

以擴增實境為基礎的互動式 3D 建模系統

陳柏鋁

元智大學資訊傳播學系
s986424@mail.yzu.edu.tw

鄧進宏

元智大學資訊傳播學系
chteng@saturn.yzu.edu.tw

摘要

3D 模型的建立是 3D 數位內容創作中必要的一環，然而由於 3D 建模的高進入障礙使得目前只有專業人士才有能力進行 3D 建模的工作。3D 建模之所以不易，主要是傳統的建模軟體必需透過平面的操作介面來操作三度空間中的物體，如此不但不直覺且在操作上也不方便。為了降低 3D 建模的門檻，在本研究中，我們發展了一個以擴增實境為基礎的互動式 3D 建模系統，此系統運用 SIFT 特徵及比對技術，搭配攝影機自我校正技術重建場景中特徵點的三度空間位置，再以此為基礎，利用一個有效的 Perspective-n-Point 演算法即時追蹤攝影機的位置與方向。利用此擴增實境技術，我們可以讓使用者在一個虛實相間的環境中對物件進行調整，進而達到 3D 建模的目的。

關鍵詞：攝影機追蹤、互動式 3D 建模、擴增實境

Abstract

3D modeling is an important and sometimes difficult task in 3D digital content creation. Its difficulty arises from the fact that most modeling systems create 3D models from unfriendly user interface based on 2D views. This unfriendly user interface creates a barrier for non-expert users who do not have professional knowledge in 3D modeling. In view of this, in this paper we develop an interactive 3D modeling system based on the technology of Augmented Reality (AR) and hope this system can reduce the barrier of 3D modeling. We employ SIFT features to establish image correspondences and use these correspondences to calibrate camera and therefore reconstruct the 3D positions of these correspondences. Following this, an efficient Perspective-n-Point algorithm is applied to track the camera so that we can create a primitive-based 3D modeling toolkit on this AR platform. By this system, users can operate and edit the 3D object via an interactive and intuitive

approach in this AR environment and finally create the required 3D models.

Keywords: Camera tracking, interactive 3D modeling, Augmented reality

1. 前言

隨著電腦 3D 顯示技術的進步，各種 3D 應用如 3D 電影及遊戲等也隨之發展，這些成功的應用讓 3D 數位內容成為一個熱門的產業。

在 3D 數位內容的製作中，3D 建模是必要的環節，目前大多數的建模方式都是使用專業的 3D 建模軟體如 3ds Max 或 Maya 等來進行，但是這些軟體大多昂貴且不易上手，對於非專業人士來說有相當困難的進入障礙。這些建模軟體之所以不易操作的一個主要原因在於它們以平面圖的方式來編輯及操作 3D 物件，誠如學者 Pierkarski 在其博士論文[28]中提到，人類在平面環境中雖然可以理解 3D 的構造，但是要從平面環境構築 3D 物件並不如想像中容易。有鑑於此，學者 Clark 等人[11]製作了一個完全虛擬的 3D 空間環境，讓使用者可以在此虛擬空間中進行建模工作。然而，此種方法雖可讓使用者以更自然的方式操作 3D 物件，但要建構一個這樣的虛擬環境並不容易，一般而言我們需要許多昂貴的設備才能完成此項工作。

比起完全由電腦進行模擬的虛擬實境，擴增實境(Augmented Reality, AR)是一種在真實的空間中疊加虛擬影像的技術，由於有現實的環境作為背景，有時在畫面的呈現上會比虛擬實境更加逼真，加上在建立物件 3D 模型時大多需要參考真實的物體，而現實空間的背景正好可直接作為參考，因此學者 Pierkarski 便使用了擴增實境技術來進行 3D 建模的開發[28]，配合頭戴式顯示器(head mounted display, HMD)、各式的定位儀器以及背包電腦，活動範圍可以延伸到戶外大型物體的建模，但是此種方式仍須昂貴的儀器設備，並不適用於一般普羅大眾。

實際上，依現今的電腦視覺技術，要使用擴增實境並不需要昂貴的定位設備，擴增實境需要的只是將虛擬的物件融合到現實環境中，因此只要掌握環境中的 3D 資訊，並將虛擬物件定位即可，故只要有一台攝影機以及電腦，便能實現簡單的擴增實境技術。

綜上所述，如果想要讓一般普羅大眾也能創造屬於自己的 3D 數位內容，3D 模型的建立過程必須簡化，而擴增實境技術能讓使用者以更直觀的方式進行 3D 物件的操作，進而降低平面建模所遭遇的困難。其次，在現今網路攝影機與電腦價格平價化的狀況下，建構一個簡單的擴增實境環境並不需要昂貴的建構成本。因此我們期望能在 PC 平台上建構一個以擴增實境技術為基礎的互動式 3D 建模系統。

2. 文獻探討

本研究的主要目的是發展一個以擴增實境為基礎的互動式 3D 建模系統，因此我們從以下兩方面來探討相關的研究。

2.1 互動式 3D 建模

3D 建模是 3D 數位內容製作中最基礎的部份，因此如何能夠快速且有效的建立物體 3D 模型一直是學術界研究的方向之一，目前人們最為熟知的 3D 建模技術是使用 3D 建模軟體如 3ds Max 或 Maya 來進行，這些軟體需藉由技術人員以人工方式進行創作，可以憑空建立現實中不存在的物體，因此對於遊戲與動畫產業來說，這種軟體有其必要性，但如同前言所述，這類軟體的使用與學習並不容易，而且以平面的方式進行 3D 的控制並不自然。

