

半自動化單視角影像 2D 轉 3D 技術

郭天穎、藍美欣、謝正宏
國立臺北科技大學電機工程系

摘要—隨著 3D 科技技術的發展，3D 立體內容提供人們更真實的視覺饗宴，而目前多數現存的照片皆為單視角 2D 格式，因此為了將其轉換為立體影像格式，本文藉由投影法先將影像分類為特寫影像以及戶外影像，並將戶外影像利用先前工作配置深度，我們將近焦影像透過影像分割並將相機參數以及相機模型結合算出合理數值，再透過 K-mean 分群的群心配置合理的深度，最後可利用介面，可讓使用者藉由分割資訊之協助，手動修正錯誤的區域深度，以得到逼近真實的深度資訊。¹

一、簡介

3D 影像的立體感，主要是因為左右眼接收到各自的影像透過兩眼視差而產生，因此 3D 影像需要透過雙鏡頭相機或攝影機取得左右眼的影像；或者單視角的影像加上深度，透過計算兩眼視差，模擬左右眼影像而得。現今所廣泛使用的影像或視訊仍為傳統的 2D 影像，面對目前龐大的 2D 影像資料庫，若是能開發一套 2D 至 3D 轉換技術，則能解決目前 3D 立體內容不足的問題。然而真實深度資訊需透過測距儀器量測而得，或透過兩個或多個視角的影像，計算兩眼視差而得；而單視角影像的深度估測，是一種不穩定性(ill-posed)問題。

單視角 2D 轉 3D 深度估測法有多種深度線索，包含相對高度、線性透視法(Linear Perspective)、大氣透視度(Atmospheric Perspective)、紋理梯度(Texture Gradient)、聚焦(Focus)。由於單張影像估測是屬於不穩定性問題，因此並不是每種深度線索皆適用所有類型的影像，如聚焦只適合特寫影像，大氣透視度就只適用於戶外影像。所以文獻[1]提出先將影像分類，再根據影像類型採用不同之深度估測法，歸納兩者的分類方式可將影像分為室外影像(outdoor image)[2][3]、室內影像(indoor image)[4]及特寫影像(closed-up image)[2][5]。本論文將提出適用於特寫以及室外影像的深度估測方法。

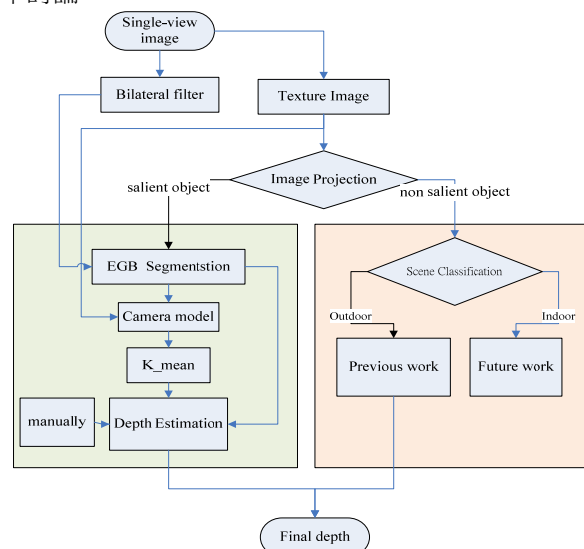
目前多數 2D 轉 3D 之文獻[2]-[5]，雖然能估測不同物件彼此的前後關係，但卻無法提供合理的相對深度。透過統計訓練雖然可以根據訓練資料庫逼近於真實深度資訊，但對於不同類型或場景的影像，會因為深度線索的泛用性不足，導致深度估測錯誤。

因單張影像深度估測技術存在不穩定性問題，以及深度估測法採用錯誤消失點或是紋理偵測，則會導致影像的深度配置錯誤。為了確保深度符合人眼視覺，本文提出深度估測結果後，提供使用者介面讓使用者及時修正深度。

二、本論文提出之深度資訊估測法

本論文之深度估測法架構，如圖一。首先將影像場景分類為特寫影像以及非特寫影像，再將非特寫影像根據天空偵測以及地面偵測將場景分類為室內影像以及室外影像，根據場景分類結果採用不同的深度估測法。

