

行政院國家科學委員會專題研究計畫 期中進度報告

數位家庭應用中內容分析與展示技術之研發(2/3) 期中進度報告(精簡版)

計畫類別：個別型
計畫編號：NSC 95-2221-E-002-332-
執行期間：95年08月01日至96年10月31日
執行單位：國立臺灣大學資訊網路與多媒體研究所

計畫主持人：吳家麟

報告附件：出席國際會議研究心得報告及發表論文

處理方式：本計畫可公開查詢

中 華 民 國 96 年 12 月 12 日

數位家庭應用中內容分析與展示技術之研發(II)

國科會補助專題研究計畫

計畫編號：94-2213-E-002-078

全程計畫：民國 94 年 08 月 01 日至民國 97 年 07 月 31 日

本年度計畫：民國 95 年 08 月 01 日至民國 96 年 07 月 31 日

計畫主持人：吳家麟 台灣大學資訊網路與多媒體研究所教授

一、摘要

(中文摘要)

本計畫以提昇數位家庭應用為目標，共分成兩大方向：(1)影片事件偵測(video event detection)與(2)複合式影音展示(composite audiovisual presentation)。分別對數位家庭中媒體接收分析與整合展示兩方面著手，適切地符合使用者對於新一代數位應用的需求。

【影片事件偵測】技術之主要目的是利用專有知識(domain knowledge)與製作原則(production rules)來輔助數位內容的分析。有別於以往的研究只能對數位資料做大致的分類或切割，本計畫希望能確切地偵測影片內事件的種類與發生的時間。本技術將從針對棒球影片開始，建立完整的分析架構。利用以規則為基礎(rule-based decision)與以模型為基礎(model-based decision)的判斷方式來實作出各種事件的偵測，並進行自動產生比賽過程的記錄與摘要。

【複合式影音展示】主要希望整合多個媒體共同展示，為不同的個別媒體增加附加價值。本技術從照片特性與音樂節奏的分析開始，建立一套架構自動產生以音樂為導向的幻燈秀。其中在照片展示方面將結合圖片分類、使用者注意模型、區域重要性模型、自動編輯與多媒體同步等功能，務求最後的展示結果能結合各照片的關連與音樂特性，以期有別於傳統賞圖軟體只能依序一張張的幻燈秀。

(英文摘要)

This project aims to enhance the digital applications at home. It focuses on two phases: (1) video event detection and (2) composite audiovisual presentation. They respectively tackle with content analysis and multimedia integration, and therefore match the needs of next-generation digital home applications.

The goal of “video event detection” is mainly on how to exploit domain knowledge and production rules to facilitate digital content analysis. Unlike conventional researches, which roughly classify or segment digital content, this project tries to exactly detect what happened and the timestamps of events. We will start with analyzing baseball videos and construct a complete framework. With rule-based decision and model-based decision methods, we implement exactly event detection and further develop the techniques of automatic scoreboard generation and game summarization.

“Composite audiovisual presentation” integrates multiple media and adds the values to isolated media. It analyzes photos characteristics and music tempo, and then builds a framework to automatically generate music-driven slide shows. We integrate the techniques of image classification, user attention model, region-based importance model, automatic

editing, and multimedia synchronization. We will present the results with correlations between photos and music characteristics, instead of sequentially showing photos one-by-one.

二、計畫緣由與目的

生活數位化的時代已隨著各種數位內容產品的普及與數位電視的全面推廣而快速來臨。在面對大量數位資訊與各種媒體來源時，許多問題因為實際的應用與使用者的需求而逐漸衍生出來。為了讓大量的多媒體資料能有效率地被運用，進而提昇數位內容的應用價值，有效的多媒體檢索與數位內容分析是急需被發展的技術。以商業附加價值大的運動影片為例，若能應用數位內容分析技術將比賽中的事件偵測出來，並自動剪輯對應的片段或進行影片摘要，將可提供使用者更多元的觀賞模式。

目前數位內容分析的研究主要面臨的課題在於如何跨越語意鴻溝(semantic gap)。以往對於數位資料的分析只著重於特徵值萃取(feature extraction)，然而，低階特徵值的特性往往跟人的感覺不同。常用的特徵值包括圖片中顏色統計(color histogram)、顏色分佈(color layout)、邊緣統計(edge histogram)等；影片中的移動(motion)、鏡頭種類(camera type)等；聲音中的過零率(zero-crossing rate)、能量(energy)、音高(pitch)、節奏(tempo)等。由於這些特徵值不直接具有人為解釋的意義，因此由這些特徵值描述而得的結果常常不如預期。此種低階特徵值與高階語意(semantics)之間的落差稱為語意鴻溝(semantic gap)。

