

影像視訊科技未來展望與願景

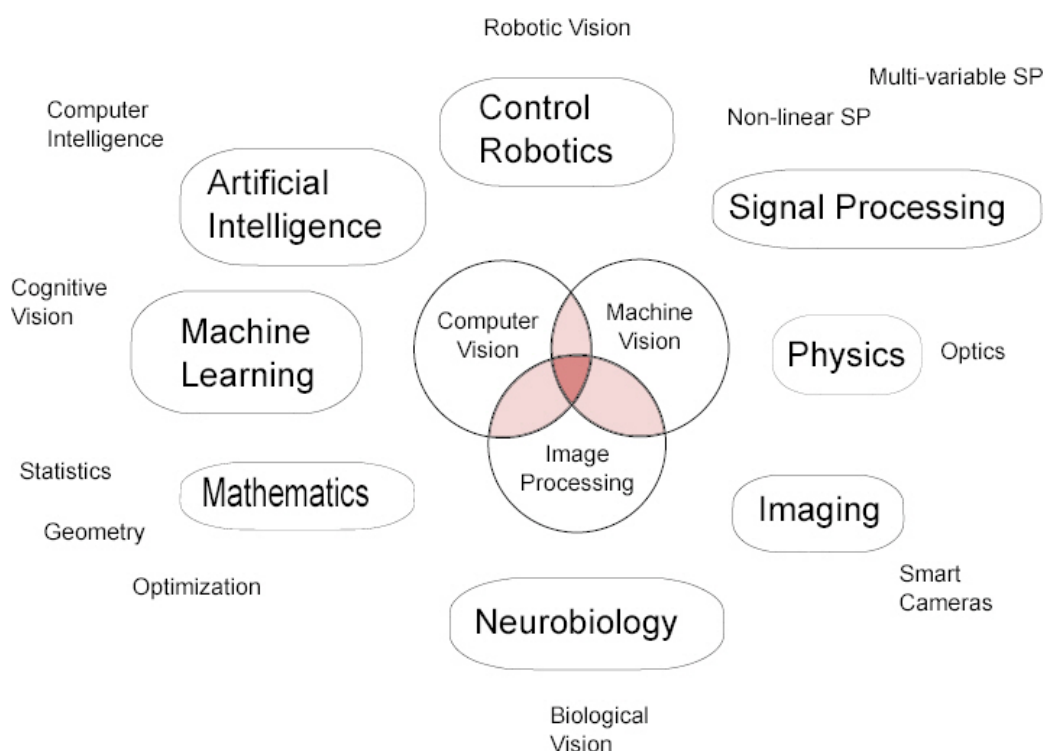
王駿發

國立成功大學電機工程學系

E-mail: wangjf@mail.ncku.edu.tw

一、影像視訊科技的背景

電腦視覺技術自從 1970 年代[1]蓬勃發展以來，就被廣泛的應用於人類生活上，不論是在工業界我們所看到的機器人手臂，亦或到現在家庭之中人手一台的行動電話，都可以看到電腦視覺技術的影子。電腦視覺技術所包含的科學知識相當廣泛[1]：從物理學、訊號處理、自動控制、人工智慧、機器學習、統計學、幾何學，甚至到人類感知領域的神經生物學、樣形識別等，是一門橫跨多領域的研究學科。



電腦視覺技術所涵蓋的學科範圍[1]

影像視訊系統大致上可分成影像擷取、儲存、處理、傳送通訊、顯示等步驟，近年來，拜硬體運算速度進步之賜，以往早期原本只能從單純的二維平面來分析影像，諸如：幾何變換、放大縮小、旋轉、色彩矯正、影像分割、邊緣偵測等，慢慢地進步到三維立體空間的處理，像是場景重建等技術；至今已能應付即時複雜的人機互動，可以運用在物體的追蹤、目標識別、監測感興趣的事件上面；這也意味著影像視訊處理技術日趨重要。

二、影像視訊科技的相關應用

在多媒體發達的今日，影像視訊處理技術的應用不勝枚舉：像是在醫療方面用以輔助醫師診斷的影像強化與影像還原技術，或者是應用於安全監控系統上的人臉辨識與車牌辨識功能等等，都是常見的範疇；此外，隨著近來高畫質的影像需求激增以及社群網路的發展，對影像/影片進行壓縮編碼以減少傳輸與儲存的資料量和多媒體資料檢索，也是目前極力研究的題材。影像視訊處理應用範疇從個人、家庭，擴大到學校、醫院、企業等社會層面，甚至是國家軍事方面上。

