

行政院國家科學委員會專題研究計畫 成果報告

於擁擠人群中以特徵點為基礎的目標物追蹤系統 研究成果報告(精簡版)

計畫類別：個別型
計畫編號：NSC 100-2221-E-216-030-
執行期間：100年08月01日至101年07月31日
執行單位：中華大學資訊工程學系

計畫主持人：連振昌
共同主持人：周智勳
計畫參與人員：碩士班研究生-兼任助理人員：王建翔
碩士班研究生-兼任助理人員：陳彥帆

報告附件：出席國際會議研究心得報告及發表論文

公開資訊：本計畫可公開查詢

中華民國 101 年 10 月 31 日

中文摘要： 在此計畫中，為了在擁擠場景中偵測個別的目標物以及分析群體移動所產生的軌跡資訊，我們提出兩個方法分別在擁擠環境中偵測、追蹤個別的目標物，以及將群體運動軌跡進行偵測分類。第一個子系統所提出的方法是以 corner 特徵點為基礎，發展由粗糙至細緻的獨立目標物切割演算法。在這個方法中，首先利用 C-Means 演算法將特徵點做粗略的分群，接著透過時空最小生成樹(spatial-temporal shortest spanning tree)在每一個移動分群內切割出更精確的個體，最後運用 corner 特徵點繼承的概念來追蹤所有移動個體目標物。第二個子系統所提出的方法為應用 longest common subsequences 演算法自動估算人群運動軌跡間之相似性，逐一地被分類到最適當的類別中。所有特徵點軌跡相似性量測是透過空間及時間上的相似性做測量。從實驗的結果發現，在擁擠環境中分割獨立移動目標物的準確度可高於 90% 以上，效能部分可每秒約處理 10 張 frame。

中文關鍵詞： corner 特徵點、時空最小生成樹、分割獨立移動目標物、運動軌跡間相似性

英文摘要： In order to detect each individual target in the crowded scenes and analyze the crowd moving trajectories, we propose two methods to detect and track the individual target in the crowd and classify the crowd motion trajectories. First, a coarse-to-fine individual segmentation approach based on the corner points' extraction and tracking is proposed. The dynamic feature points are roughly clustered by the C-means algorithm and then a spatial-temporal shortest spanning tree is proposed to segment each individual target in the moving group and each target is tracked with the concept of points' inheritance. Second, the method of the longest common subsequences is applied to automatically evaluate the similarities among the feature tracks. Then the feature tracks are classified by the similarity measured on both the temporal and spatial relationships. The experimental results show that the accuracy of individual segmentation in the crowd can be higher than 90% and the efficiency of our system can approach 10 fps.

英文關鍵詞： corner point, spatial-temporal shortest spanning tree, individual segmentation, trajectory classification

(二)中、英文摘要及關鍵詞 (keywords)。

在此計畫中，為了在擁擠場景中偵測個別的目標物以及分析群體移動所產生的軌跡資訊，我們提出兩個方法分別在擁擠環境中偵測、追蹤個別的目標物，以及將群體運動軌跡進行偵測分類。第一個子系統所提出的方法是以 corner 特徵點為基礎，發展由粗糙至細緻的獨立目標物切割演算法。在這個方法中，首先利用 C-Means 演算法將特徵點做粗略的分群，接著透過時空最小生成樹 (spatial-temporal shortest spanning tree) 在每一個移動分群內切割出更精確的個體，最後運用 corner 特徵點繼承的概念來追蹤所有移動個體目標物。第二個子系統所提出的方法為應用 longest common subsequences 演算法自動估算人群運動軌跡間之相似性，逐一地被分類到最適當的類別中。所有特徵點軌跡相似性量測是透過空間及時間上的相似性做測量。從實驗的結果發現，在擁擠環境中分割獨立移動目標物的準確度可高於 90% 以上，效能部分可每秒約處理 10 張 frame。

關鍵字: corner 特徵點、時空最小生成樹、分割獨立移動目標物、運動軌跡間相似性。

In order to detect each individual target in the crowded scenes and analyze the crowd moving trajectories, we propose two methods to detect and track the individual target in the crowd and classify the crowd motion trajectories. First, a coarse-to-fine individual segmentation approach based on the corner points' extraction and tracking is proposed. The dynamic feature points are roughly clustered by the C-means algorithm and then a spatial-temporal shortest spanning tree is proposed to segment each individual target in the moving group and each target is tracked with the concept of points' inheritance. Second, the method of the longest common subsequences is applied to automatically evaluate the similarities among the feature tracks. Then the feature tracks are classified by the similarity measured on both the temporal and spatial relationships. The experimental results show that the accuracy of individual segmentation in the crowd can be higher than 90% and the efficiency of our system can approach 10 fps.

Keywords: corner point, spatial-temporal shortest spanning tree, individual segmentation, trajectory classification.

(三)報告內容：包括前言、研究目的、文獻探討、研究方法、結果與討論（含結論與建議）等。

3.1 前言

在此計畫中，為了在擁擠場景中偵測個別的目標物以及分析群體移動所產生的軌跡資訊，我們提出兩個方法分別在擁擠環境中偵測、追蹤個別的目標物，以及將群體運動軌跡進行偵測分類。第一個子系統所提出的方法是以 corner 特徵點為基礎，發展由粗糙至細緻的獨立目標物切割演算法。在這個方法中，首先利用 C-Means 演算法將特徵點做粗略的分群，接著透過時空最小生成樹 (spatial-temporal shortest spanning tree) 在每一個移動分群內切割出更精確的個體，最後運用 corner 特徵點繼承的概念來追蹤所有移動個體目標物。第二個子系統所提出的方法為應用 longest common subsequences 演算法自動估算人群運動軌跡間之相似性，逐一地被分類到最適當的類別中。所有特徵點軌跡相似性量測是透過空間及時間上的相似性做測量。

3.2 研究目的及文獻探討

在傳統 blob-based 物件偵測系統有一些提取移動中目標的方法，例如：背景相減法 (background subtraction) [2]，光流法 (optical flow) [3, 9-10]，畫面差異分析法 (frame difference analyses) [4] 和編碼模型法 (codebook model) [5]。在 [2]，用背景相減法 (background subtraction) 在區域中可以準確的抓取移動中的目標物，但是它對於照明的變化和不斷變化的動態背景過於敏感。在 [3, 9-10]，光流法 (optical flow) 是用於獨立追蹤各別之目標，但是其計算量過於龐大。應用畫面差異分析 [4] 可以適應照明的變化，但當目標移動緩慢時，目標物的偵測會不完整。在 [5]，編碼法 (codebook model) 可以克服背景改變和照明變化的問題，但這個方法在擁擠的人群中要發現和追蹤目標物是很困難的。

