

# 以 FPGA 實現三維深度影像估算的低複雜性方法

姚長昆\*、陳育崧、楊志弘  
財團法人車輛研究測試中心

**摘要** — 本文於車用影像安全領域使用立體視覺技術，以獲得圖像的深度信息，並從二維圖像中重建道路環境的三維信息，偵測出車輛前方障礙物的 3D 位置，估算出安全可行駛空間，有效確保駕駛安全。在重建道路環境過程中，真實物體深度影像的估計是立體視覺的關鍵技術。本系統以半全域匹配演算法 (Semi-Global Matching, SGM) 為基礎，結合多個優化演算方法，提出一三維深度影像估算演算流程，並實現於 Xilinx Spartan-6 FPGA 硬體平台上；採用外部記憶體 SDRAM 來儲存影像資料，達到低成本的考量。本文演算法實現結果，可利用 FPGA 擷取 NTSC 影像輸入，實現之深度資訊估算僅使用 10K 的 LUTs 和 102KBk 的 BRAM，具有低複雜性特點。最後利用 Middlebury Tsukuba 標準測試圖與實車影像做測試驗證本文系統的準確性。

## 一、前言

由於車輛行車安全逐漸被重視，在過去的十幾年裡，各先進車廠積極投資開發主動安全系統，以利用感測器來獲得周圍環境的安全行駛道路。而車用影像安全系統是利用攝影機取得環境影像資訊，再透過 MCU、DSP、CPU 或是 FPGA 分析路面狀況及車輛四周環境，以達到即時判斷行車安全與否。目前單視覺影像系統已廣泛配備於車輛上進行前方道路偵測，如：應用於高速行駛環境的車道偏移暨前方防撞安全系統 (LDW & FCW)，但隨著需求與應用環境複雜度的提升，二維平面影像已漸漸不敷使用，故為了獲得更多的影像資訊，藉由雙攝影機所產生的三維空間深度資訊已成為研究的重點，且依據 Frost & Sullivan 的產業報告指出，立體視覺將成為未來視覺技術的發展主軸。

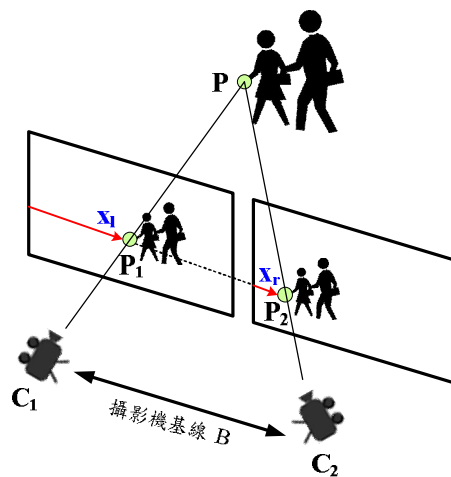
立體視覺是透過兩部不同視點的攝影機進行取像，再由所得有視差的兩個影像計算出深度影像資訊，即可由二維影像重建三維空間資訊。而深度影像的估算技術，為立體視覺準確與否的關鍵技術。該技術是把從同一環境、同一水平，但不同位置的攝影機所得的兩張影像，找出兩張影像中每個像素點之間的位置差，稱為像差 (disparity)，最後再透過極線幾何關係，求出具備三維空間資訊的深度影像 (Depth)。

一般而言，深度影像 (Depth) 估算技術可分為四大步驟，分別為：比對代價估算 (Matching Cost Computation)、比對代價聚合 (Cost Aggregation)、視差估算 (Disparity Selection) 與視差優化 (Disparity Refinement)，文獻 [1]。而其中最核心的運算為比對代價聚合與視差估算，大可分為區域演算法 (Local

Method)、全域演算法 (Global Method) 與半全域演算法 (Semi-Global Matching, SGM) 三類。本研究所使用的半全域演算法，該方法是以多個一維方向的動態規劃法影像匹配，模擬二維影像的全域匹配最佳化，該演算法運算複雜度小、硬體實現性高、環境耐受力備受肯定，且匹配準確性相近於全域演算法 [2][3]。

綜合上述，本研究以 SGM 演算法為基礎，再結合線性內插、視差優化等演算方法，提出一三維深度影像估算演算流程，利用 FPGA 和單一 SDRAM 記憶體實現其演算流程。實現結果以 Middlebury 大學的 Tsukuba 標準測試圖，進行三維空間資訊的深度影像估算正確率分析，再透過不同道路情境的實車影像測試。