美國國防部先進研究計畫署(The Defense Advanced Research Projects Agency, DARPA)曾經在 2004 舉辦第一屆的無人自動駕駛自動車「Grand Challenge」競賽 [2]，比賽的地點與目標，是在美國一處充滿崎嶇彎道的沙漠中，於有限時間內且不偏離賽道，完成總長 240 公里的路程抵達終點，可惜的是當年並沒有團隊完成任務。隔年，2005 年，儘管有五隊完成任務，首先抵達終點的車輛則是由史丹佛大學(Stanford University)的電腦視覺團隊花了將近 7 個小時成功完成賽道，抱走 200 萬美元的冠軍獎金。DARPA 於 2007 年又將競賽的困難度，從原本的沙漠賽道提升至城市的街道，賽道總長 96 公里，位於一處廢棄的空軍基地中，參賽者須在 6 小時內完成任務抵達終點；比賽途中團隊除了要遵守真實的交通規則外，還要穿越各種路障，應付大大小小的車流量，實為一項艱鉅的挑戰。此屆的冠軍卡內基美隆大學(Carnegie Mellon University)的團隊只花了約 4 小時 10 分鐘的時間抵達終點，將電腦視覺的技術又往更高一層推進。

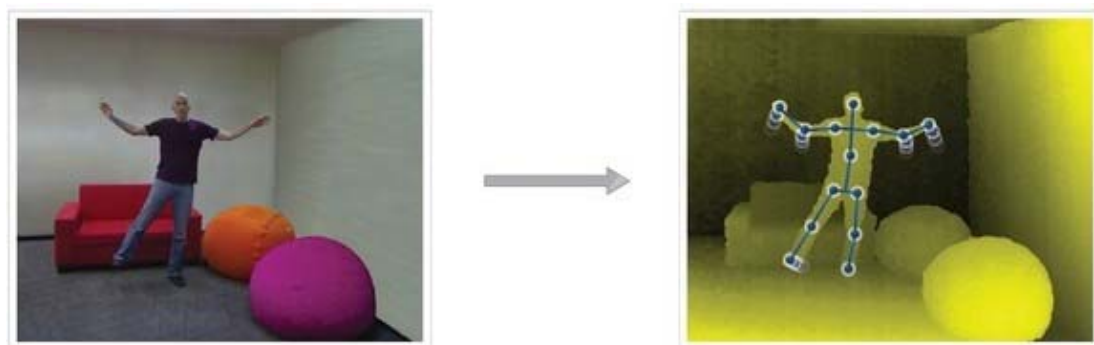
另一個則是應用在生活娛樂上面的例子，2010 年 6 月微軟公司推出 Xbox 360 遊戲機所附屬的體感裝置 Kinect[3]，它設有 3D RGB 鏡頭、紅外線距離感應器、麥克風陣列和客製化軟體發展套件等，不同於 Wii 遊戲機，使用者不需要額外的控制器便可以利用手勢、動作來對遊戲機下達指令。有趣的是，裝置所內建的影像處理系統有辦法同時偵測並追蹤到四個使用者的動作，並且適合於各種光源下應用[3][4]。



2005 年 DARPA Grand Challenge 冠軍團隊所使用無人自動駕駛自動車[2]



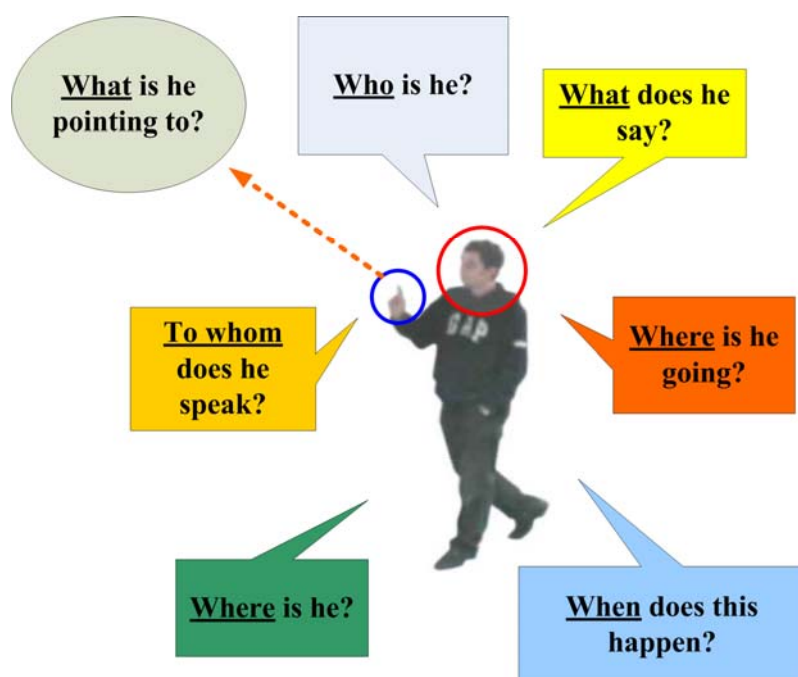
微軟 Xbox 360 遊戲機所推出的體感裝置 Kinect[3]



Kinect 可偵測到人體 48 處骨骼關節點[4]

三、影像視訊科技的未來趨勢

雖然影像視訊科技發展至今已有多餘年，目前現階段對於影像意涵了解 (Content-Based Video Understanding) 的突破仍十分有限，以下圖來說，若是在無任何先前知識 (Prior Knowledge) 下，任意給電腦一段視訊，要像人類一般成功辨識出影像中內容，似乎還有很遠的距離。