在一般的影像監控應用當中，具有挑戰性之一的問題就是在 Fig.1 這種擁擠的鏡頭中進行目標的追蹤。一般來說，嚴重的遮蔽現象會使傳統 blob-based 物件偵測/追蹤的方法失敗。例如，追蹤的目標物在擁擠的車站或廣場，被追蹤的人有部分被其他人遮住，只有部分區域可以被作為追蹤的線索。因此，在擁擠的人群中追蹤一個單獨的目標我們可能會面臨兩個主要的問題：1) 當我們在監視較大空間中人群的移動，目標物可能太小 2) 頻繁的遮蔽使得目標分割非常困難。因此，特徵點偵測及追蹤經過評估後，認為適用於解決在人群中追蹤目標的問題。即使在擁擠的場面中，可能仍然有一些區域的特徵點沒有被遮蔽。在 Fig.2-(b), 的例子中為利用 blob-based 物件偵測的方法偵測目標物的圖例。很明顯的，在前景區域幾個合併的目標難以被分割。相反的，在 Fig.2-(c) 中，顯示出特徵點 (point-based) 方法適用在人群中偵測與追蹤獨立目標的可能性。根特徵點追蹤相同的概念，Brostow 跟 Cipolla [7] 提出 Bayesian clustering algorithm 可以分割人群中每個獨立的個體與時空距離和軌跡的一致性。然而，執行效能並不高。在本計畫中，我們提出了一個新的方法來偵測和追蹤人群中的個別目標物。

在此計畫中，為了在擁擠場景中偵測個別的目標物以及分析群體移動所產生的軌跡資訊，我們提出兩個方法分別在擁擠環境中偵測、追蹤個別的目標物，以及將群體運動軌跡進行偵測分類。第一個子系統所提出的方法是以 corner 特徵

點為基礎，發展由粗糙至細緻的獨立目標物切割演算法。在這個方法中，首先利用 C-Means 演算法將特徵點做粗略的分群，接著透過時空最小生成樹 (spatial-temporal shortest spanning tree) 在每一個移動分群內切割出更精確的個體，最後運用 corner 特徵點繼承的概念來追蹤所有移動個體目標物。第二個子系統所提出的方法為應用 longest common subsequences 演算法自動估算人群運動軌跡間之相似性，逐一地被分類到最適當的類別中。



Fig. 1 (a) 在隧道擁擠場面。(b) 在中華大學的擁擠場景。(c) 測試視訊檔中之人群遮蔽景象。

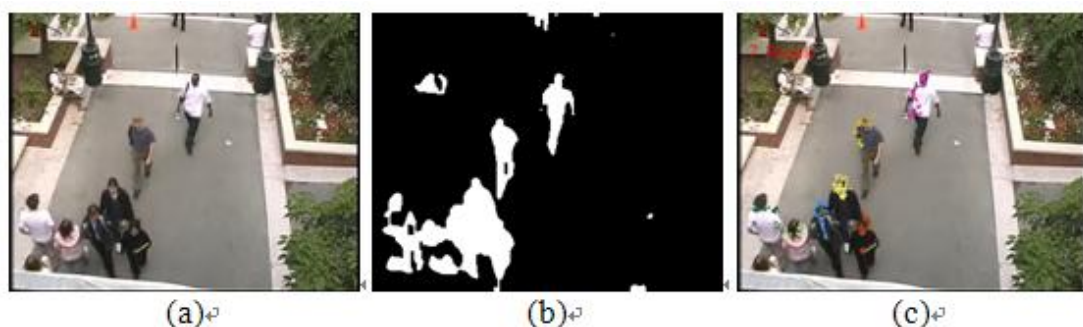


Fig. 2 (a) 測試視訊檔中之人群遮蔽景象。(b) blob-based background subtraction 方法的前景偵測 (c) 採用 point-based 特徵點之目標偵測和追蹤，每種顏色代表個別目標。

參考文獻

- [1] J. Rosen, "A Cautionary Tale for A New Age of Surveillance," *The New York Times Magazine*, Oct. 2001.
- [2] C. R. Jung, "Efficient Background Subtraction and Shadow Removal for Monochromatic Video Sequences," *IEEE Transactions on Multimedia*, Vol. 11, No. 3, April 2009, pp. 571-577.
- [3] J. Shi and C. Tomasi, "Good Features to Track," *Proceedings of the IEEE Conference Computer Vision and Pattern Recognition*, June 1994, pp. 593-600.
- [4] S. Wang, X. Wang, and H. Chen, "A Stereo Video Segmentation Algorithm Combining Disparity Map and Frame Difference," *International Conference on Intelligent System and Knowledge Engineering*, Vol. 1, Nov. 2008, pp. 1121-1124.
- [5] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, "Real-Time Foreground-Background Segmentation Using Codebook Model," *Real-Time Imaging*, Vol. 11, No. 3, June 2005, pp. 172-185.
- [6] A. M. Cheriadat and R. J. Radke, "Detecting Dominant Motions in Dense Crowds," *IEEE Journal of Selected Topics in Signal Processing*, Vol. 2, No. 4, Aug. 2008, pp. 568-581.
- [7] G. Brostow and R. Cipolla, "Unsupervised Bayesian Detection of Independent Motion in Crowds," *Proceedings of the IEEE Conference Computer Vision and Pattern Recognition*, Vol. 1, June 2006, pp. 594-601.
- [8] C. Tomasi and T. Kanade, "Detection and Tracking of Point Features," *Carnegie Mellon University Technical Report CMU-CS-91-132*, April 1991.
- [9] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, Vol. 60, No. 2, Jan. 2004, pp. 91-110.

- [10] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "Speeded-Up Robust Features (SURF)," *Proceedings of the 9th European Conference on Computer Vision*, Vol. 3951, May 2006, pp. 404-417.
- [11] O. J. Morris, M. J. Lee, A. G. Constantinides, "Graph theory for image analysis: an approach based on the shortest spanning tree," *IEE Proceedings F. Communications, Radar and Signal Processing*, Vol. 133, No. 2, April 1986, pp. 146-152.
- [12] T. Zhao and R. Nevatia. "Tracking Multiple Humans in Complex Situations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 9, Sep. 2004, pp. 1208-1221.

3.3 研究方法

一般來說，傳統在人群中目標偵測的在分割過程可能有幾個問題。第一，用背景相減法(background subtraction)[1]在擁擠的鏡頭中很難找到準確的邊界。第二，以監督式學習(supervised learning)特定的背景模型(subject-specific model)[5]需要花費較大之訓練計算量。第三，在壅擠人群中移動的物體可能有不同的移動方向，但分分合合的情形非常頻繁。為了解決這些問題，我們提出一種以 corner 特徵點為基礎發展一項從粗糙到細緻的目標物分割技術以解決在人群中的目標偵測及追蹤技術。首先利用 C-means 演算法將落在移動物體上之動態特徵點做初步概略性之分群，然後利用時空中最短的生成樹(spatial-temporal shortest spanning tree)以及行人幾何特徵執行獨立目標物之分割。獨立目標物之分割之後，持續更新特徵點與持續追蹤特徵點以追蹤每一個所切割出來之獨立目標物。

