

行政院國家科學委員會專題研究計畫 成果報告

以物件軌跡為基礎之互動式視訊內容標記及視訊資料查詢 系統(2/2)

計畫類別：個別型計畫

計畫編號：NSC91-2213-E-009-075-

執行期間：91年08月01日至92年07月31日

執行單位：國立交通大學資訊工程學系

計畫主持人：李素瑛

計畫參與人員：陳敦裕 蕭銘和 陳漪紋 陳華總 郭慧冰 吳惠君

報告類型：完整報告

處理方式：本計畫可公開查詢

中 華 民 國 93 年 2 月 12 日

行政院國家科學委員會專題研究計畫成果報告

以物件軌跡為基礎之互動式視訊內容標記及視訊資料查詢系統(2/2)

計畫編號：NSC 91-2213-E-009-075

執行期限：91年8月1日至92年7月31日

主持人：李素瑛教授 交通大學資訊工程系

計畫參與人員：陳敦裕 蕭銘和 陳漪紋 陳華總 郭慧冰 吳惠君

一、中文摘要

本計畫主要目的為視訊資料內容之分析與擷取，並利用分析所得之資訊加以推理(inference)，以偵測視訊資料內容之事件，藉此達成高階特徵之擷取並以此為基礎產生既簡短又具語意意義之視訊資料摘要(summary)，除此之外在產生視訊內容摘要之過程中，同時產生低階至高階特徵之描述(descriptor)以及描述結構(description scheme)。

此計畫主要在分析經壓縮之視訊資料，如 MPEG-1 及 MPEG-2，其中可用來分析之資訊如 I-畫面之 DCT DC 係數，P-畫面之 macroblock 的運動向量，若欲獲得更精確之分析結果亦可利用 B-畫面之運動向量作為參考。利用上述資訊可分析場景之變化，藉以做影片切割(video segmentation)，並可研發特殊之演算法來指出畫面中物件之位置，並可推廣至連續之畫面，即切割後之視訊片段，以追蹤物件位置之變化，進而獲得物件之軌跡。物件軌跡(object trajectory)即是高階特徵之一，以此高階特徵為基礎再加上 domain knowledge，即可推測出視訊資料中所發生之事件，例如運動影片之足球之射門、網球之發球上網，底線對抽等等，皆可利用上述之方法推測而來，另外亦可利用運動向量並藉 motion model 之分析可獲知視訊片段之運鏡手法，如 zoom、pan、tilt、rotation 等等，進而可推論視訊資料內容所發生之事件，如運動影片籃球之灌籃所對應之 zoom in、快攻所對應之 panning 等不同之運鏡方式。

除了低階特徵之外，經推測所獲得之高階語意特徵皆可以 MPEG-7 之描述來表

示，藉此所建構之互動式視訊資料瀏覽系統，除了資料庫之索引較為有效之外可提供使用者以高階特徵(high-level feature)或語意內容(semantic)作為查詢之依據，此系統並可符合未來 MPEG-7 之標準，藉此可達成 interoperability 之目的。

關鍵詞：場景變化、影片切割、物件軌跡、事件偵測、互動式瀏覽系統、MPEG-7

Abstract

The purpose of this project is to analyze video content and to support content-based video retrieval based on acquired high-level/semantic features. Furthermore, we can detect video events by inferring from both of original data and analyzed information of video streams. Besides, video summaries that are more meaningful can also be obtained by merging semantic video clips that are segmented according to the video events. The description/descriptors corresponding to low-level and high-level features and the description schemes can be automatically generated during the process of video summaries generation.

The goal of this project is to analyze video data in compressed domain, for example, MPEG-1 and MPEG-2 formats. The data that can be used to analyze and embedded in the video streams is like DCT DC coefficients of I-frames and motion vectors of P-frames and B-frames. Based on the information, we can detect where the scene change happens and further segment video into clips. Object location in a frame can also be indicated based on motion information of macroblocks

using some clustering algorithms. With extension to successive frames or a video clip, object trajectories can also be obtained. In general, motion trajectory is a high-level feature associated with a moving region, defined as a spatio-temporal localization of one of its representative points (such as centroid) and can be used to infer video events while applying some domain knowledge. For example, the event of shooting a ball in the football videos, serving and volley or baseline rally in the tennis videos. In the other hand, camera operations, like zoom, pan, tilt, rotation, within the video clips can also be detected based on the motion information by fitting some motion models. Since different camera operations may apply to different kinds of video events, for example, the slam-dunk in a basketball game may correspond to the zoom-in operation and the fast break may correspond to the panning operation, we can also differentiate video events based on different camera operations.

All features obtained can be described by the descriptors in MPEG-7 format. An interactive video browsing system constructed based on these descriptors and the corresponding description schemes can support queries more effectively since the extracted features are more high-level and semantically meaningful.