一般影像在拍攝時，室外影像以及室內影像皆希望所有距離景物都可清晰拍下，亦即希望是景深是無限遠，特寫影像希望著重在物件上而背景則模糊，亦即淺景深。利用上述之特性，提出場景分類法。根據分類結果採用不同的深度估測法，特寫影像使用本文提出之演算法，室外影像則使用先前工作[6]的演算法，室內影像則因為場景無特定之線索可估測深度，故留在未來的工作在目前不討論。



圖一：深度估測法架構圖

2.1 影像場景分類

單視角影像深度估測法有多種深度線索可以使用，但不同深度線索適用的場景也不同，如聚焦就只適用於特寫影像。所以本文將輸入影像通過影像場景分類，將場景分為三類；特寫影像、室外影像以及室內影像三種。首先是將影像分類為特寫以及非特寫影像，並將非特寫影像分為室外影像以及室內影像，如圖二所示三種影像場景類型。

本研究由國科會贊助，計畫編號 NSC 101-2221-E-027-078-MY2。

Corresponding author: tykuo@image.ee.ntut.edu.tw

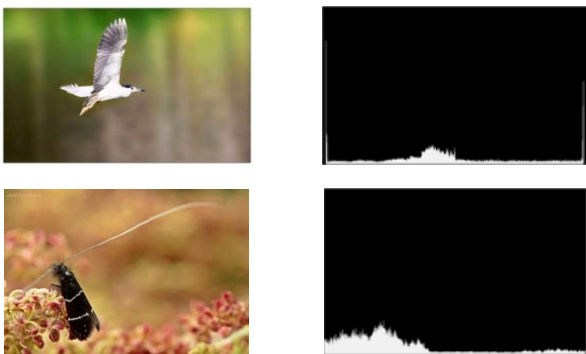
DOI:10.6159/IJSE.2014.(4-2).52



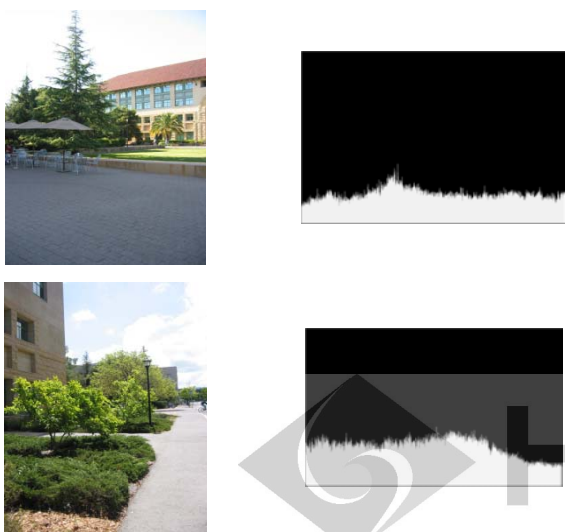
圖二：一般常見影像場景

利用紋理圖透過垂直投影法判斷其峰值與其 DC 值關係為場景分類法。紋理圖取得方法有許多，如 sobel 邊緣偵測、canny 邊緣偵測等。但影像背景的輪廓雖因模糊而不強烈但因顏色的差異等還是會影響其邊緣偵測結果，canny 只會留下強烈的包含輪廓的邊緣，利用原圖減到雙向濾波器的結果可以得到輪廓以及雜訊影響的較細部的紋理，藉由此才可觀察出影像真正因拍攝時著重物件的紋理，為了方便觀察紋理圖都提高對比顯示。

提出利用雙向濾波器之原理將原圖與通過雙向濾波器之結果相減，可得一張不受強烈紋理影響的紋理圖，在透過投影法統計影像每一行的紋理強度以及計算其平均值(DC 值)，如圖三以及圖四所示，利用特寫以及室外影像的投影結果，將紋理值排序之後統計濾掉最高 1% 排除圖框的影響，並統計最高 2% 的平均值，將平均峰值與 DC 值相減之後，結果大於 0.7 倍的 DC 值則判斷為特寫影像，反之則判斷為非特寫影像。