在此計畫中，為了在擁擠場景中偵測個別的目標物以及分析群體移動所產生的軌跡資訊，我們提出兩個方法分別在擁擠環境中偵測、追蹤個別的目標物，以及將群體運動軌跡進行偵測分類。第一個子系統所提出的方法是以 corner 特徵點為基礎，發展由粗糙至細緻的獨立目標物切割演算法。在這個方法中，首先利用 C-Means 演算法將特徵點做粗略的分群，接著透過時空最小生成樹(spatial-temporal shortest spanning tree)在每一個移動分群內切割出更精確的個體，最後運用 corner 特徵點繼承的概念來追蹤所有移動個體目標物。第二個子系統所提出的方法為應用 longest common subsequences 演算法自動估算人群運動軌跡間之相似性，逐一地被分類到最適當的類別中。本計畫所提出之系統流程圖如 Fig. 3 所示，以下逐一說明此計畫之研究步驟。

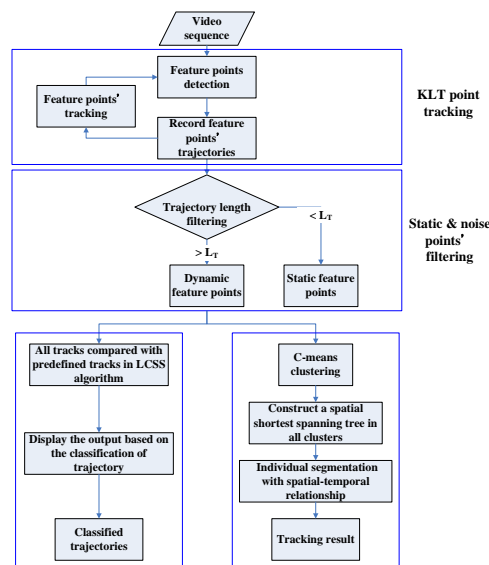


Fig. 3 系統流程圖。左半邊描述群眾運動軌跡分類，右半邊描述人群中目標物切割及追蹤技術

3.3.1 擁擠人群中之目標目切割及追蹤

此子系統之之流程圖如 Fig. 4 所描述，以下詳述之。

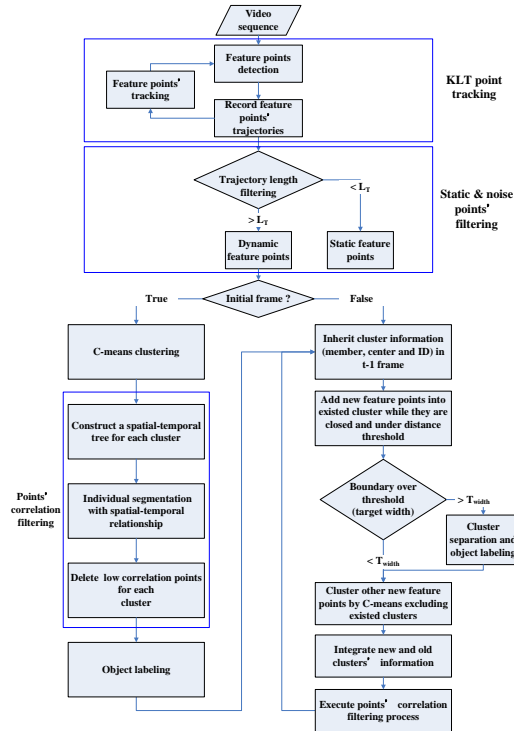


Fig. 4 擁擠人群中之目標目切割及追蹤流程圖。

(A) 特徵點偵測及追蹤

在我們的系統中，我們利用 Shi-Tomasi-Kanade detector [3] 來偵測可能落在目標物上之特徵點(corner point)。在 Kanade(KLT)的演算法中，一旦特徵點被偵測出，每一個特徵點都可以在連續兩個畫面間被 Shi-Tomasi-Kanade 光流法[8]所追蹤。假設一特徵點於時間 τ 之移動向量為 $d = (\xi, \eta)$ ，則於時間 $t + \tau$ 時之特徵點像素值可以表示為：

$$I(x, y, t + \tau) = I(x - \xi, y - \eta, t).$$

重新定義一新函數 $J(x) = I(x, y, t + \tau)$ 且 $I(x - d) = I(x - \xi, y - \eta, t)$ ，則我們可以得到：

$$J(x) = I(x - d) + n(x),$$

上式 $n(x)$ 表示雜訊。為了在一視窗範圍 W 之二者誤差最小化，我們可以定義誤差函數為：

$$\varepsilon = \int_W [I(x - d) - J(x)]^2 \omega dx$$

對上式求解，可以得到位移向量

$$Gd = e,$$

其中 $G = \int_W gg^T \omega dA$ and $e = \int_W (I - J)g \omega dA$ 。最後根據此特徵點追蹤原理，每個 corner point 被追蹤的軌跡可表示為：

$$\{(x_t^i, y_t^i), t = T_{init}^i, \dots, T_{final}^i\}, i = 1, \dots, N\}$$

(x_t^i, y_t^i) 表示 corner point i 在畫面 t 時間的影像座標， N 表示特徵點的索引， T 表示畫面的數目。

(B) 特徵點粗略的分割

經過仔細的觀察，我們發現落在每個個體上之特徵點會呈現出較高的時空關聯性。利用空間相關測量(spatial correlation)以及幾何關係，我們可以粗略分群所偵測到之特徵點是否屬於同一個目標物；另外也考量時間相關量測(temporal correlation)，也就是評估偵測到之特徵點的軌跡是否一致，從而判斷這些特徵點是否屬於同一個目標物。在本計畫裡，我們先應用 C-mean 演算法將具有空間相關性的動態特徵點做粗略的分群。

(C) 個體分割與時空最短生成樹

基於在 C-mean 群集上的目標物粗略的分割，每一群特徵點集合可能包括數個目標物。考量一般目標物(行人)幾何特徵，於一個個體上，特徵點之橫向距離會比縱向距離短。因此於這個階段，我們要建立時空最短生成樹，計算節點與節點之連接距離時，橫向距離比重需比縱向距離比重高。因此在粗略的分割群集後，我們首先建立個體分割所需要之空間最短生成樹(spatial shortest spanning tree)[11]，此方法描述如下：

1. 在每一群特徵點群集 $C^i = \{p^1, p^2, \dots, p^N\}$ 建立一個空間最短生成樹，並按下列公式計算節點與節點加權距離。

$$d_{ij} = \left\{ \sqrt{\beta_x (p_x^i - p_x^j)^2 + \beta_y (p_y^i - p_y^j)^2}, i=1,2,\dots,N, j=1,2,\dots,N \right\}$$

其中 β_x, β_y 分別為 x 和 y 方向的權重。

2. 根據在第 1 步節點與節點加權距離的計算，建構空間最短生成樹(spatial shortest spanning tree)順序。

除了考量特徵點在空間中之鄰近關係外，在時間軸上數個動態特徵點若落在同一個移動物體上，則這些特徵點所呈現出之移動軌跡原則上應該會非常一致。因此加上考量移動軌跡的相關性，我們將空間最短生成樹構建(spatial shortest spanning tree)改良變成時空最短生成樹(spatial-temporal shortest spanning tree)。