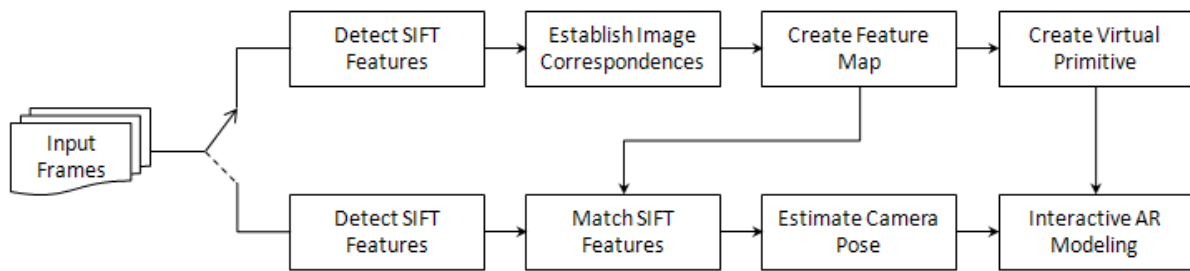
相對於以軟體進行建模，另一種傳統的建模方式是使用專業的儀器對物體進行量測以取得物體的 3D 資訊，如[22]便是一個利用結構光(structured light)以及 range finder 以即時的方式重建物體 3D 模型的方法。另外我們也可直接使用影像重建物體的 3D 模型，如[26]中作者使用了六台攝影機從不同的角度拍攝物體，使用者只要將物體放置在中央的轉台上，便可以建立精確的 3D 模型。這些方法的優點是精確度高，建立模型的過程中不需要人為介入，但是這類儀器多半極為昂貴，一般民眾不會使用這類型的技術進行建模。

為了能夠讓物體 3D 模型的建立更為容易，許多研究人員嘗試發展更簡單直覺的建模方式，其中影像建模技術(image-based modeling)

是一個相當有潛力的方式 [2][4][5][6][13][15][17][18][26]，這種建模方法只需要使用相機對物體拍攝數張影像即可建立其 3D 模型，相較於上述方法不但直覺且便宜。這些影像建模技術一般是使用 SfM(structure from motion)技術 [7][14][21]來推算影像中一些特徵點在三度空間中的座標位置，但是要如何從這些 3D 點中區分出我們所需的物體並建立其多邊形模型並不容易，因此目前的影像建模技術大多需要人工輔助才能有較佳的效果，也因此這類型的建模方式可進一步歸類為互動式的影像建模技術，例如 VideoTrace[2]是一個互動式影像建模相當成功的案例，使用者只需將一段影像中想要建模的物體的外框多邊形描繪好，系統便可以完成此物體的 3D 模型，此系統便是使用上述的 structure from motion 技術重建出影像中特徵點的三度空間座標，當使用者描繪物體外框的多邊形時，系統可利用落於多邊形的三度空間點計算每個多邊形的方向和位置，從而建立物體的 3D 模型。此技術在操作上非常容易，不過由於此技術並非即時，而是對一段影片進行處理，在使用時必需要先對物體拍攝影片後，再將其上傳至電腦進行描繪外框的處理才能完成建模。

2.2 利用擴增實境技術進行 3D 建模

擴增實境技術發展至今已有了四十年的歷史，自從學者 Sutherland 發展了第一個頭戴式顯示器後[10]，結合虛擬與現實的擴增實境應用開始被人們所注意。根據學者 Azuma 對擴增實境所下的定義[19]，一個完整的擴增實境系統應包含以下三項要素：(1)虛擬與現實結合(Combines real and virtual)，(2)即時的互動(Interactive in real-time)，(3)三度空間的定位(Registered in 3D)。由於擴增實境結合虛擬與現實的特性，加上可以容許與使用者進行即時的互動，而這正是互動式建模所需要的要素，同時在三度空間中建模也較平面建模更為直覺，建模的結果也可以跟現實的物體進行比對，也可以直接將現實物體的紋理作為貼圖使用。有鑑於此，學者 Pierkarski 發展了一套在擴增實境下的互動式 3D 建模系統[27][28]，此系統主要鎖定於戶外大型物體的 3D 建模，Pierkarski 提出了工作平面(working plane)的概念，他將建模的工作拆解成一個個的平面，分別在個別的平面上以 CSG (Constructive Solid Geometry)技術進行建模。此系統雖然可以進行



圖一、系統架構圖

戶外大型物體的建模，但在使用上需搭配頭戴式顯示器來顯示虛擬與現實結合的影像，並且需要額外的定位儀器如 GPS (Global Positioning System)、磁性以及慣性追蹤器 (magnetic and inertial tracker) 等來輔助定位，還需要一個特製的手套進行選單的操作，最後為了能在戶外運作，使用者還需背負一台背包電腦以及能維持所有設備所需的電源供應器，因此在使用上有所限制。此系統由 Pierkarsk 與其在 University of South Australia 的研究團隊所開發，他們後續還對此系統增加了一些功能，包括在擴增實境環境內一些 3D 選單的使用與建立[23]。