## 二、三維深度影像估算演算法



圖一：立體視覺極線幾何示意圖

立體視覺深度資訊的建構可以利用圖一的極線幾何示意圖來解釋，極線幾何為雙攝影機的針孔成像模型與空間中一點的幾何關係。在雙攝影機與被攝物的幾何關係中，兩攝影機置於同一水平線上，兩攝影機之間的距離為  $B$ ，此距離稱為基線距離。而在兩攝影機的光軸與影像平面的交點分別為  $C_1$ 、 $C_2$ 。 $P_1$ 、 $P_2$  分別代表空間中一點投影到兩攝影機影像平面的投影點，又稱極點。 $\Delta X_1$  與  $\Delta X_2$  代表  $P_1$ 、 $P_2$  與兩影像平面中心點的距離； $Y$  為被攝物體在兩影像的垂直位置； $Z$  則為被攝物體距離基線的距離。因此，深度資訊可由被攝物體的點  $P$  與兩攝影機中間點的三維資訊  $(X, Y, Z)$ ，藉由相似三角形關係計算而得，如式(1)和(2)。

$$\frac{X}{Z} = \frac{x-x_o}{f}, \quad \frac{Y}{Z} = \frac{y-y_o}{f} \quad (1)$$

$$Z = \frac{Bf}{|\Delta x_1 - \Delta x_2|} \quad (2)$$

上兩式中， $f$  代表攝影機的焦距； $(x,y)$ 與 $(x_0,y_0)$ 分別代表點  $P$  投影到左影像平面中的投影點以及左影像平面的中心點； $|\Delta x_1 - \Delta x_2|$ 代表像差值。

### 2.1 比對代價估算

藉由兩張相同水平的左右影像，可用比對代價估算出計算不同視差之下的像素灰階相關性。本研究以像素灰階差的絕對值(Absolute Intensity Differences, AD)作為比對代價(Matching Cost)，其公式如式(1):

$$AD = |left\_img(m+d,n) - right\_img(m,n)| \quad (3)$$

$$d = 0 \sim \max\_shift$$

其中  $left\_img$  和  $right\_img$  分別為輸入的左右影像， $d$  為像差， $\max\_shift$  為最大位移量。

### 2.2 比對代價聚合

本研究在比對代價聚合演算使用半全域比對演算法(Semi-Global Matching, SGM)，SGM 是利用代價函數及動態規劃法的觀念，以左右影像中一條水平像素序列為基準，進行水平位置的像素序列進行比對，再透過多個一維方向的動態規劃法影像匹配，模擬二維影像的全域匹配最佳化。

由於比對代價易受影像雜訊、大範圍相似區域等多種因素影響，並無法準確反映兩影像像素點的匹配，所以會造成誤匹配現象，且勿匹配的情形將影響到周圍影像像素的視差估算，進而將錯誤擴散，故於本研究的半全域比對代價聚合中加入了一平滑度限制，其能量函數如式(4)所示：

$$E(D) = \sum_p C(p, D_p) + \sum_{q \in N_p} P_1 T[|D_p - D_q| = 1] + \sum_{q \in N_p} P_2 T[|D_p - D_q| > 1] \quad (4)$$

式(4)中的  $E(D)$  是由一維的動態規劃法，所有影像像素點匹配的能量值，第一項為所有影像像素點的比對代價，為能量函數中的代價項， $N_p$  為影像像素  $P$  的近鄰點，第二項與第三項為能量函數中的平滑項，用以平滑所估算的匹配結果。而其中  $T[arg]$  為一條條件判斷式，如果條件  $[arg]$  為真，則  $T[arg] = 1$ ，若條件  $[arg]$  為假，則  $T[arg] = 0$ 。而係數  $P_1$  與  $P_2$  為深度視差變動的懲罰常數， $P_1$  為小視差變動(深度視差=1)的懲罰常數， $P_2$  為大視差變動(深度視差>1)的懲罰常數，因此  $P_1 < P_2$ ，得以要求相鄰影像像素點的深度值盡量保持一致，以保持深度影像平滑。

