

淺談 2D 至 3D 視訊轉換技術

賴文能、陳韋志

國立中正大學電機工程研究所

E-mail: ieewnl@ccu.edu.tw

<http://www.dsp.ee.ccu.edu.tw/wnlie.html>

一、3DTV 時代真的來了?!

隨著近年來 3D 電影的風行，如阿凡達 (Avatar) 與魔境夢遊 (Alice in Wonderland) 等電影的推出，3D 已在全球掀起一股熱潮，然而 3D 立體顯示技術並非最近幾年才有，早在 1928 年 John Logie Baird [1] 就已開發出第一套立體電視系統，而在 1935 年也已經出現第一部彩色 3D 立體電影，隨後在 1950 年代美國也拍攝了許多部的 3D 立體電影。然而當時的立體電影只能夠在立體劇院播放，加上當時立體播放設備並未普及，3D 立體電影未能進入一般家庭市場。時至今日，因 3D 電影又再帶起新的 3D 熱潮，許多大廠紛紛喊出 3DTV 要進入家庭的口號，這不禁讓人懷疑 3DTV 時代真的來臨了嗎？是否會重道覆轍，宛如曇花一現呢？

其實，從市場的一些趨勢分析我們可以看出其端倪，其一是立體顯示器的發展：隨著立體顯示器技術的進步，市面上已經有不少市售的立體顯示裝置，如 3D 電視、3D 筆電以及 3D 立體投影機等，且價格越趨平民化，而 3D 立體相關產品的需求量也逐年提升 [2]。隨著立體顯示裝置的普及，不難想像 3DTV 時代是真的來臨。另一個重點在於 3DTV 內容的來源，除了硬體裝置的技術要到位之外，3D 內容也是 3DTV 應用所不能缺少的一環。3DTV 廣播技術已有不少研究在進行，如日本在 2007 年 12 月開始使用數位衛星進行 3DTV 廣播，在歐洲也有 ATTEST (Advanced Three-dimensional Television System Technologies) [1] 與 3DTV 計畫，以及韓國的 3DMB 系統等，加上現今不斷推出的 3D 立體電影，顯而易見的，3DTV 時代是真的來臨了。

然而在這股 3D 風潮中其潛在主要的商機還是 3D 數位內容。在每家 3D 立體顯示器廠商不斷推陳出新的現況之下，消費者所能購買的立體顯示裝置越趨

多元，然而最重要的 3D 內容來源仍不夠充足，即使消費者購買了立體顯示裝置，3D 立體內容也只局限於立體電影與 3DTV 廣播。雖然目前市面上已經有少數的 3D 立體拍攝系統，能夠提供 3D 立體內容，不過由於其價格以及使用上不夠便利，因此尚未能普及化，對於現今的市場需求而言，3D 內容提供已經是未來最具發展潛力的市場。

二、2D 至 3D 視訊轉換的應用需求

目前市面上所能見到的 3D 立體拍攝系統主要可以分為兩種方式：(1)陣列式攝影機與(2)深度攝影機。陣列式攝影機是使用多架平行擺設的攝影機(一般而言為兩架，多架攝影機的狀況可以應用於多視域立體影像之應用)同時拍攝，在拍攝之前攝影機須進行校正，以確保攝影機之間沒有垂直的位移，而攝影機之間的距離須符合人雙眼之間的距離(一般為6.5公分)，以模擬人雙眼視物的情況。除此之外，也必須進行攝影機的同步，因此在架設拍攝環境時費時費力，此外兩架攝影機之間有可能因光影變化或是攝影機內部參數的影響，使拍攝到的立體影像有色差的問題。而深度攝影機的拍攝方式則是使用一部傳統的攝影機搭配一部深度攝影機拍攝，一般深度攝影機是使用 Time of flight (ToF) 的原理，藉由計算打出的紅外光打中物體反射後的時間，計算物體與攝影機的距離，在拍攝時傳統攝影機與深度攝影機之間的關係也需校正，以確保傳統攝影機所拍攝到的像素都能有其對應的深度值。然而目前深度攝影機尚未能普及，因此也難為 3D 內容來源提供一個好的解決方案。

然而對一般大眾而言，使用者最想看到的還是貼近於日常的內容，如平時出遊所拍攝的短片以及相片等，而對於內容供應商來說，現今所廣泛使用的視訊仍為傳統的 2D 視訊，面對目前所擁有的龐大 2D 視訊資料庫，若是能開發一套 2D 至 3D 轉換技術，不僅能解決目前 3D 立體內容不足的問題，2D 至 3D 轉換技術也提供一般使用者自行製作 3D 內容的樂趣。在應用層面上，舉凡各種傳統的 2D 視訊影片等，都可以成為 3D 內容，而透過此技術，廠商也能夠為消費者進行各種客製化的服務(如依顧客之需求加入各種虛擬 3D 立體物件或是 3D 特效等)，不僅在 3D 立體內容製作的成本較低廉，也比較有彈性。

