

基於長期關連之指定圖片物件之影像追蹤系統

葉承濤 游象甫

國立臺北教育大學資訊科學系

摘要——為利於犯罪調查或得到恐怖份子的行蹤，可以監視器取得影片，比對可疑物件於影片出現的位置及時間，幫助犯罪偵查，所以本文提出一個影片物件追蹤系統，使用者先指定一張圖片中物件，再提供欲追蹤的影片資料，本系統會自動搜尋影片中出現該物件的 frame，並標出該物件出現的位置。本系統先搜尋影片中第一個出現指定物件的 frame，再以此 frame 為基礎，搜尋出現在其他 frame 的指定物件；本系統已經實作並以 OTB-100 的影片資料庫測試其可行性及效能表現。

關鍵詞：物件追蹤、CAMShift、Long-Term Correlation Tracking。

一、緒論

近年由於電腦及網路科技的進步，為預防犯罪或得知交通資訊，許多公共場所或重要路口均安裝監視器，並經行動通訊網路或無線網路將影片資料送回控制中心。早期有關影片內容大都以人工判讀，但如果影片數目很多或時間很長時，除需花費大量時間檢驗外，也容易因為注意力降低造成準確度不佳或漏掉重要畫面，因此以程式自動化判讀影片是目前的趨勢，而其中一個自動化判讀影片的議題是追蹤物件出現在影片中的位置。

物件辨識及追蹤的方法可分為四類，第一是以連續多個 frame 中的像素是否改變做為追蹤判斷的基準，例如：時間差異法(Temporal Differencing) [1]、背景相減法(Background Subtraction) [2]、光流法(Optical Flow) [3]；第二是計算影像直方圖及機率分布的方式，例如：Mean Shift [4]、Continuously Adaptive Mean Shift (CAMShift) [5]，其優點就是較不受環境因素影響，但識別多重物件較為困難；第三是尋找影像中的特徵值，然後藉此特徵值追蹤物件，例如：Scale-Invariant Feature Transform (SIFT) [6]，此策略對物件變形或遮擋有較佳的判斷能力，但特徵值的學習可能需時較長；第四類為事先以資料庫中的資料訓練物件，例如：Support Vector Machine [7]，只要經過足夠的訓練就可以達到不錯的成功率，但資料必須事前準備，且訓練樣本的時間較長。

基於上述技術，本文提出一個影片物件追蹤系統，使用者先指定一張圖片中物件，再提供欲追蹤的影片資料，本系統會自動搜尋影片中出現該物件的 frame，並標出該物件出現的位置。本系統運作分為兩個部分，第一部分先搜尋影片中第一個出現指定物件的 frame，第二部分再以此 frame 為基礎，搜尋出現在其他 frame 的指定物件；有關搜尋影片中第一個出現指定物件的

frame 的部分，本文利用 CAMShift，透過計算圖片中指定物件 HSV 空間下的色調分量值直方圖，並透過直方圖反向投影得到目標像素的分布，再以 SIFT 演算法提取 CAMShift 處理過的資料中的特徵值，接著搜尋影片每個 frame 找出第一個出現指定物件的 frame 並框定該指定物件；第二部分則基於 Chao Ma 提出的 Long-Term Correlation Tracking [8] 物件追蹤演算法，對第一個 frame 框定的區域進行關連濾波(Correlation Filters) [9]，將目標物形變進行分析找出可靠性高的學習區域，並在跟蹤失敗之後可以在學習區域內重新檢測對象；提出的系統已經實作並以影片測試其可行性。

本文其餘內容如後：下節為相關研究探討，第三節描述系統架構及演算法，第四節為實作及效能評估；最後是簡單的結論。

二、相關研究

由於需要對視訊中移動物件進行處理，在無動量偵測的情況下須對整張影像進行分析，而存在動量偵測的情況下，則只需對處理過後的影像進行分析，所以動量偵測有助於降低每張影像所需計算量，常見的影像動量偵測方法有三種：時間差異法、背景相減法、光流法。

時間差異法又稱作影格差異法(Frame Differencing) [1]，原理是利用在時間上連續的影像做像素相減，若是兩者相減為零，就表示此像素不屬於移動物件像素，若是兩者相減不為零，則此像素為移動物件像素；時間差異法物體沒有移動但被偵測出有移動機率小，同時可快速正確偵測移動物件，對複雜背景有高辨別性；惟發生移動物件沒有被偵測到機率大，容易偵測到殘影。

