

視覺式行人偵測追蹤技術之發展

The Development of Vision-based Pedestrian Detection and Tracking Technology

胡啟恩、郭雅諒、王元凱

Chi-En Hu, Jonathan Guo, Yuan-Kai Wang

現今人工智慧受到高度矚目與快速發展，電腦視覺身為其中之關鍵技術也己成為許多研究及應用的焦點，而隨著智慧型視訊監控與無人駕駛自動車的快速發展，以電腦視覺方法進行影像視訊中的行人偵測與追蹤，更成為一個重要而關鍵的技術。行人追蹤與行人偵測在電腦視覺研究上為兩項不同的技術，但是兩者在應用面卻有十分緊密的關係。近幾年因深度學習的崛起，行人偵測正確率有顯著的進步，也使得行人追蹤有更好的效果。本文將詳細介紹行人偵測追蹤技術，包括行人偵測與行人追蹤技術的發展，並介紹在線多物件追蹤系統之技術內容。

As of today, Artificial Intelligence has gained a lot of attention and is currently being developed rapidly. As one of the key technologies, Computer Vision has become the focal point of many researches and applications, due to the rapid development of pedestrian detection and tracking methods used in Intelligent Surveillance Systems and Self-driving Cars. Pedestrian detection and tracking are two different technologies, yet inseparable when used in applications. Due to the rise of deep learning, pedestrian detection has made great progress in accuracy, which also has driven the improvement of pedestrian tracking. In this paper, we in detail go through the introduction of pedestrian detection and tracking technologies, including the development of both technologies, and also the introduction of online multiple object tracking systems.

一、人工智慧與電腦視覺之發展

人工智慧 (AI) 之目的在於使電腦具備近似人類之學習及抽象思考等能力，進而解決與人類智慧相關的常見認知問題，例如推理、感知及識別等。人工智慧應用領域也相當廣泛，在現今這個物聯網發展的時代中，人工智慧儼然成為了這一切發展的核心，例如個人語音助理、無人駕駛自動車、人臉辨識身分認證、擊

敗人類之圍棋軟體 AlphaGo 等，都是典型的人工智慧實例。隨著未來研發技術的發展，人工智慧將會更深入我們的生活中。

在討論人工智慧以及電腦是否擁有學習能力的同時，就不得不提及電腦視覺。所謂的電腦視覺便是關注於如何讓機器與計算機擁有視覺的能力，如同人類大腦與眼睛共同工作以處理視覺所得到的資訊，進而解決我們周遭複雜的問題以達成智慧應用。電腦視覺領域的一個實例如

無人駕駛汽車，該應用最大的困難在於需要隨時感知路面上的情況並計算應對行為，此即需要仰賴電腦視覺技術以提供無人駕駛汽車視覺感知能力；另外一個應用實例則如智慧視訊監控系統，可以利用監視器所回傳的影像，即時分析出各種資訊，例如行人行為、車輛類型廠牌與顏色等，這些都是透過電腦視覺進行影像辨識的結果。由此可知，人工智慧與電腦視覺已形成了密不可分的關係。

電腦視覺欲解決的問題相當多樣，應用層面也極為廣泛，因此目前電腦視覺已發展出許多重要的研究方向，其中最經典的便是識別問題。當電腦需要判斷一張或多張影像中是否包含某個特定物體、該物體擁有的圖像特徵或運動狀態時，不論此物體是簡單幾何圖形、手寫數字文字或複雜的車輛行人等，都稱之為識別問題。其中一個應用例如智慧型監控系統便是利用行人偵測確定行人的位置，接著再做後續的影像分析，例如行人追蹤、人臉辨識等，進而取得運動軌跡、行人身份等我們欲取得的資訊。其他還有如三維場景重建、虛擬實境、擴增實境等也可以包含在電腦視覺技術範疇。

二、行人偵測演算法

行人偵測是一個具有廣泛應用的重要研究題目。例如在智慧型駕駛輔助系統中，行人偵測就是一個不可或缺的功能，系統會利用雷達或攝影機偵測出現在車輛周圍的行人，在可能發生碰撞時發出警告，藉此保障駕駛人與行人的安全；類似的功能也可應用在無人機器人中，無人機器人在行進中也需要擁有避開障礙物的能力，此時也能應用行人偵測的技術，利用攝影機確定前方行人的位置，藉此

修正行進路線避免碰撞；又例如在前段提到的智慧型監控系統，也是利用監視攝影機的畫面，透過行人偵測以及追蹤確定行人的位置及移動軌跡，當發生意外或事故時，可以快速提供警方過濾可疑人物以及其動向，以上例子都是有關行人偵測的應用。

行人偵測技術可以區分為雷達感測與視覺感測兩大類，其中視覺式行人偵測為透過低成本攝影機作為硬體感測裝置再加上軟體演算法進行分析以達偵測目的，因此具有低硬體需求、低製造成本、小體積易於嵌入自動化設備等優點。視覺式行人偵測法主要透過影像取得外部環境資訊，並對影像進行解析，先抽取物體之影像特徵，接著透過分類器來確認影像中行人的存在。目前主流的視覺式行人偵測技術都是以機器學習方式來學習分類器的參數以達最佳正確率，而近年崛起之深度學習行人偵測方法則是更同時自動學習物體影像特徵與分類器參數等兩個部分，達到更高之偵測率。

1. 傳統機器學習方法

傳統機器學習方法的架構分為兩個部分：一個事先設計好的特徵擷取方法以及一個可以學習的分類器。因此其計算原理是對影像擷取指定的特徵後接著再利用不同大小的滑動視窗來對特徵影像取樣，這些樣本會送入一個已經訓練好的分類器做分類，藉此達到行人偵測的效果。舉例來說，全域搜尋偵測法就是一個典型的例子，此作法的概念是以一個固定大小的滑動視窗在影像中擷取數個子影像，再將數個子影像送入疊接 (cascade) 分類器中，藉此確認有無目標物件的存在及位置。全域搜尋偵測法的偵測流程如圖 1 所示，首先輸入原大小的影像，接著利用固定大小的滑動視窗，以由左而