Jiim (In-situ Image-based Modeling) 是由 Hengel et al. 所開發的一套擴增實境 3D 建模系統 [1][3]，此系統使用了即時相機追蹤技術及自動影像分析技術達到沉浸式的 3D 建模 (immersive modeling)，此系統概念與 VideoTrace[2] 類似，均利用 SfM 技術分析影像以獲取物體特徵點在三度空間中的資訊，再利用這些資訊來進行 3D 建模，差別在此系統建立在擴增實境的環境下，因此可達到即時的互動。

除了上述系統之外，另外還有一些應用擴增實境技術進行 3D 建模的系統，如 ARpm (Augmented Reality for Polygon Modeling) [16] 是一個結合擴增實境技術與 3dsMax 功能的系統，使用者配戴頭戴式顯示器後，可於可觸式介面 (即 AR 標記) 上看到虛擬的 3D 物體，使用者可藉由一個無線的指向設備 (pointer device) 對物件進行編輯，此可觸式介面可直接對應到 3ds Max 所提供的建模工具指令，讓使用者可直接應用 3ds Max 的強大建模能力而不需對此軟體進行任何的修改。除此之外，將 AR 標記應用於 3D 建模的系統還有 [20][24]，其中 3DARModeler[24] 可說是一個簡化版本的 3ds Max，他有著 3ds Max 的主要功能，使用者可以在擴增實境的環境中，配合使用傳統的輸入方式如鍵盤與可觸式介面，進行建模的動

作。這類系統的目標是讓非專業的技術人員能在更為直觀的環境下進行 3D 模型的製作，但是由於這些系統僅是提供了一個擴增實境的環境與建立模型的工具，在建立模型時並沒有參考實際的物體，也因此並沒有完全發揮擴增實境虛擬與現實結合的優勢。

3. 研究方法

3.1 系統架構

本研究的系統架構如圖一所示，其中可分為特徵點的偵測與比對 (Detect & Match SIFT Features)、Feature Map 的建立 (Establish Image Correspondences & Create Feature Map)、攝影機的追蹤 (Estimate Camera Pose)、虛擬物件的加入 (Create Virtual Primitive) 以及互動式擴增實境 3D 建模 (Interactive AR Modeling) 等幾個部份。

在使用擴增實境技術進行互動式 3D 建模時，我們需要得知攝影機在三度空間中的方向與位置，這些資訊的取得依靠於所謂的特徵地圖 (Feature Map) 的建立。Feature Map 是影像平面中的特徵點與其在三度空間中的對應點的集合，這些點在影像平面與三度空間的座標必需以電腦視覺技術取得。在 Feature Map 建立完成之後，針對每一張攝影機取得的畫面，只要我們找出影像中的特徵點，再將之與 Feature Map 中的特徵點進行比對，由於 Feature Map 中的特徵點在三度空間中的座標均為已知，因此可藉由比對成功的特徵點得知其對應的三度空間座標，最後根據影像平面座標與三度空間座標的對應關係，可以計算出當前畫面中攝影機的位置與方向，隨後即可將虛擬物體放置於場景中的指定位置，以便於後續的建模作業。下面我們將對本系統的幾個主要模組進行更詳細的說明。

3.2 特徵點的偵測與比對

特徵點的偵測與比對是本系統成功的關鍵之一，這些影像中的特徵點必需具備強韌性(robust)，即能夠在不同視角的攝影機畫面中被有效的偵測與比對。在本研究中，我們所採用的影像特徵點為 SIFT(Scale Invariant Feature Transform)特徵點。SIFT 特徵點是由學者 Lowe 所發展[8]，是一種已經被證明能有效克服影像旋轉、縮放與亮度變化等影響的特徵點，並且已經成功的應用於無標記式擴增實境攝影機追蹤上[9]。

SIFT 特徵點的基本概念是使用每個特徵點周圍的梯度(gradient)大小以及方向的統計值來進行點的特徵描述和比對。為了找出具有強韌性的特徵點，首先需要將影像與不同 scale 的高斯函數(Gaussian function) 做迴旋積分(convolution)，從而得到一系列的高斯影像。相鄰的高斯影像間進行相減可得到高斯差值影像(difference-of-Gaussian, DOG)，在 DOG 影像中的區域最大值即是 SIFT 特徵點的候選點。接著我們必須給予每個候選點一個特徵向量以及此點的影像方向(image orientation)。特徵向量的計算方式主要是將一個 circular Gaussian-weighted window 疊加在候選點上，接著計算此 window 內 gradient 數值的大小及方向的統計直方圖，此直方圖即是 SIFT 特徵向量，特徵向量的維度視直方圖的精細度而定，通常維度如果設為 128 可以達到相當不錯的結果。

決定了 SIFT 特徵點的特徵向量後，我們便可以使用特徵向量進行 SIFT 特徵點的比對。由於 SIFT 特徵向量高維度的特性，不正確的對應點其 SIFT 特徵向量的歐幾里德距離(Euclidean distance)與影像中其他不正確的特徵點均會相當接近，因此可藉由搜尋另一影像中 SIFT 向量的歐幾里德距離最近的兩個特徵點，如果兩個點的 SIFT 向量的歐幾里德距離數值的比例相當接近 1，則表示此點沒有對應的特徵點，反之若是距離的差距很大，則表示可能找到了對應的影像特徵點。由於 SIFT 特徵點的數量龐大，若是直接對 SIFT 特徵點進行比對需要耗費大量的運算時間，因此有學者使用 Best-Bin-First 演算法[9]，以 k-d tree 的方式[12]進行影像特徵點的搜尋及 SIFT 特徵點的比對，這種方式可以加快特徵點的搜尋與比對速度。

