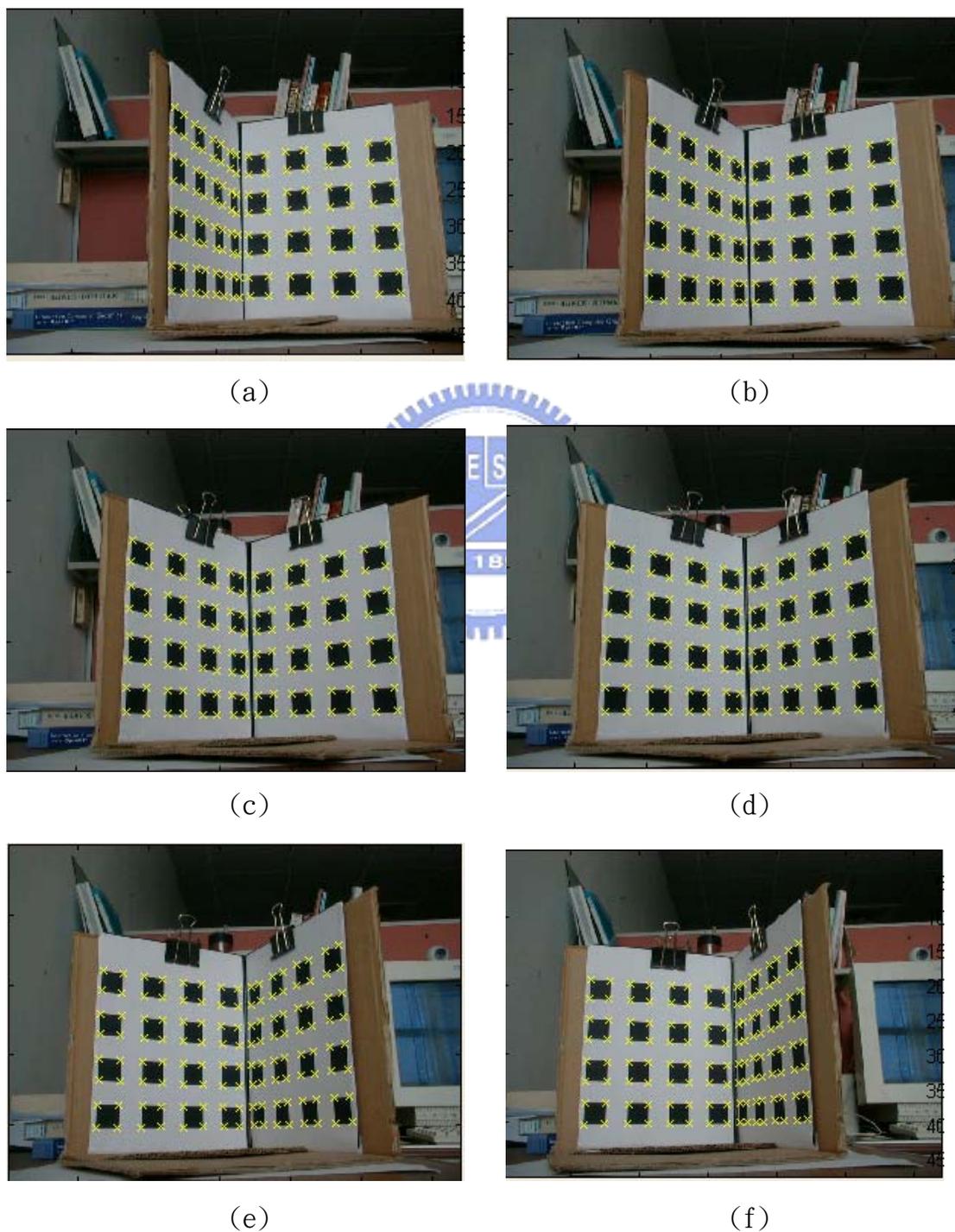


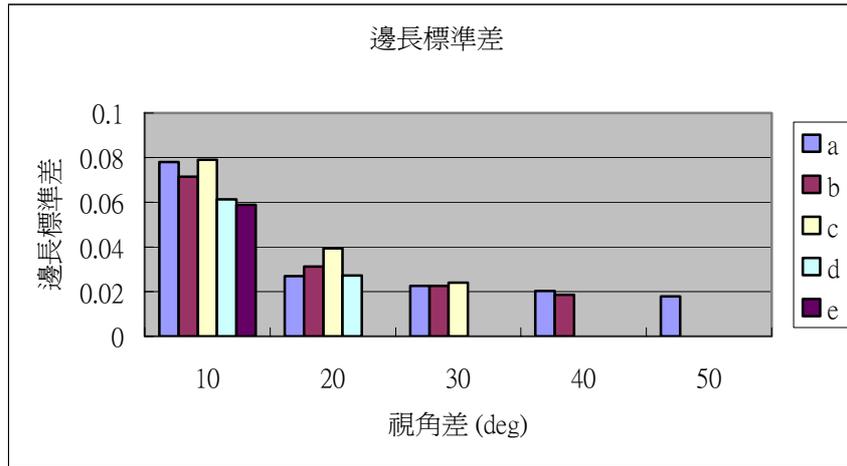
圖5.4 (a)-(f)不同角度對同一校正板所拍攝的六張影像及其特徵點。

表5.1 邊長與邊夾角重建結果分析

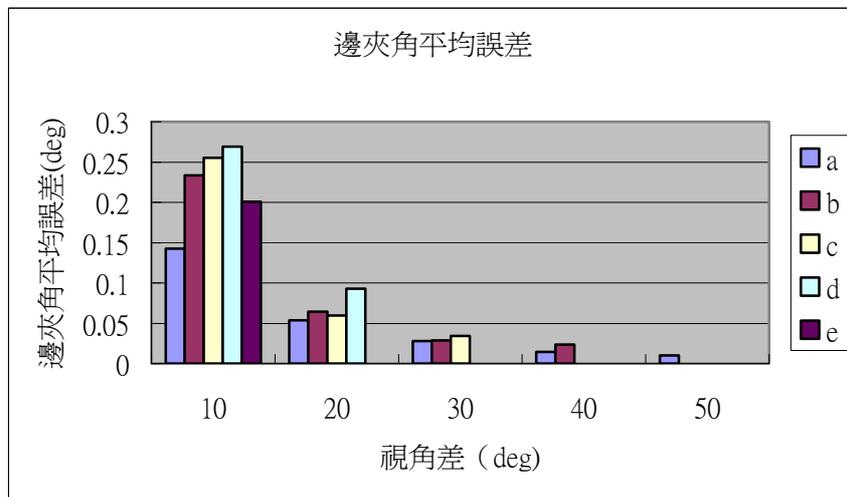
邊長: 標準差. 邊夾角(degree): a. 平均誤差 b. 標準差	1	2	3	4	5	6
1		0.0781	0.027	0.0225	0.0203	0.0178
2	a. -0.1428 b. 8.3932		0.0715	0.0313	0.0226	0.0185
3	a. -0.0537 b. 2.6499	a. -0.2334 b. 5.4565		0.0791	0.0394	0.024
4	a. -0.0279 b. 2.1463	a. -0.0644 b. 2.6724	a. -0.2554 b. 7.0105		0.0614	0.0273
5	a. -0.0146 b. 1.7376	a. -0.0287 b. 1.8981	a. -0.0596 b. 2.8088	a. -0.269 b. 5.1616		0.0589
6	a. -0.0103 b. 1.7726	a. -0.0238 b. 1.7194	a. -0.0342 b. 2.196	a. -0.0927 b. 2.4759	a. -0.2007 b. 4.5442	

表5.2 平面重建與相機定位分析

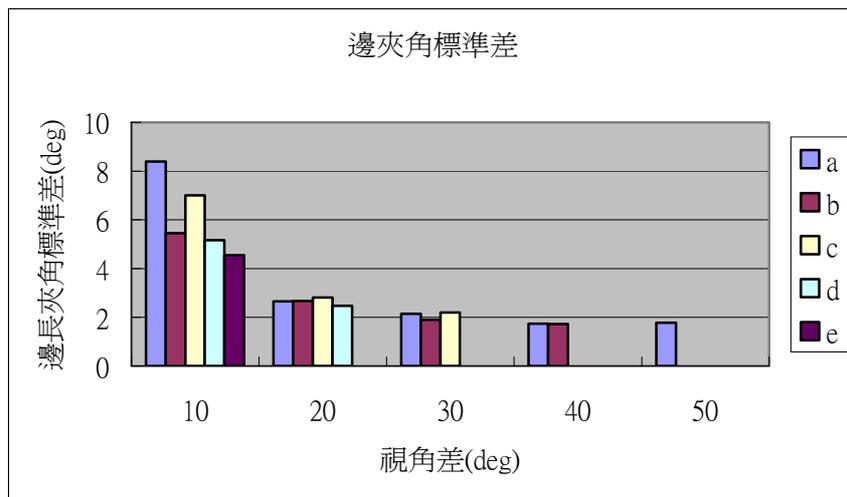
