行政院國家科學委員會補助專題研究計畫 □成果報告 ☒期中進度報告

主動式多攝影機視訊監控系統之研究

計畫類別：☒ 個別型計畫 □ 整合型計畫

計畫編號：NSC 97-2221-E-009-132-MY3

執行期間： 98 年 8 月 1 日至 99 年 7 月 31 日

計畫主持人：王聖智
共同主持人：
計畫參與人員： 黃敬群、周節、林瑋國、戴玉書

成果報告類型(依經費核定清單規定繳交)：☒精簡報告 □完整報告

本成果報告包括以下應繳交之附件：
□赴國外出差或研習心得報告一份
□赴大陸地區出差或研習心得報告一份
□出席國際學術會議心得報告及發表之論文各一份
□國際合作研究計畫國外研究報告書一份

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、列管計畫及下列情形者外，得立即公開查詢
□涉及專利或其他智慧財產權，□一年□二年後可公開查詢

執行單位：交通大學電子工程系

中 華 民 國 99 年 5 月 25 日

可供推廣之研發成果資料表

□ 可申請專利　☒ 可技術移轉　　日期：99 年 5 月 25 日

國科會補助計畫	計畫名稱：主動式多攝影機視訊監控系統之研究 計畫主持人：王聖智 計畫編號：NSC 97-2221-E-009-132-MY3 學門領域：影像處理
技術/創作名稱	應用於視訊監控之多運動物體偵測與追蹤技術
發明人/創作人	黃敬群、王聖智
技術說明	中文： 我們提出一套有效率的攝影機監控系統，透過串接一階層式結構以將多物體的偵測與追蹤能在同一架構中完成，藉由這樣的設計，物件偵測與物件追蹤的結果可以相互分享結果，因而達到更好的系統效能。首先，對於每一個物體我們會建立一動態模型來描述其運動的方式以達到物件追蹤的效果，在系統中，透過將上一時刻不同物件的偵測結果當成新的量測，每一個物件的動態模型可以自動地更新，以即時地修正追蹤模型。另一方面，更新過後的動態模型則提供物體移動的預測，這個預測代表物件層級的資訊，讓系統得以預知不同物體可能出現的區域以及可能的遮蔽現象，此外，根據物體偵測結果以及動態模型，系統也得以自動更新目標物與背景的外貌模型，這些外貌模型代表像素層級的資訊，有助於讓系統更準確的分類屬於目標物的區域，最後透過所提的貝氏階層式推論結構將所有模型與目前的觀察影像結合並進行最佳化推論，系統可以求得新的物件偵測結果。實驗結果顯示，我們的系統在攝影機搖晃，物體相互遮蔽，以及前景區域與背景區域因為外貌相似而混淆的情況下，仍然可以得到好的效果。 英文： We propose a cascaded hierarchical framework for object detection and tracking. We claim that, by integrating both detection and tracking into a unified framework, the detection and tracking of multiple moving objects in a complicated environment become more robust. Under the proposed architecture, detection and tracking cooperate with each other. Based on the result of moving object detection, a dynamic model is adaptively maintained for object tracking. On the other hand, the updated dynamic model is used for both temporal prior propagation of object labels and the update of foreground/background models, which step further to help the detection of moving objects in the next frame. The experiments show that accurate results can be obtained even under situations with camera shaking, foreground /background appearance ambiguity, and object occlusion.

可利用之產業 及 可開發之產品	安全監控產業
技術特點	可自動偵測與追蹤監控場景中的多個目標物
推廣及運用的價值	運用此技術以增強當前視訊監控軟體的智慧型功能，並協助監控端的管理者有效率的掌握監控環境。

※ 1.每項研發成果請填寫一式二份，一份隨成果報告送繳本會，一份送　貴單位研發成果推廣單位（如技術移轉中心）。

※ 2.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。

※ 3.本表若不敷使用，請自行影印使用。

主動式多攝影機視訊監控系統之研究

Visual Surveillance System with Multiple Active Cameras

計畫編號：NSC 97-2221-E-009-132-MY3

執行期限：98年8月1日至99年7月31日

主持人：王聖智 (交通大學電子工程系教授)