三、2D 至 3D 視訊轉換的基本原理

3D 視訊之所以能讓觀看者感覺到立體，主要是基於人眼視覺感知 (Human visual system) 的原理。由於人在視物時，左、右眼看到的景像中存在著視差 (disparity) 資訊，因此 2D 至 3D 視訊轉換的關鍵即為估測視差資訊，並用以產生左、右眼立體影像對。不過，光有立體影像是足夠的，要讓觀看者能夠看到立體感還需搭配立體顯示裝置，以下我們介紹目前市面上常見的數種立體顯示器原理，並分析其使用的輸入格式。

3.1 立體顯示器

立體顯示器技術已發展多年，由早期的大型立體劇院設備，漸漸走向家庭劇院的應用。在早年，立體顯示技術主要採用偏光式 (polarized)、交錯式 (interlaced)、和分色式 (anaglyph) 三種顯示原理，雖然有一定的立體效果和品質，但由於設備使用上較不方便，以及觀看上的限制 (如人數限制)，因此未能普及。但隨著顯示器技術的逐漸成熟，近年來開發的立體顯示器已改良了傳統顯示器不足的部分，在立體效果與觀看舒適度上有很大的進步，且價格也越來越低廉，依據目前主要使用的立體顯示原理，可以分為偏光式、分時式以及裸眼式三種，表 3.1 列出各種顯示器的比較。

表3.1 常用立體螢幕比較

使用原理	是否需配戴眼鏡	顯示影像之解析度	特色
偏光式 (polarized)	Y	水平或垂直方向減半	利用正交的極化左、右眼影像
分時式 (time interlaced)	Y	同原影像	顯示裝置的更新頻率為一般的兩倍
裸眼式 (auto stereoscopic)	N	水平方向減少	目前唯一支援多視角顯示的顯示方式

3.2 立體視訊資料格式

傳統立體視訊 (Conventional Stereo Video, CSV) 是最簡單的一種立體視訊表示法 [3]，此種表示法主要是模擬人眼所看到的左、右眼影像，其拍攝方式在第二章有提到，可以使用兩架或多架平行擺設之攝影機 (攝影之距離一般為相距約

6.5公分，模擬人類雙眼之距離)。由於此種格式僅使用像素的色彩資訊，使用一般傳統攝影機即可拍攝，而拍攝得到的立體影像再經由後處理(如影像之正規化以及色差校正等)後即可用於立體視訊之播放，在此值得注意的是，CSV 格式の後處理與攝影機拍攝時的幾何參數(如攝影機內、外部參數)無關，相較於其它的 3D 視訊格式而言，CSV 在壓縮處理上也是較為簡單的，可將其左、右眼視訊視為一般 2D 視訊，並使用現有的編碼標準(如 H.264、MPEG 等)進行處理，不過 CSV 的資料量為傳統視訊的兩倍。

另一種較常使用的立體影像的格式為 2D 彩色影像搭配其對應的深度影像(2D + Depth) 格式。如第二章所提到的，此種影像格式在拍攝時使用一架傳統的彩色攝影機，負責擷取整體場景的色彩資訊，並配合一部 ToF 深度攝影機，負責記錄場景中的物體至攝影機的距離，此物體與攝影機的深度距離經由量化之後記錄為深度圖(depth image、depth map)，式子(3.1) 為微軟公司所使用的距離與深度值轉換公式：

$$z(r, c) = \frac{1.0}{\frac{P(r, c)}{255.0} * (\frac{1.0}{MinZ} - \frac{1.0}{MaxZ}) + \frac{1.0}{MaxZ}} \quad (3.1)$$

其中 $z(r, c)$ 為深度攝影機所擷取到的深度距離， $P(r, c)$ 為經由量化後得到的深度值， $MaxZ$ 與 $MinZ$ 分別為深度攝影機可拍攝的最遠 (Z_{far}) 與最近 (Z_{near}) 的值。因此深度圖為一張灰階圖，圖中每一像素的值介於 0~255 之間，分別代表此像素點所記錄的距離，其中越小(暗)表示離攝影機越遠，越大(亮)表示離攝影機越近。而深度攝影機拍攝前須與 2D 攝影機進行校正，以確保兩者所拍攝到的影像是對應的。圖3.1畫出了兩種不同的立體影像格式，其中圖3.1(a)使用的是將兩張 2D 彩色影像對在水平方向上分別擺放的型式 (side by side)。由於目前立體顯示器使用的格式尚未有共通的標準，因此除了 side by side 的儲存型式外，尚有上/下 (Above/below) 以及交錯 (Interlace) 等方式，另外日本 CIPA 組織 (Camera & Imaging Products Association) [4] 也已釋出了一個用於記錄多視域影像與視訊的標準 – MP format (Multi Page format)，其中即是記錄多組彩色影像/視訊的方式，並記錄拍攝時的幾何資訊等，不僅定義了立體影像/視訊格式，也定義了多視角影像/視訊的格式。