再以運動影片為例，目前的研究成果主要在於利用顏色與邊緣資訊進行畫面分類(shot classification)(如棒球或網球影片)，或者是利用球的軌跡(ball trajectory)與移動(motion)進行比賽持續或暫停(play-break)的分析(如足球影片)。然而，這些初步的成果與使用者真正需要的相去甚遠。以球迷的觀點來說，他們關心的是場上球員的表現：是否有安打？是否有全壘打？投手今天有幾次三振？此場比賽的射門與得分畫面等等。目前的研究成果只能大致找出比賽中“可能比較精采的片

段”，而不是“比賽中確切發生了什麼事情”。許多受歡迎且有意義的應用雖然已經開始出現(如美國大聯盟網站上的線上影片)，它們仍需依靠人為的剪輯，需要大量的時間與人力。因此，在數位媒體大量被創造的同時，發展一套能有效偵測影片裡事件的技術已經是刻不容緩。

此外，從數位電子產品的普及來看，一般的使用者越來越有能力錄製或創造自己的數位內容。這些數位資料雖然可以很輕易地被創造與儲存，但由於一般使用者常疏於管理或因為拍攝品質不佳，降低了這些數位資料的價值與實用性。在這項課題當中，以往的研究主要提供使用者一個大致的分類結果，例如，以場景的相似性將類似的照片歸為同一類，或以相機或攝影機附帶的時間資訊做時間上的分類。這樣的做法在近年來也開始往內容加值的方向進行。例如，與其觀賞一段長達一兩個小時的家庭影片，不如將其重要的片段配上音樂，剪輯成一段五六分鐘的影音摘要。同理，數位相機拍得的大量照片也可在偵測畫面中物件的重要性與場景分類之後，配上適當的音樂製作成動態的幻燈片展示(slide show)。此種摘要式的後製資料更能被人們接受，其相對的觀賞價值也跟著提高。

為提昇數位家庭應用，本計劃將目標集中在兩大方向：影片事件偵測(event detection)與複合式影音展示(composite audiovisual presentation)。

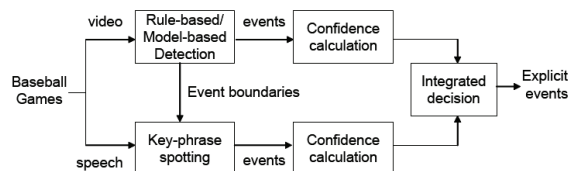
三、研究方法與成果

本計畫預計以三年時間完成，第一年著重低階特徵值萃取與以規則為基礎的事件偵測。第二年進一步發展以模型為基礎的事件判斷模組與圖片、音樂分析模組。第三年整合多項模組完成影片摘要與幻燈秀的實作。目前本計畫已完成第二年之階段性目標，以下簡介本年成果。

1. 比賽規則與慣例於以模型為基礎(model-based)的事件判斷模組

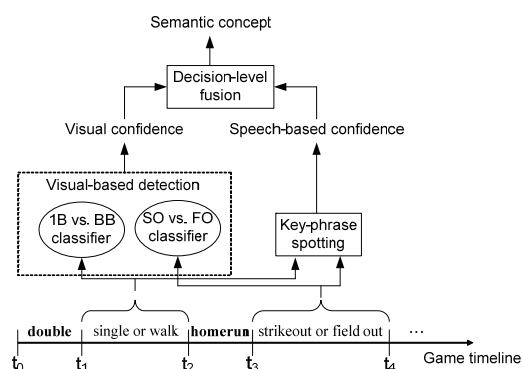
在第一年中利用比賽規則與慣例的方法雖然可偵測球賽中的大部份事件。但有些事件光靠字幕資訊與規則仍然無法明確區隔，例如一壘安打與四壞球

保送；三振與內外野出局等。因此，我們現階段再進一步地加入畫面切換與語音的資訊來區分這些混淆的狀況。



圖一、視覺跟語音資訊的偵測整合

如圖一所示，我們首先分別從視覺跟語音資訊做偵測，之後再結合這兩方面的資訊以得到一個整合偵測的結果[1]。如圖二所示，在時間 t_1 到 t_2 之間，經由字幕變化與規則判斷，我們僅能大略知道這段期間出現一壘安打或四壞球。在這種情況下，此區段的資料將分別進行以視訊為基礎的偵測 (Visual-based detection) 與以語音為基礎的偵測 (Speech-based detection)。以視訊為基礎的偵測模組以“是否有投手畫面緊接內外野畫面”、“在打擊出去前投手總共投了幾球”、以及“攝影機移動程度 (motion magnitude)”等為特徵值，利用 K-nearest modeling 來描述一壘安打或四壞球的特性。以語音為基礎的偵測模組利用“關鍵字詞偵測(key-phrase spotting)”[2]來評估某個概念發生的機率。