計畫參與人員：黃敬群、周節、林瑋國、戴玉書(交通大學電子所研究生)

中文摘要

在本計畫中，我們提出一套有效率的攝影機監控系統，透過串接一階層式結構以將多物體的偵測與追蹤能在同一架構中完成，藉由這樣的設計，物件偵測與物件追蹤的結果可以相互分享結果，因而達到更好的系統效能。首先，對於每一個物體我們會建立一動態模型來描述其運動的方式以達到物件追蹤的效果，在系統中，透過將上一時刻不同物件的偵測結果當成新的量測，每一個物件的動態模型可以自動地更新，以即時地修正追蹤模型。另一方面，更新過後的動態模型則提供物體移動的預測，這個預測代表物件層級的資訊，讓系統得以預知不同物體可能出現的區域以及可能的遮蔽現象，此外，根據物體偵測結果以及動態模型，系統也得以自動更新目標物與背景的外貌模型，這些外貌模型代表像素層級的資訊，有助於讓系統更準確的分類屬於目標物的區域，最後透過所提的貝氏階層式推論結構將所有模型與目前的觀察影像結合並進行最佳化推論，系統可以求得新的物件偵測結果。實驗結果顯示，我們的系統在攝影機搖晃，物體相互遮蔽，以及前景區域與背景區域因為外貌相似而混淆的情況下，仍然可以得到好的效果。

關鍵詞：影像標記、圖學模型、物件偵測、物件追蹤、影像切割。

Abstract

In this project, we propose a cascaded hierarchical framework for object detection and tracking. We claim that, by integrating both detection and tracking into a unified framework, the detection and tracking of multiple moving objects in a complicated environment become more robust. Under the proposed architecture, detection and tracking cooperate with each other. Based on the result of moving object detection, a dynamic model is adaptively maintained for object tracking. On the other hand, the updated dynamic model is used for both temporal prior propagation of object labels and the update of foreground/background models, which step further to help the detection of moving objects in the next frame. The experiments show that accurate results can be obtained even under situations with camera shaking, foreground /background appearance ambiguity, and object occlusion.

Keywords: Image labeling, Graphical models, Object Detection, Object Tracking, Image Segmentation

1. INTRODUCTION

Recently, intelligent surveillance systems are getting more and more popular. For a typical surveillance system, most cameras are kept static and several background subtraction algorithms, like [1], [2] and [3], can be used to detect

foreground objects. These background subtraction methods focus mainly on the modeling of background information, like the usage of the GMM model in [3] and many others. Even though this type of approach works pretty well for scenes with stationary background, it has difficulty in handling the appearance ambiguity between the foreground objects and the surrounding background. Moreover, in an outdoor scene, occasional camera shaking caused by strong wind may also seriously degrade the performance of detection.

On the other hand, many object tracking algorithms focus on foreground modeling. For example, the color-based mean-shift tracking method [4] tries to find the image patch that best matches the target model. Since the background information hasn't been considered, this approach suffers from the foreground /background ambiguity problem, and the tracked result may easily get distracted. To improve the performance, a few methods try to properly take into account the background information. For example, in [5], the authors adopted an online training process to select discriminative foreground features with respect to the surrounding background. With this mechanism, the foreground/background ambiguity problem can be relieved. However, those methods mostly focus on the tracking of single object. Some complicated situations, like the occurrence of new comers, the disappearance of tracked objects, and the inter-occlusion among multiple objects, are still big challenges to a practical tracking system.

In this paper, instead of individually performing detection and tracking, we propose a new scheme to integrate both detection and tracking into a unified framework. The proposed framework adopts a temporal prediction mechanism to provide object-level prior knowledge, which is learned from the previous decision, and to continuously update the pixel-level foreground model and background model. Based on this scheme, the object labeling, foreground modeling, and background modeling are effectively fused together to better handle the foreground/background ambiguity problem. Moreover, with the estimated depth order based on the labeling result, the inter-occlusion problem can be better solved. Besides, the emergence of new comers and the disappearance of tracked objects are also handled in the proposed system.