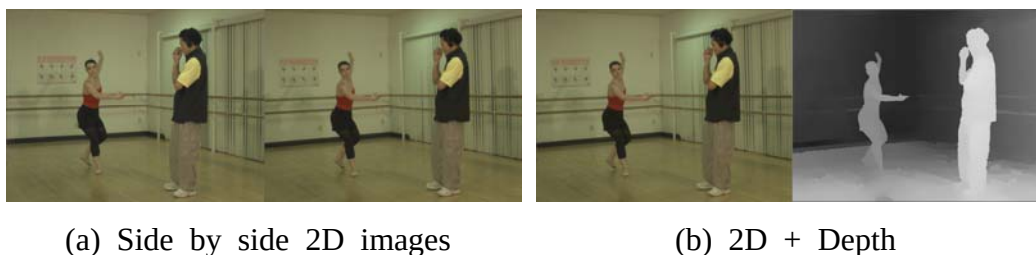


圖3.1 3D立體影像格式

3.3 深度估測技術需求

承上面所述，由於立體影像拍攝在拍攝上成本較高，而現今所使用的視訊多是 2D 視訊，因此開發一套 2D 至 3D 轉換技術是一個最實用的解決方式，其中轉換技術的核心即為估測出 2D 影像對應的深度圖，深度圖在經由轉換之後即可得到視差資訊 (立體影像最主要的關鍵在視差資訊)。

在人類視覺系統中，感知深度訊息之線索很多，例如雙眼視差、動態視差等，除了雙眼感知到的深度線索之外，從單眼也能夠感知到深度，我們稱之為單眼深度線索 (Monocular depth cue)，主要是受心理因素與後天學習而得。一般來說，深度感知的心理因素較生理因素還多且更複雜，文獻 [5] 針對影響深度感知的因素進行實驗分析，其結果如圖3.2 所示。根據圖3.2 的結果我們得知，人眼在距離景物很遠的情況下，心理因素依然有其影響力，甚至超越生理因素。

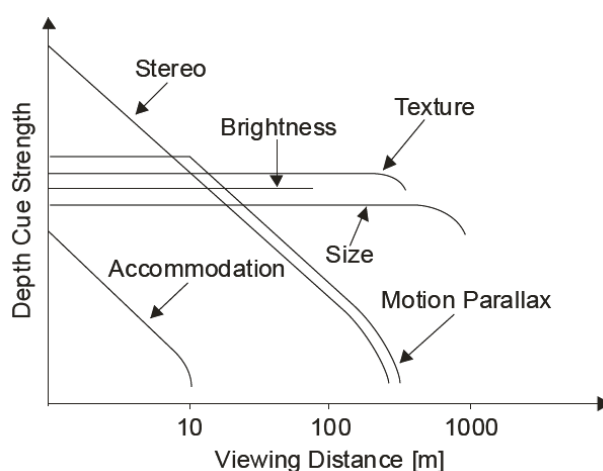


圖3.2 影響深度感知之因素關聯圖 [5]

影響深度感知的因素中又以動態視差 (motion parallax) 為最，動態視差的成因主要是由於物體的移動量與其距離觀察者的距離關係成正比，舉例來說，以同樣速度行走的兩部車子，在觀察者看來，距離較遠的車子其移動較距離近的慢，

因此觀察者可以藉由此一現象判別出兩部車的相對距離，在文獻 [6] 中也指出，動態視差對於一般人而言是一個很重要的深度線索，人們就算只用單眼也可以很容易的經由動態視差來感知深度。除了動態視差之外，其他重要的深度感知線索如線性透視 (Linear perspective，平行的鐵軌看起來在遠處相交)、大氣透視 (Atmospheric perspective，因空氣中微小粒子的影響使得遠處的物體看起來較模糊)、紋理梯度 (texture gradient，相同的紋理的物體，在遠處的物體紋理看起來較密集)、相對高度 (Elevation，物體的高度較高的感覺起來較遠)、重疊 [7] (Overlapping，經由物體間相互遮蔽的關係可以推論出其相對位置) 以及相對大小 (同樣的物體，在遠處的物體看起來較小，近處的物體看起來較大)。表3.2 列出了數個常見的影像特徵分析與深度線索之關係。

表3.2 深度線索定量分析之關係表

深度線索	線索之特徵	影像特徵之分析
移動視差	移動量愈大的物體，深度越深	可由移動向量之大小計算而得
大氣透視	越清晰的物體，深度越深	可藉由計算區域對比度與變異性而得
線性透視	消失點的位置其深度最深，此外的深度成梯度變化	可使用 Hough transform 找出消失點
紋理梯度	區域紋理梯度變化跟深度變化相關	可計算紋理能量大小而得
高度	物體垂直高度跟深度相關	物體影像垂直高度計算

透過計算各種不同的單眼深度線索，我們能夠得到場景深度圖，此即為目前最常用的 2D 至 3D 轉換技術原理，而求得的深度圖再搭配原本的 2D 影像/視訊，經由以下所述的立體影像合成技術即能產生立體影像/視訊。