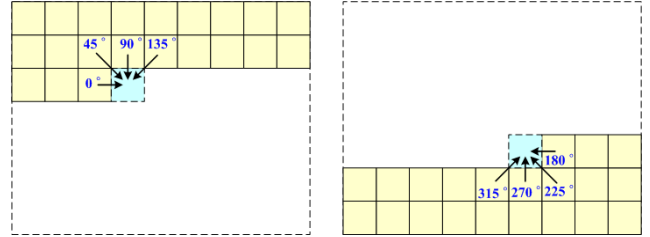
因使用平滑度限制會使一些邊緣及深度正常變化區域強迫其深度保持不變，故造成了深度邊界的模糊化，故在此引入了像素點的亮度資訊，如式(5)所示：

$$P_2' = \frac{P_2}{|I_{bp} - I_{bq}|} \quad (5)$$

其中  $P_2'$  為修正前大視差變動(深度視差>1)懲罰常數， $P_2$  為修正後的懲罰常數， $I_{bp}$  為影像像素點的亮度， $I_{bq}$  為相鄰影像像素點的亮度。由於邊界多為亮度不平滑區域，透過亮度差調整懲罰常數，將有效解決深度正常變化區域過大懲罰常數問題，避免了深度邊界模糊化。

雙視覺立體視覺的全域匹配(Global Matching)[4]是給予一能量函數，再搭配優化方式以求出能量函數的最小值，此種全域匹配計算有著良好的匹配準確度，但在實

現上相當困難且耗時，故本研究使用 SGM 演算法以多個一維方向的動態規劃法影像匹配，模擬二維影像的全域匹配最佳化，半全域的比對代價聚合的方向，如圖二所示，圖二左圖為一四個一維方向( $0^\circ$ 、 $45^\circ$ 、 $90^\circ$ 、 $135^\circ$ )動態規劃法影像匹配；圖二右圖為另四個一維方向( $180^\circ$ 、 $215^\circ$ 、 $270^\circ$ 、 $315^\circ$ )的動態規劃法影像匹配掃描方式。透過八方向的聚合，即可模擬二維影像的全域匹配最佳化，得到匹配效果相當佳的深度影像。



圖二：半全域的比對代價聚合(8方向)

本研究的 SGM 半全域比對代價聚合的路徑(Path)代價聚合公式，是參考一維的動態規劃法影像匹配的平滑度限制，如式(6)與式(7)所示：

$$L_r(p,d) = C(p,d) + C_s(p-r,d) - \min L_r(p-r,i) \quad (6)$$

$$C_s(p-r,d) = \min(L_r(p-r,d), L_r(p-r,d-1) + P_1,$$

$$L_r(p-r,d+1) + P_1, \min L_r(p-r,i) + P_2) \quad (7)$$

其中  $C(p,d)$  為影像像素點  $p$  賦予深度  $d$  的匹配代價， $C_s(p-r,d)$  為當前路徑  $r$  上影像像素點的上一像素點  $p-r$  的包含懲罰常數的最小匹配代價，第三項則對該路徑匹配代價聚合無任何影響，僅為了防止  $L_r$  過大的問題。最後各個方向的路徑匹配代價聚合再被加總，如式(8)所示：

$$S(p,d) = \sum L_r(p,d) \quad (8)$$

其中  $S(p,d)$  為半全域的比對代價聚合， $L_r$  為路徑(Path)代價聚合。

### 2.3 視差估算與視差優化

在視差估算上，為選取灰階相關性最大的匹配點，故採用贏者全拿法(Winner Take All, WTA)作為代價比對的結果，以初步估算出三維深度影像。

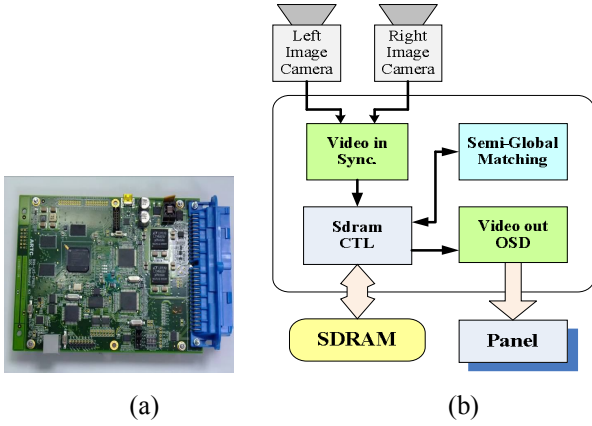
而由於一般道路紋理特徵過於稀疏，易造成低紋理區域產生誤匹配情形，為了避免誤匹配情形，故本研究導入了低紋理特徵濾波器，以得有效的高紋理特徵。該演算法流程為：首先針對輸入影像進行梯度估算，然後估算一特徵區塊內平均梯度，再透過預設閾值比較，以判定是否為高紋理特徵區域，最後再將所得深度影像，與原始影像透過低紋理濾波器所得高紋理特徵影像，進行邏輯運算，即可濾除低紋理區域。