背景相減法[2]將背景資料與欲追蹤的物件影像相減，若差值為零，則無物件移動，若不為零則有物件產生，藉此偵測出原本不在此背景資料的物件；背景相減法偵測物件之資訊完整，具有低誤判率，殘影出現機率較低，但不適用於動態背景以及運算量較時間差異法高。

不同於前述兩種方法以像素值為判斷依據，當影像中有異物進入時，其亮度會產生變化，光流則是指影像亮度的表面變化分布的情況，光流法[3]運用像素的梯度變化去判斷畫面物體的移動；光流法的偽陽性失誤率低，可快速的偵測移動中物件，同時對複雜背景具有高容忍度，但不易正確偵測平緩移動的物件，也需要大的運算量。

物件偵測及追蹤時需要比較其特徵，而特徵表示法主要分為兩種：哈爾特徵法(Haar-Like Feature) [10]，以及方向梯度直方圖(Histogram of Oriented Gradient)

[11]。哈爾特徵法提出積分圖的方式來表示物件，當分析物件特徵值時，可降低計算物件像素和所需時間，以圖 1 為例，圖中每個方格代表一個像素值，圖 1(a)藍色區域的特徵值為所有像素值和，即為 67，以哈爾特徵法可將其轉為圖 1(b)中藍色的部分。

2	3	5	4	9
10	5	9	6	3
10	5	4	9	3
7	2	5	9	6
3	2	5	4	4

(a)

2	5	10	14	23
12	20	34	44	56
22	35	53	72	87
29	44	67	86	107
32	49	73	105	133

(b)

圖 1：(a)為一 5x5 之灰階影像；(b)為以(a)計算而得之積分圖
方向梯度直方圖則利用影像的局部梯度提取特徵，梯度計算通常會利用 $[-1,0,1]$ 與 $[-1,0,1]^T$ 進行迴旋積 (Convolution)。

以下為常見物件追蹤的技術，CAMShift 以均值飄移為基礎，均值漂移主要用在視覺跟蹤，作法為從反投影直方圖的概率圖，得到目標影像出現在原始影像各位置的機率，假設已知物體的初始位置，從這位置迭代移動來得到最大概率位置，以此得到目標影像的精確位置，於是初始位置設定會影響計算結果和花費時間，CAMShift 主要步驟為：

1. 轉換為 HSV 色彩空間，其中 H(Hue)為色調、S(Saturation)為飽和度、V(Value)為亮度。
2. Back Projection：根據 HSV 色彩空間濾去背景。
3. 利用 Mean Shift 進行計算：透過投影計算出兩圖之目標物中心點並學習其旋轉角度。
4. 集中收斂區域並標示。
5. 重複進行步驟 3 及 4。

關連濾波器(Correlation Filters) [9]的原理是由於每一個 frame 被檢測的目標都提供描述該目標的資料，因此可以用每一個 frame 的目標區域作為訓練樣本；實做方法是透過追蹤的若干個 frame 的目標位置，提取有興趣的特徵，訓練得到一個濾波器模板，對於新的一個 frame 中可能的目標區域，計算該區域特徵，與濾波器模板做相關性判斷，並得到一關連值，根據關連值得到新 frame 中目標的預測位置，並以該位置中心得到特徵值，回饋訓練濾波器模型，重複上述步驟可實現線上

訓練與目標即時追蹤。

三、系統架構

本系統分為兩部分，如圖 2，第一部分先搜尋影片中第一個出現指定物件的 frame，第二部分再以此 frame 為基礎，搜尋出現在其他 frame 的指定物件；有關搜尋影片中第一個出現指定物件的 frame 的部分，本文利用 CAMShift，透過計算圖片中指定物件 HSV 空間下的色調分量值直方圖，並透過直方圖反向投影得到目標像素的分布，再以 SIFT 演算法提取 CAMShift 處理過的資料中的特徵值，接著搜尋影片的每個 frame 找出第一個出現指定物件的 frame 並框定該指定物件；第二部分則基於 Chao Ma 提出的 Long-Term Correlation Tracking 物件追蹤演算法，對第一個 frame 框定的區域進行關連性濾波，計算框定區域之特徵，並由此訓練得到濾波器模板，藉此對目標物做關連性判斷，再透過關連值進行分析找出可靠性高的學習區域，並在追蹤失敗後可以在學習區域內重新檢測。