3.4 立體影像合成技術

立體影像合成技術可以分為兩大類，影像繪圖法 (Image based rendering, IBR) 以及深度影像繪圖法 (Depth image based rendering, DIBR)，其中影像式繪圖法不需使用深度資訊，但由於其使用上的限制 [8] 因此較少被使用，目前常應用於 2D 至 3D 轉換的影像合成技術為深度影像繪圖法。在文獻 [9] 中提到的深度影像繪圖法是以一般立體系統為對象所衍生出來的立體影像合成方法，其主要流程包括

了三個部分：(1) 前處理(pre-processing)，(2) 3D image warping 以及(3) 空洞填補(Hole filling)。前處理的用途主要是將深度圖經由濾波器(smooth filter等) 處理之後，將雜訊消除，並決定 zero plane (左、右眼影像的視差為 0 的位置) 的位置，一般常見的 zero plane 多數設定為深度值 128 或 255 之處。圖3.3 為一個平行擺設的攝影機架攝與立體影像合成之關係圖，圖中 C_l 、 C_r 各別為左眼與右眼攝影機的光學中心 (optical center)， C_c 則為我們所拍攝的攝影機光學中心。

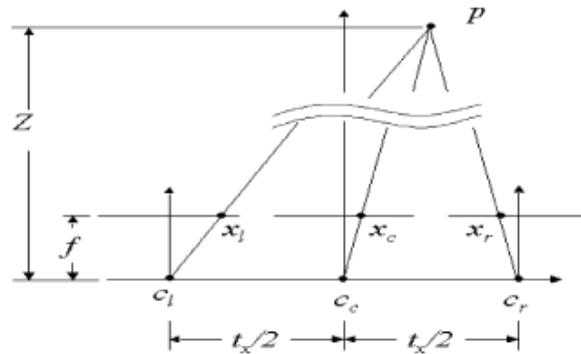


圖3.3 攝影機架設與立體影像合成示意圖 [9]

將 C_c 攝影機的影像作為左、右眼的中間影像，並計算往左與往右的水平位移後，便能合成出左、右攝影機所拍攝的影像，其計算公式如下式 (3.2)、(3.3) 所示：

$$x_l = x_c + \frac{t_x}{2} \frac{f}{Z}, \quad (3.2)$$

$$x_r = x_c - \frac{t_x}{2} \frac{f}{Z}, \quad (3.3)$$

其中， x_l 、 x_r 為欲合成出的像素在左、右眼影像中的位置； x_c 為在中間影像中的座標，即輸入影像之位置， f 為焦距，實驗中設為固定常數； t_x 為 baseline 之長度； Z 為 p 點深度值，其數值與深度灰階值相反 (即 Z 愈小，深度灰階值愈大)。由公式可知，深度愈近，視差愈大，輸入影像上座標位置將右移產生左影像，左移產生右影像。

四、現有深度估測技術分析

一般的 2D 至 3D 轉換流程如圖 4.1 所示，主要是藉由單一 2D 彩色影像/視訊，透過各種不同的深度估測方式求得深度圖，再經由上述的立體影像合成法產生立體影像/視訊。目前的深度估測方式主要可分為三類：(1) 人工指派、(2) 半

自動深度估測法，以及(3)全自動深度估測法。其中人工深度指派方式是由使用者依據數個主觀的 3D 線索以及拍攝技巧的結構繪製規則，針對欲進行處理之影像序列逐張進行畫面的深度分析與評估，再配合繪圖工具 (如PhotoShop) 進行深度序列之繪製，此種方式能夠提供最佳的深度品質，但也是三種類型中最耗成本的方式，因此許多文獻紛紛探討全自動的深度估測演算法。目前所見的自動深度估測探討文獻中，多是計算影像/視訊中所能得到的深度線索以求得深度圖，以下就數種不同的深度估測演算法進行介紹。

