

# 知識天地

## 淺談電腦視覺與影像特徵點比對

林彥宇副研究員（資訊科技創新研究中心）

### 摘要

電腦視覺旨在研究如何讓電腦能有效地「看」這個世界的科學。由於是看的研究，我們讓電腦透過分析視覺資料如影、錄像等，從中獲取場景裡的資訊，並據此作出判斷或決策，旨在建立與人類視覺系統相仿的機器視覺，其可觀察、識別甚至可以解釋所看到的世界。影像比對是電腦視覺領域中的一個核心研究議題，旨在偵測影像間是否有共同物件的存在，具有很高的應用性。在本篇文章中，我們將簡介一個透過影像特徵點進行比對的演算法，目前面臨的挑戰，以及我們近期的研究成果。

### 什麼是電腦視覺？

電腦視覺（computer vision）是電腦科學中的一個重要領域，旨在研究如何讓電腦能有效地「看」這個世界的科學。由於是看的研究，我們讓電腦透過分析與辨識由照相機或攝影機取得的影、錄像資料，從中獲取場景裡的重要資訊，並據此作出判斷或決策，例如我們在使用數位相機拍照對焦時，它會判斷是否有人臉出現在影像中與其位置？甚至是預測該人臉的表情、年齡、性別、以及身份；又例如計算監視器的錄像中有多少行人？是否有人作出危險或異常的行為？作為一個科學領域，電腦視覺研究的理論與技術，旨在建立與人類視覺系統相仿的機器視覺，它可觀察、識別甚至可以理解並解釋所看到的世界。依上述的介紹，我們可以發現電腦視覺是一個跨領域的學科，它與許多研究領域緊密相關，例如能提供影、錄像訊號強化和特徵萃取的影像處理（image processing）領域；能提供統計理論與模型作為資料分析的圖形識別（pattern recognition）領域；探討人腦如何運作與認知的腦科學（brain science）領域。

### 電腦視覺研究現況、挑戰以及應用

電腦視覺目前是一個高度熱門且亟待發展的研究領域，它同時存在許多待開發的技術與應用，以及隨之產生的挑戰。近年來取像設備的蓬勃發展，許多次世代攝影機，如彩色深度攝影機（RGB-D camera）、高速攝影機（high speed camera）、雙鏡頭攝影機（binocular camera）等等，這些攝影機可以取得與傳統彩色攝影機資訊互補的影像，供電腦「看」並作出更準確的決策；另一方面，電腦運算能力大幅提升，讓我們可以發展更強大有效的影像處理和圖形識別技術來看這些影像，近來劃時代的技術當屬深度學習（deep learning），以2012年由Krizhevsky等人[1]所發展的Alex Net為例，它使用一百二十萬筆訓練影像同時對六千萬個模型參數進行最佳化，它在該年的ILSVRC（ImageNet Large Scale Visual Recognition Challenge）影像分類比賽中，取得冠軍，正確率並大幅領先第二名10%，值得一提的是ILSVRC參與者眾且競爭激烈，從未有前二名的正確率差距達10%之多。然而即使有近年來的這些重要進展，在絕大多數電腦視覺應用中，人類視覺系統仍是功能最完善的視覺系統，電腦視覺的研究還有許多困難與挑戰待突破。

西方諺語有云：“A picture is worth a thousand words”，它指出視覺涵蓋資訊之廣泛，以及對決策的重要，據此亦反應出電腦視覺研究在學術界、產業界與醫療界的廣泛應用性，例如電腦視覺技術常用來取代昂貴的人工進行產品瑕疵的自動化檢測；它亦可作為醫學影像的輔助判讀、機器人對週遭環境的認知；而另一個近來很熱門的應用就是自動駕駛車系統的開發，電腦視覺即是核心技術的其中一項。

### 影像比對

電腦視覺因應用性廣泛，因而有許多基礎研究議題，例如人臉偵測、物件辨識、影像分割、影像比對、三維重建、人流計數、動作識別等，這些議題各自均有不少的後續應用，他們雖然基礎，但不意味簡單，因此許多電

腦視覺的研究團隊仍嘗試更妥適地處理這些議題。本篇文章以下將介紹其中一項議題「影像比對」。影像比對旨在偵測影像間是否有共同物件或區域的存在，通常藉由影像之結構、材質、紋理、形狀、顏色等特徵的分析，從兩張或多張影像中，判斷是否有共同物件的出現，及其出現位置。影像比對是電腦視覺與圖形識別領域的核心問題，透過比對我們可以評估影像間的相似性，具有十分廣泛的應用，例如影像內容分析、物件辨識、多媒體資料檢索與推薦、深度估測與物件追蹤。圖一呈現一個影像比對的實例，子圖（a）中為兩張待比對之影像，讀者可以看出這兩張影像中有哪些共同物件或區域嗎？相信並不是十分的容易，子圖（b）顯示此例所包含三個共同物件：玩偶、書和杯麵，這也就是影像比對試圖自動化求得的正確答案。