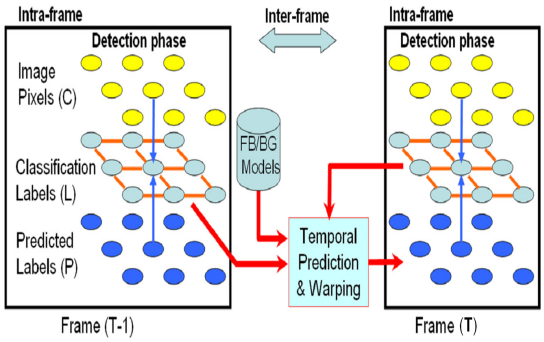

Fig. 1: Proposed scheme for object labeling and tracking.

2. PROPOSED SCHEME AND FOREGROUND/BACKGROUND LABELING

The proposed scheme is illustrated in Fig. 1. This scheme contains two major parts: the inter-frame part and the intra-frame part. The inter-frame part handles how a temporal message is propagated between successive frames; while the intra-frame part deals with object labeling. In this section, we focus on the description of the intra-frame part, which involves foreground model, background model, spatial MRF (Markov random field) constraints, and temporal prior

message. In the next two sections, we'll explain the details of the inter-frame part.

In this paper, we assume cameras are static but may suffer from slight shaking caused by winds or other factors. Hence, the background in the captured images may be trembling all the time. To handle this non-stationary background, we adopt Sheikh and Shah's approach [6] with some modifications to construct a joint spatio-chromatic probability distribution of multiple foreground/background objects based on kernel density estimation. By combining the spatial location (x,y) and the pixel color values (r,g,b) into a five-dimensional random vector $\vec{c}$ =(x,y,r,g,b), the conditional joint spatio-chromatic probabilities are defined as

$$\begin{cases} P(\vec{c}\,|\,\Omega_B) = \dfrac{1}{n}\sum_{i=1}^{n}\phi(\vec{c}-\vec{c}_{Bi}) \\ P(\vec{c}\,|\,\Omega_F^g) = U^{-1},\ if\ g=0 \\ P(\vec{c}\,|\,\Omega_F^g) = \dfrac{1}{m}\sum_{i=1}^{m}\phi(\vec{c}-\vec{c}_{Fi}^{\,g}),\ if\ g=1\sim G \end{cases}, \qquad (1)$$

where Ω_B and Ω_F^g denote the background and the g^{th} foreground, respectively. Ω_F^0 is especially designed for new comers. In (1), $\phi(.)$ is a symmetric and normalized kernel function, $\vec{c}_{Bi}$ denotes one of the n background samples, $\vec{c}_{Fi}^{\,g}$ denotes one of the m foreground samples of the g^{th} target, G is the number of foreground objects in the current image, and U^{-1} describes a uniform distribution over the five-dimensional domain. Based on the above definition, the spatial uncertainty caused by camera shaking and the chromatic uncertainty caused by lighting change can be properly modeled.

In our approach, object detection is treated as a classification problem. Besides background model $P(\vec{c}\,|\Omega_B)$ and foreground models $P(\vec{c}\,|\Omega_F^g)$, we also take into account current observation, spatial smooth constraint, and temporal prior knowledge. As shown in Fig. 1, we adopt a 3-layer structure at each time instant. The top layer **C** represents the observation layer at that time instant. In our approach, we assume **C** contains the spatio-chromatic information of the observed image data. The middle layer **L** contains the classification label for each image pixel. Basically, we aim to assign to each labeling node Li a suitable ID from the set $\{\Omega_B, \Omega_F^0, \ldots, \Omega_F^G\}$. The bottom layer **P** represents the predicted label messages propagated from the previous time instant. To find out a suitable classification label **L** under the given image observation **C** and the predicted labels **P**, we solve the following MAP optimization problem:

$$\begin{aligned} L^* &= \arg\max_L p(L\,|\,C,P) = \arg\max_L p(C\,|\,L)p(L,P) \\ &= \arg\max_L [\ln(p(C\,|\,L)) + \ln(p(L\,|\,P)) + \ln(p(P))], \end{aligned} \qquad (2)$$