若有兩個特徵點所產生之移動軌跡 T_u 和 T_v ，其移動軌跡的相關性可由下式描述：

$$\text{Correlation}(T_u, T_v) = \frac{1}{1 + \text{Variance}(T_u, T_v)}$$

結合空間中之鄰近關係，我們可以建立一個空間與時間一致性的測量標準：

$$\text{Conformance} = \frac{\sqrt{\beta_x (p_x^i - p_x^j)^2 + \beta_y (p_y^i - p_y^j)^2}}{\text{Correlation}(T_i, T_j)}, i=1,2,\dots,N-1, j=i+1.$$

如果上述公式所計算出之數值大於所設定之臨界值，表示兩點不屬於同一個目標。但同時，我們可能面臨一個當兩目標物交錯經過時，目標 A 的特徵點可能錯位於目標 B 的情況。因此，在下一節我們將計畫研發一個軌跡投票(voting)的方法來驗證特徵點的屬性。

(D) 目標物交錯移動之錯植特徵點過濾方法

雖然以特徵點為基礎之(point-based)物體偵測追蹤的方法，可以解決一些在擁擠鏡頭中追蹤的問題，但是仍存在一些例外的情形，例如當當兩目標物交錯經過時，目標 A 的特徵點可能錯位於目標 B，此情況如 Fig. 5 所顯示。當兩個人交錯的時候，很多特徵點可能會被做錯誤的分群。為了克服上述這個情形，我們提

出一個以特徵點時間訊息(移動軌跡)的投票方法，以確保群集中追蹤的特徵點之軌跡一致性(trajjectory conformance)，投票規則敘述如下：

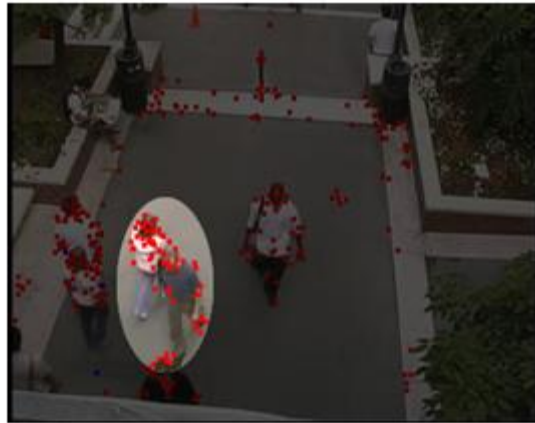


Fig. 5. 當兩個人位於交錯的位置，很多點可能會被誤設

- 從每個特徵點連續 10 個畫面的移動方向紀錄取得向量 $\mathbf{d}_{(x,y)}(F_t, F_{t-10})$
- $\mathbf{d}_{(x,y)} = (d_x, d_y)$ 的組成是基於 x 跟 y 方向上，顯示在 Fig. 6 中有四個可能標示配對來表示其可能的移動方向。分別定義為 $(+, +)$, $(+, -)$, $(-, +)$, 和 $(-, -)$ 。我們會記錄所有點屬於每個象限的動態方向。
- 通過觀察記錄的方向列表，取得其方向統計之最高得票方向。由此假設是在此群集中主要的移動方向，然後刪除其他不同的方向動態點。
- 重複過程 1~3，直到所有的群集被處理完畢。
- 一旦群集的方向列表被修改，群集的中心必須被重估。

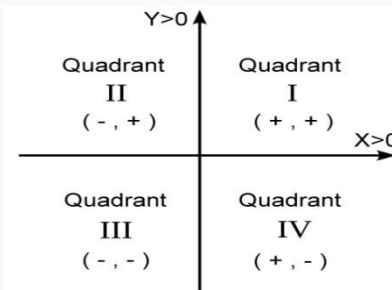


Fig. 6. 每個特徵點可能移動的方向

(E) 目標物追蹤

為了準確地追蹤所有的目標，研發穩定和可靠的基礎點目標跟踪方法是必要的。根據特徵點追蹤的機制，我們知道特徵點可以在兩個連續的畫面之間做有效的追蹤。因此，KLT 追蹤演算法的特徵是需要連續畫面內，採用存在動態的追蹤與遺傳特徵點的屬性。另一方面，新的特徵點也必須重新偵測以補償原存在特徵點的追蹤失敗或消失。因此，以特徵點為基礎之目標物追蹤重要的是**連結新舊點之間的追蹤關係**。因此此計畫所發展之以特徵點為基礎之目標物追蹤敘述如下：

1. 如果特徵點的連續追蹤成功，它將會繼承每個單一個體之屬性。反之，如果被偵測到之動態特徵點是不屬已有之物體，我們將會將此特徵點分類為 $P_{NEW} = \{p^1, p^2 \dots p^N\}$ 。此外，新的特徵點和群集中心點 C_j 間的距離計算為：

$$Distance = \left\{ \left\{ \sqrt{\alpha_x (p_x^i - C_x^j)^2 + \alpha_y (p_y^i - C_y^j)^2}, i = 1, 2, \dots, N \right\}, j = 1, 2, \dots, N \right\},$$

其中 α_x 和 α_y 為相似於 C-mean 群集過程 x 軸跟 y 軸的權重。

2. 當我們增加新的點到最近的群集之後，群集的中心需要重新計算。若數個目標物一起移動時，個體分割必須再次執行。獨立個體分割是基於一般正常人身體的寬度。首先，我們計算集群的寬度：

$$ClusterWidth = RightBoundary_y - LeftBoundary_y.$$

如果群集的寬度大於預定的大小 T_{WIDTH} (評估正常人身體的寬度)，那麼群集需要被切割。新的切割群集將會給它一個新的編號。Fig. 7 描述需要重新分割的情況。

3. 如果一些新的點沒有被分類到任何存在的族群，然後我們會將這些點分類成新的族群 $P'_{NEW} = \{p'_1, p'_2 \dots p'_N\}$ 和設成新的移動目標。Fig. 8 描述已存在的群集 (存在的目標) 和新的群集 (新目標)。
4. 整合新舊群集資訊。
5. 檢查每個群集中所有點之一致性和執行時空最短生成樹 (spatial-temporal shortest spanning tree) 之個體分割。Fig. 9 顯示最終個體分割結果。



Fig. 7 紅線寬度表示群集，藍線表示正常寬度的人體。



Fig. 8 (a)綠色點是從這些遠離其他存在紅群集新生成的點。(b)群集的結果。



Fig. 9 目標追蹤的最終結果

3.3.2、人群運動軌跡分類

第二個子系統所提出的方法為應用 longest common subsequences 演算法自動估算人群運動軌跡間之相似性，逐一地被分類到最適當的類別中。在子系統，我們計畫應用 longest common subsequences 演算法於自動分析人群運動軌跡。此子系統之流程圖如 Fig. 10 所示。於特徵點軌跡分類中，軌跡與系統所預定之群眾運動軌跡比對 $M(A^i, T)$ 之計算如下列工式鎖定義：

$$M(A_n^i, T_n) = \begin{cases} 0, & \text{if } A_n^i \text{ or } T_n \text{ is empty} \\ 1 + M(A_{n-1}^i, T_{n-1}), & \text{if } \|A^i(n) - T(n)\|_2 < \varepsilon \text{ and } |n - n| < \delta, \\ \max(M(A_{n-1}^i, T_n), M(A_n^i, T_{n-1})), & \text{otherwise} \end{cases}$$