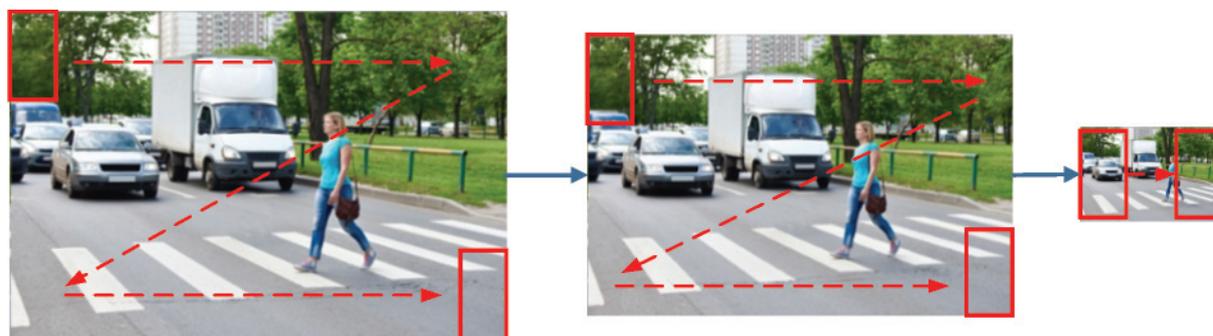


圖 1. 全域搜尋偵測法示意圖。



圖 2. HOG SVM 行人偵測流程圖。

右，由上而下的方式掃描整張影像，過程中每個滑動視窗會送入疊接分類器中，用以判斷視窗內是否有無目標物件的存在；完成一次掃描後，接著會將輸入影像做縮小的動作，再對縮小後的影像以同樣大小的滑動視窗進行同樣方式的掃描並判斷是否有無目標物件，重複以上動作直到縮小影像的寬或高與滑動視窗的寬或高相同為止。

傳統做法中另外一個經典演算法是由 N. Dalal 與 B. Triggs 提出的梯度方向直方圖 (histogram of oriented gradient, HOG)⁽¹⁾ 行人偵測演算法，整個方法也是採用全域搜尋並也分為二大步驟：特徵擷取與線性分類器。第一步驟先對影像計算特徵描述，此處即為影像的邊緣梯度方向，後再利用滑動視窗來對特徵影像作取樣，第二步驟會將樣本影像送入分類器做分類，藉此確認是否有行人的存在及位置。整個演算法的流程如圖 2 所示，首先輸入欲偵測行人的影像，在前處理進行 gamma 值校正以降低因攝影機不同造成的影像像素值誤差；接著計算影像的梯度方向直方圖，此部分會將滑動視窗影像切分為數個區塊及細胞，每個區塊內包含 4 個細胞，在取得影像中每個像素的邊緣資訊後，便可將不同強度的邊緣資訊轉換為向量資訊。為了簡化特徵值的數量及範圍，該方法將向量分為 9 個方向區間；接著統計細胞內各像素的梯度方向，便可得出有 9 個 bins 的直方圖，最後將區塊內 4 個細胞的直方圖依順序連接成為一個 36 個 bins 的區塊直方圖；計算完滑動視窗內的所有區塊直方圖即可得到滑動視窗的特徵資訊。接著與全域搜尋偵測法相同，滑動視窗會由左而右，由上至下擷取整張影像

的特徵資訊，每個滑動視窗都將擷取特徵再送入一個訓練後的線性分類器進行分類。此處使用的線性分類器為支持向量機 (support vector machine, SVM)，其為一種機器學習常見的分類器，主要目的為找出一個超平面將多維度的資料進行分類，訓練時採用監督式的學習方式，當我們給予一群已分類完成的資料，SVM 可經由訓練得到一組模型，之後當有新的未分類資料進入時，SVM 便可依據先前已訓練好的模型，預測此筆資料所屬的分類。此梯度方向直方圖演算法便是利用 SVM 來判斷滑動視窗內是否有行人存在。

可變形部件模型 (deformable part model, DPM)⁽²⁾ 也是傳統做法中經典的演算法之一。由於影像中的物件經常會因其動作或角度而有多種姿態，以行人為例，行人的雙手及雙腳姿勢變化都會導致物件偵測上的困難，因此 DPM 希望能透過使用多個模型組成的混合模型來對目標進行特徵擷取，藉此適應物件在影像中形變的問題。整體而言，DPM 與 HOG-SVM 演算法類似，依然是使用 HOG 擷取物件的特徵，並使用 SVM 分類器進行分類。特徵截取的部分，DPM 除了原本完整的物件特徵外，還會另外將物件分割為多個部位，計算物件部位的 HOG 特徵。整個偵測流程一樣使用滑動視窗對影像做掃描及特徵截取，擷取 DPM 模型時會對影像進行特徵金字塔上的採樣，特徵金字塔由一定範圍內尺度變化的特徵圖所構成，藉此作法不僅可以取得解析度更高的部位特徵，同時也能增加辨識的準確度，接著根據每個部位的位置給予其分數權重，最後根據分數加權將所有的部位特徵與原始完整特

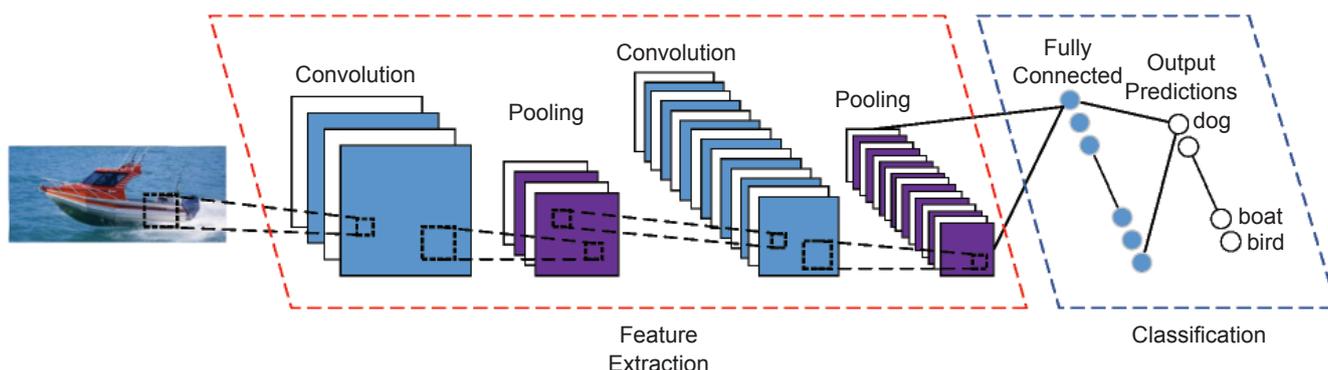


圖 3. CNN 分類器概念圖。

徵結合成一個混合模型，最後同樣使用一個經過訓練的 SVM 對 DPM 特徵做行人及非行人的判斷。DPM 的作法雖然增加了對物件形變的容忍度，但對於更大範圍的形變，如旋轉或拉伸，效果依然不佳，且因特徵模型相對複雜導致計算速度也較慢。