平面重建: 左平面誤差變異量 ( $cm^2$ ) 右平面誤差變異量 ( $cm^2$ ) 平面夾角誤差 (degree) 相機定位: 左距離誤差 (cm) 右距離誤差 (cm) 視角誤差 (degree)	1	2	3	4	5	6
	1		0.0187 0.0458 -3.8988	0.0095 0.0124 0.673	0.0083 0.0077 0.1858	0.0068 0.0067 -0.0289
2	0.0752 1.6799 -0.5054		0.0281 0.0339 -0.8696	0.0132 0.0108 -1.0583	0.0077 0.0073 0.7948	0.0051 0.0062 -1.1255
3	4.8943 4.5716 -0.0442	2.7642 2.7764 -0.2993		0.0453 0.031 -2.0109	0.0119 0.0136 1.035	0.0066 0.0078 1.2834
4	5.1239 4.8521 0.2825	4.7635 4.429 -0.3005	1.5521 2.3787 -0.1049		0.0355 0.0241 2.2979	0.0112 0.0082 1.0988
5	5.5413 4.8065 0.833	5.1524 4.1819 -0.0036	4.6258 4.425 0.0917	4.03 4.1489 -0.0777		0.042 0.016 -3.1683
6	5.3002 3.2326 0.7438	4.9649 2.7747 -0.2783	4.7345 3.0085 -0.2437	4.588 2.8204 -0.5258	5.539 4.3594 -0.8022	



(a)

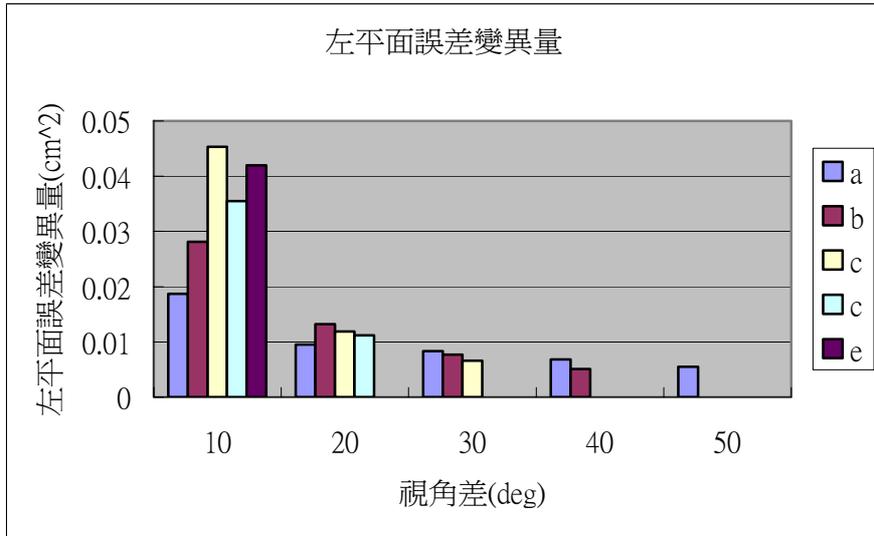


(b)