此外，我們定義一二維陣列 Q 來計算最常相似序列值 (longest common subsequence)，比對程序如 Fig. 11 所描述。Fig. 12 為目前我們所模擬之成果。

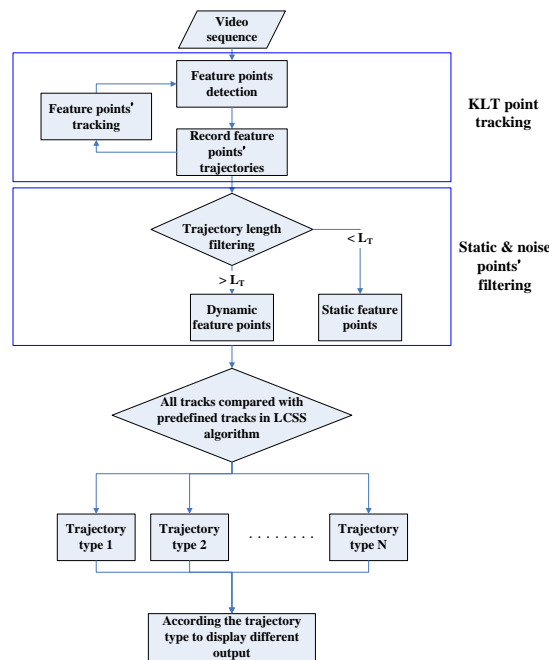


Fig. 10 The flowchart of the crowd moving trajectories classification using the LCSS algorithm.

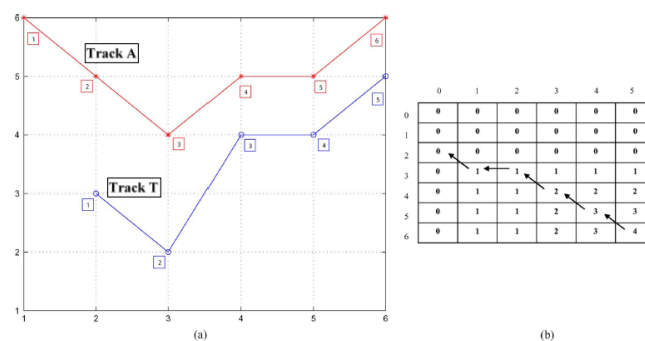


Fig. 11 Example of matching cost M . (a) Two tracks are compared in term of matching cost. (b) The 2D array for computing the matching cost.

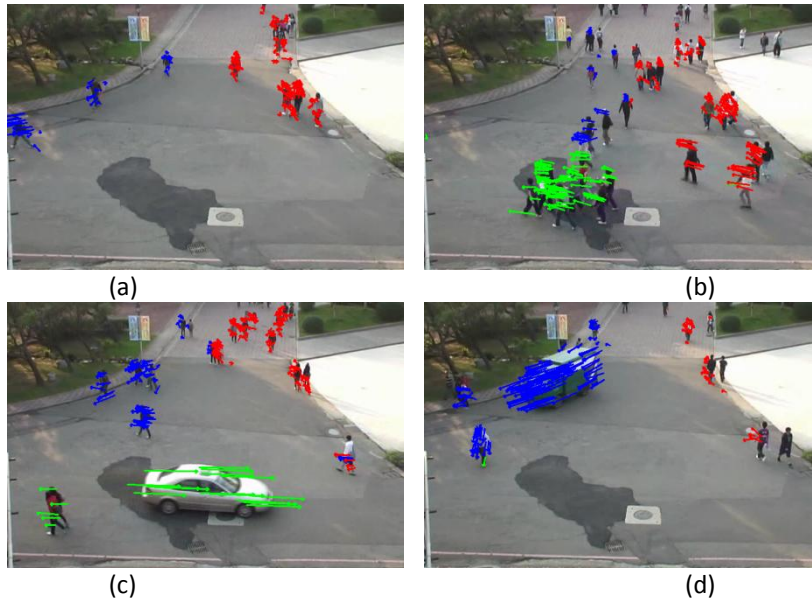


Fig. 12 Classification of the points' trajectories. (a) The case of few people walks separately. (b) Crowds walk closely. (c), (d) The trajectories of the moving car is detected.

3.4 結果與討論

3.4.1 移動目標物切割分析

兩個測試影片“Commons01”和“Tunnel-A125”是用於獨立個體的分割和目標追蹤的效果評估。Fig. 13 顯示特徵點目標追蹤的實驗結果。在 Fig. 13，有數個人靠近走在一起且有嚴重的遮蔽問題。這個提出的方法可以分割每個單獨的人甚至遮蔽問題。為了評估提出系統的正確性，我們把提出的方法與最近幾年提出的方法做比較。Table 1 列出這些方法之比較，可以看出我們的方法優於其他的方法。



Fig. 13. 分割與追蹤目標群集

Table 1 The accuracy analysis for the methods of Brostow & Cipolla, Zhao & Nevatia, and ours.

	Brostow & Cipolla	Zhao & Nevatia	Ours
distinct detections	144	8466	1319
correctly detected	136	7881	1254
missed detections	8	585	65
false detections	33	291	56
detection rate	94%	93.09%	95.07%
miss detection rate	22.9%	6.91%	4.92%
false detection rate	5.6%	3.43%	4.25%

