使用深度資訊與主動式形狀模型之人臉定位技術

A Novel Face Detection Technique Using Depth Information and Active Shape Models

洪國銘
開南大學資訊管理系
hkming@mail.knu.edu.tw

賈勝文
開南大學資訊管理系
m10207006@mail.knu.edu.tw

高嘉男
開南大學資訊管理系
m10107003@mail.knu.edu.tw

陳鼎文
開南大學資訊管理系
m10007004@mail.knu.edu.tw

摘要

電腦視覺研究中人臉偵測的方法相繼被提出，一般僅使用彩色影像定位人臉的方法，只依靠彩色影像做定位，容易受環境與光線影響，產生錯誤人臉位址。本文使用 KINECT 控制器捕捉的彩色影像與深度資訊，提出人臉定位技術，以 ASM 定位出影像中人臉的區域，並以深度資訊做為輔助，去除形狀模型偵測錯誤的區域，增加定位的精確度。結果顯示，此方法成功消除錯誤偵測區塊，使得 ASM 檢測更為正確。

關鍵詞：人臉偵測、主動式形狀模型（Active Shape Models，ASM）、深度影像

Abstract

In this paper, we employ Kinect to capture color and depth image. After then we detect the face blocks by ASM. When faces were detected, we compare the face object size between detection and table we count distance in depth image. Finally, remove the error blocks to increase the detection accuracy from ASM. The result shows our method not only removes the error blocks successfully, but also makes face detection with higher accuracy.

Keywords : face detection, active shape models(ASM), depth image

1.緒論

近年來，電腦視覺的研究拓展了人類的視覺空間，在電腦運算速度提升至可以短時間處理圖片和影像等大量資料後，學者開始關注電腦模擬人類的視覺效果。以電腦視覺達成非監督式的監控系統，減少人員浪費和成本支出，並有效監督特定區域，達到更高的安全性。

生物偵測方法中，可分為察覺性與非察覺性兩種偵測類型。察覺性的偵測，會影響到生物正常行為，始得取得辨識所需要的特徵，此方式容易因目標偽裝而形成誤判。非察覺性的偵測，則在沒有影響目標的情況下，擷取偵測目標所需要的特徵進行處理。

人臉偵測則以非察覺性的方式，從自然行為中，取得人臉影像，擷取所需要的特徵進行偵測定位。方法大致以統計其特徵紋理的方式作為指標，過去利用主動式形狀模型（Active Shape Models，ASM）[1][2]擷取影像中人臉位置，但在擷取時，常因光源等環境因素而產生誤差，故本篇將提出使用深度資訊做為輔助，偵測錯誤的區域，使影像中人臉定位的正確率更為提升。

本文包含五個章節，第一章介紹研究的背景、動機與目的；第二章探討相關文獻，介紹

本研究所使用的技術；第三章詳盡說明本研究提出的方法；第四章呈現實驗的結果，說明實驗的環境規格，並提出相關研究數據；第五章為結論與未來展望。

2.文獻探討

在影像處理技術發展之下，能夠發現人臉越來越多種特徵類型，定位人臉的方法也日新月異，且更趨於成熟穩定。以下介紹關於本文的相關研究文獻與方法。包括主動式形狀模型與深度影像的建立方法。

2.1 主動式形狀模型

在目前的臉部定位方法之中，局部定位與整體定位是最常被使用的兩種方法，局部定位使用的方法如樣板比對法（Template Match）[5]，將臉部五官，如眼睛、鼻子、耳朵和嘴巴，分別進行偵測後加以定位，不過這個方法容易受臉的外部配件影響而形成雜訊，且在五官分別偵測狀況下，缺少位置的對應關係，在定位上容易形成較大的誤差，而產生錯誤的判斷。整體定位的方法則考慮了臉部五官互相對應的位置，以 T.F.Coots 和 C.J.Taylor 等人在 1992 年提出的主動式形狀模型（Active Shape Model，ASM）[1]為主，以統計方式來蒐集人臉各個特徵與器官相對關係的特徵位置，加以訓練後建立成臉部形狀模板，再將此模板套用至目標上。