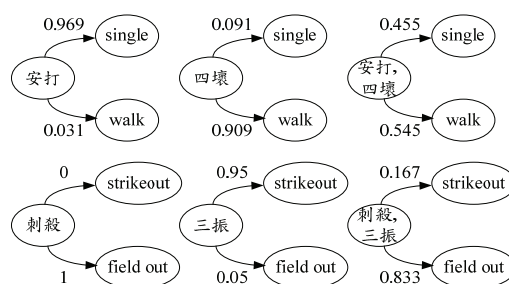
圖二、結合語音關鍵字句的結果來進行混淆概念的區分

由於特定概念的發生將伴隨著特定的主播語音說明，如表一所示，因此我們可根據資料庫中的

比賽來評估各個關鍵字詞出現時某特定概念出現的機率。據本研究所統計的結果，關鍵字詞與相關概念的機率關係如同圖三所示。另外，三振與內外野出局的區分狀況亦同。

Concepts	Corresponding Key-phrases
Single (C_1)	$R_1 = \{\text{安打(hit), 一壘安打(single)}\}$
Walk (C_2)	$R_2 = \{\text{觸身球(hit by pitch), 保送(walk), 四壞球(four balls)}\}$
Strikeout (C_3)	$R_3 = \{\text{三振(strikeout), 三振出局(strikeout)}\}$
Field out (C_4)	$R_4 = \{\text{刺殺('touch out' or 'out before reaching bases'), 接殺(catch out)}\}$

表一、常見的關鍵字詞與相對應的概念



圖三、關鍵字詞與相關概念的機率關係

在分別進行完兩種不同的偵測之後，我們應用結合不同分類器的概念將兩者的意見融合起來。我們可選擇利用以下規則來整合多個分類器的結果。相加規則(sum rule)：

$$\text{assign } Z \rightarrow C_j \text{ if } \sum_{i=1}^2 P(C_j | \mathbf{x}_i) = \max_{k=1}^2 \sum_{i=1}^2 P(C_k | \mathbf{x}_i)$$

相乘規則(product rule)：

$$\text{assign } Z \rightarrow C_j \text{ if } \prod_{i=1}^2 P(C_j | \mathbf{x}_i) = \max_{k=1}^2 \prod_{i=1}^2 P(C_k | \mathbf{x}_i)$$

取大值規則(max rule)：

$$\text{assign } Z \rightarrow C_j \text{ if } \max_{i=1}^2 P(C_j | \mathbf{x}_i) = \max_{k=1}^2 \max_{i=1}^2 P(C_k | \mathbf{x}_i)$$

取小值規則(min rule)：

$$\text{assign } Z \rightarrow C_j \text{ if } \min_{i=1}^2 P(C_j | \mathbf{x}_i) = \max_{k=1}^2 \min_{i=1}^2 P(C_k | \mathbf{x}_i)$$

經過這些步驟之後，本研究所發展的系統可以將棒球概念準確地偵測出來。

此研究的成果可精確偵測十三種不同的棒球概念而非僅只於大約找出比賽中的精采片段。這十三

種棒球概念分別是：一壘安打(single, 1B)、二壘安打(double, 2B)、三壘安打(triple, 3B)、全壘打(home run, HR)、盜壘(stolen base, SB)、盜壘阻殺(caught stealing, CS)、內外野飛球接殺(fly out, AO)、三振(strikeout, SO)、四壞球(base on ball, Walk, BB)、犧牲觸擊(sacrifice bunt, SAC)、犧牲高飛(sacrifice fly, SF)、雙殺(double play, DP)、以及三殺(triple play, TP)。表二列出五大類最常出現於棒球比賽中的概念，以 CPBL(Chinese Professional Baseball Leagues)[3]與 MLB(Major League Baseball)的比賽為例，其偵測結果具有非常高的準確度。

