

多角度人臉之偵測與追蹤

Multi-Pose Face Detection and Tracking

江政杰
Cheng-Chieh Chiang
德明財經科技大學
資訊科技系
pluto@takming.edu.tw

周嘉榮
Chia-Jung Chou
德明財經科技大學
資訊管理系
joe@takming.edu.tw

摘要

近年來人臉偵測與追蹤在電腦視覺領域中已經成為備受重視的研究主題。在眾多影片相關的多媒體應用中，常見各種需要先判定人臉位置以進行後續處理的技術需求。在本論文中，我們將結合視覺追蹤與臉部偵測兩個方向的技術，針對影片資訊設計一套多角度之人臉偵測與追蹤方法。本研究所設計的架構主要包含四個模組：一是初始化模組以設定整個系統所需的初始狀態；二是一個稱為 Condensation 模組的粒子過濾追蹤方法，用以估計影片畫面中人臉區域的位置；三是範本模組以紀錄在 Condensation 模組中需要的人臉特徵資訊；以及最後的校正模組以便在追蹤發生錯誤時能夠重啓人臉追蹤機制。這四個模組透過不同的合作機制，能夠正確定位出影片中移動的人臉位置。此外，本論文同時包含多個實驗內容，顯示在不同角度與不同移動模式的人臉區域，本研究所設計的方法仍能正確的偵測追蹤出人臉的位置。關鍵詞：人臉偵測與追蹤、人臉特徵、粒子過濾追蹤、範本模組。

Abstract

Face detection and tracking has recently been a hot research topic in computer vision. Many applications need to determine the position of human faces in videos. In this paper we propose a method for multi-pose face detection and tracking by combining the visual tracking and face detection in video sequences. Our proposed model consists of four modules: (i) an initialization module for setting the initial status of the whole system, (ii) a Condensation module for estimating the position of the human face for tracking, (iii) a template module for recording the face features for the observation process in Condensation, and (iv) a correction module for restarting the tracking if the tracked face has been lost. These four modules cooperate to detect the face area and track the face moving in video sequences. This paper also demonstrates several experiments to show the performance of our proposed methods.

Keywords: face detection and tracking; face feature; particle filter; template model.

1. 緒論

人臉偵測(face detection)是電腦視覺中重要的資訊技術,意指在圖片或影片中判斷人臉的位置。臉部區域是辨識一個人最重要的資訊;而確認臉的位置,同時也可以確認有人的存在,因此人臉偵測廣泛應用於偵測圖片或影片內人的位置,以便發展更進一步的應用。例如監視系統要自動判斷有什麼人出現在監控區域、門禁系統以人臉判斷身份、相片管理系統自動判斷照片中人的身份而進行管理,諸如此類應用都需要先對圖片或影片中進行人臉偵測,找出臉部位置後依據系統目的而進行後續的處理。

在圖片或影片這兩種不同媒介中處理人臉偵測的問題,最大的差異是圖片與影片本身的不同。一個影片的播放可以視為連續的圖片所組成;因此在影片中進行人臉偵測,可以簡化成在連續圖片中個別進行人臉偵測,然後將個別的偵測結果組合即可。問題是個別圖片中人的姿勢位置或光線條件等變動,會影響到人臉偵測的結果,因此連續的圖片在偵測後很可能出現不同的人臉位置判斷。因此除了在個別圖片的判斷結果外,還需要加入連續畫面的變動資訊,以輔助偵測結果之正確性與一致性。

我們的想法是使用視覺追蹤(visual tracking)的技巧來輔助解決連續畫面的人臉偵測問題。視覺追蹤的問題是指針對一個特定的目標,在影片中不斷的定位此目標所在位置。視覺追蹤的定位過程大約依據兩種資訊所影響:此目標在各畫面中出現的信心程度,以及連續畫面中此目標變動的狀態。將視覺追蹤的技術應用至影片中人臉偵測的問題,我們可以將問題轉換成:『在影片開始處偵測人臉的位置後,開始對人臉區域進行追蹤』。

由於人在影片中會有不同的動作,人臉的角度與位置也會有不同的變化,因此單純的在各圖片中進行人臉偵測會有很大的困難。在影片內連續的圖片之間人臉的變動狀況並不會太大,視覺追蹤的技術可以正確的克服各圖片人臉偵測正確度不足的問題。藉由視覺追蹤技術的引進,本研究所設計的人臉偵測技術,可以處理在移動過程中人臉所呈現的多角度變動,提高人臉偵測的判斷正確度。