Cootes 提出「Active Shape Models – Smart Snakes」[1]，此方法原應用於搜尋電子電路影像中電阻的形狀，使用人為的方式將需要的形狀標示後，加以訓練建立模組，再將此形狀模組套用於相似的目標上，藉以找出影像中相似的物件。

2.2 深度影像建立方法

現今深度影像有很多種的建立方法，一般常見的方法有利用雙眼視覺的三角測距法[9]、結構光掃描法[10]、時差測距[11]以及 Light Coding[12]技術。

3.系統架構

本研究以傳統主動式形狀模型為基礎，並以深度資訊做為輔助，去除形狀模型偵測錯誤的區域，增加人臉定位的精準度，其架構可分為：主動式形狀模型的模板建置、臉部偵測定位以及利用深度資訊消除非人臉區塊三個階段。第一階段同樣將多張臉部影像以人工方式標示特徵點，經對齊訓練之後產生的人臉特徵資訊，用於主動式狀模型的人臉模型建置。第二階段利用 Kinect 擷取目標空間的彩色影像與深度影像，使用 ASM 偵測其彩色影像的人臉加以定位。第三階段為利用深度影像提供的資訊計算於在該深度下，ASM 定位的物件是否為人臉，藉此去除非人臉的物件，如圖 3. 1。以下將詳細介紹本研究如何運用上面三個階段達成人臉定位。

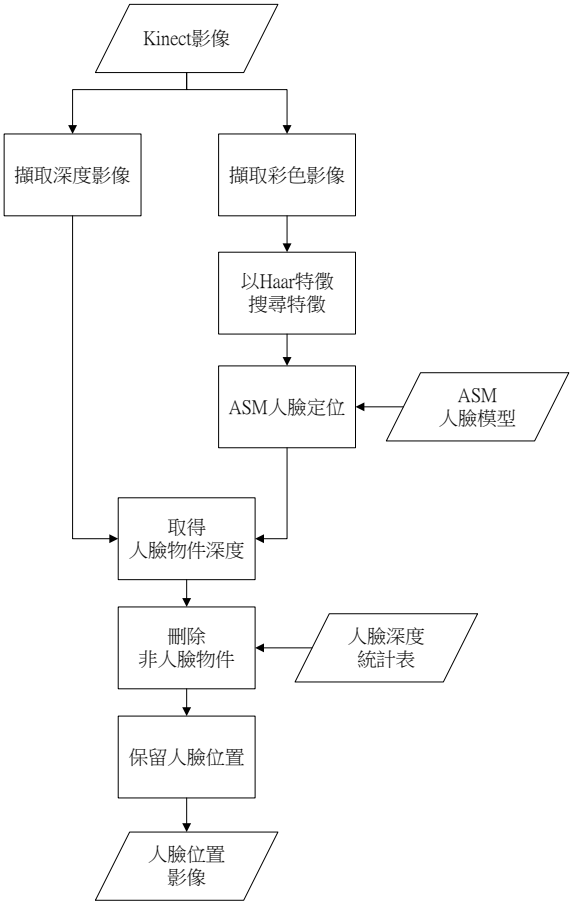


圖 3. 1 人臉定位系統流程圖

3.1 臉部模型建置

在使用主動式形狀模型定位人臉之前，必須建置一個擁有人臉特徵的模型，使其在定位的時候能依據人臉模型，套用在相似外型的物件上[1]。

首先將多張臉部影像以人工方式標註特徵點，特徵點標記方式依照 Dryden et al[6]所提出的三個準則，優先標示影像中的邊角（Corner）、高曲率（High Curvature）和接合處（Junction）。

每張影像在標示特徵點的時候必須依照一定的順序標註，於臉部標註 75 點特徵，如此建立的特徵集合，便能依據編號來標記臉部的器官。標註多張人臉影像的特徵點後，將各點分別累加求平均值，取得平均形狀，產生臉部模型。