where *p(C|L)* is the likelihood terms and *p(L|P)* denotes the label messages form temporal prior. Since *p(P)* is a constant, it can be ignored. In our approach, once if **L** is given, we assume the conditional probability density function of the observation data at two different pixels are independent of each other. We also assume the data $\vec{c}_i$ at Pixel i doesn't depend on the labels at other pixels. With these two assumptions, we define

$$p(C\,|\,L) = \prod_{i=1}^{K} e^{-E_D[\vec{c}_i, L_i]} e^{-E_A[\vec{c}_i, L_i; N_p]}, \qquad (3)$$

where K is the total number of image pixels. $E_D[\vec{c}_i, L_i]$ is the "classification energy" for the labeling node L_i and the feature data $\vec{c}_i$ at the i^{th} pixel. Here, we define $E_D[\vec{c}_i, L_i]$ as

$$E_D[\vec{c}_i, L_i] = \begin{cases} -\ln(p(\vec{c}_i \mid \Omega_B)) & \text{if } L_i = \Omega_B \\ -\ln(p(\vec{c}_i \mid \Omega_F^g)) & \text{if } L_i = \Omega_F^g \end{cases}. \quad (4)$$

On the other hand, we define the "adjacency energy" $E_A[\vec{c}_i, L_i; N_i]$ based on a 4-neighbor MRF model [7], where N_i denotes the 4-connectivity neighborhood of Pixel i. That is,

$$E_A[\vec{c}_i, L_i; N_i] \equiv \sum_{j \in N_i} (\beta \times (1 - \delta[L_i, L_j]) / (\|\vec{c}_i - \vec{c}_j\| + \alpha)), \quad (5)$$

where β is a normalized constant, α is a small constant to avoid division by zero, and $\delta(\cdot)$ is defined as

$$\delta[p, q] = \begin{cases} 1 & \text{if } p = q \\ 0 & \text{otherwise} \end{cases}. \quad (6)$$

Basically, $E_A[\vec{c}_i, L_i; N_i]$ denotes the spatial correlation between pairs of classification labels (L_i, L_j). This energy softly forces neighboring pixels to share the same label, especially when they have similar spatio-chromatic features.

In addition, $p(L|P)$ represents the expected labeling map based on the previous prediction. Here, we assume the predicted image location of each foreground object at the current instant t could be modeled as probability $P_g(x,y;t)$; the g^{th} object at time instant t-1 is bounded by a compact rectangular box $RB_{g;t-1}$ around the g^{th} object. The extraction of $P_g(x,y;t)$ and $RB_{g;t-1}$ are to be explained later. To model $p(L|P)$, we adopt the Monte Carlo based method to draw many expected labeling samples and approximate $p(L|P)$ in a sample-based manner. To generate a labeling sample, we draw a location sample $(x_s,y_s)_g$ from $P_g(x,y;t)$ for each object, and warp the center of $RB_{g;t-1}$ to $(x_s,y_s)_g$. While the rectangular boxes get overlapped, inter-occlusion is expected to occur and the depth order is needed to determine the occlusion pattern. Here we adopt the Bhattacharyya coefficient (BC) based metric [4] to determine the depth order. Basically, if a predicted target region is more similar to its target model in appearance, that target has a higher possibility to be the object that occludes the others. In detail, for a target g, we measure the Bhattacharyya coefficient at location $(x_s,y_s)_g$ as

$$\rho_{x_s,y_s}(g) = \int \sqrt{h_{x_s,y_s}(z;g)\, p(z;g)}\, dz, \quad (7)$$

where $h_{x_s,y_s}(z;g)$ is the normalized color histogram of the image region inside the warped $RB_{g;t-1}$ centered at $(x_s,y_s)_g$, P(z;g) is the normalized color model of target g, and z denotes a possible (r,g,b) color feature. Here, P(z;g) is derived from the foreground model of the target g based on

$$P(z;g) = \iint p(\vec{c} \mid \Omega_F^g)\, dxdy. \quad (8)$$

By comparing the BC values among inter-occluded objects, the depth order is determined and an expected labeling sample is generated. By accumulating the occurrence number of different labeling IDs at each pixel from many expected labeling samples, we can model $p(L|P)$ to well handle occlusion. In Fig. 2 we show an example of $p(L|P)$.