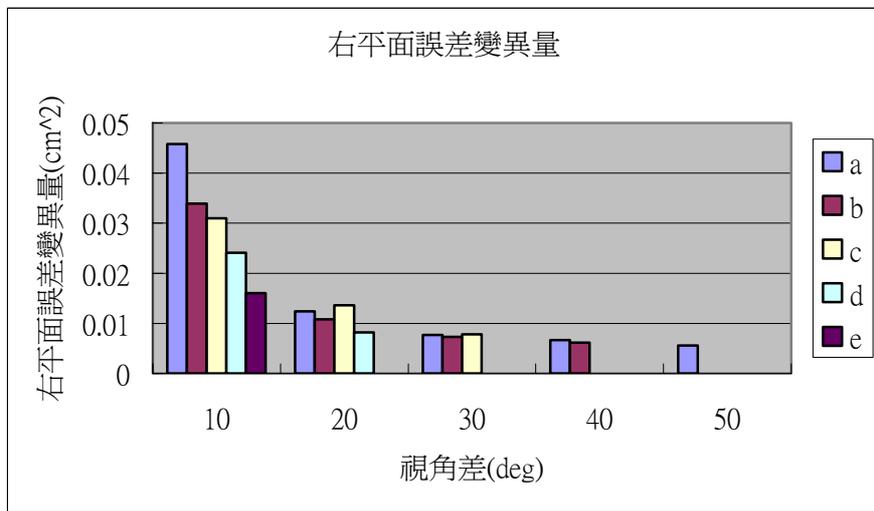


(c)

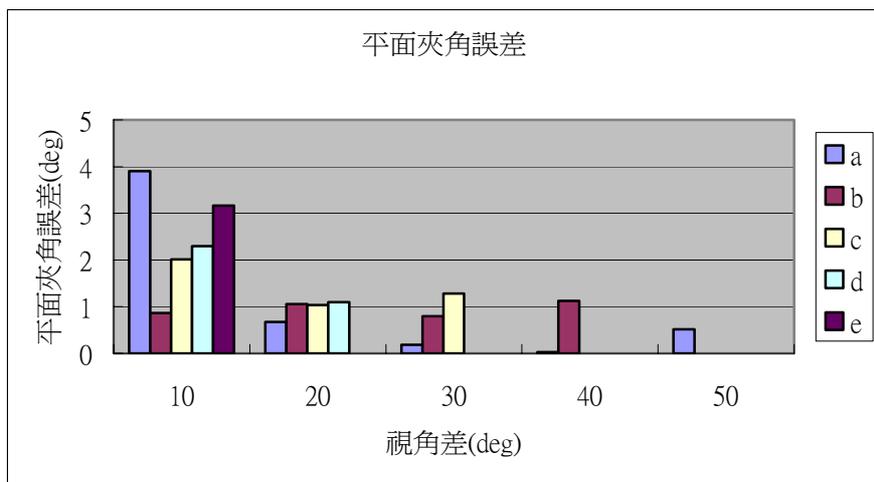
圖5.5 (a)邊長標準差 (b)邊夾角平均誤差 (c)邊夾角標準差 之長條圖。



(a)

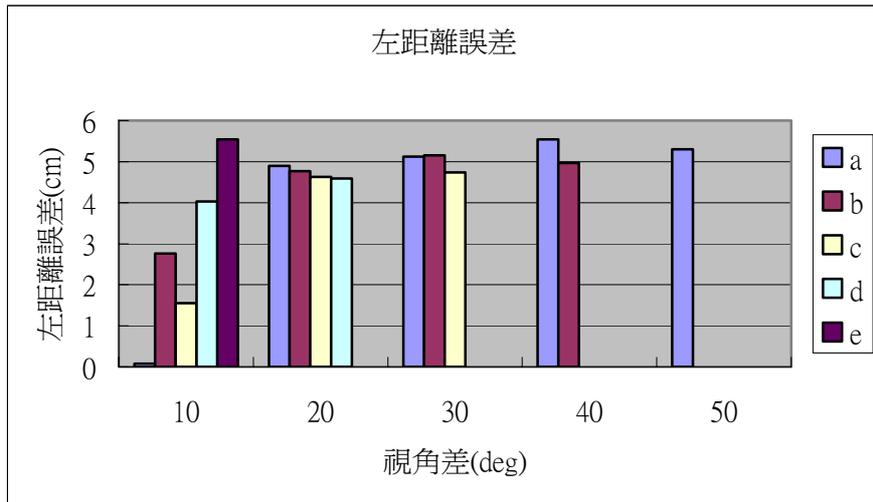


(b)

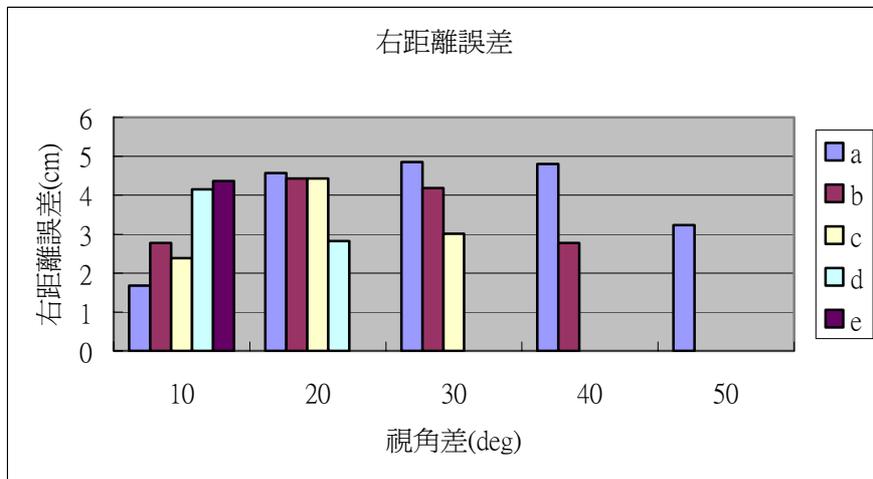


(c)

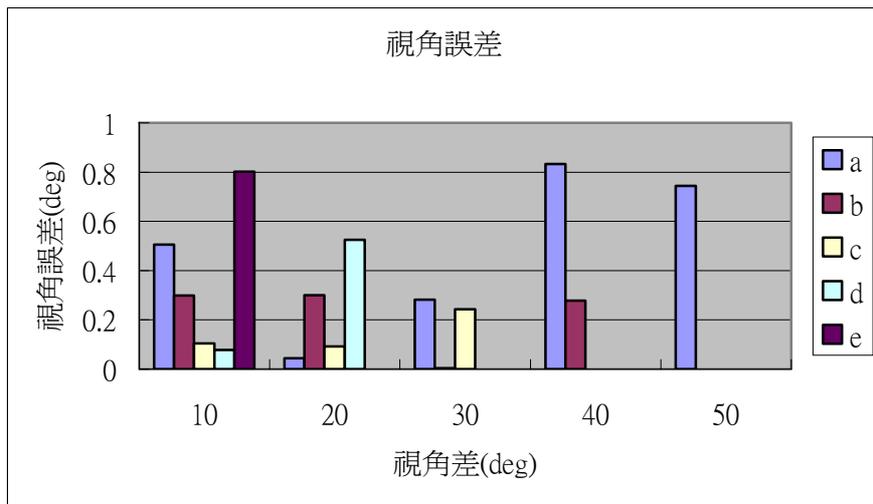
圖5.6 (a)左平面誤差變異量 (b)右平面誤差變異量 (c)平面夾角誤差 之長條圖。