3.2 主動式形狀模型的人臉定位

主動式形狀模型的人臉建立後，即可藉由此模型尋找於影像中類似與相同的形狀。本研究使用 Haar-like[7]的方法搜尋目標影像中多個相似於人臉的特徵位置，隨後將主動式形狀模型放置於目標處，透過迭代將模型參數收斂至符合物件大小。

ASM 以特徵點互相連結形成後，再經過數次的迭代將模型收斂，藉以找出符合形狀的物件。為了尋找目標物件邊界，每個特徵點在收斂時會產生垂直於形狀模型邊緣的法線向量，做為修正方向的參考點，經由搜尋其法線向量，能取得每個特徵點將要移動的方向 dX，如（3.1）所示。

$$dX = (dX_0, dX_1, dX_2, \dots, dX_{m-1}) \qquad (3.1)$$

其中，x 為特徵點集合的矩陣，X_c是主動形狀模型初始化的中心座標，記為（X_c, Y_c）。而經過調整的主動式形狀模型，公式如（3.2）所示。

$$X + dX = M(s(1 + ds), (\theta + d\theta))[x + dx] + (X_c + dX_c) \qquad (3.2)$$

ASM 使用形狀模板對目標物件進行搜尋，再調整模型參數，使形狀模板能夠吻合目標物件的形狀。經過數次的迭代調整，形狀模板中每個特徵點會漸漸的收斂並求出最佳值，找出最符合目標物件特徵的位置。主動式形狀模型的搜尋，將透過收斂目標函式所產生的結果來判斷其變化，如沒有變化，即已經搜尋出目標物件位置；如果變化量大，則須再次計算模型參數，直到模板收斂至不再產生變化。當模板與目標物件匹配完成後，即可透過特徵點的分布，找到目標物件形狀或特徵位址。

3.3 以深度資訊排除非人臉物件

經由上述步驟，取得影像中人臉的位置後，需要將偵測錯誤的物件消除，圖 3. 2 為因偵測錯誤，而在影像非人臉部部分上框出臉部物件之情況。此時可以藉由物體與攝影鏡頭距離的大小關係判斷，在物體與鏡頭越接近的情況下，影像中偵測到的該物體面積會越大，反之，物體與鏡頭距離越遠時，影像中偵測到該物體的面積則越小。

圖 3. 2 含有偵測錯誤物件的結果

本研究使用該影像的深度圖協助辨識物件的前後距離，藉由 Kinect 快速取得影像的深度資訊，並統計每個深度區間人臉的面積，藉

此區分出人臉物件在每個深度區間的面積大小，製作成參考表格。在相同的解析度與深度平面下，臉部呈現的面積應為相近似值。如圖圖 3.3 所示。

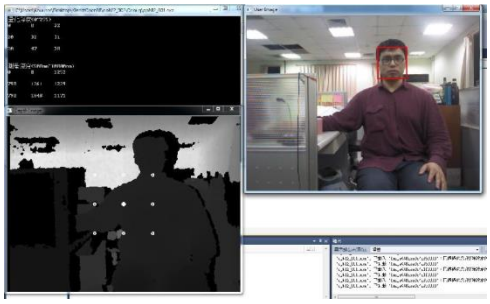


圖 3.3 人臉深度測定

在研究測試的統計下，圖 3. 4 水平方向的刻度為深度正規化的距離，將機器最大深度 10 公尺，分為 255 個區間，在深度影像上以灰階亮度表示，垂直方向的刻度表示偵測到臉部物件之面積，其單位為*pixel*。使用人臉於各深度所呈現面積之統計表格，配合 Kinect 擷取影像，能夠獲得 RGB 的彩色影像與相對應的深度影像，將彩色影像以主動式形狀模型（Active Shape Model，ASM）取得人臉位置後，即能對照深度圖將每個人臉物件所對應的深度值取出，先以統計之臉部深度面積表格做比對，再以同深度的周圍物件比對其面積大小，以此兩個維度的交叉比對，濾除影像中因為偵測錯誤所框選出過大或過小的物件，保留正確的人臉位置。