Game		Hit/BB	Double	Home Run	Out	Sacrifice	Double Play
CPBL1	Prc.	1 (15/15)	1 (6/6)	1 (2/2)	1 (35/35)	1 (5/5)	1 (3/3)
	Rec.	1 (15/15)	1 (6/6)	1 (2/2)	0.95 (35/37)	1 (5/5)	1 (3/3)
CPBL2	Prc.	1 (15/15)	1 (3/3)	1 (2/2)	1 (34/34)	1 (4/4)	0.75 (3/4)
	Rec.	0.83 (15/18)	1 (3/3)	1 (2/2)	0.89 (34/38)	1 (4/4)	1 (3/3)
CPBL3	Prc.	1 (17/17)	1 (3/3)		0.98 (43/44)	1 (2/2)	1 (2/2)
	Rec.	0.89 (17/19)	1 (3/3)		0.91 (43/47)	1 (2/2)	1 (2/2)
MLB	Prc.	1 (18/18)	1 (3/3)		1 (25/25)	0.67 (4/6)	0.75 (3/4)
	Rec.	0.95 (18/19)	1 (3/3)		0.81 (25/31)	1 (4/4)	0.75 (3/4)
Total	Prc.	1 (65/65)	1 (15/15)	1 (4/4)	0.99 (137/138)	0.88 (15/17)	0.85 (11/13)
	Rec.	0.92 (65/71)	1 (15/15)	1 (4/4)	0.90 (137/153)	1 (15/15)	0.92 (11/12)

表二、四場不同比賽的概念偵測結果，以 precision 與 recall 表示

另外，對於混淆概念的區分(discrimination)，由表三可以看到，在大多數的情況下，同時整合視訊與語音偵測的結果會有最好的區分效能。在此圖中我們以 F1 值來表示效能，它代表的是同時考慮 precision 與 recall 的一種效能指標：

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Games	Decision	Precision / Recall	F1
Lions vs. Bears	Visual	0.88 / 0.82	0.85
	Visual + speech	0.96 / 0.89	0.92
Bulls vs. Lions	Visual	0.76 / 0.68	0.70
	Visual + speech	0.85 / 0.74	0.79
Lions vs. Whales	Visual	0.77 / 0.73	0.75
	Visual + speech	0.93 / 0.88	0.90

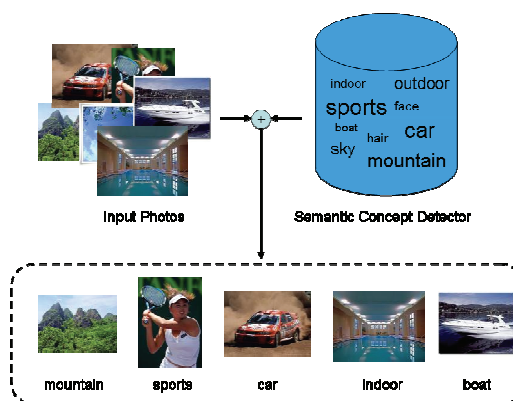
表三、概念區分的整體效能分析

2. 實作自動圖片分類模組

在人手一台相機的時代，照片和我們的生活緊緊相連，但大量的照片不僅整理不易且無從瀏覽，因

此，自動圖片分析技術也日漸重要。

我們根據第一年所作的圖片叢集(photo clustering)的分析，實作完成圖片自動分類的模組，我們利用第一年所提的方法萃取出圖片的特徵值[4]，除了利用時間資訊來作使用者的圖片叢集以及利用內容資訊的圖片叢集外，我們另外發展一種分類模組可供使用者選擇，此模組為採用將網路上收集到的大量照片配合機器學習的方法，如 SVM、HMM 等，得到圖片的語義概念集(Semantic Concept Set)。最後，使用訓練好的分類器將使用者照片依序分類，實作的圖片分類模組的流程如圖四所示。



圖四、圖片分類流程

圖片分類完成後，使用者即可使用關鍵字查詢或瀏覽想要的照片，這種分類方式幫助使用者更有效地管理個人照片。同時分類完的照片主題相關性更強，更適合複合式影音的呈現。

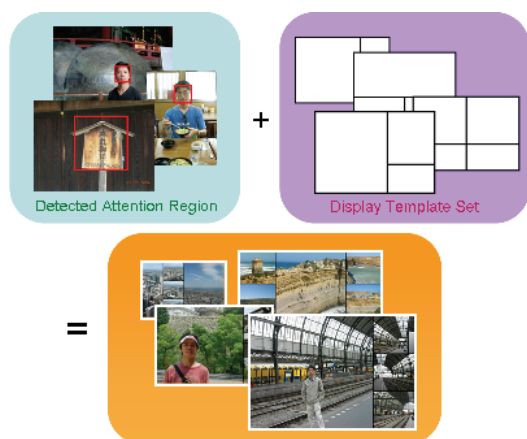
3. 使用者注意模型與音樂節奏分析

3.1 實作使用者注意模型分析模組

目前影像處理中最待解決的問題，即是同一張影像在不同裝置上的呈現，如個人電腦及手機等。直接的影像大小的縮放，經常會導致照片中的人物或景物變形失真，我們利用第一年中所分析的使用者注意特徵值[5]，實作出可以有效決定照片中使用者感興趣的視覺中心的模組，並在指定的區域大小(例如手機的螢幕比例等)為限制下，我們可進一步得到適合指定區域大小的照片區域。同時，我們的方法可確保照片中的人事物保持正常的比例，使用