然而，即使使用了上述的方法來加快 SIFT 特徵點的搜尋和比對，但是由於 SIFT 特徵點的高複雜度特性，因此仍不足以支持 SIFT 特

徵點在需要即時運算的擴增實境應用下的場合。為了提昇運算效率，有學者嘗試使用圖形加速卡的 GPU 來進行 SIFT 特徵點的運算，利用 GPU 的平行運算能力大幅提昇 SIFT 特徵點計算與比對的效能[30]。在本研究中，我們也採用此方法來進行 SIFT 特徵點的計算與比對。

3.3 Feature Map 的建立

如同之前所述，本系統的實現需要建立一個 Feature Map，在本研究中，我們使用了攝影機自我校正(camera self-calibration)演算法[7]來取得影像平面中的特徵點在場景中的三度空間座標。實際運作的流程如下，首先對場景拍攝數張不同視角的影像，然後對這些影像進行 SIFT 特徵點偵測，再利用這些特徵點進行 SIFT 特徵向量的比對來建立這些特徵點在影像間的對應關係(即圖一的 Establish Image Correspondences)，有了這些影像特徵點的對應關係後，我們即可透過攝影機自我校正演算法來估算當時的攝影機參數以及這些影像特徵點所對應的三度空間座標。

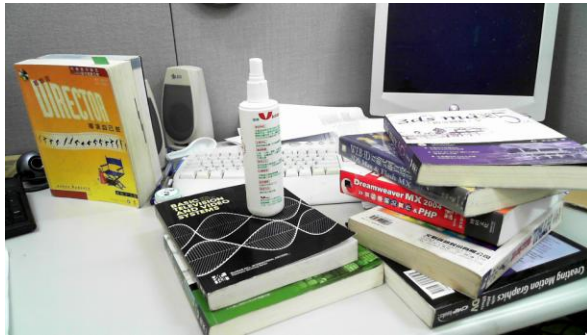
具體而言，假設 $\mathbf{m}_j^{(i)} = (m_{j,x}^{(i)}, m_{j,y}^{(i)})$ 是第 i 張影像上的第 j 個對應點，則攝影機自我校正便是找出一組攝影機內外部參數和一組與之相對應的三度空間點，使其能儘量滿足投影方程式。換言之，攝影機自我校正便是透過最小化以下的目標函數來達成：

$$\sum_{i,j} \left[\left(m_{j,x}^{(i)} - \frac{\mathbf{P}_1^{(i)T} \mathbf{M}_j}{\mathbf{P}_3^{(i)T} \mathbf{M}_j} \right)^2 + \left(m_{j,y}^{(i)} - \frac{\mathbf{P}_2^{(i)T} \mathbf{M}_j}{\mathbf{P}_3^{(i)T} \mathbf{M}_j} \right)^2 \right] \quad (1)$$

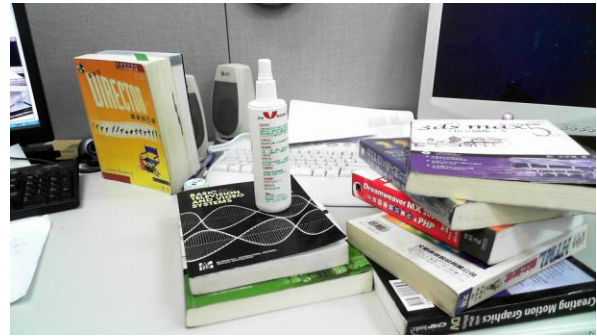
其中 \mathbf{M}_j 為第 j 個三度空間點的齊次座標(homogeneous coordinates)表示， $\mathbf{P}_k^{(i)}$ 代表第 i 張影像中攝影機的投影矩陣 $\mathbf{P}^{(i)}$ 的第 k 列，其中 $\mathbf{P}^{(i)} = \mathbf{K}^{(i)} [\mathbf{R}^{(i)} | -\mathbf{R}^{(i)} \mathbf{t}^{(i)}]$ ， $\mathbf{R}^{(i)}$ 為旋轉矩陣， $\mathbf{t}^{(i)}$ 為平移向量，表示攝影機在三度空間中的位置及方向，即攝影機的外部參數。 $\mathbf{K}^{(i)}$ 為第 i 張影像的攝影機校正矩陣(camera calibration matrix)，亦即攝影機的內部參數。此攝影機校正矩陣為一 3×3 的矩陣，如下所示：

$$\mathbf{K}^{(i)} = \begin{bmatrix} \varphi^{(i)} & \gamma & u_0 \\ 0 & f^{(i)} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

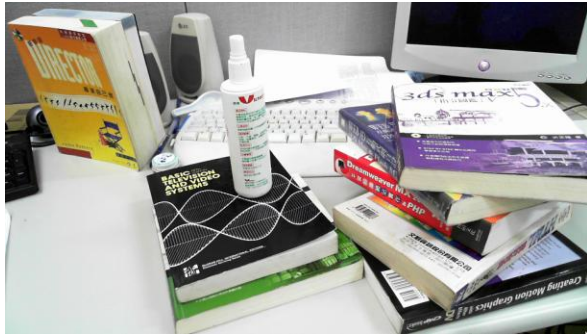
其中 $f^{(i)}$ 為攝影機的有效焦距(effective focal



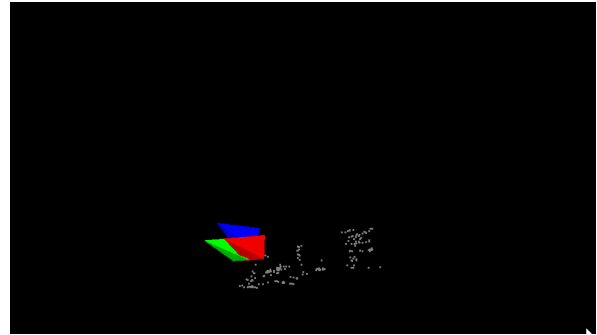
(a)