2. 深度學習方法

深度學習的方法中以 Y. LeCun 等人提出了 convolutional neural network (CNN) 為最典型的演算方法，並常見於電腦視覺之研究。CNN 的概念圖如圖 3 所示，此方法之架構主要分成兩部分的多層網路：特徵擷取網路層以及分類網路層。特徵擷取網路層大致由兩種網路層組成，分別是卷積層 (convolution) 與池化層 (pooling)。卷積層會將原始圖片與特定的濾波器做卷積運算，進而產生多張特徵圖 (feature map)，這些特徵圖目的是幫助我們擷取出圖片中的特徵。接著通過池化層減少數據的空間大小，目的為降低計算量與參數量。經過多次反覆的卷積層與池化層的計算後，網路便可擷取出最有效的特徵。此時網路進入第二部分，也就是分類網路層。分類網路層由全連接層 (fully connected) 組成，全連接層會連接前部分特徵擷取網路的結果，判斷該特徵屬於何種類別，以圖 3 為例，分類網路層輸出即為分類預測類別，此範例即為經典的使用 CNN 實現物件分類器。在深度學習方法中，特徵擷取網路層即為整個架構的核心技術，此概念同樣也可應用於行人偵測器中，它的功能等同梯度方向直方圖或 DPM 等擷取特徵的方法，唯獨不同的地方是 CNN 所擷取的特徵並非人工設計的，而

是藉由訓練網路而學習到的，也因此 CNN 能擷取到更為關鍵的特徵進而達到更好的效果。分類網路層則等同於 SVM 等分類器，透過學習利用前一個網路所提供的資訊來分類影像是否為物件。以上方法在近年研究中效果較佳也較為穩定，因此是目前該領域的研究主軸。

近年來的行人偵測演算法中，以 CNN 為架構之深度學習法成為了主要的研究方向，其中一個較有代表性的方法為 R. Girshick 等人提出的 Region-based CNN (R-CNN)⁽³⁾。此方法利用目標建議 (object proposal) 來事先找出物件可能出現位置，然後利用 CNN 擷取特徵後，最後再將特徵樣本送入 SVM 進行分類以及使用偵測框回歸 (bounding-box regression) 的方法來使偵測框更為準確。

繼 R-CNN 之後，Fast R-CNN⁽⁴⁾ 將分類器與偵測框回歸改用類神經網路來實現，並且提出特徵共享 (feature sharing) 以及感興趣區域池化 (ROI pooling) 的概念。S. Ren 等人所提出的 Faster R-CNN⁽⁵⁾ 則更進一步的以一個 region proposal network (RPN) 取代原本目標建議的方法來加快運算速度。然而 Faster R-CNN 中因為 RPN 取樣的方式仍然過於固定使得過大或過小的物件都難以偵測。Z. Cai 為了解決這個問題提出了 multi-scale CNN (MS-CNN)⁽⁶⁾ 的架構，MS-CNN 會針對不同大小的物件選用不同深度的 CNN 所擷取出來的特徵影像，藉此希望利用較淺的 CNN 特徵解析度高這樣的特性，來偵測較小的物件，而較深的 CNN 特徵則比較適合偵測較大的物件，使整體可偵測的物件大小範圍更廣更大。

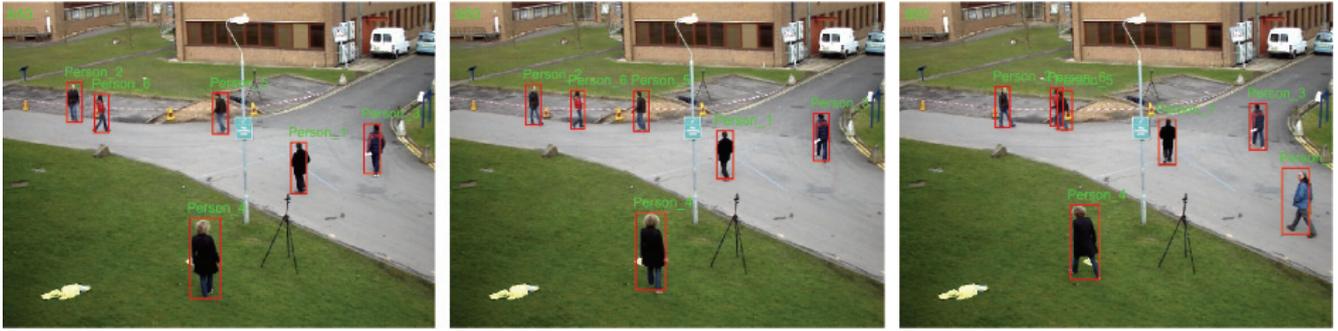


圖 4. 多物件追蹤示意圖。

三、多物件追蹤演算法

在上述提到的汽車的自動駕駛或監控系統的應用中，多物件追蹤也是其中重要的核心技術之一。多物件追蹤即是在一部影片中找出所有獨立的指定物件並且取得各個物件的移動路徑資訊。多物件追蹤的目的大致上可分為三個動作：定位感興趣的物件、給予物件一個標記編號並持續追蹤物件、取得物件於連續影像中的移動軌跡。圖 4 為多物件追蹤示意圖，圖中的追蹤物件以行人為例，演算法需標出所有出現在影像中的行人，並同時對多個行人進行追蹤，隨著時間推進最後取得這些行人的移動軌跡。

在解決多物件追蹤問題上，tracking-by-detection 是一種常見的追蹤策略。在 tracking-by-detection 中，首先會使用物件偵測器偵測當前時間影像中的物件，取得物件的位置；接著與過去時間中的物件做比對，其中一類較常見的做法是對物件