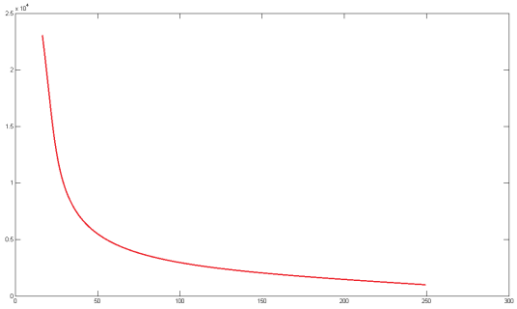


圖 3.4 深度與面積關係圖

本研究使用該影像的深度圖協助辨識物件的前後距離，藉由 Kinect 快速取得影像的深度資訊，並統計每個深度區間人臉的面積，藉此區分出人臉物件在每個深度區間的面積大小，製作成參考表格。在相同的解析度與深度平面下，臉部呈現的面積應為相近似值。

4.實驗結果

本研究依照上一章節所提出的系統架構預先對人臉做主動式形狀模型的模型建置以人工標定特徵點後，介面提供形狀對齊功能，並將結果人臉模型輸出，取得主動式形狀模型，用於系統定位人臉。

另一方面，研究中以 Kinect 擷取並統計各深度人臉面積，製成統計圖區分每個深度的範圍及符合該深度的臉部面積，如圖 4. 1。

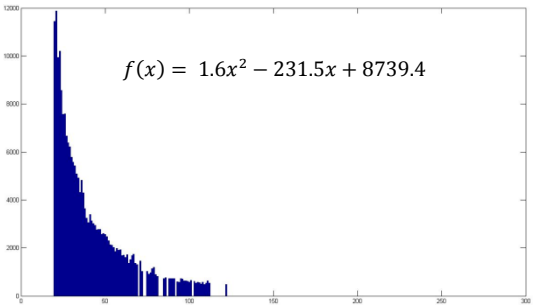


圖 4. 1 各深度區間面積統計圖

由圖 4. 1 中水平方向的刻度為深度正規化的距離，以 OpenNI 擷取 Kinect 所能讀取的最大深度，Kinect 控制器的回報值為 10,000 公厘（mm），將 10,000 的數值分為 255 個區間，在深度圖中以亮度值的方式顯示，而垂直方向的刻度表示偵測到臉部物件之面積，其單位為*pixel*。圖中顯示深度 0 至 20 的範圍因距離攝影機太過於接近，無法偵測到臉，並且在過近的距離下，深度影像中顯示的深度值均為 0，深度值於 20 之後，臉部面積在近距離為最大，而後依距離增加面積逐漸減小，近距離時面積落差較大，若深度落於 20 至 30 的區間每一個深度均做紀錄，當目標物件深度落於此區間，

必須依照表 4.1 做面積深度比對，在深度大於 30 之後，臉部面積的變化逐漸減少，此時以 5 個深度值為一群組統計，物件離攝影機越遠，則包含的範圍越大，深度值於 50 之後以 10 個深度值分為一個區間。

表 4.1 中，樣本為 Kinect 各深度下顯示的人臉面積，p 為各深度下人臉面積的平均值，且計算出每個區間人臉面積的標準差，於統計深度區間內，人臉面積皆為常態分佈，與表格中人臉平均值相差 2 個標準差之內的面積範圍，均接受為該深度的人臉。