圖三：垂直投影結果(特寫影像)



圖四：垂直投影結果(室外影像)

2.2 特寫影像深度估測法

特寫影像深度估測法架構，首先利用到雙向濾波器的特性，保留物件強烈的邊緣並模糊其他細微紋理，如呈現卡通化現象，本論文就是利用此特性，將原始影像與通過雙向濾波器再將原圖與其結果相減，可得到留下細微紋理，並透過影像分割、相機參數算出數值(往後稱 q 值)，並以 K-mean 分群將數值分為若干群；再利用分群之後的群心(Cluster Center)以及該張影像的 q 值的最大值以及最小值的相對位置配置合理相對深度，最後，透過手動修正區域錯誤深度以獲得最終估測深度。

2.3 雙向濾波器以及影像分割

在影像分割之前，為防止物件紋理或雜訊影響，導致物件分割錯誤，通常會增加一個平均濾波器，但過度平滑會導致分割不足之現象，所以使用雙向濾波器，可有效移除物件中的細節，並可保留物件之輪廓。

為了計算利用相機模型所需要的紋理影像，且該影像需要不受輪廓影響的細部紋理，所以利用雙向濾波器的特性將原圖以及雙向濾波相減後的圖相減得到一張紋理影像。

執行影像分割的目的，主要是希望將各物件分為不同的區塊，採用分割方法[7]，因為此方法比採用的平均值移動法(Mean Shift)來的快速，本論文採用分割方法[7]，經實驗測試後分割參數 k 為 600，並將小於影像大小的分割區塊，將其與周圍外部差異最小的區塊合併，利用影像分割法將影像分為若干區塊以利後續以區塊為一單位計算 q 值，圖五為影像分割結果。

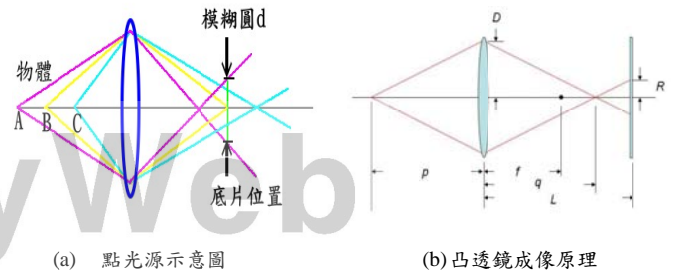


(a) 原始影像 (b) 影像分割結果

圖五：影像分割結果

2.4 相機模型

透過凸透鏡成像原理可看出當點光源與凸透鏡距離大於焦距時，光源會匯聚成一點，當物體越遠模糊圓的光圈越大，如圖五(a)所示，因此本論文採用文獻[8]提出之模型如圖五(b)，推導出以區塊為單位，利用先前產生之紋理圖，計算每一區塊的 FV 值並套用公式算出 q 值後並記錄。



(a) 點光源示意圖 (b) 凸透鏡成像原理

圖五：相機模型

根據文獻提出之理論，推導後如式(1)，FV 可利用影像紋理梯度求出，D 則利用相片中的 f#(f number)以及焦距 f 求出 D， $f\# = f/D$ 而 L 本論文假設為 $L = f * 1.3$ 。

$$q = \frac{2 * L * FV * D * D - \sqrt{(2 * L * FV * D * D)^2 - 4(FV * D^2 - 1)(FV * D^2 * L)}}{2(FV * D^2 - 1)} \quad (1)$$

利用此式(1)可以計算出每一區塊所代表之 q 值，因每張影像紋理強度不一樣，所以後面利用到分群將 q 值做分群結果以配置該區塊深度。

2.5 K-mean 分群法

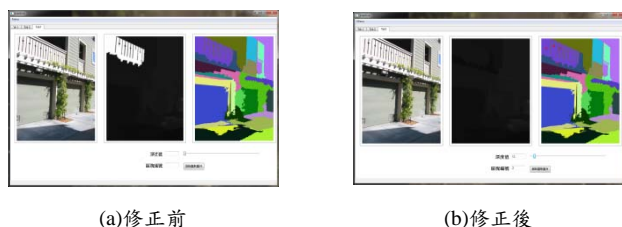
對於深度影像，通常用 0 到 255 之間的值表示，因此欲將代表該區塊的深度值合理的配置，需透過分群將先算出的 q 值，分為指定群聚的數目。