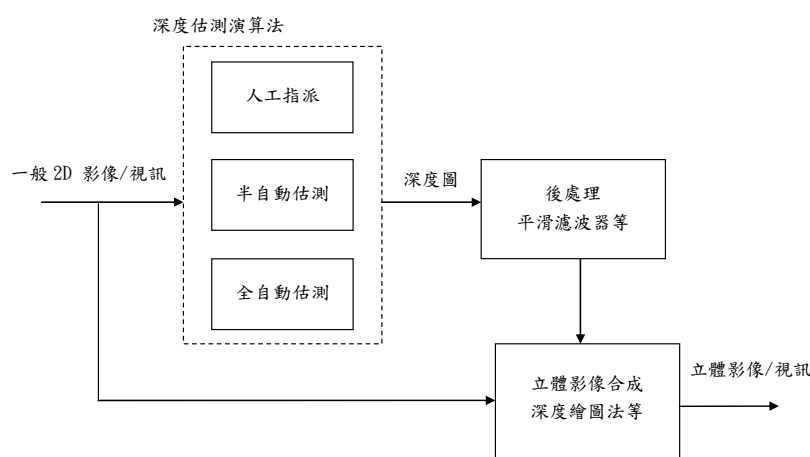


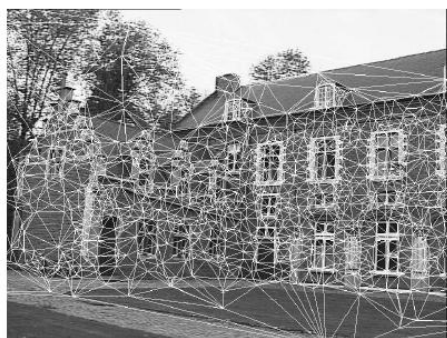
圖4.1 深度估測流程

4.1 SfM 法 (structure from motion)

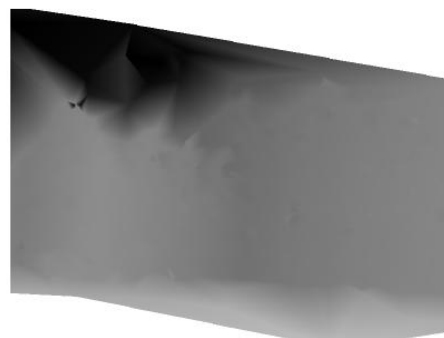
SfM 方法主要利用了影像中物體的移動量、相機運動以及物體在三維空間中運動等物理關係。SfM 最大的好處在於可以極佳的模擬針孔相機模型 (pin-hole camera model) 與 epipolar geometry 等物理現象。然而 SfM 不足的地方在於針對某些拍攝狀況 (如僅有攝影機旋轉的運動) 無法進行深度之估測。此外, SfM 也無法應用於非靜止畫面且場景中的物體會形變的狀況。但在一般的狀況下 SfM 法可以提供一個精確且不錯的深度圖。

SfM 主要的目的為經由一組影像計算拍攝時的攝影機運動參數以及 3D 場景幾何資訊，影像可經由校正或非校正的攝影機拍攝。SfM 為電腦視覺的熱門研究領域，在文獻 [10] 中提出了一個基於 SfM 的深度估測演算法，其流程主要可以分為三個部分：(1)特徵追蹤 (feature tracking)、(2)場景幾何重建(motion and structure recovery) 與 (3)深度估測(dense depth map creation using geometry fitting)。由於 SfM 是利用連續的幾張畫面估測場景的運動以及攝影機的運動等

幾何關係，這些關係可以透過計算各畫面中的特徵（特徵點、線、表面資訊等）之間的相關性，而特徵擷取與特徵追蹤的演算法已有許多的文獻可以參考，計算出畫面間特徵的相關性後，便可應用於計算攝影機之移動以及場景的結構，最後再經由三角化 (Delaunay triangulation) 演算法計算出深度圖，如圖 4.2 所示。



(a)三角化的結果



(b)場景深度圖

圖4.2 文獻 [10] 所使用的 SfM 深度估測結果

SfM 的演算法如前所述，受限於某些拍攝時的場景，因此無法應用於所有的情況，由其是在攝影機不動的狀況之下，可能因場景的資訊不足而無法計算深度，且特徵追蹤的運算複雜度高，對於即時轉換是一大挑戰，因此短期內較難實際商品化。

4.2 DfC法 (Depth from cues)

DfC 深度估測法主要是藉由計算單一通道視訊中的深度線索為主，其中最常被使用的深度線索為畫面的移動量，在 3.3 節中我們也提到人眼深度感知中以畫面移動量最為重要，在文獻 [11] 中即提出一個以畫面中物體的移動量為深度估測的基準，其中物體移動量的計算使用的是 MPEG 4 編碼標準所使用的移動向量搜尋法 (motion estimation)，其求得的移動向量 (motion vector, MV)，作者將應用定為一般經 MPEG 4 壓縮過的視訊資料，由於其 MV 已經求得，在 [11] 中直接以每一個編碼區塊的 MV 強度作為該區塊之深度值。

但是單純使用移動量估測難以適用於各種不同的視訊類型，有研究提出使用前景切割的方式，將前景物體與背景物體分離，再各別進行前景物體以及背景物體的深度值估測，在文獻 [12] 中提出了以影像切割為基礎，並配合消失點產生的梯度平面，來估測出深度圖的方式。在文獻 [12] 中，作者使用影像中的色彩資訊，以 mean shift 演算法 [13] 將影像中色彩相似的區域進行群組，得到不同

的影像區域切割，如圖4.3 (a) 與 (b) 所示，之後分析各區域的色彩資訊 (R、G、B 資訊)，將各切割區域分類為天空、遠處的山、近處的山、土地與其他區域等，並依據區域分配的結果進行深度值的指派，其結果如圖 4.3(c) 所示。