進行表面特徵截取，接著與過去時間中的物件做表面特徵的比對，另一類作法會紀錄物件的運動狀態，這些狀態包含物件的位置、移動速度、移動加速度…等，接著以過去的移動狀態為依據預測當前時間的物件位置並與偵測器結果做比對；做完比對後將被認為是同一物件的偵測框做連結，最後便可取得物件的移動軌跡。整個追蹤概念如圖 5 所示，圖 5(a) 為時間 $t-1$ 的影像，而圖 5(b) 為時間 t 的影像，影像中的紅色圓代表目標物件，分別為編號 1, 2, 3 三個物件，灰色矩形代表場景遮蔽物，一開始先取得目標物件位置，接著利用表面特徵或運動資訊預測物件接下來的位置，接著在下個時間點與偵測框做比對，若一致則做連結維持其編號，持續取得物件軌跡。此策略能自然地處理新物件的出現與既有物件的消失，但相當依賴物件偵測器的效能，若是物件偵測器的效果不好，則會直接影響追蹤的效果。

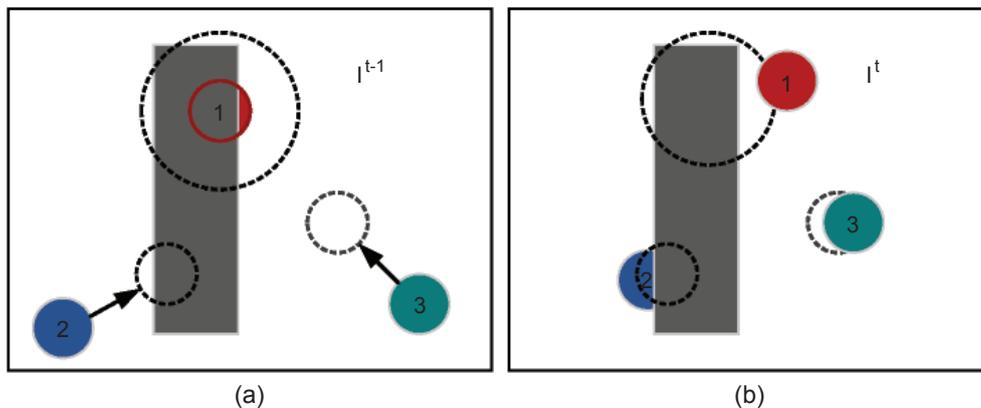


圖 5. 追蹤原理示意圖。(a) 時間 $t-1$ 之影像，(b) 時間 t 之影像。

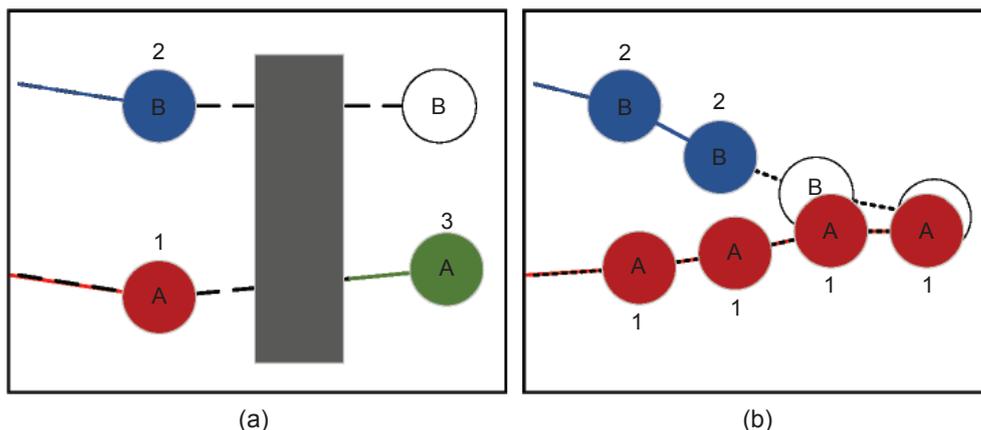


圖 6. 多物件追蹤挑戰示意圖。(a) 景物遮蔽、(b) 物件互相遮蔽。

目前的多物件追蹤演算法可以分為離線與在線兩大類。離線作法是指演算法會取得所有時間的偵測資訊後才會開始分析各個物件的路徑，這種方法因為在運行時已經持有完整的物件位置歷史資訊，因此可以使用一些同樣為離線作法的路徑最佳化演算法來使取得的路徑更為準確，但是這樣做的缺點在於運行時間長而且無法應用於需要即時的情境中。在線的方法則是指每當有新的偵測框，演算法就會利用這些新的資訊來做進一步的追蹤，雖然準確性不會比離線作法來的好，但是在應用上會相對較為廣泛。

多物件追蹤有許多必須要克服的困難與挑戰，有單物件追蹤就必須面對的物件外觀改變、尺度變化、物件旋轉、因背景物件而造成的遮蔽等。多物件追蹤除了以上挑戰之外，還必須面對其他挑戰，例如：當人群擁擠或行人交錯時造成的遮蔽、必須能應對隨時會有新的行人出現或既有的行人離開場景的情況、如何區分具有相似外觀的行人等。例如圖 6 便說明了 2 種可能發生追蹤失敗的情況。圖 6(a) 出現 2 個物件，分別為 A 與 B，A 與 B 出現時的 ID 分別為 1 與 2，此 2 個物件移動時將因場景障礙物造成短暫遮蔽，此時追蹤器常見的做法為記錄物件的外觀特徵，以便在物件脫離遮蔽後做重新連結。若物件的外觀改變或追蹤器紀錄的外觀不夠精確，則就有可能發生如圖 6(a) 的結果，圖中可看到 A 在脫離遮蔽後並未重新連結，反而被辨認為新出現的物件而給予新的 ID；而 B 則因偵測器失誤而不被辨認為物件，進而造成追蹤失誤。另一

情況如圖 6(b) 所示，途中同樣有 A 與 B 兩物件，ID 分別為 1 與 2，圖中原為 2 條獨立的軌跡因 2 物件距離相近，導致 B 被 A 大範圍遮蔽，進而導致 B 的追蹤失誤，此情境在追蹤人群時相當容易發生，如何追蹤人群中的各個物件也是多物件追蹤最重要的挑戰之一。

另外，在 tracking-by-detection 策略下需要利用行人偵測器來確定行人的位置，若行人偵測器的準確率不夠好，無論如何改善追蹤演算法，演算法依然會難以取得完整的軌跡連結進而影響最終的追蹤效果。

多物件追蹤目標在於如何將每張偵測中的偵測框 (bounding box) 做最佳的連結，因此大多數方法都可以定義為一個最大後驗機率問題。最大後驗機率問題的解決方法可以分為兩大類：第一類是將問題當成一個確定性的優化問題解決；第二類是使用機率性推論的方法。第一類作法適合用於離線的追蹤方法，因為這類做法會希望使用全部時間的偵測資訊，來取得全局的最佳連結軌跡。