本論文執行 K-mean 分群的目的是將利用這些數值，找出多個具有代表性的資料點，這些資料點可以稱為是群心(Cluster Center)，利用群心數值來代表相對深度值，降低資訊量以及計算量，並可避免雜訊之影響。

在使用 K-mean 分群法(K-means Clustering)之前，需先指定群聚數目 k，經實驗結果本論文將群聚數目設為 7，然後找尋下列公式的極大值，以達到分群的最佳化之目的，其中是群體的平均。

2.6 手動修正介面

對於全自動深度估測結果，因為深度線索不一定能完全適用於所有影像，所以本論文提供 QT 撰寫的介面，利用初始深度以及分割結果圖，利用介面將配置錯誤之深度值使用人工修正。如圖六所示，左為原圖，中為初始深度，右為分割結果，利用滑鼠點到分割結果上，可對應其座標取得該深度值以及標籤編號，並藉由使用者輸入新的深度值，以修正錯誤深度，並重新顯示於深度結果，直至結束。如圖六(a)範例將錯誤估測區塊修正為圖六(b)。



圖六：手動修正介面

三、實驗結果

本論文提出之方法主要採用 C/C++ 語言，並配合 OpenCV2.0 函式庫編寫而成；而使用的硬體設備為 Intel(R) Core(TM)2 Quad CPU Q8200 @2.33GHz 2.34G, 3.50GB RAM，本論文室內外影像結果比較資料庫為史丹佛大學的 Saxena[9]所提供，每一組資料包含原始影像與 3D 雷射掃描器所測得之真實深度資訊(Ground Truth)，此影像資料庫也是目前唯一提供真實深度資料庫，在此使用其提供 400 張測是影像。測試之影像格式。特寫影像因目前資料庫沒有提供 Ground Truth，此測試僅此在網路

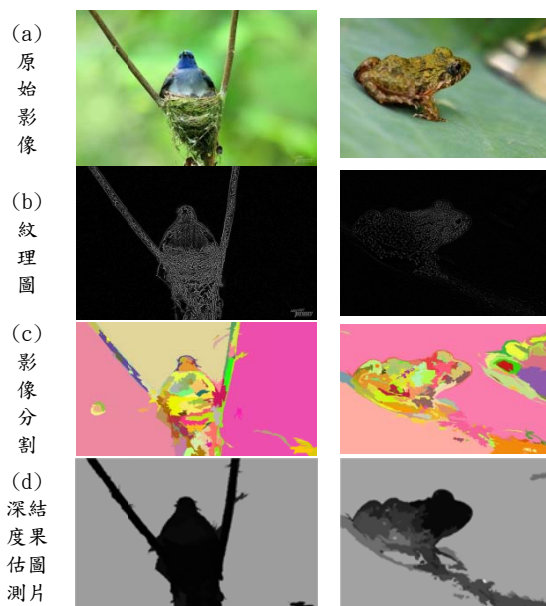
下載圖片以及日常生活照提供共 100 張測視圖。

由於本論文將輸入影像依照輸入影像之場景判斷該使用何種演算法估測深度，所以場景分類的結果非常的重要，因對不同的場景適用的深度線索不同，深度結果也會不同。本文利用輸入影像與通過雙向濾波器之結果相減後的紋理圖，透過投影法統計影像每一行的紋理強度以及計算其平均值(DC 值)，並利用數學式判斷，判斷影像場景為特寫影像或非特寫影像，測試結果如表 I。