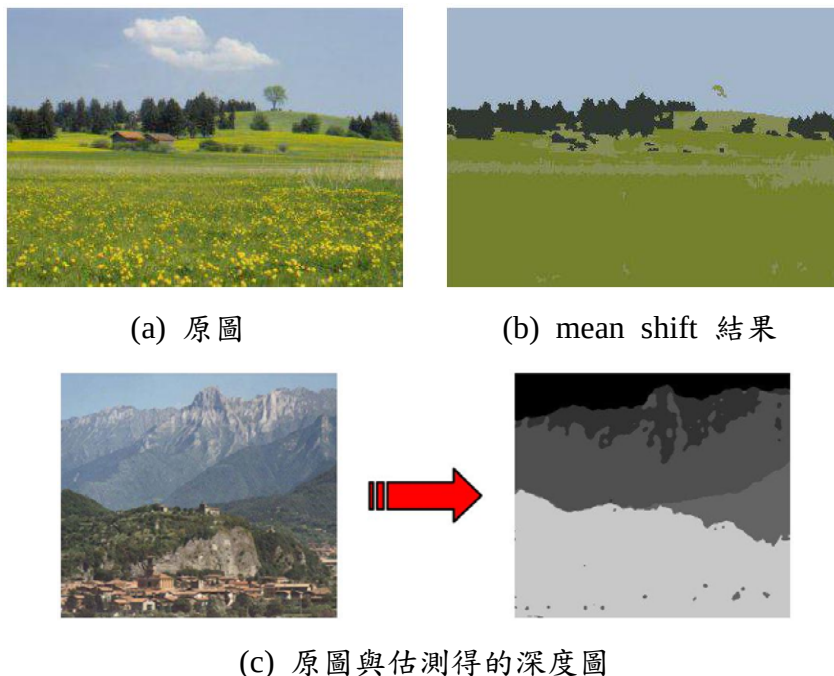


圖4.3 使用色彩資訊的影像切割與深度估測結果 [12]

除了針對區域切割的結果進行深度值指派外，文獻 [12] 中也使用了線性透視的深度線索，藉由計算影像中的消失點進行整體場景的深度指派，如圖4.4 所示，經由結合區域深度值與場景深度值後，得到最後的深度圖。而在文獻 [14] 中也使用了類似的概念，作者事先定義了數個不同的場景深度類型，利用計算影像中的特徵選取對應的深度類型，最後經由 bilateral filter 濾波後得到最後的深度圖，文獻 [15] 則是使用了影像中的邊緣資訊 (edge information)，將影像切割成許多個 4x4 的小區塊再進行區塊結合，最後搭配不同梯度方向的場景深度漸層，產生深度圖。

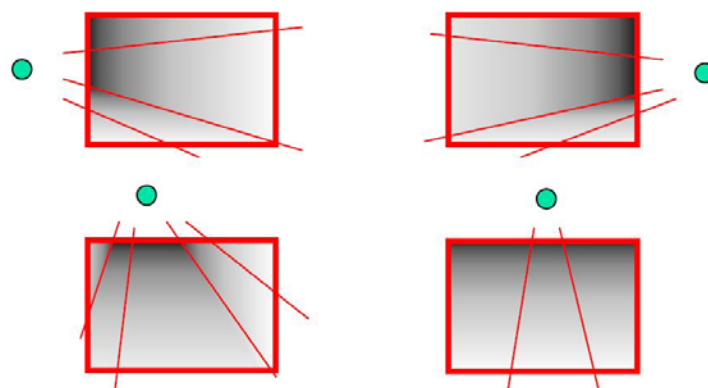


圖4.4 各種消失點與其深度梯度圖指派

DfC 的方式相較於 SfM 的方式較不受影像內容的限制，不過由於影像特徵與深度的關係並不是絕對的，因此容易有誤判的情形，加上各種深度線索的計算只能提供相對的深度資訊，且各深度線所之間該如何結合也是一個問題，另外由於深度估測的複雜度受影像特徵計算所影響，使用較多的深度線索有助於更準確的深度估測，但其運算複雜度也高，因此需在之間取得一個平衡。

4.3 MLA 法 (Machine Learning Algorithm)

在文獻 [16] 中提出使用 MLA 的方式進行深度估測，由於前述的 SfM 法受限於攝影機與物體的移動，而 DfC 法在深度線索的計算上較費時且容易有誤判的狀況，因此文獻 [16] 使用數個簡單的影像特徵，以 MLA 演算法將影像特徵與深度圖進行連結，找出影像特徵與場景深度的關係，用於深度值指派。

圖4.5(a) 為文獻 [16] 所使用的深度估測演算法流程，使用的影像特徵為影像中像任一素點的座標 (x, y) 以及其色彩資訊 (r, g, b) ，而其對應的深度圖則假設已知（真實拍攝或是使用人工指定）。藉由挑選數張 key frame 進行訓練之後可以得到分類器，如圖4.5(b)，之後此分類器可以應用於其他畫面的深度指派。MLA 的方式相較於前面所述的方式，其使用上不受視訊類型的限制，另外由於可事先進行特徵訓練，因此可用於即時轉換，不過其難處如同 DfC 的方式，影像特徵的挑對估測的深度準確性影響極大。