第二類使用機率性推論的方法會將目標的狀態視為一個不確定性的機率分佈，基於現有偵測資訊來估計目標的機率狀態分佈，較適合用於在線的追蹤方法。實際分為 2 步動作，首先預測目標的位置；接著更新目標的狀態。常見的作法例如⁽⁷⁾ 便是使用卡爾曼濾波器作為目標的運動模型，結合匈牙利演算法做數據連結實現多物件追蹤；而 MDP Track⁽⁸⁾ 則是使用 markov decision processes (MDPs) 模型來決定目標的狀態以及追蹤策略。

四、在線多物件追蹤系統

在線的多物件追蹤系統，系統採用 tracking-by-detection 架構，整個運作流程如圖 7 所示。第一部分為物件偵測器；第二部分為追蹤器。行人偵測器可以分為三個步驟：首先取得行人可能出現的區域；接著擷取特徵；最後分類辨識是否是行人。而追蹤器採取分成四個步驟：首先預測目標位置；接著將目標資訊與偵測框做比對並做連結；接著更新追蹤器的狀態；最後判斷是否有物件出現或消失。這樣的架構能即時處理新物件的出現與舊物件的離開，但偵測器失誤則會直接影響追蹤器的效果。舉例來說，false negatives (FN) 可能使得追蹤器無法連結到目標，false positives (FP) 則可能使得追蹤器連結錯誤目標。

擁有偵測廣大範圍大小物件的能力對於一個偵測器來說是非常重要的，應用在追蹤上也更是如此，畢竟追蹤演算法大多都仰賴偵測演算法來準確的提供下一步運算所需的資訊。若行人走遠了、變小了都還是希望能繼續追蹤下去。MS-CNN 的架構主要來自 Faster R-CNN 是由 RPN 與 CNN 所組成，利用一個主幹 CNN 來擷取影像特徵，不同的深度的特徵會送入 RPN 來取得各種大小的區域建議 (region proposal)，最後送入子網路取得偵測

框。MS-CNN 的子網路擁有的輸出有兩組，第一組類別機率為一個類別的機率分佈輸出。第二組則是經過偵測框回歸所得到的偵測框。網路訓練採用了 multi-task loss 作為訓練網路時的評估依據。損失函數從兩個方面來計算誤差，分類的正確性與偵測框的準確性，再透過隨機梯度下降 (stochastic gradient descent) 即可得到最佳參數。

MDP-Track 將多物件追蹤視為多個物件的狀態決策問題，每個物件都會被建成一個 MDP 模型。MDP 模型有四個狀態：active、tracked、lost、inactive。首先當物件偵測器偵測到物件時，Active 狀態 MDP 使用一個經過訓練的 SVM 來對偵測框做分類判斷，如果偵測框是目標物件，會將狀態轉移至 tracked；反之則會將狀態轉移 inactive。tracked 狀態使用一個表面特徵模型持續對目標進行追蹤並維持狀態；當遇到遮蔽等原因無法進行追蹤時，會將狀態轉移至 lost。lost 狀態可以視為一個資料連結問題，MDP 將此使用一個訓練過的相似度方程式來解決，如果再次連結到目標時會將狀態轉移至 tracked，否則維持為 lost；當狀態持續為 lost 超過一段有效時間，會將狀態轉移至 inactive，此時 MDP 將中止所有動作且不會再做任何的狀態轉移。

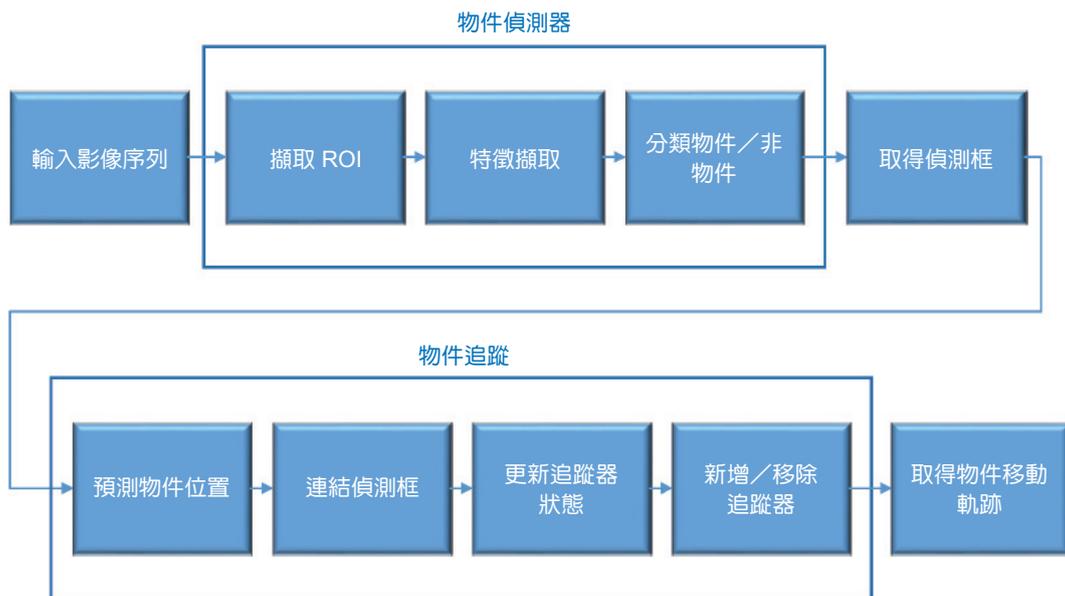


圖 7. 多物件追蹤系統流程圖。



圖 8. 實驗資料庫。(a) PETS09-S2L2⁽⁹⁾、(b) TUD-Crossing⁽¹⁰⁾。

五、實驗分析

實驗分析分為兩大部分，首先第一部分我們將對行人偵測的參數與行人追蹤的參數分別進行調整，並觀察各參數對最後追蹤結果的影響；接著第二部分將對多部實驗影片進行追蹤，並與其他追蹤方法進行比較。

實驗部份皆以行人作為追蹤物件，使用了 PETS2009 中的 S2L2，以及 TUD-Crossing 作為測試影片。PETS2009 拍攝視角屬於固定攝影機監控視角，S2L2 場景中會有大量行人團體行走的情況，也會有許多行人之間的交互遮蔽。TUD-Crossing 拍攝視角屬於固定攝影機平視視角。影片中會有多位行人於街道場景中行走，行人與行人間得遮蔽會比監控視角更多且面積更大。