者可以在指定區域中(例如手機螢幕)，清楚地瀏覽自己的照片。

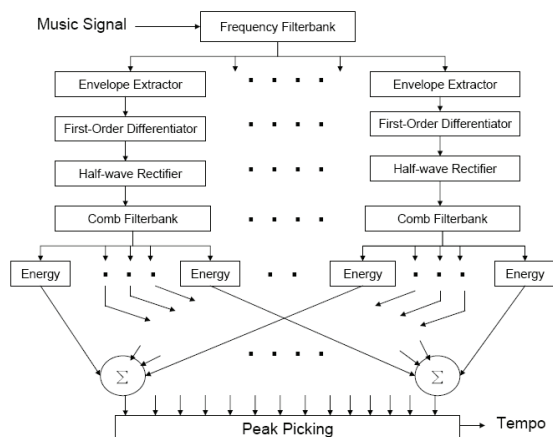
利用相同的技術，我們未來還可以將使用者的照片，配合預先設計好的拼貼範本，產生創意十足的照片拼貼，結果如圖五所示。



圖五、使用者注意模型分析與應用

3.2 實作音樂節奏分析模組

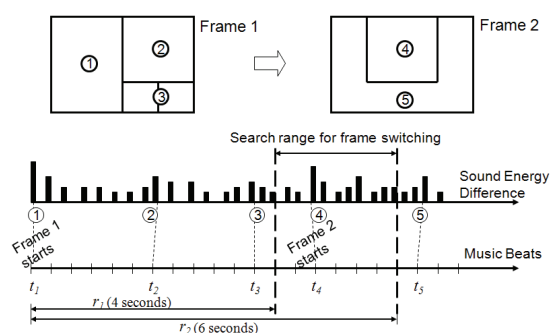
基於第一年在音樂特徵值萃取的研究方法，我們實作出音樂節奏分析的模組，其演算法如圖六所示。首先我們將音樂訊號分成六個不同的頻帶(frequency band)來分析，為了避免在聲音上突然的變化所造成的干擾，這些不同頻帶中的信號會進一步轉變為主要能呈現出信號趨勢的包絡型式(envelope form)。緊接著對於不同頻帶的包絡信號，利用一階微分器以取得聲音振幅(amplitude)的最大改變量的位置，並透過半波整流(half-wave rectified)以作為進一步的週期性脈波分析。



圖六、音樂節奏偵測演算法

與傳統的方法不同，我們採用梳狀濾波器(comb filter)作進一步的分析[6]，如果梳狀濾波器的週期與輸入訊號相同，則會有比週期不相同的輸入信號還大的能量輸出值。接著將對於不同的頻帶中，相同週期的能量輸出加總後，有最大能量加總輸出值的週期，即為音樂訊號的節奏資訊。因此將輸入的音樂信號作節奏的偵測與分析之後，我們便可以利用音樂的節奏資訊來輔助取得整首音樂的節拍位置。

未來利用音樂的節拍資訊便可進一步作為拼貼畫面切換的依據。由於音樂長度有限，為了能讓音樂和照片能更緊密的結合，我們將以音樂為主並將音樂分成幾個段落，而後根據音樂的分段數來選擇合適的照片來作展示播放的搭配。同時，我們也將考慮音樂節拍的強弱，選擇強拍發生的時機來做畫面的切換，藉由此方式加強複合式影音的瀏覽效果，照片和音樂節拍搭配的流程如圖七所示。



圖七、音樂節奏分析與應用

四、結論與討論

本計畫目前的成果主要分兩部份。第一部份是以棒球影片的事件分析為目的，已發展的模組除了第一年的畫面分類、字幕資訊辨識以及基於棒球規則的事件偵測模組外，今年更進一步地實作出以比賽規則與慣例於以模型為基礎的事件判斷模組，完成準確的事件偵測。第二部份是實作出將來可用於複合式展現的基礎模組，包括自動圖片分類模組、使用者注意模型與音樂節奏分析模組。棒球比賽中的事件是球迷了解比賽過程與球員表現的重要依據。不同事件所造成的效果對於球員與觀眾有不同的意