3.4.2 人潮移動偵測

應用 3.3.2 節的原理，我們於中華大學的場景時機實際偵測並分類人群移動方向。圖 14 顯示出人群移動方向分類的結果。

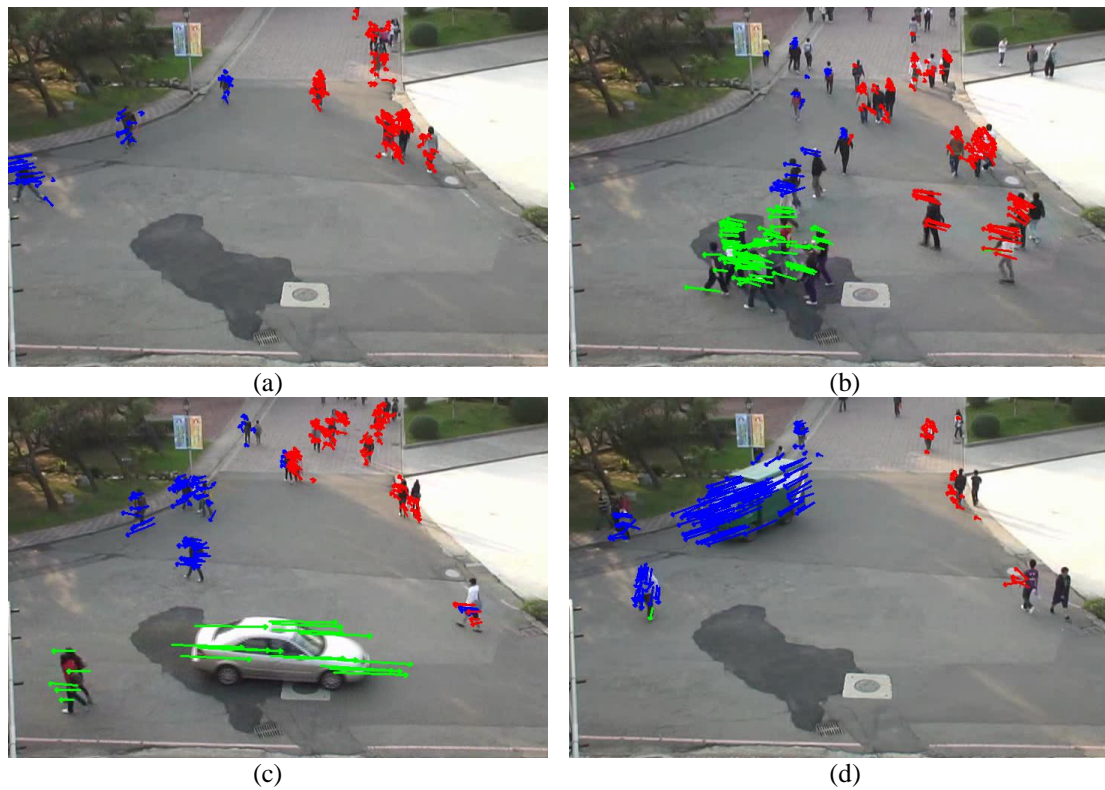


Fig. 14 Classification of the points' trajectories. (a) The case of few people walks separately. (b) Crowds walk closely. (c), (d) The trajectories of the moving car is detected.

國科會補助專題研究計畫出席國際學術會議心得報告

日期：__年__月__日

計畫編號	NSC 100-2221-E-216 -030 -		
計畫名稱	於擁擠人群中以特徵點為基礎的目標物追蹤系統		
出國人員姓名	連振昌	服務機構及職稱	中華大學資訊工程學系 副教授
會議時間	101年7月17日至 101年7月19日	會議地點	希臘 雅典
會議名稱	(中文)第八屆國際智慧資訊隱藏及多媒體訊號處理會議 (英文) The Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing		
發表題目	(中文)使用最低相關之LBPH特徵之瞌睡辨識 (英文) Drowsiness Recognition using the Least Correlated LBPH		

一、參加會議經過

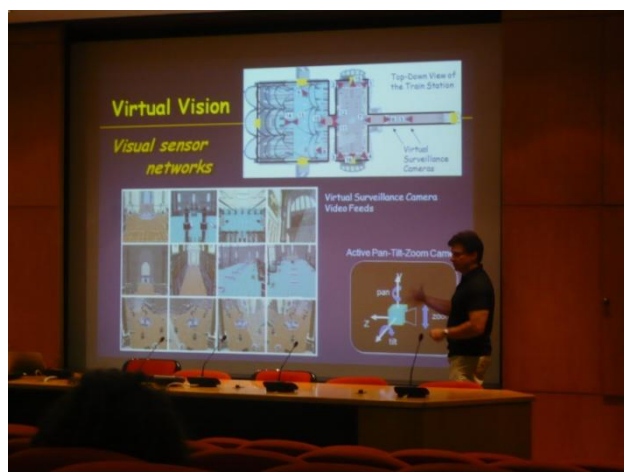
這一次赴希臘參加 I IHMSPP 會議，首先聆聽 Prof. **Demetri Terzopoulos** 所演講之”Virtual Vision: Computer Vision in Virtual Reality”，對於日後於電腦視覺方面之研究有所啟發。以下為其演講之摘要及影像紀錄。

Prof. **Demetri Terzopoulos**, **IEEE Fellow**, Computer Science, University of California, USA
(<http://www.cs.ucla.edu/~dt/>)

Title: Virtual Vision: Computer Vision in Virtual Reality

Abstract:

Realistic virtual worlds can serve as software laboratories within which vision researchers may efficiently develop and evaluate sophisticated, active machine perception systems. Known as "Virtual Vision", this unorthodox philosophy posited at the intersection of the fields of computer vision and computer graphics, enables virtual reality to subserve computer vision research and development. In the context of the virtual vision paradigm, this talk will focus on the rapid development and evaluation of distributed smart-camera sensor networks and intelligent surveillance systems that can persistently monitor humans in large-scale urban environments. The visually realistic virtual environments exploited in this work are populated by autonomous virtual humans, which are the product of a comprehensive, artificial life approach to multi-human simulation.



之後本人於會議中發表兩篇論文，題目分別為

1. Drowsiness Recognition using the Least Correlated LBPH
2. Kodály Hand Signs Recognition without Background Modeling

會議中與國外學者交流相關研究之心得。



二、與會心得

此次 IJHMSP 會議，於希臘舉辦。參加國際會議之目的在於了解吸收國際最新研究方向及技術發展新知，對於規劃日後研究方向有甚大助益。此會議之論文也會被 EI 收錄，增加研究之國際能見度。

三、發表論文全文或摘要

限於篇幅關係，以下僅列出於此次會議所發表之二篇論文首頁，

Drowsiness Recognition using the Least Correlated LBPH

Cheng-Chang Lien

Dept. Computer Science & Information Engineering
 , Chung Hua University, Taiwan, R.O.C.
 E-mail: cclien@chu.edu.tw

Pei-Rong Lin

Dept. Computer Science & Information Engineering
 Chung Hua University, Taiwan, R.O.C.
 E-mail: M09702010@chu.edu.tw

Abstract—In recent years, the drowsiness recognition is widely applied to the driver alerting or distance learning. The drowsiness recognition system is constructed on the basis of the recognition of eye states. The conventional methods for recognizing the eye states are often influenced by the illumination variations or hair/glasses occlusion. In this paper, we propose a new image feature called “least correlated LBP histogram (LC-LBPH)” to generate a high discriminate image features for recognizing the eye states robustly. Then, the method of independent component analysis (ICA) is applied to derive the low-dimensional and statistical independent feature vectors. Finally, support vector machines (SVM) are trained to recognize the eye states. Furthermore, we design four rules to classify three eye transition patterns which define the normal (consciousness), drowsiness, and sleeping situations. Experimental results show that the eye-state recognition rate is about 0.08 seconds per frame and the drowsiness recognition accuracy approaches 98%.

Keywords—drowsiness recognition, eye state, LBPH, ICA, support vector machine

I. INTRODUCTION

The vision-based drowsiness recognition systems can be categorized into the feature-based and template-based methods. In the template-based methods [1, 2], the eye templates are constructed to search and identify the eye states. Then, the correlation and thresholding are often adopted for recognizing the eye states. However, these methods can be influenced by the variations of illumination and appearance (hair or glasses).