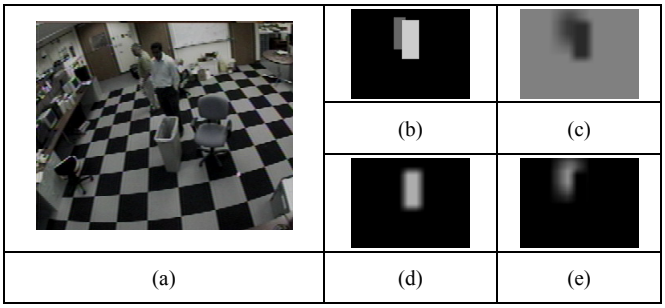

Figure 2. (a) Test image. (b)An expected labeling sample of (a). (c) Estimated p(L|P) for $L=\Omega_B$ or Ω_F^0. (d) Estimated p(L|P) for $L=\Omega_F^1$. (e) Estimated p(L|P) for $L=\Omega_F^2$.

Based on (1)~(8), we form the formulae for MAP optimization. We adopt the Graph Cuts method [7] to find the optimal label $\mathbf{L}^*$ that maximizes (2). Based on the classified labels, we detect foreground objects. Moreover, for each non-occluded foreground object, the rectangular box $RB_{g;t}$ at the current time t is estimated from the vertical and horizontal projection histograms [8] of its foreground region. For both vertical and horizontal directions of $RB_{g;t}$, we search for the minimum continuous-valued ranges that can cover 95% energy of the projection histograms. For occluded objects, the size of rectangular box remains the same value at the previous time instant but the center of the box is shifted to the new object center. Besides, we also identify the new comers by evaluating the vertical projection histogram of the foreground region with the ID Ω_F^0.

3. OBJECT TRACKING

As mentioned above, the predicted temporal prior **P** at the current frame is warped from the classification results $\mathbf{L}^*$ at the previous frame. To provide the temporal message and to model the inter-frame relation, a dynamic tracking model is maintained for each foreground object. Moreover, since there could be some errors in the prediction of foreground movement, the result of classification labeling is fed back to update the dynamic models of foreground objects. Under the proposed architecture, object tracking is actually treated as the temporal prediction and update of object labels.

To design a tracker for each foreground object, the Bayesian-based filters are widely used. In this work, we adopt the Kalman filter for the sake of computational complexity. Here, we define $\mathbf{S_t}=(\mathbf{x_t},\mathbf{v_t})$ as our motion state, including object center $\mathbf{x_t}=(x,y)$ and object velocity $\mathbf{v_t}=(v_x,v_y)$. Based on the Kalman filter updating rule, F(.), for each object, the optimal estimation of object motion state $\mathbf{S_t}$ is determined by

$$S_{t|t} = F(S_{t|t-1}, K_t, z_t), \quad (9)$$

where $\mathbf{S_{t|t-1}}$ is the optimal prediction of $\mathbf{S_t}$ based on its previous motion state $\mathbf{S_{t-1|t-1}}$; $\mathbf{z_t}$ is the observed object center determined by the object detection in (2); K_t is the Kalman gain. With the Kalman filter, the probability of the predicted location of g^{th} object $P_g(x,y;t)$ is modeled by a Gaussian distribution $N(\mathbf{x_{t|t-1}},\mathbf{Q}_k)$, with the covariance of the noise process $\mathbf{Q}_k$.. The detail of the Kalman filter is not stated here.