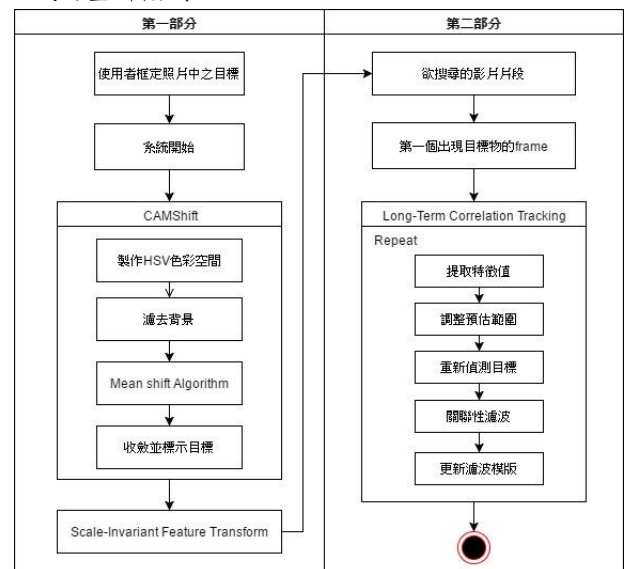


圖 2：本文提出之系統架構

有關系統第一部分的詳細說明如下，首先取出參考照片並以 CAMShift 進行該照片資料分析以減少資料量，取出目標物件的 HSV 色彩空間，再以目標物的 HSV 色彩空間為模板濾去該張目標物照片的背景，接著利用 MeanShift 計算與學習目標物件旋轉後的特徵。MeanShift 在初始化時需要人工標定待追蹤的目標區域，以目標位置為起始中心位置，目標長寬的一半分別作為搜索窗口的長寬，計算窗口區域內的顏色直方圖分佈作為候選目標，利用相似度量 Bhattacharyya [12] 係數判斷目標模型和當前候選目標的相似程度，該係數越大表明候選目標與目標模型越相似，通過不斷疊代計算，當該係數達到最大時，目標中心即收斂到目標的真實位置。得到物件位置後便利用 SIFT 提取影片的 frame 的參考物件邊緣特徵與照片中的資料進行比對由於 SIFT 是基於物體的局部外觀的特徵點而與影像的大小和旋轉無關，因此對於光線、雜訊、視角改變的容忍度也相當高。

系統第二部分主要針對第一個框定目標的 frame 進行分析及追蹤，基於 Chao Ma 提出的 Long-Term Correlation Tracking 物件追蹤演算法進行追蹤，特徵值採用方向梯度直方圖，計算特徵值的過程會先將目標灰階化並採用 Gamma 校正 [13]對輸入的目標物進行顏色空間標準化，以降低陰影和光照造成的影響；接著對圖像內每個像素進行梯度計算，藉此得到輪廓資料，並進一步減少光照變化對目標物的干擾；將目標物分成多個 cells(本文將目標物分成 6*6 像素/cell)，統計每個 cell 的梯度直方圖，即得到對每個 cell 的描述，再將 3*3 個 cell 組成一個 block，每個 block 內所有 cell 的特徵描述合起來便可得到該 block 內的方向梯度直方圖特徵描述，將目標物內所有 block 的方向梯度直方圖特徵合起來得到該目標的方向梯度直方圖特徵描述；利用此特徵描述作為分類的特徵向量並製作濾波模版，針對此物件的中心建立關連性濾波點並據此更新下張 frame 的濾波模版，以此重複即可準確追蹤該目標物位置直至影片結束。

四、效能分析

表 1：系統開發及實驗環境

硬體設備	
CPU	intel core i5-4590 CPU @ 3.30ghz
RAM	12.0 GB
系統開發環境	
作業系統	Microsoft Windows 10 64-bits
開發軟體	MATLAB R2016a 64-bits
使用函式庫	OpenCV Library

本系統主要以 MATLAB 開發，並使用 OpenCV 函式庫，其他硬體及作業系統規格如表 1。物件追蹤的資料庫是利用 Object Tracking Benchmark (OTB)-100 [14]進行檢測，其中提供 89 種不同影片及其預先訓練的數據。本文實驗取 OTB-100 中 4 種影片進行追蹤，並與 OTB-100 中對此 4 種影片預先訓練的數據進行精準度比較，主要針對影片中每張 frame 所框出之位置差異做為比較依據。