(b)



(b)



(d)

圖二、(a)(b)(c)建立 Feature Map 的三張影像。
(d)攝影機自我校正所得到的三度空間點以及攝影機位置。

length), τ 為影像像素水平與垂直方向的比例, 一般來說會接近 1。 γ 為攝影機的傾斜因素(skew factor), 代表的是攝影機成像面 x 方向與 y 方向的夾角關係, 一般來說會接近 0。 (u_0, v_0) 代表影像平面主點, 代表光軸通過攝影機鏡頭中心點與影像平面的交點, 一般來說 (u_0, v_0) 會接近影像的中心點。通常我們可以合理的假設 τ 為 1、 γ 為 0 以及 (u_0, v_0) 的數值為影像的中心點, 這樣在矩陣 $\mathbf{K}^{(i)}$ 中便只有焦距會隨著影像的不同而改變。

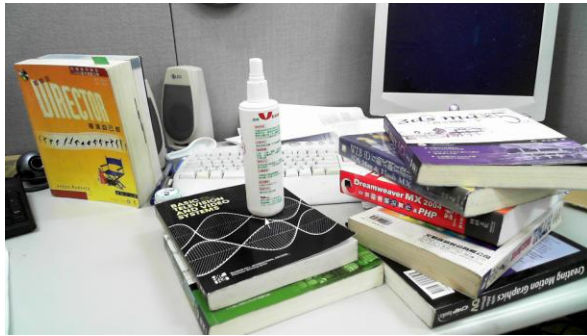
由於需要最小化的目標函數內有大量的未知參數 (\mathbf{M}_j 、 $\mathbf{K}^{(i)}$ 、 $\mathbf{R}^{(i)}$ 及 $\mathbf{t}^{(i)}$), 若直接對其進行最小化會非常沒有效率, 實際上攝影機的內外部參數與這些三度空間點有著密切的關係, 因此我們可利用此關係來降低函數中的參數個數[7]。我們可先給定一組攝影機內部參數, 接著根據影像間的同軸幾何(epipolar geometry)關係(即 fundamental matrix)計算出其對應的攝影機外部參數, 接著使用三角定位法計算出最符合此攝影機內外部參數的三度空間點。藉由此種方式, 我們可將(1)式的目標函數轉化為只與攝影機內部參數有關, 最後再以數值最佳化演算法找出可使(1)式最小化的解並找出其對應的三度空間點的坐標。這些三度

空間點和其對應的影像平面座標點便可做為我們所需要的 Feature Map。

圖二所示為我們使用上述方法建立 Feature Map 的一個範例, 首先我們先對場景拍攝了三張不同視角的圖像如圖二(a)(b)(c)所示, 接著對這些圖片進行 SIFT 特徵點的偵測, 並透過特徵點的比對建立影像間的對應點。由於特徵點的比對難免會有比對錯誤的情形發生, 若是沒有去除這些不正確的特徵點(outliers)有可能影響到相機的自我校正結果。本研究使用了影像間的同軸幾何關係, 即 fundamental matrix 來去除不正確的對應點[21]。我們使用了一個強韌式的演算法(robust estimation)來估算 fundamental matrix 的數值, 並利用此方法去除不正確的對應點, 最後利用剩下的正確對應點進行攝影機自我校正並重建對應的三度空間點。最後的重建結果如圖二(d)所示, 紅綠藍三個角錐分別代表拍攝照片時的攝影機位置與方向, 角錐體頂點為攝影機的中心位置, 角錐體底面為攝影機面對的方向, 白色的點為重建完成的三度空間點。

3.4 攝影機的追蹤

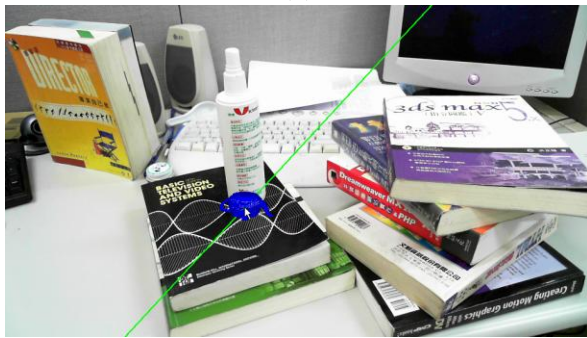
在 Feature Map 建立完畢後, 我們即可藉



(a)



(b)



(c)



(d)