(a)



(b)



(c)

圖5.7 (a)左距離誤差 (b)右距離誤差 (c)視角誤差 之長條圖。

## 5.2 使用Orite MY120進行三維重建

Orite MY120 為價位 NT. 500~1000 的 webcam。是平價數位相機中，不良影響較輕微的，因此我們首先以此相機做實驗，與 Agfa 1680 做一比較。事先取得相機的內部參數為：

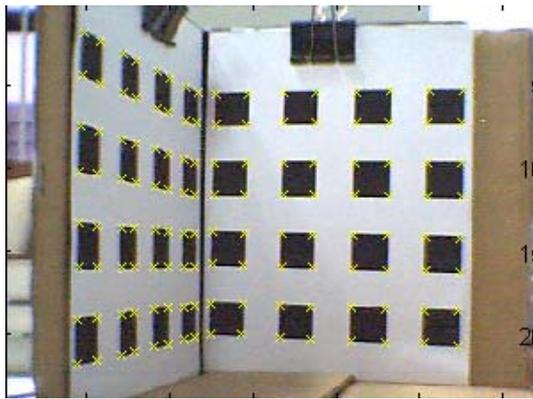
$$\begin{bmatrix} 519.3762 & -0.6419 & 158.7845 \\ 0 & 519.6614 & 124.4745 \\ 0 & 0 & 1 \end{bmatrix}$$

由Orite MY120取得的影像，解析度為320×240。經過 Harris 角點偵測，並人工刪除不必要的點後，結果如圖5.8。接著進行後續的重建步驟，同5.1節所述，算出邊長與邊夾角資訊，列於表5.3，平面重建結果則列於表5.4。

由表 5.3、表 5.4 可看出，取第一張和第二張影像所做重建結果，其平面夾角與實際上之角度有很大誤差。這可能是因為第一張影像之視角過斜，而且第一張影像與第二張影像之間的視角差太小所造成的結果。其他影像組之重建，雖然沒有 AGFA 1680 之重建結果那樣精確，卻也在可接受範圍內。整體來說，視角差若在二十度以上，都可以得到另人滿意的重建結果。在相機定位方面，視角差除一、二影像組誤差為3°外，其他誤差都在2°以內。而相機到物體旋轉中心軸的距離，受相機中心不確定性所影響，誤差在4公分以內。

除了視角對重建結果有影響以外，特徵點的多寡、待重建物體在影像中所佔的比例也都會影響到重建結果。一般來說，特徵點越多，且在影像中分布越廣，即待重建物體在影像中所佔比例越大，重建結果越精確。

若不考慮特徵點擷取、特徵點對應的時間，此重建實驗，在 CPU 為 Pentium3 800MHZ，隨機存取記憶體為 256MB 的平台上，以 Window XP 為作業系統，執行 MATLAB 6.5，跑完十五組重建結果需時 6 秒。平均一組重建結果約為 0.4 秒，此效能應已足夠做某些應用，未來程式若改為 C 語言撰寫，執行效率可望提升。



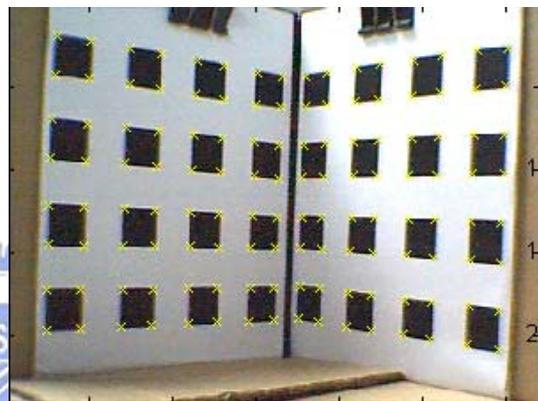
(a)



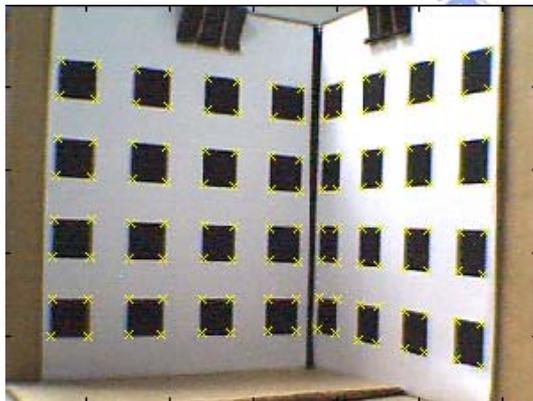
(b)



(c)



(d)



(e)



(f)

圖5.8 (a)-(f)不同角度對同一校正板所拍攝的六張影像及其特徵點。

表5.3 邊長與邊夾角重建結果分析

邊長: 標準差 邊夾角(degree): a. 平均誤差 b. 標準差	1	2	3	4	5	6
1		0.3057	0.0476	0.0413	0.0306	0.0284
2	a. -0.1943 b. 10.0656		0.0962	0.0556	0.0368	0.0336
3	a. -0.0933 b. 4.492	a. -0.4748 b. 8.2089		0.0994	0.043	0.0344
4	a. -0.0594 b. 3.1281	a. -0.1538 b. 4.1744	a. -0.4476 b. 9.0951		0.0763	0.042
5	a. -0.0215 b. 2.4006	a. -0.041 b. 2.4857	a. -0.119 b. 3.6268	a. -0.2983 b. 6.5376		0.0641
6	a. -0.0154 b. 2.0954	a. -0.0321 b. 2.2837	a. -0.0566 b. 2.8541	a. -0.087 b. 3.5031	a. -0.1751 b. 6.2563	