義。因此，有別於以往只大約偵測精采畫面或場景分類的研究，本系統已能確切地偵測棒球比賽中發生的各種事件。

為了使視覺畫面與聽覺刺激有明顯的調和，我們必須分別對視訊(旅遊照片)與音訊(音樂)進行分析。有了自動圖片分類的模組之後，我們才能將相近的照片同時展現，藉以提昇整個視覺的強度與資訊的廣度。有了音樂節奏分析模組，我們才能配合音樂的節拍作圖片的展現，藉以同時滿足視覺與聽覺上的感官享受。本計畫第三年將進一步利用圖片分類模組、使用者注意模型和音樂節奏分析模組，整合視訊與音訊以完成自動幻燈秀的實作。

五、相關論文之發表情況

1. W.-H. Cheng, C.-W. Wang, and J.-L. Wu, "Video Adaptation for Small Display Based on Content Recomposition," IEEE Transaction on Circuits and Systems for Video Technology, vol. 17, no. 1, pp. 43-58, January 2007.
2. J.-C. Chen, W.-T. Chu, J.-H. Kuo, C.-Y. Weng, and J.-L. Wu, "Tiling Slideshow," Proceedings of ACM Multimedia Conference, pp. 25-34, 2006.
3. W.-T. Chu and J.-L. Wu, "Explicit Semantic Events Detection and Development of Realistic Applications for Broadcasting Baseball Videos," revised for Multimedia Tools and Applications, 2006.
4. W.-T. Chu, C.-W. Wang, and J.-L. Wu, "Extraction of Baseball Trajectory and Physics-Based Validation for Single-View Baseball Video Sequences," accepted by IEEE International Conference on Multimedia & Expo, 2006.
5. W.-T. Chu and J.-L. Wu, "Development of Realistic Applications Based on Explicit Event Detection in Broadcasting Baseball Videos," Proceedings of International Multimedia Modelling Conference, pp. 12-19, 2006.

6. W.-T. Chu and J.-L. Wu, "Integration of Rule-based and Model-based Methods for Baseball Event Detection," Proceedings of IEEE International Conference on Multimedia & Expo, pp. 137-140, 2005.
7. W.-H. Cheng, W.-T. Chu, and J.-L. Wu, "A Visual Attention based Region-of-Interest Determination Framework for Video Sequences," IEICE Transactions on Information and Systems Journal, vol. E-88D, no. 7, pp. 1578-1586, 2005.

六、參考文獻

- [1] Kitter, J., Hatef, M., Duin, R.D.W., and Matas, J., "On combining classifiers," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 3, pp. 226-239, 1998.
- [2] Chen, B., Wang, H.-W., Chien, L.-F., and Lee, L.-S., "A*-admissible key-phrase spotting with sub-syllable level utterance verification," Proceedings of IEEE International Conference on Spoken Language Processing, 1998.
- [3] Chinese Professional Baseball League, <http://www.cpbl.com.tw>
- [4] S.-F. Chang, T. Sikora, A. Purl, "Overview of MPEG-7 standard," IEEE Transactions on Circuits and Systems for Video Technology, vol. 11, no. 6, pp. 688-695, 2001.
- [5] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," IEEE Trans. PAMI, vol. 20, no. 11, pp. 1254-1259, Nov. 1998.
- [6] E.D. Scheirer, "Tempo and beat analysis of acoustic musical signals," Journal of Acoustical Society of America, vol. 103, no. 1, pp. 588-601, 1998.
- [7] 其他參考文獻請參考計畫書

ACM Multimedia 2007 國際會議與會報告心得

報告人：網媒所 吳家麟

會議地點：德國 Augsburg 大學

會議時間：2007 年 9 月 24 日至 29 日

一. ACM Multimedia 國際會議簡介：

ACM Multimedia〔ACM-MM〕自 1993 年首次舉辦以來，由於其 program committee 陣容堅強，審稿嚴謹，立即成為多媒體領域中最權威、影響最大的學術會議。ACM-MM 之投稿分為長篇論文與短篇論文兩種。被接受之長篇論文在會議中每篇以長達 30 分鐘之口頭報告方式發表，而短篇論文則以 Poster 方式與出席會議者進行互動。為促使研究與實用可同時交流，ACM-MM 還包括有 doctoral symposium，Brave New Topics Session，Technical Demonstration，Video program，Open Source Contest 及最近頗受重視之 Interactive Arts Program 等不同性質的節目。