圖三、虛擬物件的加入流程。

(a)在畫面上選定欲放置虛擬物件的位置(滑鼠游標處)。

(b)在另一張影像上顯示 epipolar line(綠線)。

(c)在 epipolar line 上選定一點之後即可將虛擬物件疊加至場景中指定的位置。

(d)調整虛擬物件的方位後可將虛擬物件正確的疊加至指定的平面上。

此進行攝影機的追蹤，亦即我們必需取得攝影機所拍攝的每張畫面當下的位置及方向，有了這些資訊我們才能將虛擬物件疊加至真實空間中指定的位置上。一般來說，傳統的擴增實境應用如 ARToolKit[29]是假設攝影機的內部參數(即(2)式的矩陣 \mathbf{K})為已知，因此需要利用一些離線的攝影機校正手續，即讓攝影機拍一些特殊圖案以估算攝影機的內部參數。但是在本研究中，我們已經透過攝影機自我校正技術取得了攝影機的內部參數，因此不需要事先行離線校正手續即可直接進行攝影機的追蹤。

在得知攝影機內部參數後，接著只要取得一些三度空間中的座標點及與之對應的影像平面座標點，即可計算出攝影機的外部參數，也就是攝影機在三度空間中的位置與方向。此問題在文獻上被稱為 Perspective- n -Point (PnP) problem，目前已經有許多學者對此問題提出解法，其中 Lepetit 等所提出的 EPnP 演算法[25]在效率與精確度上均有一定的水準，因此本研究採用此方法來進行攝影機的追蹤。

EPnP 的基本概念是將所有的三度空間點表示為四個控制點的線性組合，PnP 的問題即

轉換為找出這四個控制點在攝影機座標系下的座標，此可藉由解一個 12×12 矩陣的特徵向量(eigenvector)以及數個二次方程式來求出這些座標。此方法只需四組以上的三度空間點和與其對應的影像平面投影點即可進行計算，加上不論是平面與非平面的點均可正常運作，因此我們只需給其一組三度空間點和與其對應的影像平面座標點，即可求出當下攝影機的方向及位置，有了這些資訊便可以將虛擬物件疊加於影像中的特定位置。

為了將三度空間點與影像平面上的投影點建立對應關係，我們先將攝影機取得的畫面進行 SIFT 特徵點的偵測，再將這些特徵點與 Feature Map 中的影像特徵點進行 SIFT 特徵點的比對，由於我們已經知道 Feature Map 中每一個影像平面特徵點對應的三度空間座標，因此只要比對成功的點，我們便可成功建立其在影像平面與三度空間中的對應關係。同時如 3.3 節所述，為了避免特徵點比對錯誤造成對應不正確，我們利用比對成功的對應點進行強韌式演算法計算目前影像與當初攝影機自我校正的影像之間的 fundamental matrix，藉此去除不



圖四、互動式 3D 建模過程

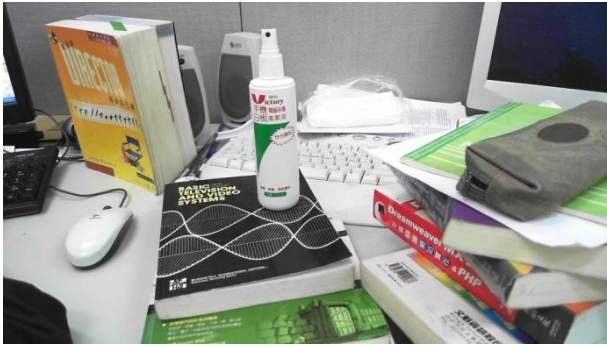
正確的對應點，最後將剩下的正確對應點代入 EPnP 演算法便可以求出當下攝影機的旋轉矩陣 \mathbf{R} 與平移向量 \mathbf{t} 。

3.5 虛擬物件的加入

為了能夠在擴增實境的環境下進行互動式 3D 建模，我們必需在真實的場景中疊加虛擬物件，但是由於我們所擁有的資訊只有 Feature Map 中 3D 點的集合，因此無法知道場景中平面的位置，加上我們並沒有使用標記 (marker) 進行輔助定位，因此我們需要使用者的幫助來設定虛擬物件在場景中的位置及方向。

本研究採用了一個簡單的方式來加入虛擬物件。在 Feature Map 建立完畢後，我們會在螢幕上顯示當初使用於攝影機自我校正的

其中一張圖像(如圖三(a)所示)，使用者需在圖像上進行點擊以選擇虛擬物件欲放置的位置，隨後我們可藉由影像之間的 epipolar 關係 [21]，在另外一張影像上產生所謂的 epipolar line(如圖三(b)所示)，此線代表了在第一張影像中所選擇的點反投影至三度空間中的射線，再將之投影到第二張影像上的結果，之後使用者只要在此線上對當初在第一張影像欲放置虛擬物件的位置再度進行點擊，系統即可利用攝影機自我校正的結果以三角定位法算出此點的三度空間座標，也因此可以將虛擬物件放置於此點上(如圖三(c)所示)，但是此時僅有座標已知，虛擬物件的方向還不确定(如圖三(c)中倒轉的茶壺)，因此還需要一些人為調整。此部份我們使用簡單的鍵盤操作，使物件可以在原地



圖五、互動式 3D 建模的結果。上排圖為攝影機所取得的畫面。下兩排圖為本系統所建立的 3D 虛擬物件經過貼圖後在不同視角下的繪製結果。

進行旋轉，使用者可將之轉到正確的方向(如圖三(d)中與書面貼齊的茶壺)，這樣便完成了虛擬物件的加入，隨後系統便可進入擴增實境的環境，讓使用者進行互動式 3D 建模。