本研究以粒子過濾(particle filtering)[4]的方法為基礎,設計一套影片的人臉偵測機制。我們在圖1呈現本研究規劃的處理流程。整個流程大略分成四個部分:初始設定(Initialization)、Condensation、範本資訊(Template)、以及校正(Correction)。一開始系統先進入初始設定區,取得影片中人臉區域的初始位置,以及後續追蹤作業所需的人臉範本資訊。影片進行中的人臉偵測工作主要在Condensation這個模組內完成,透過我們設計的動態程序(dynamic process)與觀察程序(observation process)整合出每個時間點的人臉位置。決定影片中每個人臉位置後,我們設計一套校正機制,剷除掉可能錯誤的判斷後,可以避免追蹤的過程被其他類似人臉的資訊引導產生錯誤的結果。

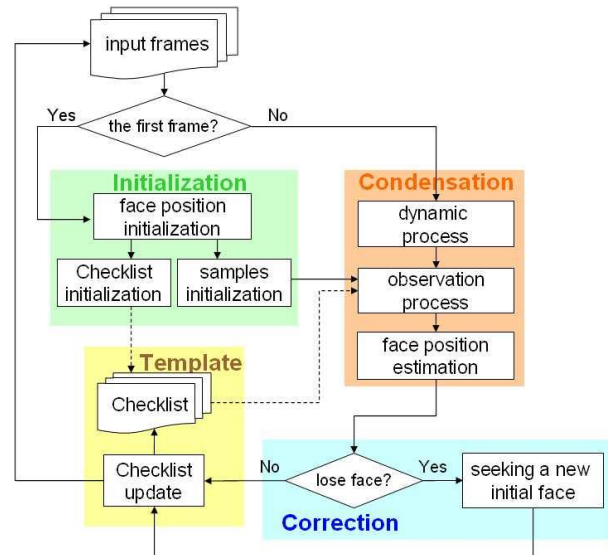


圖 1 本研究所規劃的流程圖

本論文的組織架構如下。除了本節的緒論外,第二節將對過去相關研究提供文獻供讀者參考,第三節詳述我們的研究架構與方法,在第四節內提供數個實驗結果,以驗證本研究所達成的成效,最後我們在第五節提出本研究的結論。

2. 相關文獻

視覺追蹤的問題可以視為在連續影像畫面中進行連續的臉部區域定位,並且將這些定位資訊組合成連續的人臉追蹤資訊。學者 Zheng 和 Bhandarkar [10]採用適性化粒子過濾(adaptive particle filtering)方法,此方法藉由臉部輪廓資訊來強化人臉偵測的工作。學者 Yao 和 Li [9]根據臉部特徵資訊設計一套動態程式規劃(dynamic programming)的方法以完成人臉偵測的工作。學者 Frank 等人[3]所設計的人臉偵測方法,首先對眼睛的位置定位,透過兩眼相對的位置決定臉部的角度,藉此推估出臉部區域位置。學者 Ross 等人[6]提出一種漸進式方法來修正傳統的 Eigenface 方法以偵測影片中的人臉位置。學者 Kim 等人[5]設計一套機率模型,透過粒子過濾機制[4]處理影片中人臉特徵資訊的改變狀態,以偵測並追蹤人臉的位置。

3. 研究架構與方法

在第一節中,我們已經針對圖1的處理流程給予簡單的介紹,這一節將對整個架構中的各單元詳細說明。

3.1 範本資訊

考慮同一個人的臉部在不同角度下會有不同的外觀，因此我們設計一套內涵個角度臉部特徵的清單(Checklist)資訊，提供人臉偵測與追蹤的過程中所需的範本比對工作。

假設系統可以處理 p 個不同的人臉角度，清單中將會有 p 組資訊，如圖 2 的例子所示，每一組包含對應角度的人臉特徵與此特徵所代表的信心分數。所謂的特徵代表的信心分數，是個別判斷人臉區域時所計算的特徵距離，此分數數字越小表示對應的臉部特徵有越大的信心是正確的人臉資訊。例如圖 2 的例子中，五個角度內頭尾兩個臉部特徵還沒出現，中間的臉部分數最好，代表未來追蹤判斷時可以用此範本當成此角度人臉位置判定時，將有更好的信心。清單中的臉部特徵分數，由第 3.3 節所描述的校正程序所計算出。






P_1	P_2	P_3	P_4	P_5
				
2.5	1.79785	0.66423	1.95621	2.5

圖 2 五個臉部角度的清單內容範例

3.2 Condensation 程序

Condensation 是 M. Isard and A. Blake [4] 兩位學者在 1998 年發表的一種蒙地卡羅抽樣方法。在視覺追蹤的研究主題上，已經有很多文獻採用 Condensation 的方法 [7][10]。在典型的 Condensation 的方法中包含三個主要部分：(1)動態程序針對下一個時間點決定可能的抽樣位置、(2)觀察程序計算每個抽樣位置與追蹤目標的相似程度、以及(3)追蹤目標位置估計。