效能評估部分使用多個評估標準，包含 multiple object tracking accuracy (MOTA)、multiple object tracking precision (MOTP)、switch of identities (ID switch)、FP、FN、mostly tracked (MT)、mostly lost (ML)。

FP 代表將非物件標示為物件。FN 將物件標示為非物件。追蹤期間物件編號發生變更的情況則稱為 ID switch。MOTA 考慮了所有目標編號上的配置錯誤，計算方式為 $1 - (FN + FP + ID\ switch)$ 。MOTP 顯示了估計目標位置的能力，計算方式為統計全部目標估計位置的總誤差再取平均。如果軌跡

被追蹤了 80% 以上，那就被分類為 MT；如果軌跡被成功追蹤不到 20%，那就被分類為 ML。

實驗第一部分將對行人追蹤系統的參數分別進行調整，觀察各參數對最後追蹤結果的影響，並從中找到最佳的參數。此部分實驗使用的測試影片為 TUD-crossing。首先，行人偵測的第一個參數為可信度閾值 (confidence threshold)，我們只留下可信度高於閾值的偵測框。此閾值調整影響如圖 9(a)，當閾值高於一定程度時，MOTA 與 MT 會下降，而 ML 則會上升。在此我們選擇的最佳參數為 0.5，使結果保持著最高的 MOTA。第二個參數為 non-maximum suppression 閾值 (NMS Threshold)，偵測器會將重疊率高於閾值的多個偵測框合併取得最佳偵測位置。此閾值調整影響如圖 9(b)，當閾值較高時，MOTP 會下降，MT 會上升。在此影片，我們選擇的最佳參數為 0.7，保持著最高的 MOTP。

接著，行人追蹤的參數為偵測框閾值 (bounding box threshold)，當預測位置與偵測框的重疊率高於此閾值時，便會作連結。閾值調整影響如圖 10，當閾值較低時，追蹤器較容易發生連結錯誤與 ID switch，因此 MT 與 MOTA 都較低，當閾值升高到一定程度後，追蹤器就會因閾值過高而無法連結到任何偵測框，MT 與 MOTA 都會開始下降。在此我們選擇的最佳參數為 0.5，可以保持著最高的 MOTA。

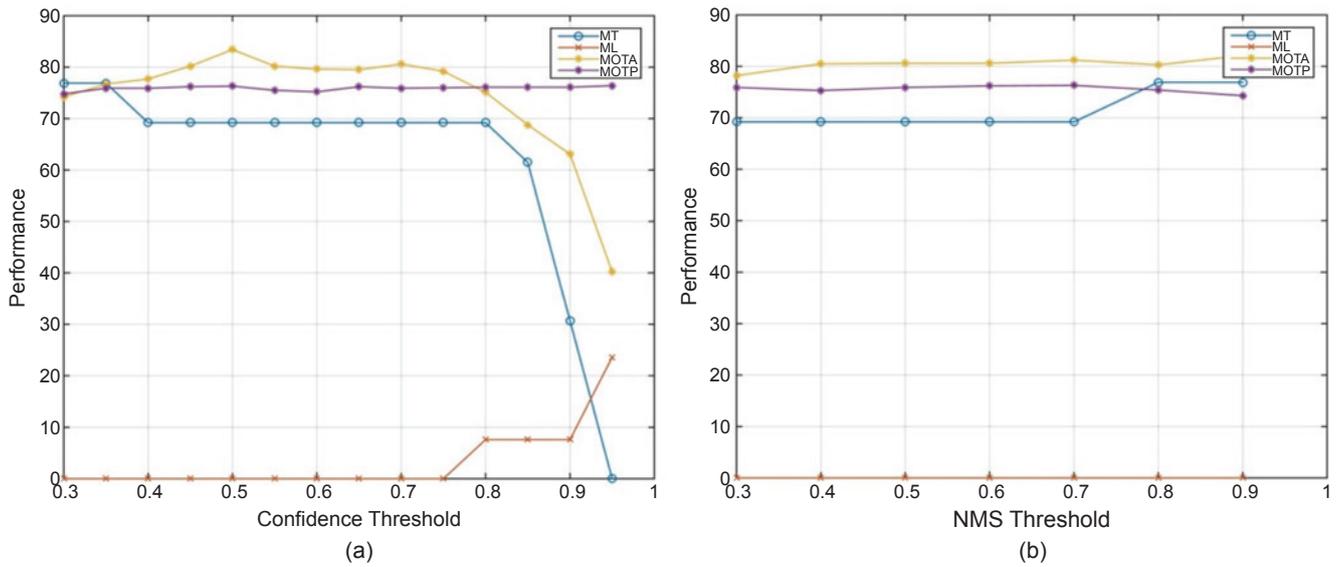


圖 9. 行人偵測參數調整之影響。(a) Confidence Threshold、(b) NMS Threshold。

在經過最佳化參數調整後，接下來將實際觀看我們的方法在測試影片中的效果，此部分實驗使用的測試影片為 PETS09-S2L2 與 TUD-crossing 2 部影片。比較方法的部分，除 MDP 外，還會與五個 MOT benchmark 中排名前 20 名的方法做比較：JointMC⁽¹¹⁾ 為第 17 名、DMT⁽¹²⁾ 為第 11 名、HybridDAT⁽¹³⁾ 為第 20 名、AMIR⁽¹⁴⁾ 為第 15 名、APRCNN⁽¹⁵⁾ 為第 14 名、MDP⁽⁸⁾ 為第 41 名。而偵測器的部分，DMT 使用了 Fast R-CNN 作為其偵測器，而 JointMC、HybridDAT、AMIR、

APRCNN、MDP 皆使用集合通道特徵 (ACF) 作為其行人偵測器。

在 PETS09-S2L2 上的比較結果如圖 11 所示。可以看出我們的方法與其他方法相比有著最好的 FP、MOTA 與 MOTP。在 TUD-crossing 上的結果如圖 12 所示。可以看出我們的方法與其他方法相比有著最好的 FN、ID switch、MOTA 與 MOTP。整體結果如預期的，我們的方法在 2 部實驗影片情境都取得了最好的 MOTA 與 MOTP。