3.6 互動式擴增實境 3D 建模

本研究所使用的 3D 建模方式我們稱之為 primitive-based 建模法，即我們會先從一個基本的 3D 物件(primitive)開始，然後藉著擴增實境互動的方式對此基本物件進行塑形來達到 3D 建模的目的地。我們首先會讓使用者選擇一個基本的 3D 物件並設定其基本參數如圖四中圓柱體的底面半徑、高度及多邊形面數，使其儘量接近欲建模的真實物體，此基本物件會依前一節虛擬物件的加入所述的方式疊加至畫

面中指定的位置，之後使用者可透過選擇此 3D 物件的一個或一組頂點使用鍵盤進行移動，藉此來達到塑形的目標。3D 物件上頂點的選擇方式可利用先前使用攝影機自我校正所求出的攝影機內部參數(及上述的矩陣 \mathbf{K})搭配 EPnP 演算法所求出的攝影機外部參數(即攝影機當下的旋轉矩陣 \mathbf{R} 與平移向量 \mathbf{t})將 3D 物件上所有的頂點投影到影像平面上，然後藉此判定哪些點位於使用者所選取的範圍內。換言之，若 \mathbf{M} 是 3D 物件中的一個頂點， \mathbf{P} 是當下畫面的攝影機投影矩陣，則此 3D 點的影像座標 \mathbf{m} 可用以下的方程式計算：

$$\mathbf{m} = \begin{bmatrix} u \\ v \\ w \end{bmatrix} \sim [\mathbf{KR} \mid -\mathbf{KRt}] \mathbf{M} \sim \mathbf{PM} \quad (3)$$

其中符號 \sim 表示此方程式等號兩邊只相差了一個倍率(scale)，3D 點 \mathbf{M} 及影像平面座標點 \mathbf{m} 皆為齊次座標表示，即若影像座標點為 (x, y) ，則 $(x, y) = (u/w, v/w)$ 。

圖四為本系統中互動式 3D 建模的一個範例，在此範例中我們希望能建立一個清潔液瓶子的 3D 模型。首先使用者選擇圓柱體作為基本物件並將之放置於真實物體的平面上，接著使用者選取了圓柱體頂面的頂點，被選中的點會呈現綠色(如圖四(a)所示)，接著使用者將圓柱體的頂面拉高，使圓柱體與真實的物體高度貼齊(如圖四(b)所示)。之後使用者可調整圓柱體切片的高低並將圓柱體的半徑縮小以符合真實物體的外觀，讓此基本物件和真實的瓶子外觀越來越接近。本系統為了讓使用者能夠隨時看到建模的狀況，因此我們將虛擬物件以半透明的狀態顯示，讓使用者可以觀察模型與真實物體間的差異以便使用者調整模型的外型。因此在圖四中的淡藍色物件便是完整的 3D 模型，此模型再經過貼圖後更能完整呈現此 3D 建模的結果。

4. 實驗結果

為檢視本論文所提互動式擴增實境 3D 建模的可行性，我們依照前述方式以 primitive-based 建模法，對真實物體進行 3D 建模。圖五所示即為本系統所完成的 3D 建模案例，從此圖中可看出不論是圓柱狀的清潔液瓶子或是立方體的茶葉罐，在模型經過貼圖輔助後均可達到不錯的效果。從這些實驗結果可得知，對於構造簡單的物體(如圓柱狀或立方體)，本系統可以成功的建立其 3D 模型，且由於本系統係以擴增實境方式進行建模，與時下常用的建模軟體如 3ds Max 相比，由於使用者可用真實的物體做參考即時操作 3D 物件的外型，因此在使用上更為方便且自然。

5. 結論與未來展望

本研究發展了一套以擴增實境為基礎的互動式 3D 建模系統，此系統使用 3D 基本元件進行建模，利用 SIFT 特徵及比對技術，搭配攝影機自我校正技術來重建場景中的特徵點，再利用 Perspective-n-Point 演算法求出攝影機的位置與方向，以達成在擴增實境的環境中進行 3D 建模的目的。

本研究雖然成功地發展了一個在擴增實

境下的 3D 建模系統雛形，但做為一個完整的 3D 建模工具仍有不足之處，以下是我們未來的努力方向：

1. 新增更多的基本元件。由於本系統目前的基本元件僅有球、圓柱體及立方體三種，在選擇上有所限制，若加入更多的基本元件有助於在一開始就建立更接近真實物體外觀的 3D 元件。

2. 加入更多的建模技術。由於本系統目前在修改模型上僅有頂點移動的方式，若能加入更多的建模技術如旋轉面(surface of revolution)、擠壓面(extrusion)以及自由形變(free-form deformation)等將有助於節省使用者的時間與勞力。