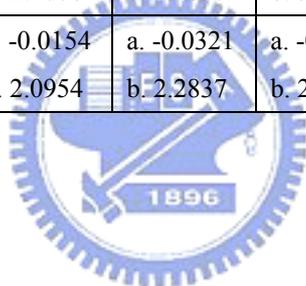


表5.4 平面重建與相機定位分析

平面重建: 左平面誤差變異量 ( $cm^2$ ) 右平面誤差變異量 ( $cm^2$ ) 平面夾角誤差 (degree) 相機定位: 左距離誤差 (cm) 右距離誤差 (cm) 視角誤差 (degree)	1	2	3	4	5	6
	1		0.2587 0.9997 -60.201	0.017 0.0223 1.8638	0.0134 0.0158 2.0096	0.0116 0.0102 2.0794
2	-1.9229 -2.26 3.1959		0.0741 0.0724 0.2574	0.0335 0.0248 0.6501	0.0206 0.0106 1.0063	0.0161 0.0116 -0.303
3	2.9518 4.9282 0.8861	1.2501 2.5807 0.0454		0.0666 0.0769 1.9779	0.0224 0.0171 -1.2224	0.0127 0.0149 -0.0346
4	3.1548 4.7055 1.4725	3.4322 3.9591 0.3541	-0.0888 -1.4653 0.7567		0.079 0.038 1.1035	0.0304 0.0169 1.419
5	4.0309 5.6175 1.6699	4.5614 5.2459 0.8172	3.8632 3.6408 0.6233	3.6671 3.9518 0.0533		0.0774 0.0334 1.3187
6	4.4257 5.1527 1.7188	4.9604 4.5655 1.0262	4.3736 2.9837 0.5413	4.4392 3.1691 0.0895	4.0004 1.5598 -0.8825	

### 5.3 使用Orite MC310進行三維重建

此相機於近距離取像時，物體外圍會有扭曲 (distortion) 的情形，如圖 5.9。觀察影像外圍，可發現直線邊緣在影像中呈弧形。原本[17]所用的相機校正方法並無校正 distortion，取得相機的內部參數為：

$$\begin{bmatrix} 400.1749 & -1.4156 & 142.4034 \\ 0 & 397.7272 & 107.4276 \\ 0 & 0 & 1 \end{bmatrix}$$

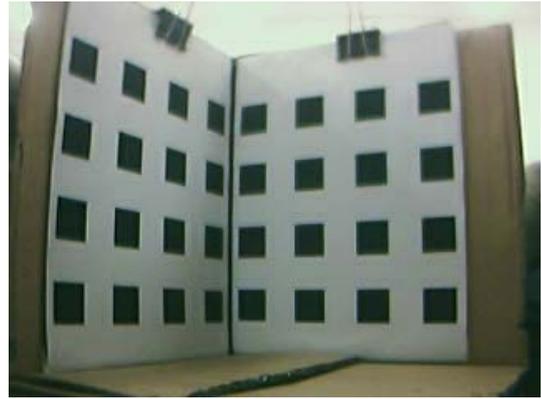
參考[19]中所提之校正工具，將扭曲影像還原，結果如圖 5.10。同時將校正板影像還原，算出還原影像相機內部參數為：

$$\begin{bmatrix} 404.3329 & -0.0057 & 187.5377 \\ 0 & 405.7930 & 108.9445 \\ 0 & 0 & 1 \end{bmatrix}$$

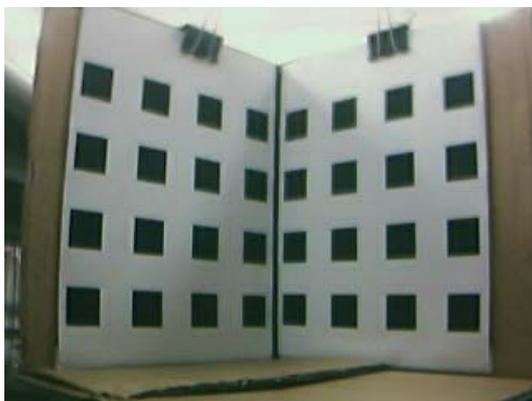
將圖 5.9、圖 5.10 兩組影像分別進行後續的重建步驟，同前一節所述，算出邊長與邊夾角資訊，列於表 5.5，平面重建結果則列於表 5.6。由表中可看出，經過扭曲校正的影像，其重建結果，在方格邊長和方格邊夾角的精確度皆有提升。經過扭曲校正後，最顯著的改善是平面的夾角，重建的平面夾角都趨近 90°。



(a)



(b)



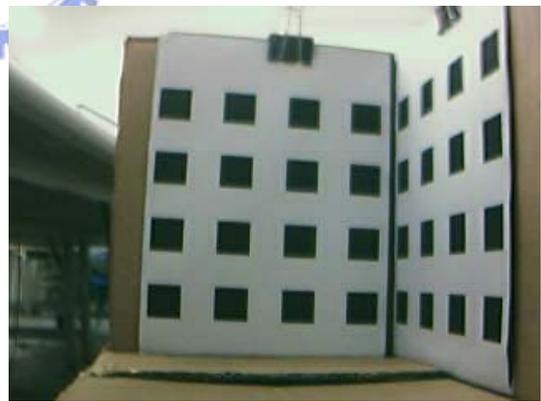
(c)



(d)

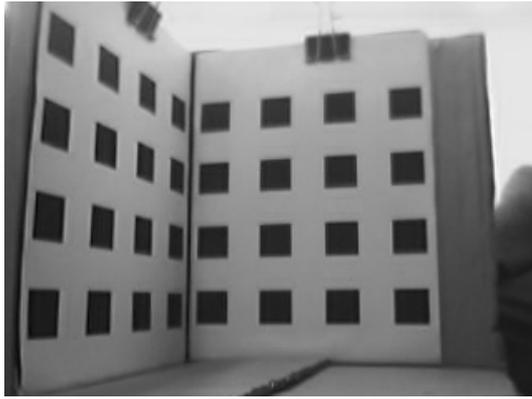


(e)

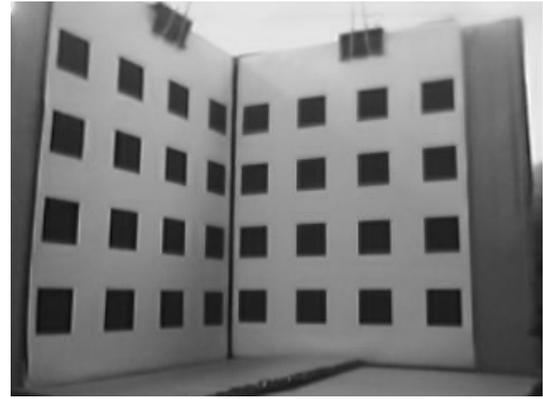


(f)

圖5.9 (a)~(f)由相機Orite MC310所取得的立體校正板影像。



(a)