表 4.1 深度區間與面積關係統計表

深度值	距離(cm)	p	標準差
0-19	0≦ x＜75	p＞15000	
20	75≦ x＜78	13453	877
21	78≦ x＜82	12687	849
22	82≦ x＜86	11586	804
23	86≦ x＜90	10846	765
24	90≦ x＜94	9655	713
25	94≦ x＜98	8851	679
26	98≦ x＜102	8388	648
27	102≦ x＜106	7365	623
28	106≦ x＜110	6735	592
29	110≦ x＜114	6160	568
30	114≦ x＜118	5458	537
31-35	118≦ x＜137	5065	607
36-40	137≦ x＜157	3816	543
41-45	157≦ x＜176	3053	427
46-50	176≦ x＜196	2657	376
51-60	196≦ x＜235	2043	365
61-70	235≦ x＜275	1558	234
71-80	275≦ x＜314	1075	227
81-90	314≦ x＜353	748	101
91-100	353≦ x＜392	641	84
100-255	392≦ x＜1000	p<550	

接著求出距離與面積的變化關係方程式，將樣本距離定值（X：深度為 20 至 100）與深度點臉部面積帶入後，計算出距離與面積近似的線性方程式，如（4.1）。

$$f(x) = 1.6x^2 - 231.5x + 8739.4 \quad (4.1)$$

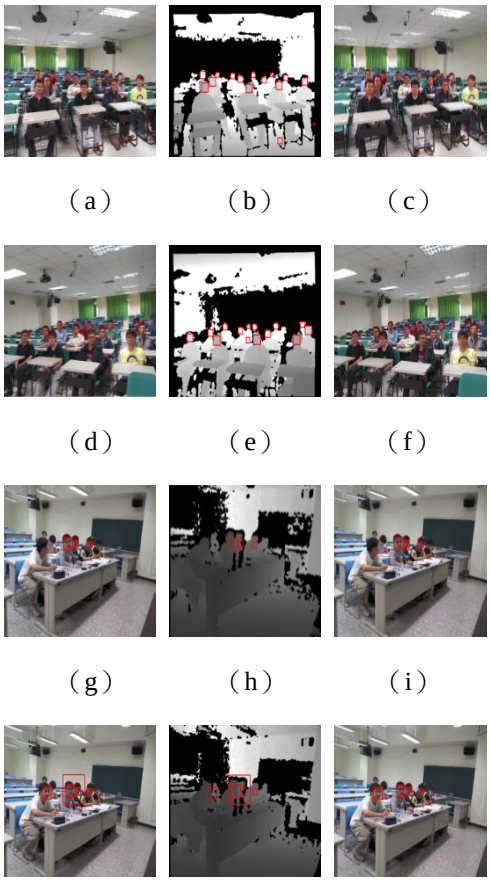
由於每個深度區僅使用兩個標準差進行人臉深度面積的判斷，使得濾除錯誤人臉定位物件仍然存在 5%的誤差。

Kinect 官方的網頁[14] 所提出的使用距離建議為 1.2 公尺至 3.6 公尺，代表它的深度感應功能，能在一定距離內提供較高的精確度，而較遠的地方，則無法分辨細微的差距。以表 4.1 來說，深度值從 20 開始有深度的數值表現出來，20 以前的值基本上在測量時都顯示為 0，這時候能抓到的資料量並不穩定，所以不建議擷取距離小於 20 的資料。而深度到達 50 後，取得的資料就不再是每個深度的連續，偶爾會有缺值，這表示精確度開始降低了。缺資料的深度會依據距離越遠而越高，密度也就越稀疏，也就是表格中的深度距離越遠，統計所使用深度範圍越大的原因。

本篇使用自然影像實驗，一般的人臉定位後容易出現誤抓的狀況，此時將彩色影像同時擷取的深度影像輸入，提取人臉物件相對應的影像深度，將每個物件的深度資訊參照表 4.1 比對其面積後，可以發現非人臉區域的定位物件面積並不符合相對的深度，將錯誤物件消除後取得正確的人臉定位影像結果。

本實驗將其他自然影像以相同方式做定位如（j）　　（k）　　（l）

圖 4.2，圖中左半部為使用 ASM 定位影像中人臉的位置，中間的影像是套入深度資訊做面積比對，右半邊影像顯示如發現錯誤的深度面積物件，將其消除，保留定位正確的人臉位置。