In the feature-based methods, the Gabor wavelet features and AdaBoost classifier are used to detect the fatigue [3]. Furthermore, the LBP codes are usually employed to extract the texture features and combined with AdaBoost or SVM methods [4, 5] to recognize the drowsiness. In addition to the gray image analyses, Wu et al. [6] use particle filter to track the eye positions and apply PCA to extract image features. Then, they use logistic regression to estimate the eye blinking. Hence, how to generate the highly discriminative image features for recognizing the eye states is crucial for the drowsiness recognition. Furthermore, the statistical independent requirements are often demanded for generating the highly discriminative image features.

To this end, we propose a novel feature-based method to recognize the drowsiness. First, by studying the methods for establishing the LBPH image features, we propose a highly discriminative image features called least correlated LBP

histogram (LC-LBPH) to recognize the eye states. Then independent component analysis (ICA) method is used to generate the low-dimensional and statistical independent feature vectors. Finally, Support Vector Machines are trained to identify the eye states. According to literatures of human blink [7] and our observations, we define three eye states transition patterns to describe the normal (consciousness), drowsiness, and sleeping states that can be detected with four proposed rules. Experimental results show that our system can recognize the drowsiness with accuracy 98% within 2 seconds. The eye state recognition rate is about 0.08 seconds per frame.

II. LEAST CORRELATED LBPH FEATURE

In our eyes state recognition system, we apply the AdaBoost algorithm in Intel’s open source computer vision library to detect both of the face and mouth regions such that the accuracy of eyes locating can be increased. Some results of face and mouth detections are shown in Fig. 1.

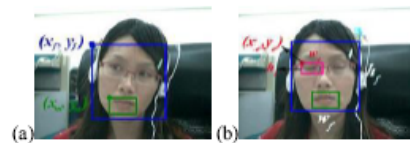


Figure 1. (a) The detected face and mouth regions in a video sequence. (b) Eyes position is located according to the rule in Eq. (1).

Based on the careful measurements, we found that the eyes will locate in the certain facial region denoted in Fig. 1-(b) and the eye region can have a fixed geometrical relationship to the facial width and length. Hence, the eyes location can be located according to the following simple rules:

1. With the facial region detected, the top-left corner point (x_f, y_f) , face width w_f and face height h_f are determined.
2. By using the detected top-left corner point (x_m, y_m) on mouth region, the top-left corner point (x_e, y_e) of the eye region can be determined by the following formula:

$$(x_e, y_e) = (x_m - 0.16 w_f, y_f + 0.16 h_f + 20) \quad (1)$$

3. The width and height of eye region are determined as $w_e = 1.6 \times 0.16 \times w_f$ and $h_e = 0.16 h_f$ respectively.

To describe the texture distribution over a large region, a new image feature called LBP histogram (LBPH) [9] is proposed. There are 59 dimensions in the LBPH feature

Kodály Musical Hand Signs Recognition without Visual Background Modeling

Chun-Yuan Lee
General Education Center
Chung Hua University
Hsinchu, Taiwan, R.O.C.
Maggie@chu.edu.tw

Cong-Wei Huang, Cheng-Chang Lien
Department of CSIE
Chung Hua University
Hsinchu, Taiwan, R.O.C.
cclien@chu.edu.tw

Abstract—In this study, we develop a novel vision-based Kodály musical hand signs recognition system to recognize the gestures of the musical notes. Vision-based gesture recognitions often face the following problems. First, the illumination change can influence the hand detections. Second, the hand tracking will become difficult under the complex background. To overcome the aforementioned problems, we propose several novel technologies to overcome these problems. The first one is the block-based foreground detection method in which the difference between consecutive frames of moving hand can be identified. The second one is the dual foregrounds fusion method that can generate the precise hand regions. The third one is the texture-based fist tracking method that can locate the fist position precisely without the influence of illumination variations. After the fist locating, the skin color detection is applied to extract the complete hand region and then the various kind of Kodály musical hand signs can be recognized with the moment invariants and support vector machines. The experimental results show that the hand can be tracked with the accuracy 95.71% and efficiency 20 fps under the complex background. The recognition accuracy for the Kodály musical gestures is about 97%.

Keyword: *Kodály musical hand signs, Dual foregrounds fusion, Moment invariants, Texture-based fist tracking.*

I. INTRODUCTION

In Kodály musical teaching method [1], musical skills are introduced according to the capabilities of the child. Children are first introduced to musical concepts through experiences such as listening, singing, or movement. In this study, we develop a novel vision-based Kodály musical hand signs recognition system to recognize the gesture movements of the musical notes shown in Fig. 1.



Fig. 1 Kodály musical hand signs [1].

In general, the moving hands could be detected by three kinds of methods: motion-based [2], background modeling [3-5], and temporal difference [6] approaches. In the motion-based approaches, the optical flow method [2] utilizes the motion flow segmentation to separate the background and foreground regions. By applying the optical flow method [2], the moving hands can be extracted even in the presence of camera motion. However, the high computation complexity makes the real-time implementation difficult.

In the background modeling methods, the construction and updating of background models [3-5] often is time-consuming. For example, in [3,5], the Gaussian Mixture Model (GMM) is frequently adopted to model the intensity variation for each pixel within a time interval and then high computing cost is required to calculate the GMM parameters. Furthermore, the foreground detection with background modeling method is extremely sensitive to the rapid illumination variation or the dynamic background changing. In [4], the Kalman filter is used to update the background model with less computational complexity. But this method can't solve the problem of serious scene change which can make the system unable to update the background model accurately.

The advantage of temporal difference method [6] is less susceptible to the scene change, i.e., it has capability to detect the moving hands in dynamic environments but the regions of the moving objects can't be extracted completely when the objects move slowly. Here, we propose a novel moving hands detection method without background modeling to overcome the aforementioned problems and then develop the hand tracking and Kodály musical hand signs recognition system. First, a modified block-based frame differential method is established to quickly detect the moving hands without the influences of rapid illumination changes. Second, the precise hand regions are extracted by the dual foregrounds fusion method. Third, the texture-based fist tracking method is proposed to track the moving