To explain the interaction of object detection and tracking, we assume the classification label L^* at time instant t-1 has been determined. Based on the classified label L^*, a few foreground objects are detected. For each foreground object, we calculate its $RB_{g;t-1}$ and measure its object-mass-center (OMC_{t-1}) as the observation data $\mathbf{z_{t-1}}$. At the current time instant t, we track the location of each foreground object based on its OMC_{t-1} and $RB_{g;t-1}$.This object tracking process consists of the following 4 major steps. An illustration of this object tracking process is shown in Fig. 3

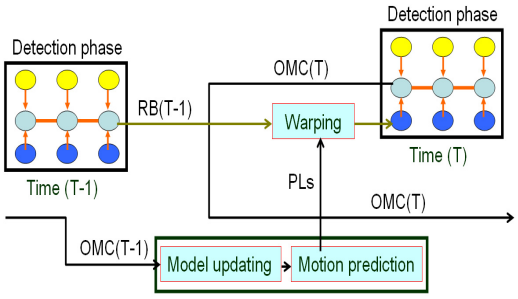

Fig.3: Illustration of the tracking process.

Step1:Creation/Update/Deletion of Tracking Model

For new comers, their Kalman trackers are created. Next, for each foreground object, its OMC_{t-1} is used to update the tracking model by (10). Based on the updated model, we draw 255 predicted locations PL's from $P_g(x,y;t)$. For objects having no motion for a long enough period or moving out the scene, their tracking models are deleted.

Step2: Temporal Propagation of Foreground Labels

Based on the PL's, the $RB_{g;t-1}$'s at time t-1 are warped to their new location at time t to construct the expected labeling map $p(L|P)$ at time t.

Step3: Update of object models

Based on classification label L^* at time t-1, we update both foreground and background classification models. Moreover, we predict the location and appearance of each foreground object and update its foreground model before detection. The detail is described in Section 4.

Step4: Foreground/Background Labeling

At time t, we deduce the optimal classification label $\mathbf{L}^*$ based on the optimization of (2). From the optimal $\mathbf{L}^*$, we detect a set of foreground objects at the current time t.

4. UPDATE OF OBJECT MODELs

To adapt to a varying environment, the foreground model and background model in (1) should be updated all the time. Traditionally, model updating is performed after the detection stage. This makes it very difficult to handle the foreground/background ambiguity problem. On the contrary, we update the foreground model ***before*** we perform object labeling. That is, if the foreground object is currently at $\mathbf{x}_{t|t}$ and we predict the optimal location of this object will move to $\mathbf{x}_{t+1|t}$, we adjust the foreground model accordingly so that $P(\vec{c} \mid \Omega_F^g)$ will be high around both $\mathbf{x}_{t|t}$ and $\mathbf{x}_{t+1|t}$. With this mechanism, if the foreground object happens to move into some background region with a similar appearance, both $P(\vec{c} \mid \Omega_B)$ and $P(\vec{c} \mid \Omega_F^g)$ will be high within the ambiguous region. The update of foreground model will reduce the probability that a foreground region being mistakenly classified as a background region. Moreover, since the prediction layer **P** also provides useful prior knowledge about the predicted location of foreground objects, the foreground/background ambiguity problem can be more effectively solved.

On the other hand, we update the background model $P(\vec{c} \mid \Omega_B)$ based on the result of foreground/background labeling. In our approach, only those pixels labeled as background pixels will be considered in the update of background model. Occasionally, a foreground object may become a part of the background, like the situation that a car parks in the scene for a long time. For this kind of situation, we may simply check whether the foreground object has been motionless for a long enough period. If so, the features of the foreground object can be added into the background model.

5. EXPERIMENTS RESULTS

We test our system over the IBM datasets [9], OVVV datasets, and our own datasets. We

also do comparison with the GMM method [3], as shown in Fig 4. In Fig 4(b,c), due to the appearance ambiguity between foreground object and background, the GMM method generates fragmented results. Instead, our method well adopts the object prior from temporal and can still robustly detects the whole foreground object. In Fig 4(a,b), our labeling results clearly identify the inter-object occlusion and the depth order. Moreover, in Fig 4(d), owing to camera shaking, the GMM method generates lots of false detections. With the use of the kernel function in (1), the proposed method generates reliable detection result. Besides, our system can detect the new comers or the vanishing objects automatically. In fact, an object is leaving in Fig 4(a). To quantitatively evaluate our system, we use the ground truth and the metrics proposed by IBM [9]. The evaluations are listed in Table 1. Currently, the whole system is implemented in Visual C++ on a PC with a 2.4 GHz CPU. It takes about 1 second to perform the detection and tracking for a 320x240 color image frame. For more experimental results, please visit our website at http://140.113.238.220/~chingchun/projects.html. In the future, we plan to move our system to a GPU based platform to boost the system speed.