（a） （b） （c）

（d） （e） （f）

（g） （h） （i）

（j） （k） （l）

圖 4.2 四張自然影像偵測結果

5.結論與未來發展

人臉偵測的方法在電腦視覺研究領域中相繼被提出，使用彩色影像能獲取的資訊越來越多樣化，但僅依靠彩色影像處理，容易受環境與光線影響，產生雜訊或是錯誤的資訊。微軟公司近年推出的 KINECT 控制器，能夠利用紅外線感測的功能擷取目標與攝影機間距離之深度影像，而深度是人類產生立體感覺的重要資訊，在影像處理的領域裡，有著越來越多研究朝向影像的第三維度，深度來探討並開發新的技術。

本研究提出結合主動式形狀模型 ASM 與深度資訊的方法，定位人臉於影像之位置，使用空間中深度遠近，在視覺上造成影像面積關係，有效地將錯誤的非人臉資訊去除，使得系統於人臉定位更加精確。在實驗顯示出，該方法成功提升了人臉定位的準確度。本文以 ASM 為基礎，加上深度資訊所提出的人臉定位技術，能夠依照人臉距離鏡頭的遠近，消去影像中不屬於人臉的物件。

微軟公司開發的 Kinect 控制器，雖然能夠測量鏡頭向後 10 米的距離，但實際使用時僅有約 4 米能捕捉到較精確的深度距離，4 米之後則不一定產生深度值，使技術在測量上受到限制，故在未來研究上，針對能精確探測到各種距離的深度空間建立技術，做更進一步的研究與探討。

參考文獻

[1] Cootes, T.F., Taylor C, J., “Active shape models – smart snakes”, Proceedings British Machine Vision Conference, Springer, Berlin, pp.266-275,1992.

[2] T.F.Coots, C.Taylor, D.Cooper, and J.Graham, “Active shape models - their training and application”, Computer Vision and Image Understanding, 61（1）:38-59, January 1995.

[3] A. Hill, T. F. Cootes, and C. J. Taylor. Active shape models and the shape approximation problem. In 6th British Machine Vison Conference, pages 157-166. BMVA Press, Sept. 1995.

[4] R.H.Davies, T.F.Cootes, C.Twining and C.J.Taylor, ”An Information theoretic approach to statistical shape modelling”, Proc.British Machine Vision Conference, pp.3-11, 2001.

[5] ZHANG Baizhen and RUAN Qiuqi, “Facial feature extraction using improved deformable templates”, The 8th International Conference on Signal Process Volumn4, 2006.

[6] I.L.Dryden and K.V.Mardia, Statistical shape analysis, John Wiley&Sons, 1998.

[7] Viola, Paul and Michael J. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Volume: 1, pp.511–5,182001.

[8] J. Liu and J.K. Udupa, “Oriented active shape models,” IEEE Transactions on Medical Imaging, vol. 28, no. 4, pp. 571-584, 2009.

[9] Redert. A., de Beeck. M.O., Fehn. C, Ijsselsteijn. W., Pollefey. M., Van Gool, L., Ofek. E., Sexton. I., Surman. P., “Advanced

three-dimensional television system technologies," 3D Data Processing Visualization and Transmission, 2002. Proceedings. First International Symposium on , vol., no., pp. 313-319, 2002.
[10] http://www.stockeryale.com/i/lasers/structured_light.htm
[11] Pasi Palojarvi," Integrated electronic and optoelectronic circuits and devices for pulsed time-of-flight laser range finding," Department of Electrical and Information Engineering and Infotech Oulu, University of Oulu,2003.
[12] http://www.bb.ustc.edu.cn/jpkc/guojia/dxwlsy/kj/part2/grade3/LaserSpeckle.html
[13] http://www.xbox.com/zh-TW/kinect.
[14] http://msdn.microsoft.com/zh-tw/hh367958.aspz