ACM-MM 之所以能保證期論文品質主要原因是因其論文審查係以“Double-Blind”方式進行，每一篇論文至少送給三位委員進行評審，其後 TCP-Members 還要齊聚一堂，依據“Double-Blind”之審查結果進行最後之篩選。ACM-MM 今年之長篇論文接受率僅有 19%〔298 篇投稿，錄取 57 篇〕而短篇論文之接受率亦僅有 27%〔255 篇投稿，錄取 70 篇〕遠較一般之國際學術會議為低。

除了上述 Main-Conference 之外，為使各專門領域之最新研究成果可達成充分交流的目標，ACM-MM 還在 Main-Conference 之後安排了 Multimedia Information Retrieval (MIR)，Mobile Video (MV)，Trecvid Video Summarization (TVS)，Human-Centered Multimedia (HCM)，The Many Races of Multimedia Semantics (MS)，Educational Multimedia and Multimedia Education (EMME) 等 6 大 workshop。

二. 旅程

由於今年大會期間與世界聞名之慕尼黑啤酒節日期相接，為能親身體會此一世界知名的盛會，本人特於台北時間 9 月 20 日晚上搭乘泰國航空 9 時 10 分之班機經泰國曼谷再轉德國法蘭克福國際機場。原本預訂 4 個小時左右可到曼谷機場

的行程，卻因當地天候不佳，在空中多盤旋了近 40 分鐘才安全降落。原本有大約 1 個小時的轉機時間也因此只剩下不到 20 分鐘的時間，在泰航人員的引導下，趕上了飛往法蘭克福的班機，此時我十分擔心隨行行李是否也能順利趕上這班飛機。經過 11 個半小時的飛行，飛機終於於當地時間 9 月 21 日上午 6 時 40 分安全降落，不幸的是，正如本人所擔心的，行李未能趕上本班飛機仍留在曼谷，所幸德國人辦事效率很高，經過大約 1 個小時，便取得了行李延誤的相關證明並將本人在 Augsburg 的旅館資訊留交機場行李追蹤部門。為不影響往後的行程，本人便由法蘭克福搭火車前往目的地：Augsburg。到達 Augsburg 市區，先購買了一些生活必需品，再搭電車於下午 3 時左右到 Quality Hotel Augsburg，一家在網路上評比相當好的旅館。

經過一夜的休息，22 日上午 11 時多被延誤的行李終於由慕尼黑機場送達旅社，午飯後便至 Augsburg 市區熟悉一下環境，並試搭電車前往此次大會舉辦單位：The University of Applied Sciences Augsburg。此次 ACM-MM 會場所在地為該校新完成的 Institut Fur Physik—Horsaalzentrum，是一個設備新穎場地寬拓，十分適合舉辦國際會議的場地。大會亦提供以大會註冊證即可免費搭乘市區公車與電車的服務，使與會者均感到十分便利。

9 月 23 日，為親身感受啤酒節盛會，特於 9 時左右便由 Augsburg 搭火車到慕尼黑火車站，再搭一站地鐵到達啤酒節會場。去年預估啤酒節吸引了超過 6 百萬人次，今年預估到訪人數只會多不會少。為防止旅客因酒醉而發生意外，啤酒節大會提供了不少的防範措施，例如：充分的醫療及警衛人員，為數相當多的急救站與廁所，離開會場地鐵之流動售票人員及定時巡邏之警察。此外，大會規定，旅客必須進到帳棚內且有座位才能買到啤酒，在帳棚外則只能買到不含酒精的飲料。此一規定使外來且人數較少的旅客比較不便（不易擠進帳棚，搶到位子，喝到啤酒），但相較之下，也使大會秩序不至於完全無法掌控。雖然，憑我一個人無法順利擠進帳棚買到啤酒，但在大會會場中吃到著名的德國豬腳，香腸，烤雞及烤魚等，也算不虛此行。

24 日 ACM-MM 大會安排了 6 場不同的 Tutorial，由於收費不低，在完成報到手續後便回到 Augsburg 市區參觀了路易維格二世國王留下的宮殿及一些著名的教堂。

25 日至 27 日參加了 ACM-MM Main Conference 之各項議程（詳見三(a)），28 日至 29 日則參加了 ACM-MM MIR Workshop（詳見三(b)），並發表三篇文章，29 日下午由 Augsburg 搭火車趕到慕尼黑車站再轉搭郊區列車到達慕尼黑國際機場。9 時多搭乘泰航班機飛往曼谷，再轉搭當地時間下午 6 時 15 分之班機於台北時間 9 月 30 日晚上 10 時 45 分到達桃園國際機場。