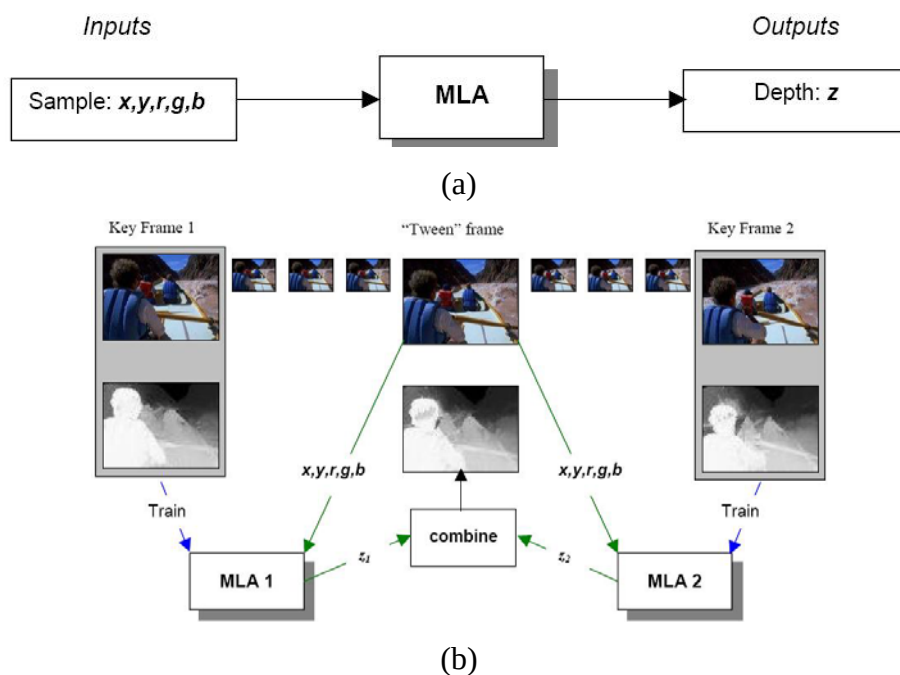


圖4.5 MLA 深度估測演算法 [16]

4.4 市場已有產品使用分析

目前市面上已有廠商推出即時 2D 轉 3D 軟體，使用者只需輸入一般 2D 影像/視訊即可轉換為 3D 影像/視訊，其中最常見的即為 DDD (Dynamic Digital Depth) 公司 [17] 推出的 TriDef 3-D Experience。另外訊連科技 (Cyberlink) [18] 也在 CES 2010 (Consumer Electronics Show) 中展出了 3D 版的 PowerDVD 10。TriDef 支援各種的壓縮格式，包括了影像 (如 JPEG、BMP、PNG 以及 TIFF 格式等) 以及視訊 (MPEG、WMV 等格式)，而 PowerDVD 10 則是支援 DVD 視訊格式的 3D 轉換，兩款 3D 轉換軟體皆支援即時 (real-time) 轉換，其中 TriDef 支援數種型式的 3D 顯示裝置 (如 SHARP LL-151-3D、Acer 5738DG 等)，而 PowerDVD 10 已知可支援垂直交錯式的 3D 顯示器 (Acer 5738DG)。

就 3D 轉換效果實際看起來，兩套軟體確實能讓畫面感覺出深度的變化，但是深度感覺起來較沒有“層次”。以 TriDef 來說，雖然對於各種場景皆能適用，但是使用者皆會感覺到深度是固定地由畫面底端向畫面上方“斜躺”的變化型式，亦即不管影像的內容為何，使用者僅能感覺出越靠近畫面底部離使用者越近，反之在畫面上方的物體則遠離使用者，雖然在大部份的情況下這種深度感是可以接受，但是某些場景下會讓使用者感覺到深度的不自然。例如畫面中的物體從底部一直往上延伸的狀況 (像是樹，柱子等)，則使用者會感覺物體是斜躺著；而

PowerDVD 10 在深度轉換的效果上頗類似 TriDef，不過其深度感是讓使用者覺得畫面由左向右“斜躺”，同樣的在某些狀況之下，使用者很容易感覺出錯誤的深度感。

縱觀目前市面上所發售的 2D 轉 3D 軟體，雖然可以進行即時轉換，但其深度效果差強人意，不過由於市面上的競爭對手少，因而形成少數廠商寡佔的市場，因此未來 2D 至 3D 視訊轉換技術的重點除了能夠支援即時轉換外，最重要的還是能夠讓使用者有好的深度感，且不易凸垂。

五、其他可能後續技術發展

由於 3D 拍攝技術不斷的進步，未來的發展趨勢勢必改為 3D 直接拍攝方式，因此 2D 至 3D 轉換技術最主要的應用還是在轉換傳統拍攝的影片(如早期拍攝的電影、KTV影帶等)，而目前的全自動深度估測技術仍難以估測出完美的深度資訊。但在家用市場上，消費者所希望看到的仍是能提供良好深度感的立體影片。基於追求良好立體品質與轉換效率上的考量，未來最可能的發展方式為結合人工轉換的半自動方式，在文獻 [19] 中即提出一個半自動轉換技術，文中提出了深度漫延 (Depth Propagation, DP) 的概念，假設視訊中的幾張 key frame 使用人工的方式進行深度圖的繪製，其他畫面的深度圖則由 key frame 漫延而得，由於此種方式只需繪製 key frame 的深度圖，而 non-key frame 的深度圖在漫延後也可由人工修改以確保其深度圖的品質。