表 I
影像測試結果

類型	精確率	召回率	F_measure
非特寫影像	98%	94%	95%
特寫影像	74%	94%	82%

為了能清楚辨別影像細節，紋理圖皆需經過亮度強化。由圖七可看到特寫影像深度估測流程結果圖，包含紋理圖、影像分割結果、以及估測深度之結果。

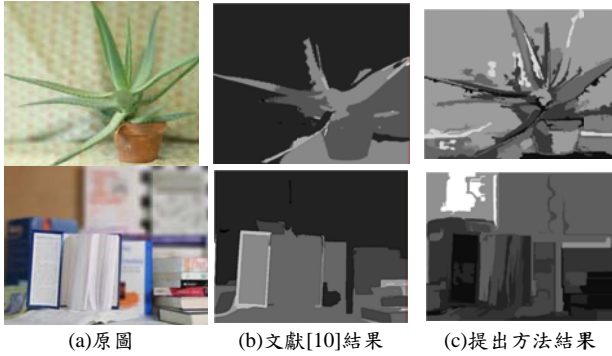


圖七：提出方法之結果

經由結果發現，當影像具有前景深以及聚焦點在於距離相機最前方時，都會有不錯的深度配置結果，由圖七右邊序列可看出，雖然青蛙之屁股理應離相機最近，但因為聚焦在青蛙的頭部，所以會導致因為紋理的走向而導致深度配置與人眼視覺觀差不同。但最後都可利用手動修正即時將錯誤深度修正正確。

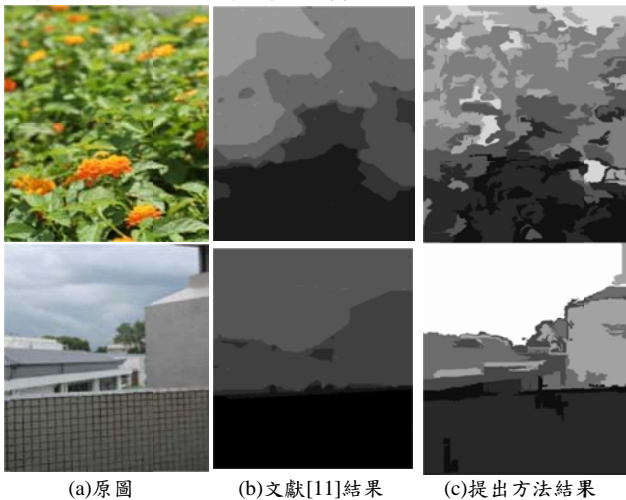
為了評論提出之自動的單張影像深度估測結果的正確性，將提出方法結果與文獻[10]-[12]之深度估測結果比較，本論文提出之演算法深度值是介於 0~255，為配合與先前工作結合以及統一性，本文使用的灰階值是 0~255 對應最近以及最遠。如圖八為文獻[10]比較，經由比較發現本論文提出之特寫影像深度估測法，在文獻[10]的比較發現，花盆上的葉子後面會消失不見，而導致估測錯誤，而本論文因取不到原圖，且 JPEG 壓縮後

Ringing Effect 影響，導致雖分割過細，但可以藉由介面修正錯誤之區塊。



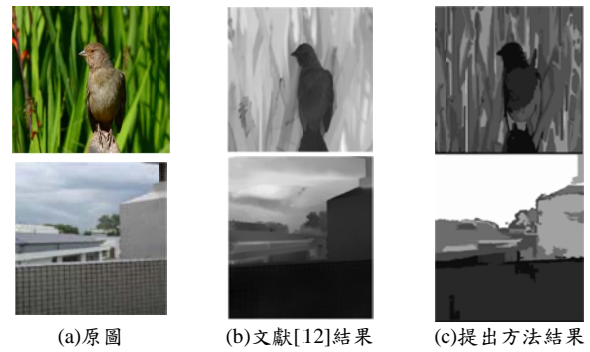
圖八：文獻[10]與提出方法之比較

如圖九為文獻[11]比較，在文獻[11]的比較發現，首先在聚焦程度相比雖說文獻的距離差逐漸層較明顯，但因切割深度太大反而造成落差，而本文提出方法較能看出緩慢的變化較無跳動，在牆壁那張圖片，雖然在本論文提出之演算法，其影像是分類為室外影像，但用特寫影像的深度估測法，雖無法取得原圖，且壓縮過後會有方塊效應，雖無法公平比較，但仍然可得到不錯的效果，且不需經過複雜的數學式子運算。