(b)



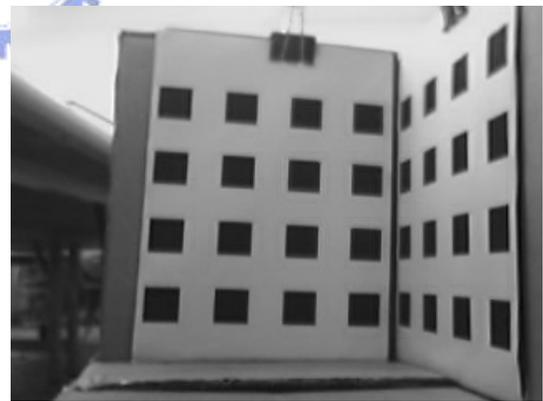
(c)



(d)



(e)



(f)

圖5.10 (a)~(f)之影像，分別為圖5.10中(a)~(f)影像還原扭曲後的結果。

表5.5 邊長與邊夾角重建結果分析

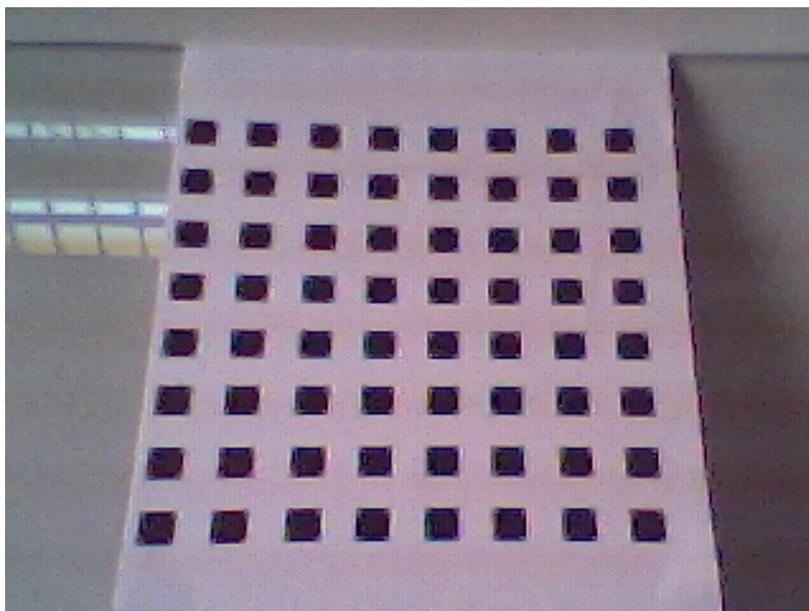
邊長標準差: a. 校正前 b. 校正後  邊夾角 平均誤差 (degree): 校正前 校正後  標準差: 校正前 校正後	1	2	3	4	5	6	
	1		a. 0.1239 b. 0.1063	a. 0.0755 b. 0.05	a. 0.0554 b. 0.0452	a. 0.0563 b. 0.0387	a. 0.0524 b. 0.0419
	2	-0.6651 -0.6162 8.3927 7.5033		a. 0.1002 b. 0.0609	a. 0.0625 b. 0.0544	a. 0.0665 b. 0.0433	a. 0.0596 b. 0.0423
	3	-0.0659 -0.0977 3.8651 3.8356	-0.2162 -0.2108 5.7072 4.8032		a. 0.0983 b. 0.0867	a. 0.0944 b. 0.0545	a. 0.0802 b. 0.053
4	-0.0281 -0.0514 3.6336 2.8756	-0.0763 -0.047 4.2176 3.2948	-0.1978 -0.2834 5.8072 5.6937		a. 0.13 b. 0.1083	a. 0.079 b. 0.0789	
5	-0.0396 -0.038 3.423 2.3709	-0.0572 -0.0735 3.6399 2.9787	-0.1293 -0.1295 4.096 3.3737	-0.7479 -0.5201 7.5004 5.9005		a. 0.1212 b. 0.1888	
6	-0.0349 -0.0219 3.5286 2.7541	-0.0305 -0.0275 3.6023 2.8364	-0.085 -0.0632 3.9277 3.2686	-0.4468 -0.279 5.3488 4.1066	-1.607 -1.4304 8.0664 10.1593		

表5.6 平面重建分析

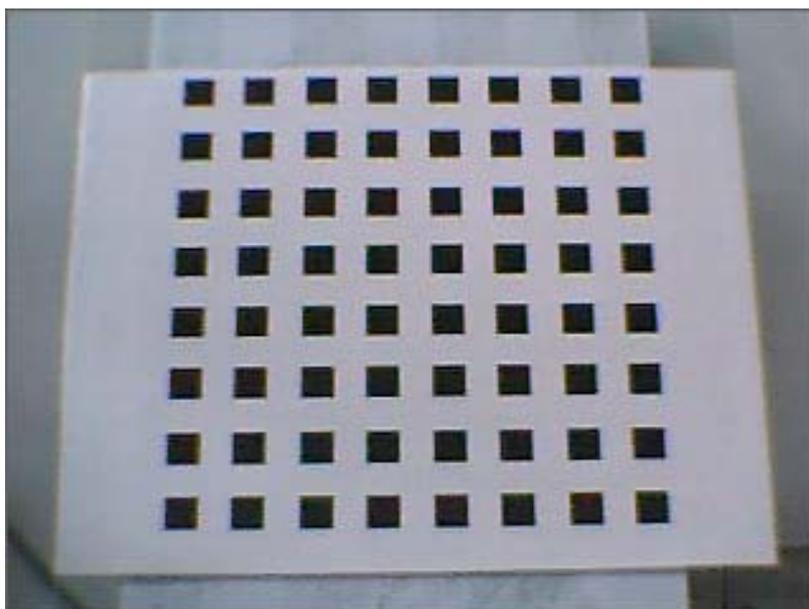
平面夾角誤差(degree): a. 校正前 b. 校正後  平面誤差 (cm <sup>2</sup> ): 左平面誤差: 校正前 校正後  右平面誤差: 校正前 校正後	1	2	3	4	5	6	
	1		a. 5.3445 b. -1.9161	a. 7.9167 b. 0.483	a. 7.7861 b. -0.8241	a. 6.59 b. -1.0247	a. -4.9659 b. -1.0474
	2	0.0648 0.0623 0.1456 0.1264		a. 9.0783 b. -0.6441	a. 8.2031 b. -0.2138	a. 6.4451 b. 0.2372	a. -4.3666 b. -0.263
	3	0.0378 0.034 0.025 0.0275	0.0653 0.0496 0.0603 0.0676		a. -5.9822 b. 0.0274	a. 3.9207 b. 0.5146	a. 1.3742 b. 0.366
4	0.033 0.0227 0.0157 0.018	0.0377 0.0257 0.0211 0.0247	0.079 0.0469 0.0436 0.0643		a. 1.5812 b. 1.6049	a. -3.5979 b. 1.1061	
5	0.0313 0.0193 0.0168 0.0134	0.0345 0.0215 0.0194 0.0179	0.0526 0.0218 0.0229 0.025	0.1892 0.0883 0.0535 0.0575		a. 5.0241 b. 3.4963	
6	0.0306 0.0154 0.0174 0.0231	0.0304 0.0163 0.0183 0.0263	0.0417 0.0138 0.0196 0.0332	0.0975 0.0521 0.0217 0.0449	0.2709 0.2266 0.0594 0.1472		