影像意涵了解

因此，未來的發展仍朝向理解影片中 5W (Who、What、Where、When、Why) 邁進，亦即：

- 影像中的人物手勢為何?指的方向為何?所指又為何物?(Where is he/she pointing to?)—任意角度下手勢辨識、姿勢辨識技術。
- 影像中的人是誰?(Who is he/she?)—任意角度下人臉辨識、身分辨識。
- 影像中的人在做什麼?(What is he/she doing?)—行為辨識。
- 影像中的人要去哪裡?(Where is he/she going?)—物體追蹤。
- 影像中的人說了甚麼?(What does he say?)—免持麥克風語音辨識、語意分析。
- 影像中的人正在跟誰說話?(To whom does he/she speak?)—聲源定位、身分辨識。
- 影像中的人身處何地?(Where is he/she?)—場景辨識。

影像視訊處理不再單純只是依賴單一影像輸入資料，許多研究已結合多重異質感應接受器(Heterogeneous Sensor)，諸如：多維影像攝影機、麥克風陣列、溫度感測、壓力感測器等，可以同時接收多型態(Multimodal)的資訊。如此一來，便可結合多方面資訊，建立強健性(Robustness)的軟硬體應用。

2010 年的今天，影像視訊處理已走向數位化、網路化、無所不在(Ubiquitous)化，如何提出適合的方法並採用智慧型化影像視訊處理，以應付未來的需求，便是一項值得研究的問題。下面內容中，我們列出了三項未來可能會面臨的威脅與挑戰，以及它們與影像視訊處理的討論。

身心靈健康照護：隨著高齡化社會的來臨，老年人口不斷的增加，社會福利政策勢必逐漸向銀髮族傾斜。為了提高他們的身心靈健康，如何將智慧型影像視訊處理與便利的人機介面融入人們的生活裡還需要相當大的努力；在生理照護方面，諸如建立醫療照護網，使得遠距醫療照護、無線行動醫療照護得以延伸至社區，並提供社區醫院、診所方便醫護人員與老年人進行會診、視訊會議、病歷影像傳輸的功能，亦是研究的課題之一；再者，由於老年人行動不便，為了提供老年人心靈的娛樂，智慧型影像視訊處理還可融入社群網路分析功能，方便老年人

與朋友、兒女聯絡，以豐富他們的精神生活；此外，機器人視覺結合機器人輔具的研究，對於改善行動不便或特殊疾病患者的起居生活，亦有很大的發展性。

安全監視·社會治安：自從 911 恐怖攻擊以來，美國政府已在各公共場所大量部屬相當多的監視攝影機，用以打擊並遏阻犯罪。世界各國亦紛紛投入許多經費加強設置攝影裝備。雖然這些攝影監視畫面有助於案件的偵查，但是卻需要相當大的人力資源逐一去過濾錄影畫面找尋相關線索，實為一項費時又費力的工作。再加上錄影環境的問題，例如解析度不佳、光影影響、雜訊干擾等，使得辦案困難重重。智慧型化影像視訊處理在此方面可以著眼的地方在於：(1)如何找出冗餘且過長的錄影畫面，以方便辦案人員迅速找出感興趣的畫面；(2)重組多張模糊、遠距離影像使得人臉部或車牌特徵得以強化；(3)對影像中的物件進行自動註解，以方便辦案人員對於線索的搜尋。

上述這些問題僅是影像視訊科技在安全監視的冰山一角，未來若能朝向影片理解邁進且開發出自動分析方法，想必對偵查速率的改善與犯罪的遏止有很大的潛力。

防災·救災：由於全球暖化，全球各地平均氣溫逐漸上升、海洋水氣蒸發旺盛，導致氣候異常，使得各地頻傳風災、水災、土石流等災害。指揮中心為了可充分掌握救災狀況與進度，可以利用單點或多點視訊會議，將救災、急難救助等災害現場的畫面即時傳輸回指揮中心，方便人員分析、迅速調派資源；未來，還可將此應用在救護車上，於必要時即時提供病患支援。在災害的預防方面，還可運用地理圖形資訊搭配遙測影像技術，即時監測異常影像狀況，提早偵測出不穩定的山坡地、或是河川，以便達到預警的效果。此外，智慧火災影像偵測、早期火災預警、淹水偵測、防盜保全等系統，也是智慧型化影像視訊處理發展的目標。

四、結語

影像視訊處理科技已與人們的日常生活密切的結合，並且改善了我們的生活

品質，可惜的是目前它們離智慧型處理還有一段距離，因此未來結合人性與科技的智慧型化影像視訊處理勢必是大家一同努力的目標。

參考資料

- [1] http://en.wikipedia.org/wiki/Computer_vision
- [2] http://en.wikipedia.org/wiki/DARPA_Grand_Challenge
- [3] <http://en.wikipedia.org/wiki/Kinect>
- [4] <http://www.primesense.com/>