6. REFERENCES

[1] T. Horparasert, D. Harwood, L. A. Davis, "A Statistical Approach for Real-time Robust Background Subtraction and Shadow Detection," *IEEE Conf. ICCV*, 1999.

[2] Chris Stauffer, W. Grimson, "Adaptive Background Mixture Models for Real-time Tracking," *IEEE Conf. CVPR*, 1999.

[3] P. Power and J. A. Schoonees, "Understanding Background Mixture Models for Foreground Segmentation," *Image and Vision Computing*, 2002.

[4] D. Comaniciu, V. Ramesh, P. Meer, "Real-time Tracking of Non-rigid Objects Using Mean Shift," *IEEE Conf. CVPR*, 2000.

[5] R.T. Collins, Y. Liu, and M. Leordeanu, "Online Selection of Discriminative Tracking Features," *IEEE Trans. PAMI,* 2005

[6] Y. Sheikh, M. Shah, "Bayesian Modeling of Dynamic Scenes for Object Detection," *IEEE Trans. PAMI*, 2005.

[7] T. Boykov, O. Veksler, R. Zabih, "Markov Random Fields with Efficient Approximations," *IEEE Conf.e CVPR,* 1998.

[8] I. Haritaoglu, and L. Davis, "W4:Real-time surveillance of people and their activities," *IEEE Trans. PAMI*, 2000.

[9] H. Merkl and M. Lu, "Performance evaluation of surveillance systems under varying conditions," *IEEE PETs Workshop*, 2005

Table. 1. Evaluation of 5 tested sequences. (a)(b)(c)IBM "Line_Circle", "Splite", and "Circle" sequences. (d)(e)Two OVVV sequences. The adopted IBM metrics are frames number (FraN), true positive (TP), false positive (FP), false negative (FN), Track TP (TTP), Track FP (TFP), and Track FN (TFN) [9].

Seqs	FraN	TP	FP	FN	TTP	TFP	TFN
(a)	415	377	4 / 415	8 / 377	2	0	0
(b)	352	372	8 / 352	12 / 372	3	0	0
(c)	371	657	17 / 371	11 / 657	3	0	0
(d)	300	750	1 / 300	0 / 750	3	0	0
(e)	1000	2746	0 / 1000	32 / 2746	17	0	1