四、建議

無

五、攜回資料名稱及內容

論文光碟片

六、其他

國科會補助計畫衍生研發成果推廣資料表

日期:2012/10/31

國科會補助計畫	計畫名稱: 於擁擠人群中以特徵點為基礎的目標物追蹤系統	
	計畫主持人: 連振昌	
	計畫編號: 100-2221-E-216-030-	學門領域: 圖形辨識
研發成果名稱	(中文) 於擁擠人群中以特徵點為基礎的目標物追蹤系統	
	(英文) Feature-Point-Based Target Tracking in the Crowd	
成果歸屬機構	中華大學	發明人 (創作人) 連振昌
	<p>(中文) 在此計畫中, 為了在擁擠場景中偵測個別的目標物以及分析群體移動所產生的軌跡資訊, 我們提出兩個方法分別在擁擠環境中偵測、追蹤個別的目標物, 以及將群體運動軌跡進行偵測分類。第一個子系統所提出的方法是以corner特徵點為基礎, 發展由粗糙至細緻的獨立目標物切割演算法。在這個方法中, 首先利用C-Means演算法將特徵點做粗略的分群, 接著透過時空最小生成樹(spatial-temporal shortest spanning tree)在每一個移動分群內切割出更精確的個體, 最後運用corner特徵點繼承的概念來追蹤所有移動個體目標物。第二個子系統所提出的方法為應用longest common subsequences演算法自動估算人群運動軌跡間之相似性, 逐一地被分類到最適當的類別中。所有特徵點軌跡相似性量測是透過空間及時間上的相似性做測量。從實驗的結果發現, 在擁擠環境中分割獨立移動目標物的準確度可高於90%以上, 效能部分可每秒約處理10張frame。</p> <p>(英文) In order to detect each individual target in the crowded scenes and analyze the crowd moving trajectories, we propose two methods to detect and track the individual target in the crowd and classify the crowd motion trajectories. First, a coarse-to-fine individual segmentation approach based on the corner points' extraction and tracking is proposed. The dynamic feature points are roughly clustered by the C-means algorithm and then a spatial-temporal shortest spanning tree is proposed to segment each individual target in the moving group and each target is tracked with the concept of points' inheritance. Second, the method of the longest common subsequences is applied to automatically evaluate the similarities among the feature tracks. The experimental results show that the accuracy of individual segmentation in the crowd can be higher than 90% and the efficiency of our system can approach 10 fps.</p>	
技術說明		
產業別	電機及電子機械器材業	
技術/產品應用範圍	電子、資訊	
技術移轉可行性及預期效益	本技術已與123視科技完成簽約, 依其需求修改方法, 以符合該公司之應用情境。	

註: 本項研發成果若尚未申請專利, 請勿揭露可申請專利之主要內容。

100 年度專題研究計畫研究成果彙整表

計畫主持人：連振昌		計畫編號：100-2221-E-216-030-				計畫名稱：於擁擠人群中以特徵點為基礎的目標物追蹤系統	
成果項目		量化			單位	備註（質化說明： 如數個計畫共同 成果、成果列為 該期刊之封面故 事...等）	
		實際已達成 數（被接受 或已發表）	預期總達成 數(含實際已 達成數)	本計畫實 際貢獻百 分比			
國內	論文著作	期刊論文	0	0	100%	篇	
		研究報告/技術報告	0	0	100%		
		研討會論文	0	0	100%		
		專書	0	0	100%		
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	1	1	100%	件	本技術已與 123 視 科技完成委託研究 簽約，依其需求修改 方法，以符合該公司 之應用情境。
		權利金	0	0	100%	千元	
	參與計畫人力 （本國籍）	碩士生	2	2	100%	人次	
		博士生	0	0	100%		
博士後研究員		0	0	100%			
專任助理		0	0	100%			
國外	論文著作	期刊論文	1	1	100%	篇	Cheng-Chang Lien and Ming-Hsiu Tsai, ' Vehicle Counting without Background Modeling,' accept to publish in Journal of Marine Science and Technology. (SCI, EI)
		研究報告/技術報告	0	0	100%		
		研討會論文	2	2	100%		[1]Chun-Yuan Lee, Cong-Wei Huang, and Cheng-Chang Lien, ' Kodaly Musical Hand Signs Recognition without Visual Background

							Model,' the Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP-2012), Piraeus-Athens, Greece, July 18-20, 2012. (EI)
							[2]Cheng-Chang Lien and Pei-Rong Lin, ' Drowsiness Recognition Using the Least Correlated LBPH,' the Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP-2012), Piraeus-Athens, Greece, July 18-20, 2012. (EI)
		專書	0	0	100%	章/本	
專利		申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
技術移轉		件數	0	0	100%	件	
		權利金	0	0	100%	千元	
參與計畫人力 (外國籍)		碩士生	0	0	100%	人次	
		博士生	0	0	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		

<p>其他成果 (無法以量化表達之成果如辦理學術活動、獲得獎項、重要國際合作、研究成果國際影響力及其他協助產業技術發展之具體效益事項等,請以文字敘述填列。)</p>	<p>(1)初期成果已發表如下: Shin-Ji Lin, Cheng-Chang Lien, Wei-Hsin Kan and Hsiao-Hu Tan, ' Feature-Point-Based Target Tracking in the Crowd' , APSIPA Annual Summit and Conference, December 14 - 17, 2010, Biopolis, Singapore . (EI)</p> <p>(2)本年度研究成果已發表 SCI 期刊一篇, EI 國際研討會論文二篇</p> <p>(3)本技術已與 123 視科技完成委託研究簽約, 依其需求修改方法, 以符合該公司之應用情境。</p>
--	--

	成果項目	量化	名稱或內容性質簡述
科教處計畫加填項目	測驗工具(含質性與量性)	0	
	課程/模組	0	
	電腦及網路系統或工具	0	
	教材	0	
	舉辦之活動/競賽	0	
	研討會/工作坊	0	
	電子報、網站	0	
	計畫成果推廣之參與(閱聽)人數	0	

國科會補助專題研究計畫成果報告自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現或其他有關價值等，作一綜合評估。

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估

達成目標

未達成目標（請說明，以 100 字為限）

實驗失敗

因故實驗中斷

其他原因

說明：

2. 研究成果在學術期刊發表或申請專利等情形：

論文： 已發表 未發表之文稿 撰寫中 無

專利： 已獲得 申請中 無

技轉： 已技轉 洽談中 無

其他：（以 100 字為限）

(1) 初期成果已發表如下：

Shin-Ji Lin, Cheng-Chang Lien, Wei-Hsin Kan and Hsiao-Hu

Tan, ' Feature-Point-Based Target Tracking in the Crowd' , APSIPA Annual Summit and Conference, December 14 - 17, 2010, Biopolis, Singapore . (EI)

(2) 本技術已與 123 視科技完成簽約，依其需求修改方法，以符合該公司之應用情境。

3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）（以 500 字為限）

在一般的影像監控應用當中，具有挑戰性之一的問題就是在擁擠的場景中進行目標的追蹤。一般來說，嚴重的遮蔽現象會使傳統 blob-based 物件偵測/追蹤的方法失敗。例如，追蹤的目標物在擁擠的車站或廣場，被追蹤的人有部分被其他人遮住，只有部分區域可以被作為追蹤的線索。因此，在擁擠的人群中追蹤一個單獨的目標我們可能會面臨兩個主要的問題：1) 當我們在監視較大空間中人群的移動，目標物可能太小 2) 頻繁的遮蔽使得目標分割非常困難。因此，特徵點偵測及追蹤經過評估後，認為適用於解決在人群中追蹤目標的問題。即使在擁擠的場面中，可能仍然有一些區域的特徵點沒有被遮蔽。根據特徵點追蹤相同的概念，Brostow 與 Cipolla 提出 Bayesian clustering algorithm 可以分割人群中每個獨立的個體與時空距離和軌跡的一致性。然而，執行效能並不高。在本計劃中，我們提出了一個新的方法來偵測和追蹤人群中的個別目標物。