本研究以 4 個實驗測試提出方法的效果，實驗一及實驗二中測試提出的物件追蹤方法是否會因為距離，即尺度的改變影響框定效果；實驗三及實驗四主要針對目標物進行旋轉對提出方法框定效果的影響。

實驗一(圖 3)中框定的範圍會隨著目標物距離增加而變小，由於本系統對於中心點的計算會將原本顏色相似的背景陰影加入濾波資料中，因此可看到實驗一中提出追蹤方法的圖 3(g)及(h)將背景色塊也加入濾波資料內，所以出現框定範圍大於 OTB-100 資料庫所提供的圖 3(c)與(d)範圍，但目標物仍舊落在所框定之範圍內。

實驗二(圖 4)中除了目標物變小外也從影片左側移動

至右側，由圖 4(e)至(h)可知提出方法對於快速移動目標可得到良好效果；同時圖 4(g)及(h)對顏色相近區域部分相對 OTB-100 資料庫所提供的圖 4(c)及(d)仍有改善之空間，惟本法的目標物仍位於框定範圍內。

實驗三(圖 5)中目標物會進行旋轉，比較本法得到的圖 5(e)至(h)與 OTB-100 資料庫的圖 5(a)至(d)，可得知提出方法在目標進行旋轉後仍可抓取目標物之位置並框定，與 OTB-100 資料庫的結果相差可不遠。

實驗四(圖 6)中目標物除進行旋轉、移動外也會有些微遮擋現象，對照 OTB-100 的圖 6(c)，本法求得的圖 6(g)出現誤判，這是因為目標物快速反轉造成提出方法無法透過濾波器正確偵測目標物邊緣，但透過濾波器中心點模版更新後，當目標物件再次反轉後則可正確追蹤，如圖 6(f)所示。



圖 3：測試目標物尺度變化對追蹤物件的影響，其中(a)(b)(c)(d)為 OTB-100 影像，(e)(f)(g)(h)為本法得到影像

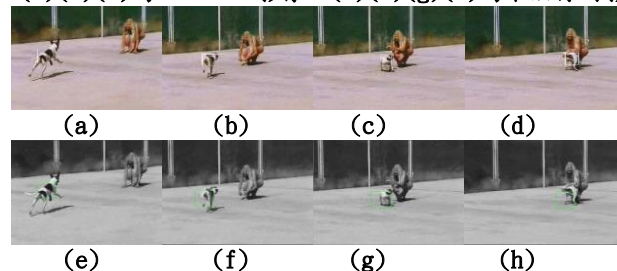


圖 4：測試目標物尺度變化對追蹤物件的影響，其中(a)(b)(c)(d)為 OTB-100 影像，(e)(f)(g)(h)為本法得到影像

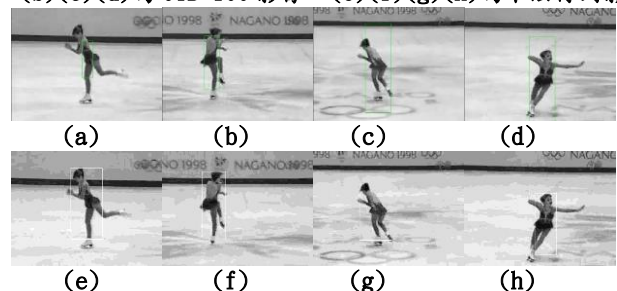


圖 5：測試目標物旋轉形變對追蹤物件的影響，其中(a)(b)(c)(d)為 OTB-100 影像，(e)(f)(g)(h)為本法得到影像

圖 7(a)至(d)分別是實驗一至實驗四影片的準確度(y 軸)對失誤容忍閾值(x 軸)的關係圖，其中準確度為提出方法與 ODB-100 結果比對之差距，而閾值單位為框定像素的像素範圍差，圖 7(a)至(d)顯示當容忍閾值愈高，其準確度會提升，當失誤容忍閾值升到 20 至 30 像素框定範圍差距時，提出方法表現良好；另圖 7(a)及(c)的結果揭露本方法對較大目標物表現較好，由圖 7(b)可知提出方法對較小目標物也達到一定準確度；圖 7(d)表示出本文對於快速形變之物體仍有改善空間。