六、結論

總結，本文詳細介紹了在線多物件追蹤系統的技術內容，除了有系統方法的原理概念外，還包含如何調整行人偵測與多物件追蹤之參數，並顯示其後續對追蹤結果之影響。由於做了參數最佳化，行人偵測器可給出更準確的效果及行人位置，也因此減少了後續多物件追蹤的錯誤，其結果對比於其他追蹤方法，明顯有著更好的追蹤準確率，在 MOTA 與 MOTP 二個評測指標上尤其明顯，由此可見行人偵測在多物件追蹤系統上扮演著重要的角色。在未來多物件追蹤系統的發展上，可以最佳的行人偵測器為基礎，進一步改良追蹤演算法，以取得最佳的追蹤效果。

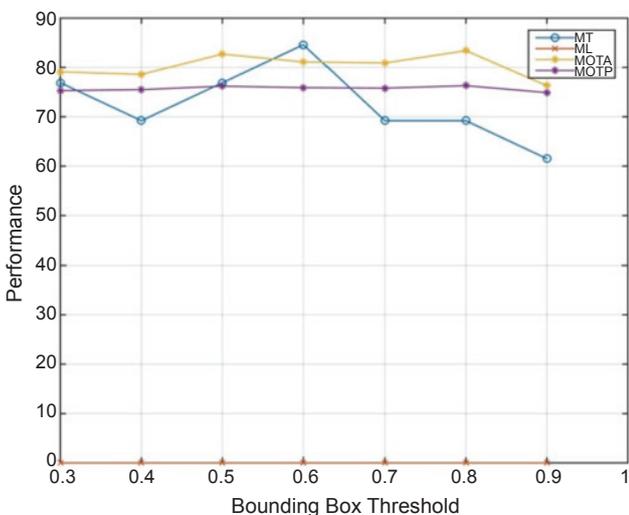


圖 10. 行人追蹤參數-Bounding Box Threshold 之影響。

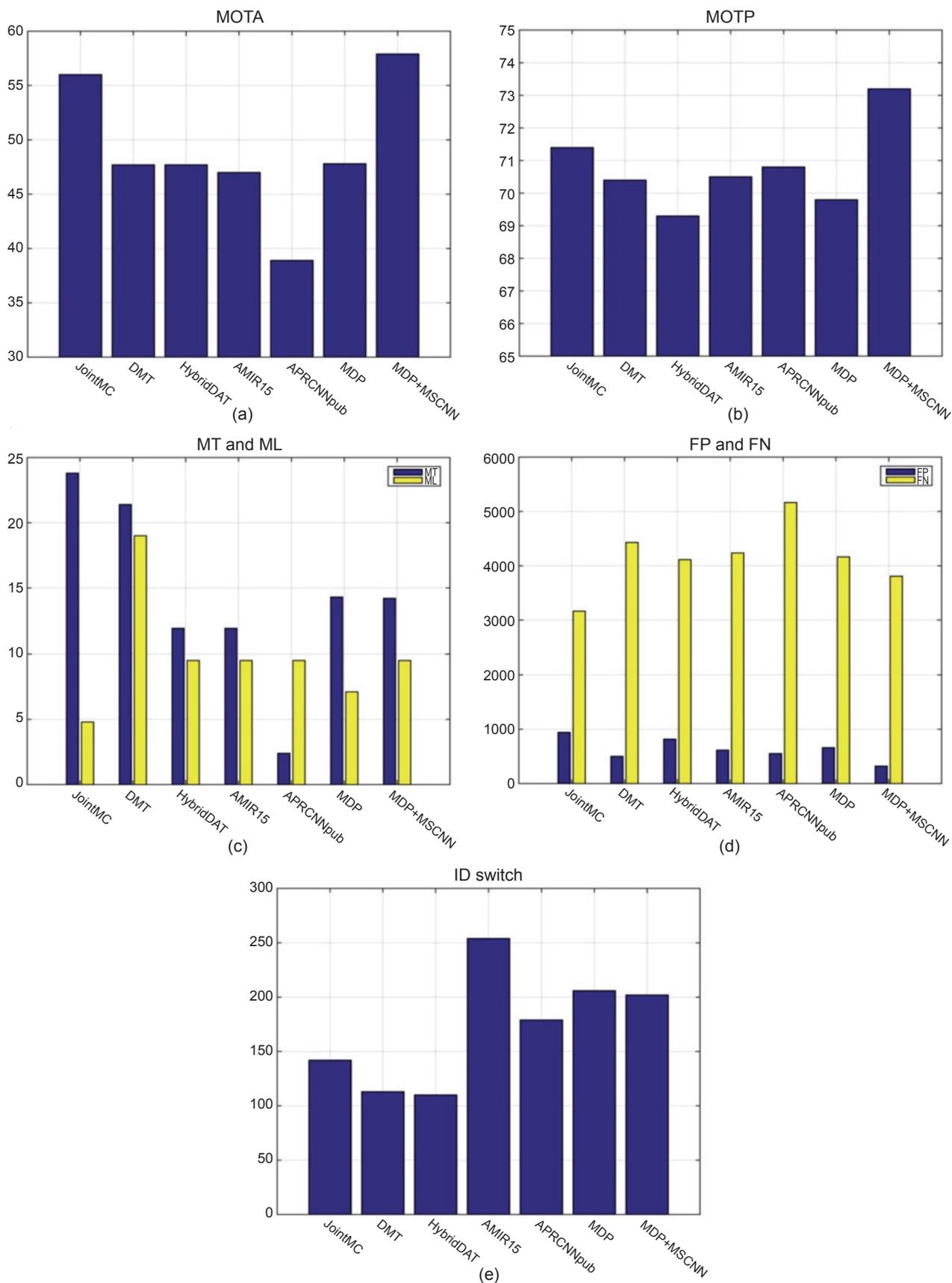


圖 11. PETS09-S2L2 分析圖。(a) MOTA、(b) MOTP、(c) MT and ML、(d) FP and FN、(e) ID switch。