最後所得到的高紋理深度自訊再透過中間值濾波器，將由強的突峰訊號分量組成的高頻雜訊去除，取得新的採樣，並保持所得視差影像的邊緣銳度。

## 三、立體視覺 FPGA 平台與架構

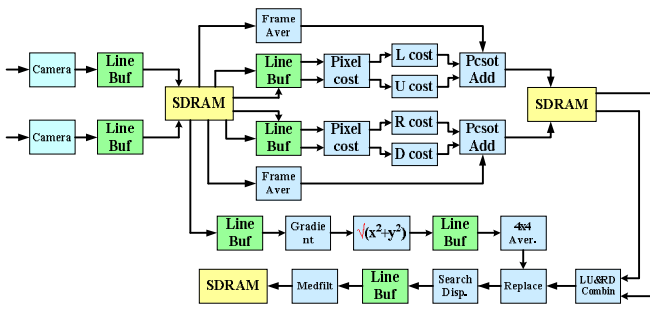
圖三(a)為本文使用之 FPGA 硬體平台，此硬體平台是使用 Xilinx Spartan-6 FPGA 來運算三維深度估算演算法；深度資訊實現流程如圖三(b)所示，影像輸入端使用雙 CCD 攝影機，此端採用類比影像訊號輸入至解碼晶片轉換成數位訊號後，以影像格式 BT656 的八位元格

式傳至 FPGA。兩支影像攝影機會有影像不同步的情況發生，所以在 FPGA 中藉由左右攝影機最相近時間的影像訊號作運算。而影像運算過程需要用到記憶體，在此採用單一顆 SDRAM 來儲存影像資料，且透過資料流的排序來減少 SDRAM 的使用，適用於非 real time 的系統應用。而運算完的影像再透過編碼晶片輸出類比影像訊號到外部顯示器上。



圖三：(a) FPGA 硬體平台；(b)立體視覺系統架構圖

在 FPGA 所實現的三維深度資訊估算架構-如圖四所示，左右影像由影像解碼晶片解碼成 YCbCr 422 格式後，利用影像明度 Y 資訊進行運算。每條影像 Y 資訊皆會暫存在 line buffer，再藉由 SDRAM Controller 寫入 SDRAM。當要計算比對代價估算時，SDRAM Controller 將從 SDRAM 讀取左右影像同一條水平的 Y 資訊進行運算。運算完比對代價估算後會得到相對應的像素比對代價，利用四種方向的 pixel cost 運算 Semi-Global Matching。四個方向分別為：Left、Right、Up and Down。為了讀取資料的方便性，將順向讀取出 Y 資訊進行 Left cost and Up cost 運算；而再反向讀取出的 Y 資訊進行 Right cost and Down cost 運算。運算完四個方向的比對比價估算後，再互相加總。求得新的代價值後，為了濾除道路的稀疏紋理，即對運算 SGM 後的代價值進行梯度估算，估算一特徵區塊內的平均梯度，再透過預設閥值比較，以判定是否為高紋理特徵區域，最後在利用中值濾波將由強的突鋒訊號分量組成的高頻雜訊去除，並保持所得視差影像的邊緣銳度。

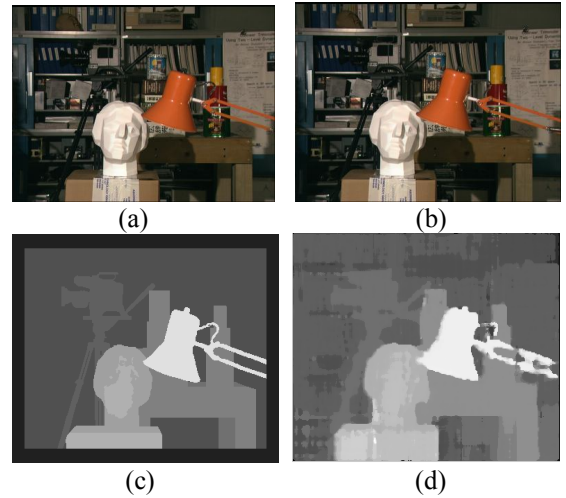


圖四：FPGA 內部 SGM 運算架構圖

四、 FPGA 實現結果與立體深度影像結果

本研究於三維深度影像估算是採用不同場景，且基線 B 經過校正，並提供標準視差圖來和演算法的執行結

果進行比較，以得演算法的精度與適用性。圖五為最常用的 Tsukuba 測試圖，由圖五(a)至圖五(d)分別為左攝影機影像、右攝影機影像、Tsukuba 標準視差圖以及最後由 FPGA 輸出之標準視差影像比較。由圖五(c)與圖五(d)可看出，本研究所實現的三維深度資訊能明確表示物體深度關係。