表 1 本研究所設計的 Condensation 演算法

Procedure: Condensation
Input
$(s_{t-1}^i, w_{t-1}^i)_{i=1}^N$: N samples with weights at time $t-1$.
TR_{t-1} : the target position at time $t-1$.
Output
$(s_t^i, w_t^i)_{i=1}^N$: N new samples with weights at the current time t .
TR_t : the target position at the current time t .
// Dynamic Process
for $i = 1$ to N
compute the cumulative probability of each sample: $c_{t-1}^i = \sum_{j=1}^N w_{t-1}^j$.
end for
for $i = 1$ to N
generate a random number r and get the smallest k such that $c_{t-1}^k \geq r$.
$s_t^i \sim N(s_{t-1}^k, \sigma_t^2)$ // new samples are generated with a displacement.
end for
// Observation Process
for $i = 1$ to N
$w_t^i = \text{EuclideanDistance}(s_t^i, f_p)$
// f_p : the p -th face pose (classified at time $t-1$) in the checklist.
end if
Normalize $\{w_t^i\}_{i=1}^N$.
// Target Estimation
$TR_t = \sum_{i=1}^N w_t^i \times s_t^i$.
End Procedure

表 1 詳述本研究所設計的 Condensation 演算法，基本程序是在時間點 $t-1$ 的追蹤目標決定後，首先會根據目標位置以高斯分布方式取出時間 t 的抽樣位置 (動態程序)，每個抽樣位置取出固定大小區域後與範本內清單資訊比對，以計算抽樣位置是人臉的可能性 (觀察程序)，最後將所有的抽樣位置與對應的可能性計算期望值，就可以得到時間 t 的追蹤目標位置估計。

對於動態程序內的抽樣位置來說，我們會決定一個影像範圍代表人臉區域，我們以圖 3 為例來說明。在圖 3(a)中代表一個抽樣位置的決定後，會得到長寬值以決定要抽樣的影像區域，圖 3(b)代表此影像區域所呈現的人臉畫面。當然這只是一個完美的範例，大部份的抽樣位置都不會剛好是人臉區域，因此將取出各式各樣的影像區域，這些抽樣區域中只有非常接近臉部區域的樣本，才能在觀察程序中得到高分，進而決定整體追蹤目標之期望值位置。

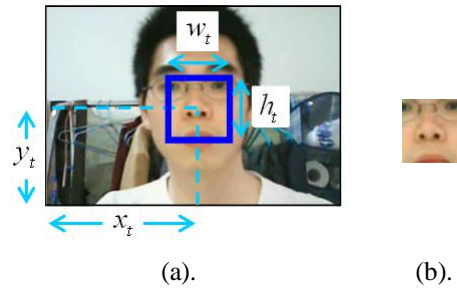


圖 3 臉部區域截取 (a): 每個抽樣位置決定一長寬值以取出影像區域 (b): 對應的臉部影像

3.3 人臉偵測結果校正

受到人臉位置移動、臉部角度不同、或者是四周環境光線的改變等因素，影片中人臉的呈現狀態有各種可能的變化，實務上要能夠有完全正確的人臉偵測方法是不可能。因此我們額外設計一個校正機制，以輔助前述偵測機制的成果。

針對每次所計算出的人臉區域估計位置，我們採用 LDA (Linear Discriminant Analysis) [2] 分類器另外設計一個人臉偵測器，對於 p 個不同的人臉角度產生 p 個對應的 LDA 分類器。在每個人臉追蹤的估計位置產生後，所推估的人臉區域透過 LDA 分類器計算此區域是人臉的相似分數，當分數過低時，就判定此次的人臉追蹤發生錯誤，系統將放棄這次追蹤的成果而重新啟動初始設定的機制，重新定位人臉的。

我們所設計的校正機制的在本質上是採用另一種方式，彌補原先人臉追蹤過程中可能產生的錯誤，這樣的做法可以避免單一方法的偏差。此外，在移動的過程中如果人臉被其他物體遮蔽時，不論哪一種追蹤方式都將失效，此時校正機制可以在人臉重新出現時再度啟動追蹤程序。

3.4 初始設定

系統剛開始運作時，或者是校正機制確認人臉追蹤失敗後，都將進行設定系統初始設定。初始設定的主要工作有二。首先，系統將在開始的畫面上執行單張影像的人臉偵測工作，這裡我們採用的是opencv函數庫[1]內所提供 AdaBoosting 方法 [8]，如果這張影像中無法找到人臉區域，就捨棄此影像而取得影片中下一張影像。接著系統將設定範本內的清單資訊，以便啟動 Condensation 的追蹤程序。