3. 加入 Feature Map 動態更新機制。由於本系統目前在特徵點不足或是視角變化太大時會有抖動的現象，因此若能動態更新特徵點，應有助於解決此問題。

參考文獻

- [1] A. Hengel, "Image-based modeling for augmented reality," International Symposium on Ubiquitous Virtual Reality, 2010.
- [2] A. Hengel, A. Dick, T. Thormahlen, B. Ward, P. H. S. Torr, "VideoTrace: Rapid interactive scene modeling from video," ACM Transactions on Graphics, Vol. 26, No. 3, 2007, Special issue: Proceedings of ACM SIGGRAPH 2007.
- [3] A. Hengel, R. Hill, B. Ward, and A. Dick, "In situ image-based modeling," in Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR'09), pp. 107-110, 2009.
- [4] A. Reche, I. Martin, G. Drettakis, "Volumetric reconstruction and interactive rendering of trees from photographs," ACM Transactions on Graphics, Vol. 23, No. 3, pp. 720-727, 2004, Special issue: Proceedings of ACM SIGGRAPH 2004.
- [5] B. Neubert, T. Franken, O. Deussen, "Approximate image-based tree-modeling using particle flows," ACM Transactions on Graphics, Vol. 26, No. 3, 2007, Special issue: Proceedings of ACM SIGGRAPH 2007.
- [6] C.-H. Teng and Y.-S. Chen, "Image-based tree modeling from a few images with very narrow viewing range," The Visual Computer, Vol. 25, No. 4, pp. 297-307.
- [7] C.-H. Teng, Y.-S. Chen, and W.-H. Hsu, "Camera self-calibration method suitable for

- variant camera constraints," *Applied Optics*, Vol. 45, No. 4, pp. 688-696, 2006.
- [8] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, Vol. 60, No. 2, pp. 91-110, 2004.
- [9] I. Skrypnik and D. G. Lowe, "Scene modelling, recognition and tracking with invariant image features," in *International Symposium on Mixed and Augmented Reality (ISMAR'04)*, pp. 110-119, November 2004.
- [10] I. Sutherland, "A head-mounted three-dimensional display," in *AFIPS Fall Joint Computer Conference*, pp. 757-764, Washington, DC, 1968.
- [11] J. H. Clark, "Designing surfaces in 3-D," *Communications of the ACM*, Vol. 19, No. 8, pp. 454-460, 1976.
- [12] J. S. Beis and D. G. Lowe, "Shape indexing using approximate nearest-neighbour search in high-dimensional spaces," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1000-1006, 1997.
- [13] L. Quan, P. Tan, G. Zeng, L. Yuan, J. Wang, S. B. Kang, "Image-based plant rendering," *ACM Transactions on Graphics*, Vol. 25, No. 3, pp. 599-604, 2006, Special issue: *Proceedings of ACM SIGGRAPH 2006*.
- [14] M. Pollefeys, "Self-calibration and metric 3D reconstruction from uncalibrated image sequences," Ph.D. thesis, Katholieke Universiteit Leuven, 1999.
- [15] P. E. Debevec, C. J. Taylor, and J. Malik, "Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach," in *Proceedings of SIGGRAPH 1996*.
- [16] P. Fiala and N. Adamo-Villani, "ARpm: an augmented reality interface for polygonal modeling," in *Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR'05)*, 2005.
- [17] P. Muller, G. Zeng, P. Wonka, L. V. Gool, "Image-based procedural modeling of facades," *ACM Transactions on Graphics*, Vol. 26, No. 3, 2007, Special issue: *Proceedings of ACM SIGGRAPH 2007*.
- [18] P. Tan, G. Zeng, J. Wang, S. B. Kang, L. Quan "Image-based tree modeling," *ACM Transactions on Graphics*, Vol. 26, No. 3, 2007, Special issue: *Proceedings of ACM SIGGRAPH 2007*.
- [19] R. Azuma, "A survey of augmented reality," *Presence: Teleoperators and Virtual Environments*, Vol. 6, No. 4, pp. 355-385, 1997.
- [20] R. Freeman and A. Steed, "Interactive modeling and tracking for mixed and augmented reality," in *Proceedings of the ACM symposium on Virtual Reality Software and Technology, (VRST'06)*, 2006.
- [21] R. I. Hartley, A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2nd edition, ISBN: 0521540518, 2003.
- [22] S. Rusinkiewicz, O. Hall-Holt, M. Levoy, "Real-time 3D model acquisition," *ACM Transactions on Graphics*, Vol. 21, No.3, Special issue: *Proceedings of ACM SIGGRAPH 2002*, pp. 438-446, 2002.
- [23] T. N. Hoang and B. H. Thomas, "Augmented reality in-situ 3D model menu for outdoors," in *Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR'08)*, pp. 185-186, 2008.
- [24] T. V. Do and J.-W. Lee, "3DARModeler: a 3D modeling system in augmented reality environment," *International Journal of Electrical, Computer, and Systems Engineering*, Vol. 4, No. 2, pp. 145-154, 2010.
- [25] V. Lepetit, F. Moreno-Noguer, and P. Fua, "EPnP: An accurate O(n) solution to the PnP problem," *International Journal of Computer Vision*, Vol. 81, pp. 155-166, 2009.
- [26] W. Matusik, H. Pfister, A. Ngan, P. Beardsley, R. Ziegler, and L. McMillan, "Image-based 3D photography using opacity hulls," *ACM Transactions on Graphics*, Vol. 21, No.3, Special issue: *Proceedings of ACM SIGGRAPH 2002*, pp. 427-437, 2002.
- [27] W. Piekarski, "3D modeling with the Tinmith mobile outdoor augmented reality system," *IEEE Computer Graphics and Applications*, Vol. 26, No.1, pp. 14-17, 2006.
- [28] W. Piekarski, "Interactive 3D modeling in outdoor augmented reality worlds," Ph.D. thesis, University of South Australia, 2004.
- [29] ARToolkit, Available at: <http://www.hitl.washington.edu/artoolkit/>.
- [30] SiftGPU, Available at: <http://cs.unc.edu/~ccwu/siftgpu/>.