## 5.4 使用Logitech QuickCam Express實驗

Logitech QuickCam Express 市價約為 NT. 1000左右，屬於平價 webcam，但其所攝得之影像，若用於三維重建，有雜訊過多的問題。以相機校正時所取得的影像為例，所取得校正板影像如圖5.11(a)。我們用 Orite MY120所取得的影像與之比較，如圖5.11(b)。影像解析度皆為  $320 \times 240$ 。利用 Harris coner detection 的方法自動取角點，並人工刪除不必要的點後，結果如圖5.12。



(a)

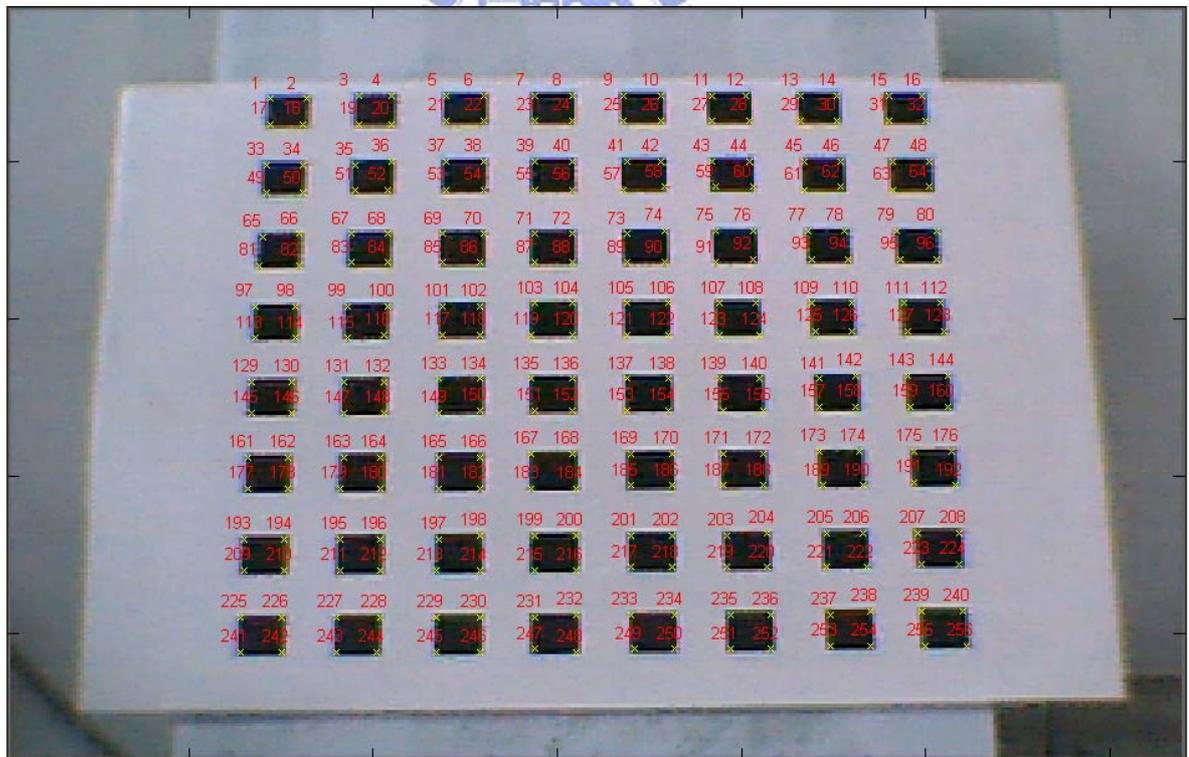


(b)

圖5.11 (a)Logitech QuickCam Express (b)Orite MY120 所取之校正板影像。



(a)



(b)

圖5.12 此圖中之(a)、(b)影像，分別為圖5.11之(a)、(b)影像，利用Harris corner detection 的方法自動取角點，並人工刪除不必要點後的結果。

比較圖中自動偵測到的點位置與實際角點位置，可發現圖 5.12(a)因雜訊較嚴重，某些自動偵測的角點與實際位置有蠻大的偏差。此一偏差將會影響到相機校正的結果，並於實際的應用時，因影像本身的品質而造成特徵點取得與對應的困難，這些不良的影響也都會反應在重建的結果上。

我們以此相機拍攝不同視角角六張影像，並進行後續的重建步驟，同5.1節所述，算出邊長與邊夾角資訊，列於表5.7，平面重建結果則列於表5.8。由這些分析的資訊可看出，使用此相機的重建結果，與使用 Orite MY120 進行重建的結果相較，的確是較差的。

相機 Logitech QuickCam Express 與 Orite MY120，價位皆為 NT. 500~1000，目前市面上低雜訊的相機已很普遍，雜訊問題可於事先的相機挑選來避免。

表5.7 邊長與邊夾角重建結果分析

邊長： 標準差						
	1	2	3	4	5	6
邊夾角(deg): a. 平均誤差 b. 標準差						
1		0.2654	0.0898	0.0582	0.051	0.0376
2	a. -0.3309 b. 22.0669		0.2894	0.0665	0.0481	0.0407
3	a. -0.1565 b. 8.188	a. -0.6071 b. 29.1214		0.1252	0.0762	0.0557
4	a. -0.0446 b. 4.3157	a. -0.1898 b. 5.4932	a. -1.107 b. 8.9104		0.101	0.067
5	a. -0.048 b. 3.2597	a. -0.0643 b. 3.5025	a. -0.1796 b. 5.0015	a. -0.701 b. 7.5539		0.1443
6	a. -0.0442 b. 3.02	a. -0.0714 b. 3.3724	a. -0.1077 b. 4.2096	a. -0.1923 b. 4.6793	a. -0.5755 b. 11.9438	