4. 實驗

為了能夠有效的驗證本研究設計的方法，我們自行拍攝一段影片，以求在不同的變因控制下觀察本研究方法的成效。我們的實驗影片所處理的人臉變動角度有 5 種 ($p=5$)，分別是正面、左右各 15 度、左右各 30 度。人臉在影片中移動狀態包含下面四種特性，其實驗結果顯示在後面的圖片中：

- 人臉正常移動：測試系統在人臉正常移動的狀態下偵測的效果，結果如圖 4 所示。
- 不同角度變化：測試系統在人臉角度改變的狀態下偵測的效果，結果如圖 5 所示。
- 前後移動造成人臉大小變化：測試系統在人臉前後移動而產生臉部區域大小改變時偵測的效果，結果如圖 6 所示。
- 遮蔽後的校正機制：測試系統在人臉被其他物體遮蔽後的偵測效果，結果如圖 7 所示。









frame id	抽樣粒子	追蹤結果
1081		
1087		
1096		
1099		

圖 4 人臉正常移動的狀態下偵測的效果









frame id	抽樣粒子	追蹤結果
114		
129		
141		
158		

圖 5 人臉角度改變的狀態下偵測的效果









frame id	抽樣粒子	追蹤結果
128		
151		
171		
187		

圖 6 人臉區域大小改變下偵測的效果








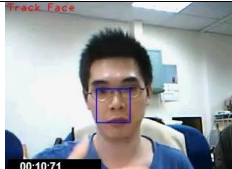
frame id	抽樣 粒子	追蹤 結果
622		
630		
639		
653		

圖 7 人臉被其他物體遮蔽後偵測的效果

圖 4-7 顯示四組實驗中四種特性的影片執行人臉偵測與追蹤的結果。大致上我們的方法都能正確追蹤到人臉的移動，只有圖 7 中人臉被遮蔽時，系統當然無法判斷人臉的位置，此時校正機制將協助系統在人臉出現後再重新啟動追蹤機制，這樣的設計可以避免在真實應用中，追蹤過程中發生錯誤時造成整個人臉追蹤的程序被中斷。

5. 結論

本研究設計一套以粒子過濾追蹤為基礎的機制，處理影片中人臉偵測與追蹤。我們所設計的方法包括四個主要部分：初始設定以便取得影片中人臉所在的初始位置並記錄相關資訊、Condensation 以完成粒子過濾追蹤的主要機制、範本資訊以紀錄人臉在移動過程中各種不同角度、以及校正程序以便對於追蹤目標錯誤或遺失時能夠重新啟動人臉追蹤程序。本論文同時完成相關的實驗，顯示本研究所設計的方法，面對不同類型的人臉移動影片時，仍能順利完成人臉偵測與追蹤的工作。

在未來研究方面，我們預期修正範本資訊所記錄的人臉角度，期望能偵測更多類型的人臉移動，同時能提升追蹤的正確度。此外，人臉偵測是一系列進階應用的第一步，我們希望能夠透過現有工作的成果，更進一步的發展後續相關應用，初步規劃的研究主題是針對追蹤的人臉區域，辨識出此人臉的身份資訊，完成更具實用性的研究內容。

參考文獻

- [1] The opencv library. <http://sourceforge.net/projects/opencvlibrary/>.
- [2] R.O. Duda, P.E. Hart, and D.G. Stork, (2001), Pattern classification. In second ed., John Wiley and Sons, Inc.
- [3] W. Frank, A. Markus, and R. Gerhard, (2006), "Multimodal face detection, head orientation and eye gaze tracking", in Proceedings of IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, pages 13–18.
- [4] M. Isard and A. Blake, (1998), "Condensation - conditional density propagation for visual tracking", International Journal Computer Vision, 29(1):5–28.
- [5] M. Kim, S. Kumar, V. Pavlovic, and H. Rowley, (2008), "Face tracking and recognition with visual constraints in real-world videos", in Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR.
- [6] D.A. Ross, J. Lim, R. Lin, and M. Yang, (2008), "Incremental Learning for Robust Visual Tracking", International Journal of Computer Vision, vol. 77, pp. 125-141.
- [7] R.C. Verma, C. Schmid, and K. Mikolajczyk, (2003), "Face detection and tracking in a video by propagating detection probabilities", IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(10):1215–1228.
- [8] P. Viola and M. Jones, (2001), "Rapid object detection using a boosted cascade of simple features", in Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR, volume 1.
- [9] Z. Yao and H. Li, (2006), "Tracking a detected face with dynamic programming", Image and Vision Computing, 24(6):573–580.
- [10] W. Zheng and S. M. Bhandarkar, (2009), "Face detection and tracking using a boosted adaptive particle filter", Journal of Visual Communication and Image Representation, 20:9–27.