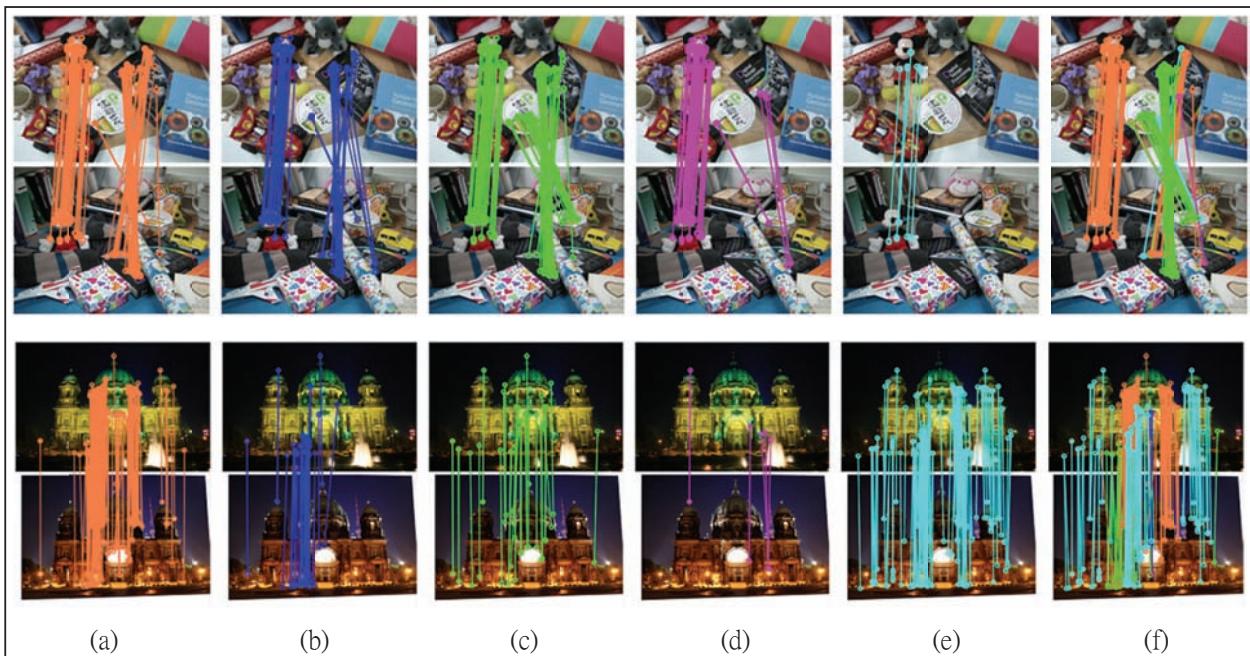
圖一、(a)兩張待比對之影像，圖片引用自[2]。(b)兩張影像相同內容之區域。(c)偵測到的影像特徵點。

(d)比對後的特徵點對應。

### 特徵點比對

目前有若干類實現影像比對的方法，我們介紹其中一個主流的方式「特徵點比對」。顧名思義，它以特徵點比對來進行，比對的單位為所謂的「特徵點」（feature point），其比對演算法可分為三個步驟：1) 特徵點偵測（detection）；2) 特徵點描述（description）；和3) 特徵點對應（matching）。好的特徵點在影像比對指的是具獨特性質，並能重複地出現影像間相同區域位置上的點。根據影像亮度梯度的分佈，我們可將影像點大致分為平坦（flat）、邊（edge）以及角落（corner）三種，一般來說角落最符合特徵點所需要的性質，圖一（c）顯示使用 Hessian Affine detector [3]所偵測出的影像特徵點。偵測出特徵點後，我們對以每個特徵點為中心的區塊(local patch)進行特徵描述，我們將描述特徵的方法稱為描述子（descriptor），不同的描述子描述不同的區塊特徵，目前最廣為使用的描述子名為SIFT [4]。在完成特徵點的描述後，我們即可進行跨影像的特徵點對應（feature correspondence or feature matching），許多現存的對應演算法在進行特徵點對應時，同時考慮到跨影像特徵點的特徵相似性，以及產生對應後，對應彼此間的幾何相似性。圖一（d）為我們所提出的特徵點對應方法，搭配Hessian Affine偵測器以及SIFT描述子後所產生的影像比對結果，從中可以發現，共同物件玩偶與書因有充足的特徵點對而能順利找出，然而杯麵上只有一個對應，很容易在後處理中誤判雜訊而去除，因此仍有改善的空間。

我們觀察到現有特徵點比對演算法大多基於單一描述子的使用，然而描述子的優劣往往取決於影像內容，最佳描述子隨著影像內容改變而改變，有鑑於此，我們提出一個非監督式融合描述子之技術，它可隨著影像區域動態調適選取合適之描述子，圖二中顯示二個特徵點比對的實例，上、下列各一，其中子圖（a）~（e）為分別使用五種不同之描述子所產生的特徵點對應結果，其中（a）~（c）之描述子著重於材質紋理的特徵萃取；（d）之描述子著重於色彩分佈的特徵萃取；（e）之描述子著重於形狀的特徵萃取，我們所提出之方法所產生的特徵點對應結果呈現於子圖（f），於圖中可看出提出方法可利用不同描述子間的互補性，產生更多與更正確的特徵點對應，增強影像比對的效能。



圖二、兩個（上、下）特徵點比對的實例。（a）～（e）分別使用五種不同之描述子所產生的特徵點對應結果。  
（f）所提出之方法所產生的特徵點對應結果。

## 參考文獻

- [1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems*, 2012.
- [2] M. Cho, Y.-M. Shin, and K.-M. Lee, “Co-recognition of image pairs by data-driven Monte Carlo image exploration,” in *Proc. Euro. Conf. Computer Vision*, 2008.
- [3] K. Mikolajczyk and C. Schmid, “Scale and affine invariant interest point detectors,” *Int. J. Computer Vision*, vol. 60, no. 1, pp. 63–86, 2004.
- [4] D. Lowe, “Distinctive image features from scale-invariant keypoints,” *Int. J. Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.