圖五：(a)Tsukuba 左影像測試圖；(b)Tsukuba 右影像測試圖；  
(c)標準影像視差圖；(d)FPGA 輸出視差圖

4.1. FPGA 硬體實現結果

本研究所使用的 FPGA 是 xilinx spartan-6 xc6xlx150。其中內部主要運算頻率為 108MHz，亦提供給 SDRAM 當操作頻率。而影像輸入是 NTSC 影像格式，影像大小為 720 x 480 的解析度，影像輸入的頻率為 27MHz。藉由 FPGA 去實現此三維深度影像估算的 logic utilization 如表 I 所示；其中使用了 102KB 的 BRAM 暫存影像資訊；而使用了 10K 的 LUTs 和 6K 的暫存器。表 II 為本文與其他文獻[5][6]在 FPGA 上的使用資源比較表，本文所實現方式具有低複雜性的硬體架構，適用於一個非高速的應用中，且不需要使用大量多顆記憶體。

表 I：FPGA 邏輯資源使用表

Logic Utilization	Used	Available	Utilization
Number of Slice Registers	6,528	184,304	3%
Number of Slice LUTs	10,467	92,152	11%
Number of used As logic	9,539	92,152	10%
Number of Occupied Slices	3,856	23,038	16%
Number of RAMB8BWERS	92	536	17%

表 II：FPGA 邏輯資源比較表

FPGA 資源	本文	文獻[5]	文獻[6]
Number of Slice LUTs	10K	60K	56K

Number of BRAM	102KB	300KB	220KB
----------------	-------	-------	-------

#### 4.2. 實際影像運算結果

本研究使用 FPGA 來實現三維深度資訊，由圖六可知，在不同道路情境的實車影像測試，皆能透過色階表示出真實環境的三維空間資訊，顯示本研究所提出的三維深度影像估算流程，具有相當佳的準確性，雖對於紋理缺乏區域，仍有誤匹配現象產生，但整體而言，本研究提出的三維深度影像估算流程，其穩定性與環境耐受性，相當適用於真實道路環境。



圖六：實際道路測試圖與深度結果

#### 五、 結論

本研究以 FPGA 實現立體視覺車用影像系統中最常用的半全域匹配演算法(Semi-Global Matching, SGM)演算法為基礎，並提出一三維深度影像估算演算 FPGA 流程，並以 Middlebury Tsukuba 標準測試圖與實車影像，分析所提出演算方法的比對效果。利用 FPGA 來實現其演算法運算運算，演算法是由像素灰階的比對代價，並結合 SGM 比對代價聚合計算，最後再以交叉驗證、中間值濾波與低紋理區域濾除，對於不同道路情境的實車影像測試。

未來將以該演算方法為主架構，並持續研究低紋理區域、積水反射與系統加速處理等議題，進行其演算法的優化，開發出精確且適用性高的三維深度影像估算方法。

#### 誌謝

本研究承蒙經濟部技術處科技專案計畫編號 102-EC-17-A-04-02-0803 的補助，得以完成此研究，特此致上

感謝之意。

#### 參考文獻

- [1] D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms," *International Journal of Computer Vision*, 2002, vol. 47 (1) 7-42.
- [2] H. Hirschmuller, "Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information," in *IEEE Conference of Computer Vision and Pattern Recognition*, 2005, vol. 2, 807-814.
- [3] H. Hirschmuller, "Stereo Processing by Semi-Global Matching and Mutual Information," in *IEEE Transactions of Pattern Analysis and Machine Intelligence*, 2008, vol. 30, 328-341.
- [4] 陳育崧，廖育昇，徐子建，立體視覺特徵點比對演算法分析與實現，中華民國第十六屆車輛工程學術研討會論文集，B2-007，台北，台灣，2011。
- [5] Gehrig S, Eberli F, Meyer T, "A real-time low power stereo vision engine using semi-global matching matching." *Proceeding of the International Conference on Computer Vision System Springer*; 2009, 134-143.
- [6] Banz C, Hesselbarth S, Flatt H, Blume, P "Real-time stereo vision system using semi-global matching disparity estimation: architecture and FPGA-implementation." *Proceedings of the International Conference on Embedded Computer System 2010*, 93-101, IEEE.