圖九：文獻[11]與提出方法之比較

如圖十為文獻[12]比較，而在文獻[12]的比較發現，雖因本論文取不到原圖，且 JPEG 壓縮後 Ringing Effect 影響，所以效果不佳，而無法公平比較。但在鳥類那張影像，提出方法雖然在鳥類軀幹的部分算出之深度值較差，但因其複雜運算式複雜度較高，本文提出方法複雜度較低且可以透過手動修正部分將其誤判區域修正。在磁磚那張圖片，在牆壁那張圖片，雖然在提出之演算法，其影像是分類為室外影像，但使用特寫影像的深度估測法，雖說文獻[12]的深度較為平滑，但還是有些區域不符合人眼視覺，而本文不需經過複雜的數學式子運算且可以透過手動修正即時將深度修正符合人眼視覺。



圖十：文獻[12]與提出方法之比較

結論

目前多數文獻對於深度估測使用的深度線索通常只適用於某一種場景類型，而本論文透過快速以及不需訓練的場景分類法將影像分類，並針對場景類型的特性，使用不同的演算法，提出之演算法只適用於室外影像以及特寫影像。但不需要經過訓練、或高複雜度的演算法，即可估測出與真實深度資訊相近的深度，同時可提供高品質之立體影像。本文將相機模型提出深度配置方法，解決目前多數文獻無法提供合理相對深度之問題。透過場景分類，依照場景使用不同演算法，可使深度資訊更加的正確。最後透過介面，可將錯誤深度之區塊，將深度重新配置，讓深度資訊更符合真實深度。

參考文獻

- [1] J. Lee, S. Yoo, C. Kim, and B. Vasudev, "Estimating Scene-Oriented Pseudo Depth With Pictorial Depth Cues," *IEEE Transactions on Broadcasting*, Vol. 59, No. 2, pp. 238-250, June 2013.
- [2] Yea-Shuan Huang, Fang-Hsuan Cheng and Yun-Hui Liang, "Creating Depth Map from 2D Scene Classification," 3rd Int'l Conf. Innovative Computing Information and Control, pp.69, 2008.
- [3] S. Battiatto, S. Curtib, M. L. Casciac, M. Tortorac, and E. Scordato, "Depth-map Generation by Image Classification," *Proc. SPIE on Three-Dimensional Image Capture and Applications*, vol. 5302, 95, April 2004.
- [4] Xiao-Ling Deng, Xiao-Hua Jiang, Qing-Guo Liu, and Wei-Xing Wang, "Automatic Depth Map Estimation of Monocular Indoor Environments," *Int. Conf. on MultiMedia and Information Technology*, pp. 646-649, 2008.
- [5] Jaeseung Ko, Manbae Kim, and Changick Kim, "2D to 3D stereoscopic conversion: depth-map estimation in a 2D single-view image," *Proc. SPIE, the International Society for Optical Engineering*, vol. 6696, 2007.
- [6] Tien-Ying Kuo and Yi-Chung Lo, "2D-to-3D conversion for single-view image based on camera projection model and dark channel model," *IEEE Conf. on ICASSP*, pp.1433-1436, 2012.
- [7] P. Felzenszwalb and D. Huttenlocher, "Efficient Graph-Based Image Segmentation," *Int'l J. Computer Vision*, vol. 59, no. 2, pp. 167-181, 2004.
- [8] L. C. Chiu and C. S. Fuh, "An efficient auto focus method for digital still camera based on focus value curve prediction model," *J. Inf. Sci. Eng.* 26, 1261-1272 (2010).
- [9] "Make3D," Available: <http://make3d.cs.cornell.edu/code.html>
- [10] P. Chan, B.-Z. Jing, W. Ng, and D. Yeung, "Depth estimation from a single image using defocus cues," in *Proc. Int. Conf. Mach. Learning Cybern.*, vol. 4, Jul. 2011, pp. 1732-1738.
- [11] S. Zhuo and T. Sim, "On the Recovery of Depth from a Single Defocused Image," *Proceedings of the International Conference on Computer Analysis of Images and Patterns*, pp. 889-897, 2009.
- [12] S. Zhuo and T. Sim, "Defocus map estimation from a single image," *Pattern Recognition*, Vol. 44, No. 9, pp. 1852-1858, September 2011.