三. 與會經過

（a）MM-Conference 部分：

ACM-MM 2007 的第一場 Keynote 是邀請德國 DFKA 的 Wolfgang Wahlster 教授以 Small Web 為主題發表演講，不但描繪出 Small Web 的光明前景，更以 BMW 機、汽車相當產品規劃，證實 Small Web 已達實用階段。今年的 Best Paper Session 共有來自微軟亞洲研究院，美國 Wisconsin-Madison，美國 Stanford 大學，印度 IIT 及德國當地學者跨國合作，及美國 Arizona 州立大學等 4 個團隊競爭，其中尤其以 Wisconsin-Madison 團隊之”Re-Cinematography”一文對 video stablization 提出一些新的做法，對我們有最直接的幫助。

25 日下午的 program 中，來自中國的”Trajectory Based Event Tactic Analysis in Broadcast Sports Video”與來自德國的”Can We Trust Digital Image Forensics？”兩篇文章因和本人研究主題密切相關，透過雙方交流，獲益良多。令人最高興的是在 Live Arts Exhibition session 中，看到網媒所同學在本所兼任教授許素朱指導下，以”One Million Heartbeats”為主題的作品入選（被接受率僅 16%）。

26 日的 Keynote speech 分別由德國 CeWe Color 公司的 Dr. Reiner Fageth

及 SAP 公司的 Prof. Lutz Heuser 以”Applied Image Science – From Consumers’ digital files to tangible products”及”The Workshop of the Future”為題闡述 MM 相關研究與產業契機的關連性，看來 MM 對產學合作是相當重視的。上午的”You’re Being Watched” session 中，三篇與 Video Surveillance 相關的論文，值得深入精讀。下午的”Brave New Topic” sessions 共有四篇考慮到 Social Computing 對 MM 影響的論文，因這些主題與資電中心分項計畫之未來的研究主題有一定之相關性，故為本人鎖定之重點所在。

27 日的 Keynote speech 係邀請 NTT DoCoMo-USA 的 Dr. Minoru Etoh 以”Insights into Future Mobile Multimedia Application”為主題，說明 3G 在 MM 應用上之前景。本日的”Computing Complexity” session 的三篇論文分別針對 JPEG2000, MPEG 及 H.264/AVC 三種不同壓縮標準，提出控制計算複雜度的有效方法。本人從事多媒體壓縮研究多年，聽取這些發表，收穫頗大。網媒所新進之徐宏民教授在今天的 Video Search session 中發表他與哥倫比亞大學張世富教授合作之論文”Video Search Reranking through Random Walk over Document – Level Context Graph”，頗受與會者重視。

(b) ACM-MM MIR 部分：

9 月 28 日，ACM-MM MIR workshop 正式開始，一早我便與中正資工朱威達教授一起趕往會場參加 MIR 的’Keynote speech’及”Image Retrieval and Multimedia Modeling”兩個 session，其內容均以如何縮小”Semantic Gap”及提升 Retrieval 效率為主軸。

下午則在張世富教授主持的”Special Session 1：Semantic Indexing of Consumer and Web Videos”中發表與朱威達教授等人共同合作之”ITEMS：Intelligent Travel Experience Management System”一文。本文與同一 session 來自 Kodak，MSRA，Yahoo 及 Google 等業界的論文主軸略有不同，反而受到張教授及業界人士的好評。

9 月 29 日在聽完來自荷蘭 Amsterdam 大學 Prof. Arnold Smeulders 有關

Concept Search 與 Semantic Gap 的 Keynote speech 之後，我們在"Video Retrieval"的 session 中分別發表了"RoleNet:Treat a Movie as a Small Society"及"Semantic-Event Based Analysis and Segmentation of Wedding Ceremony Videos"等二篇文章，由於此二論文均有相當之前瞻性，引起不少的迴響，與會人士對來自 NTU 的團隊留下了深刻的印象。下午因班機時刻的問題，我們在聽完 Panel session 後便離開會場，趕往慕尼黑機場。

四. 與會心得

ACM-MM 係多媒體領域最頂尖的國際會議，其論文被接受率較大多數第一級國際期刊為低，許多世界一流大學包括 MIT，Stanford，Berkeley，UIUC 及 CMU 等多媒體領域相關學者都以在此會議發表論文做為學術貢獻之評斷標準之一。今年 NTU 在整個 ACM-MM 的表現尚可，計有 2 篇 Full Main Conference papers，1 篇 Art Demonstration 及 5 篇 Workshop papers。但相較於中國大陸，由於 MSRA 的全力投入，我們的表現仍遜色很多，這也可以解釋為何 2009 年的 ACM-MM 將在北京舉行。如果要在這場競爭中取得優勢，國內從事多媒體研究之學者仍有很大的努力空間。