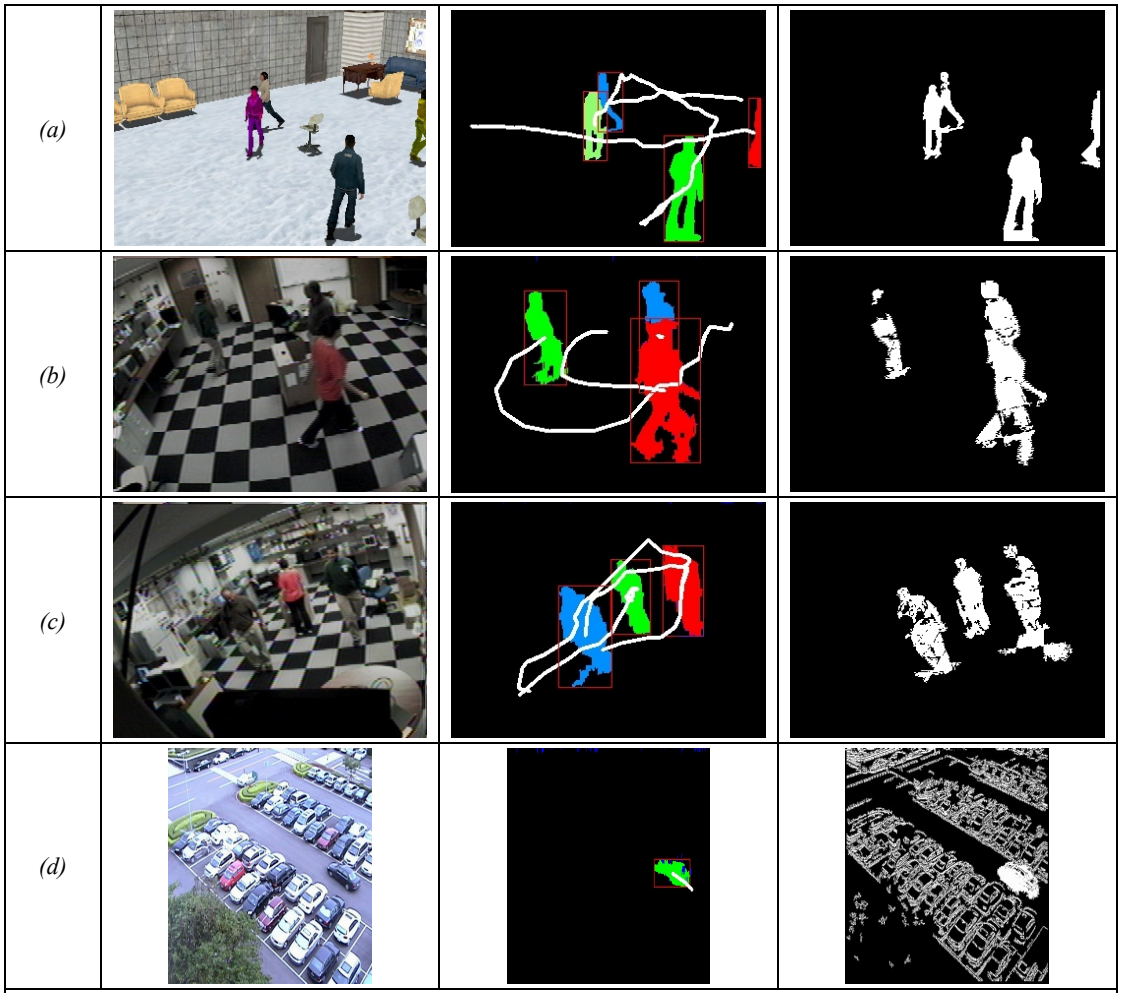

Fig. 4. Experimental results. 1st, 2nd, and 3rd rows are the tested sequences, our results, and results of [3], respectively. (a) OVVV dataset. (b)(c) IBM dataset. (d) Our outdoor dataset. (Tracking results are shown as white curves in the 2nd row.

計畫成果自評

在本計畫中，希望能建構一主動式的多攝影機監控系統，當所監控的場景內出現運動人物時，本系統一方面估測這些人物的三維運動軌跡及姿勢，一方面控制
協調各攝影機進行上下左右轉動或是鏡頭縮放，以達到較佳的監控取像。在研究發展這樣的監控系統過程中，我們利用多攝影機的二維運動偵測結果來估算出被追蹤物在三度空間中的定位與姿勢。此時，對於運動偵測與追蹤的精確度與準確度，都將有更高的要求。因此，在本年度的研究中，我們根據原計畫內容的規劃，開發出更為準確的運動物體偵測與追蹤技術。在今年的研究成果中，我們系統所偵測出運動物體的結果不再只是單純的bounding box而已，而是準確地偵測出運動物體的輪廓，從技術的層面來看，我們的研究成果提出了一套架構，在偵測多人物的同時，可以考慮物體相互遮蔽的問題，以及攝影機搖晃所造成的錯誤偵測。此外，我們的架構同時針對前景物體與背景建立不同的外貌模型，因此可降低前景區域與背景區域因為外貌相似而造成偵測結果破碎不完整的情況。透過這種準確的偵測結果，我們得以進一步推算出三維空間的姿勢，朝向原訂之主動式智慧型監控系統邁進。