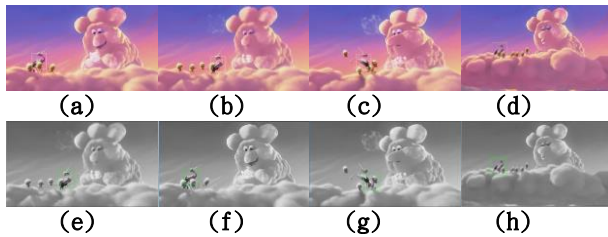


圖 6：測試目標物旋轉形變及遮擋對追蹤物件的影響，其中 (a)(b)(c)(d) 為 OTB-100 影像，(e)(f)(g)(h) 為本法得到影像

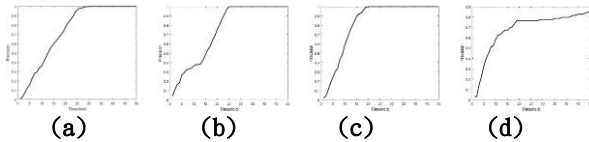


圖 7：失誤容忍閾值對提出方法準確度的影響

結論

本文提出一個影片物件追蹤系統，使用者先指定一張圖片中物件，再提供欲追蹤的影片資料，本系統會自動搜尋影片中出現該物件的 frame，並標出該物件出現的位置。本系統先搜尋影片中第一個出現指定物件的 frame，再以此 frame 為基礎，搜尋出現在其他 frame 的指定物件；有關搜尋影片中第一個出現指定物件的 frame 的部分，本文利用 Continuously Adaptive Mean Shift (CAMShift)，再以 Scale-Invariant Feature Transform (SIFT) 演算法提取 CAMShift 處理過的資料中的特徵值，接著搜尋影片每個 frame 找出第一個出現指定物件的 frame 並框定該指定物件；第二部分則基於 Chao Ma 提出的 Long-Term Correlation Tracking 物件追蹤演算法，對第一個 frame 框定的區域進行關連性濾波，將目標物形變進行分析找出可靠性高的學習區域；本系統已經實作並以 OTB-100 的影片資料庫測試其可行性及效能表現。

參考文獻

- [1] Tsitsiklis, J. N., & Van Roy, B., "An analysis of temporal-difference learning with function approximation," IEEE Transactions on Automatic Control, 42(5), 674-690, 1997.
- [2] Piccardi, M., "Background subtraction techniques: a review," 2004 IEEE International Conference on Systems, Man and Cybernetics, Vol. 4, pp. 3099-3104, October 2004.
- [3] Barron, J. L., Fleet, D. J., & Beauchemin, S. S., "Performance of optical flow techniques," International Journal of Computer Vision, 12(1), pp. 43-77, 1994.
- [4] Comaniciu, D., & Meer, P., "Mean shift: A robust approach toward feature space analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(5), pp. 603-619, 2002.
- [5] Allen, J. G., Xu, R. Y., & Jin, J. S., "Object tracking using camshift algorithm and multiple quantized feature spaces," Pan-Sydney Area Workshop on Visual information Processing, pp. 3-7. June 2004.
- [6] Lowe, D. G., "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, 60(2), pp. 91-110, 2004.

- [7] Suykens, J. A., & Vandewalle, J., "Least squares support vector machine classifiers," Neural Processing Letters, 9(3), pp. 293-300, 1999.
- [8] Ma, C., Yang, X., Zhang, C., & Yang, M. H., "Long-term correlation tracking," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5388-5396, 2015.
- [9] Bolme, D. S., Beveridge, J. R., Draper, B. A., & Lui, Y. M., "Visual object tracking using adaptive correlation filters," IEEE Conference on Computer Vision and Pattern Recognition, pp. 2544-2550, 2010.
- [10] Viola, P., & Jones, M., "Rapid object detection using a boosted cascade of simple features," 2001 IEEE Conference on Computer Vision and Pattern Recognition, Vol. 1, pp. I-I, 2001.
- [11] Zhang, G., Gao, F., Liu, C., Liu, W., & Yuan, H., "A pedestrian detection method based on SVM classifier and optimized histograms of oriented gradients feature," 2010 Sixth International Conference on Natural Computation, Vol. 6, pp. 3257-3260, 2010.
- [12] Guorong, X., Peiqi, C., & Minhui, W., "Bhattacharyya distance feature selection," 13th International Conference on Pattern Recognition, Vol. 2, pp. 195-199, 1996.
- [13] Farid, H., "Blind inverse gamma correction," IEEE Transactions on Image Processing, 10(10), pp. 1428-1433, 2001.
- [14] http://cvlab.hanyang.ac.kr/tracker_benchmark/datasets.html