半自動轉換技術的重點在於如何達到最大的減少比率(視訊的整體畫面數/key frame數)，以及 key frame 的選取規則，減少比率大則人工繪製的 key frame 數減少，可以減少更多的人工成本，而這也與繪製 key frame 的深度品質相關。另外，如何選擇 key frame 的數量也是影響半自動轉換效能的一個重要因素。由於一般視訊多是由數個不同的場景組成，而場景中又可能有攝影機移動、畫面中物體的運動或是影片的特效等因素，因此 key frame 的選擇決定了 DP 的效果。

除了深度估測技術之外，改良式的深度繪圖法也是目前研究的重點之一，由於深度影像不僅能用於立體影像之合成，也可以用於合成多視域影像 (multi-view image)，而合成的影像會因遮蔽 (occlusion) 效應的影響而產生許多小洞，因此有些研究使用不同的方式對合成影像進行後處理 [20]，以達到更好的影像品質。

六、結論

隨著近幾年 3D 電影的熱潮，加上現今立體顯示器的發展逐漸成熟，我們可以預見 3DTV 的時代真的來臨。所謂萬事俱備，只欠東風，現今 3DTV 市場最欠缺的仍是 3D 內容，而 2D 至 3D 轉換技術正是關鍵。由於目前已有的全自動 2D 轉 3D 軟體效能仍有改進之處，未來要追求高品質的深度感勢必走向半自動的轉換方式，未來半自動轉換技術開發的重點應該在如何能夠達到最有效率的轉換（最大的減少比率），以及 key frame 的選擇。對於使用者而言，深度感還是最主要的考量因素，好的深度感才能帶動消費者的購買慾，因此 3D 內容供應商所追求的主要目標將會是更好的深度品質而非轉換速度。

參考文獻

- [1] http://en.wikipedia.org/wiki/John_Logie_Baird
- [2] http://www.eettaiwan.com/ART_8800603673_480702_NT_b2859681.HTM 電子工程專輯
- [3] A. Smolic, K. Mueller, P. Merkle, P. Kauff, T. Wiegand, “An overview of available and emerging 3D video formats and depth enhanced stereo as efficient generic solution,” *IEEE Conf. on Picture Coding Symposium (PCS)*, pp.1-4, 2009
- [4] 日本攝影機映像機器工業會 Camera & Imaging Products Association (CIPA), <http://www.cipa.jp/>
- [5] C. Fehn, P. Kauff, M. Op de Beeck, F. Ernst, W. Ijsselsteijn, M. Pollefeys, L. Vangool, E. Ofek, and I. Sexton, “An evolutionary and optimised approach on 3D-TV,” *Proc. of IBC 2002, Int. Broadcast Convention*, Amsterdam, Netherlands, Sept. 2002.
- [6] Foundations of Cyclopean Perception, Bela Julesz, The MIT Press.
- [7] 陳明民，數位立體影像之理論探討與創作實驗之研究，國立嘉義大學視覺藝術研究所碩士論文，中華民國九十四年十二月。
- [8] T. Okino, H. Murata, K. Taima, T. Iinuma, and K. Oketani, “New television with 2D/3D image conversion technologies,” *Proc. of SPIE*, vol. 2653, pp. 96-103, 1996.
- [9] Liang Zhang and Wa James Tam, “Stereoscopic image generation based on depth images for 3D TV,” *IEEE Trans. on Broadcasting*, Vol.51, Issue 2, pp.191 – 199,

- June, 2005.
- [10] Ping Li, Rene Klein 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.
- [11] Ianir Ideses, Leonid Yaroslavsky, Barak Fishbain , “Depth Map Manipulation for 3D Visualization,” *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video*, pp.337-340, 2008
- [12] 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.
- [13] D. Comaniciu, 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.
- [14] Ludovic J. Angot, Wei-Jia Huang and Kai-Che Liu, “A 2D to 3D video and image conversion technique based on a bilateral filter,” *Proc. of SPIE-IS&T Electronic Imaging*, Vol. 7526, doi:10.1117/12.838571, 2010
- [15] Chao-Chung Cheng; Chung-Te Li; Liang-Gee Chen, “ A 2D-to-3D conversion system using edge information,” *Int’l Conf. on Consumer Electronics (ICCE)*, pp.377-378, 2010.
- [16] P. Harman, J. Flack, S. Fox and M. Dowley, “Rapid 2D to 3D Conversion,” *Proc. SPIE*, Vol. 4660, pp.78-86, 2002.
- [17] <http://www.ddd.com/>
- [18] <http://www.cyberlink.com/>
- [19] C. Varekamp, B. Barenbrug, “Improved depth propagation for 2D to 3D video conversion using key-frames,” *Proc. of 4th European Conference on Visual Media Production (IETCVMP)*, pp.1 – 7, 2007
- [20] Yu-Cheng Fan, Tsung-Chen Chi, “The Novel Non-Hole-Filling Approach of Depth Image Based Rendering,” *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video*, pp.325-328, 2008