表5.8 平面重建與相機定位分析

平面重建: 左平面誤差變異量 ( $cm^2$ ) 右平面誤差變異量 ( $cm^2$ )		平面夾角誤差(degree)					
		1	2	3	4	5	6
相機定位: 左距離誤差 (cm) 右距離誤差 (cm)		視角誤差 (degree)					
		1	2	3	4	5	6
1			0.0786	0.0595	0.0414	0.0407	0.0298
			0.1987	0.06	0.0298	0.0227	0.0239
			12.7112	-1.216	-0.3651	-1.1445	0.8391
2	26.6749		0.0651	0.0331	0.0281	0.0163	
	23.1149		0.158	0.0385	0.0255	0.0265	
	-0.4075		-12.4393	1.0152	1.3783	-2.0642	
3	0.4691	38.6468		0.1731	0.0676	0.033	
	-11.6375	36.685		0.109	0.0422	0.032	
	1.9216	0.1227		1.4972	0.5372	-3.9824	
4	-3.0416	-1.6348	0.0336		0.2136	0.0555	
	-18.0715	-9.6905	-4.6076		0.0835	0.0411	
	4.3772	1.7112	1.2402		-7.8456	-5.333	
5	-3.1521	-1.6014	-0.56	1.1913		0.1787	
	-20.0878	-11.6999	-7.0756	-3.3121		0.068	
	5.8493	2.4418	1.9491	-0.262		7.237	
6	-2.2803	-0.4189	0.5748	0.7269	2.6093		
	-19.0876	-9.7886	-4.7905	-1.1306	2.942		
	8.2659	4.1096	2.9875	0.8771	-0.2044		

## 第六章 結論與未來工作

本論文以一套現有的電腦視覺軟體系統，探討使用較平價的相機為機器人視覺器材，以之進行場景重建與機器人定位的可行性。文中針對了一些較低階 webcam，做三維重建與相機定位之精確度分析。除了影像解析度降為  $320 \times 240$ ，在本論文裡也探討了可能影響重建與定位結果的因素，包括影像的雜訊與影像扭曲。另外，實驗中一些與取像系統本身較無關的因素，如視角差的大小、特徵點的多寡、待重建物體在影像中所佔比例等等，也都會影響到重建結果的好壞。從實驗結果可看出，在我們所使用的平價 webcam 中，若無嚴重的雜訊干擾或是影像扭曲的情形，在大部分的情況下，若擷取視角差  $20^\circ$  以上之  $320 \times 240$  影像，再以這些影像做三維重建以及相機之定位，其結果往往是令人滿意的。

在本論文的分析實驗裡，為了避免被特徵點擷取產生的誤差所影響，因此在特徵點的選擇與對應，是以人工介入的方式來達成，未來可進一步與前端的特徵點自動擷取系統結合，分析重建與定位結果，並從影像品質、特徵點的選取與對應、視角與視角差之決定等方面設法改善，以滿足實際應用的要求。

## 參考文獻

- [1] P. Narayanan, P. Rander, and T. Kanade, "Constructing Virtual Worlds Using Dense Stereo," *Proc. International Conference on Computer Vision*, 1998.
- [2] M. Okutomi and T. Kanade, "A Multiple-baseline Stereo," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 15, no. 4, pp. 353-363, April 1993.
- [3] M. Pollefeys and L. Van Gool, "From Images to 3D Models," *Communications of the ACM*, vol. 45, no. 7, pp.50-55, July 2002.
- [4] Y.F. Li and R.S. Lu, "Uncalibrated Euclidean 3-D Reconstruction Using an Active Vision System," *IEEE Transactions on Robotics and Automation*, vol. 20, no. 1, pp. 15-25, Feb. 2004.
- [5] Fanhuai Shi, Xiaoyun Zhang, Yuncai Liu. "A New Method of Camera Pose Estimation Using 2D-3D Corner Correspondence," *Pattern Recognition Letters*, vol. 25, no. 10, pp.1155-1163, July 16, 2004.
- [6] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2001.
- [7] O. Faugeras and Q.-T. Luong, *The Geometry of Multiple Images*. The MIT Press, 2001.
- [8] M. Pollefeys, "Tutorial on 3D Modeling from Images," in conjunction with *European Conference on Computer Vision*, June 26, 2000.
- [9] O. Faugeras, "Stratification of Three-Dimensional Vision: Projective, Affine, and Metric Representations," *Journal of the Optical Society of America*, vol.12, no.3, pp. 465-484, 1995.
- [10] M. Pollefeys, "Self-Calibration and Metric 3D Reconstruction from Uncalibrated Image Sequences," Ph.D. Thesis, Katholieke Universiteit Leuven, 1999.
- [11] Z. Zhang, "A Flexible New Technique for Camera Calibration," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330-1334, November 2000.

- [12] H.C. Longuet-Higgins, "A Computer Algorithm for Reconstructing A Scene from Two Projections," *Nature*, vol. 293, pp. 133-135, 1981.
- [13] K.-L. Lee, "Optimization of Fundamental Matrix and Euclidean Reconstruction," Master Thesis, National Chiao Tung University, 2002.
- [14] R. Hartley, "In Defense of the Eight-point Algorithm," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 6, pp. 580-593, June 1997.
- [15] Z. Zhang, "Determining the Epipolar Geometry and Its Uncertainty: A Review," *Research Report, No. 2927, INRIA Sophia-Antipolis*, July 1996.
- [16] T. Huang and O. Faugeras, "Some Properties of the E Matrix in Two-View Motion Estimation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11, no. 12, pp. 1310-1312, Dec. 1989.
- [17] L.-C. Guo, "Robot Vision for 3D Scene Reconstruction," Master Thesis, National Chiao Tung University, 2003.
- [18] Chris Harris and Mike Stephens, "A Combined Corner and Edge Detector", *Proc. The Fourth Alvey Vision Conference*, pp 147-151. 1988.
- [19] Myron Z. Brown, Darius Burschka, and Gregory D. Hager, "Advances in Computational Stereo", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 25, no. 8, pp. 993-1008, August 2003.  
([http://www.vision.caltech.edu/bouguetj/calib\\_doc/](http://www.vision.caltech.edu/bouguetj/calib_doc/))