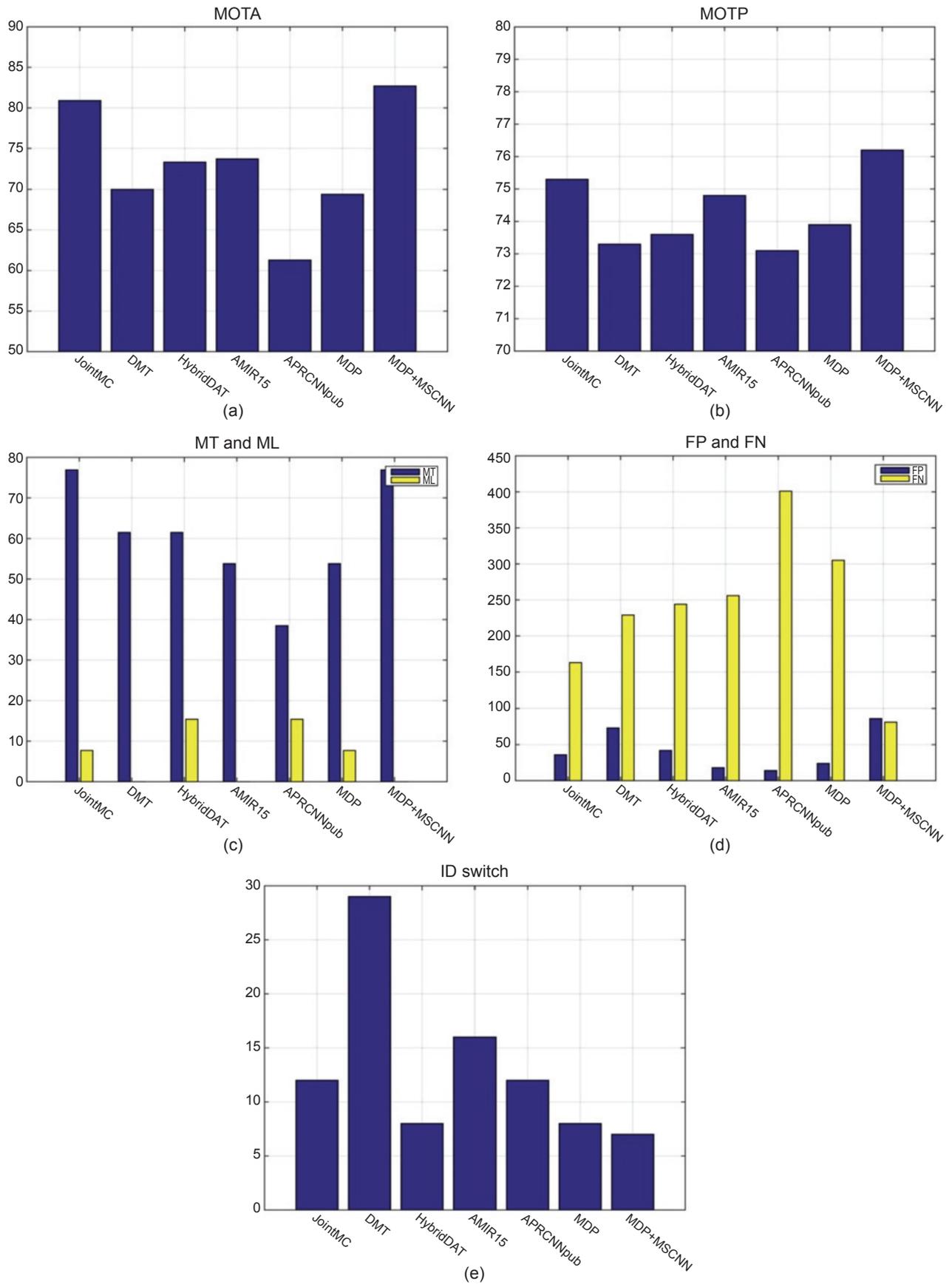


圖 12. TUD-crossing 分析圖。(a) MOTA、(b) MOTP、(c) MT and ML、(d) FP and FN、(e) ID switch。

參考文獻

1. N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005.
2. P. F. Felzenszwalb, R. B. Girshick, D. McAllester and D. Ramanan, "Object Detection with Discriminatively Trained Part-Based Models", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010.
3. R. Girshick, J. Donahue, T. Darrell and J. Malik, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **38** (1), 142 (2016).
4. R. Girshick, "Fast R-CNN", *IEEE International Conference on Computer Vision (ICCV)*, 2015.
5. S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 06 June 2016.
6. Z. Cai, Q. Fan, R. Feris and N. Vasconcelos, "A Unified Multi-scale Deep Convolutional Neural Network for Fast Object Detection", *European Conference on Computer Vision*, 2016.
7. A. Bewley, Z. Ge, L. Ott, F. Ramos and B. Upcroft, "Simple Online and Realtime Tracking", *arXiv:1602.00763*, 2016.
8. Y. Xiang, A. Alahi and S. Savarese, "Learning to Track: Online Multi-object Tracking by Decision Making", *IEEE International Conference on Computer Vision (ICCV)*, 2015.
9. Please refer to the web site: <http://www.cvg.reading.ac.uk/PETS2009/a.html>
10. Please refer to the web site: <https://www.mpi-inf.mpg.de/departments/computer-vision-and-multimodal-computing/research/people-detection-pose-estimation-and-tracking/people-tracking-by-detection-and-people-detection-by-tracking/>
11. M. Keuper, S. Tang, Y. Zhongjie, B. Andres, T. Brox and B. Schiele, "A Multi-cut Formulation for Joint Segmentation and Tracking of Multiple Objects", *arXiv:1607.06317*.
12. H.-U. Kim and C.-S. Kim, "CDT: Cooperative Detection and Tracking for Tracing Multiple Objects in Video Sequences", *European Conference on Computer Vision (ECCV)*, 2016.
13. M. Yang, Y. Wu and Y. Jia, "A Hybrid Data Association Framework for Robust Online Multi-Object Tracking", *IEEE Transactions on Image Processing*, 5667 (2017).
14. A. Sadeghian, A. Alahi and S. Savarese, "Tracking The Untrackable: Learning To Track Multiple Cues with Long-Term Dependencies", *IEEE International Conference on Computer Vision (ICCV)*, 2017.
15. L. Chen, H. Ai, C. Shang, Z. Zhuang and B. Bai, "Online Multi-Object Tracking with Convolutional Neural Networks", *IEEE International Conference on Image Processing (ICIP)*, 2017.



胡啟恩先生現為輔仁大學電機工程研究所碩士班學生。

Chi-En Hu is currently a M.S. student in the Department of Electrical Engineering at Fu Jen Catholic

University.



郭雅諒先生現為輔仁大學電機工程研究所碩士班學生。

Jonathan Guo is currently a M.S. student in the Department of Electrical Engineering at Fu Jen Catholic

University.



王元凱先生為國立中央大學資訊工程系博士，現為輔仁大學電機工程系教授。

Yuan-Kai Wang received his Ph.D. in computer science and information

engineering from National National Central University. He is currently a professor in the Department of Electrical Engineering at Fu Jen Catholic University.