Keywords: Scene change detection, video segmentation, object trajectory, video event detection, interactive video browsing system, MPEG-7

二、緣由與目的

隨著網際網路規模的不斷擴大以及資訊科技進步快速，愈來愈多的多媒體資料以數位的形式藉網路為媒介廣泛地傳播著。資訊種類之多元以及資訊量之豐富對於使用者來說哪些是有用的資訊及哪些是可忽略的資料自然成為一個重要的議題。因此對於使用者而言，在眾多資料之中如何有效且快速地獲得真正所要的內容便是我們所要努力的目標。然而除了提供快速有效的查詢方式之外，還必須讓使用者同時享有自然且友善的查詢環境，而利用資

料內容為索引及查詢 (Indexing & Retrieval) 為依據的方式 (Content-Based Access) 正能夠滿足此需求。因為所謂 Content-Based Access 即是以資料內容之特徵當成索引來建立資料庫，因此特徵描述之充分與否則直接影響查詢結果之優劣。而傳統資料內容之描述可為某些關鍵字、屬性的集合、或是某些以文字為主的抽象表示法。

過去以文字為基礎的資料庫系統已發展了一段時間並且已經相當成熟，但相反地以多媒體內容為索引及抽取依據的資料庫管理系統發展卻相當落後。最近將 meta-data 的觀念延伸至其他應用已經引起廣泛的探討，其應用可包含形形色色且較為複雜的資料型態，如壓縮的影像、audio、video 等等。因此對於大量的音樂資料、新聞、紀錄片、電影以及影像提供充分適當的描述對於不同領域及不同應用的使用者而言相當地重要。在應用上比如，數位圖書館、隨選視訊以及網路多媒體傳播等等。

因此 MPEG-7 試著定義多媒體特徵的描述方法使得以內容為基礎的資料索引及資料搜尋更為容易。此外 MPEG-7 包含更多的資料型態並且將內容的描述 (descriptor) 和描述法 (description scheme) 標準化藉以達成能夠描述更多種類之多媒體資料。然而特徵產生之方式並非 MPEG-7 欲標準化的部份，除此之外由於多媒體資料型態及內容的多樣性使得更有效、更具高階意義的特徵亟待產生。另一方面為了提供系統間之相互作用 (interoperability)，特徵抽取以及搜尋引擎兩大議題皆非 MPEG-7 欲標準化之部分。

本計畫主要目的為視訊資料內容之分析與擷取，並利用分析所得之資訊加以推理 (inference)，以偵測視訊資料內容之事件，藉此達成高階特徵之擷取並以此為基礎產生既簡短又具語意意義之視訊資料摘要 (summary)，除此之外在產生視訊內容摘要之過程中，同時產生低階至高階特徵之描述 (descriptor) 以及描述結構 (description scheme) 以符合 MPEG-7 之標準。

三、實驗結果與討論

We design a novel descriptor to classify video shots and implement a prototype system to infer tennis events from compressed videos. The experimental results of the motion activity descriptor are demonstrated in Fig. 1 and Fig. 2. Fig. 1 shows the result of query by the shot of “soccer game”. The criterion for matching the shot of “soccer game” is that multiple small objects appear in the consecutive frames and hence shot-5 is also matched and in Fig. 2, the shots with single large object are matched with the shot of “anchor person”.

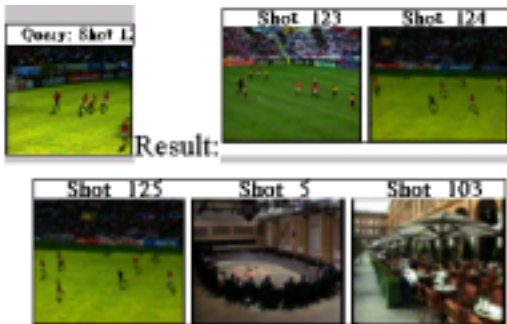


Fig. 1. Query by “Soccer game” Video Shot



Fig. 2. Query by “Anchor Person” Video Shot

We also show the experimental results of the module of semantic event detection Fig. 3. Fig.3 demonstrates the trajectory of objects. In addition, the performance is shown in Table 1 and we can see that the value of the average precision is up to 90% and the value of average recall is 89%. The results show the effectiveness of the proposed module.

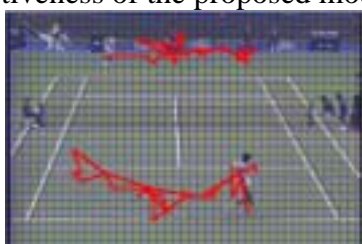


Fig. 3. An Example of Baseline Rallies

Table 1. Experiment Results of Tennis Event Inference

Scene Type \ Results	Actual Number	Detected	False Detection	Precision	Recall
Baseline Rally	91	87	5	95%	96%
Serve and Volley	115	106	6	95%	92%
Passing Shot	24	20	6	77%	83%
Average				89%	90%

Besides, background-based object detection and motion clustering method is used in I frames and P frames respectively to separate objects from background, and a GOP-based object tracking method is proposed in this paper, the proposed object tracking method uses spatial information and the color information as features to match regions. Fig. 4 shows the tracking result of walking person. In our tracking method, we also concern about the spatial relation, thus we would find the two walking people merge into one, and then we correctly match them with color after it split into two again.

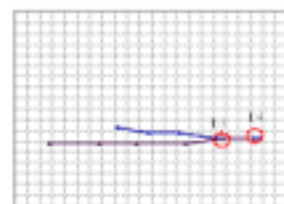
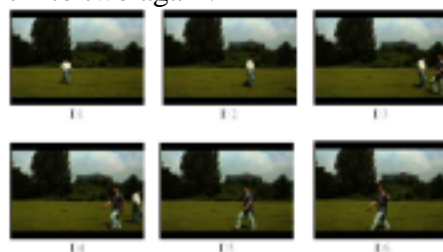


Fig. 4. Tracking result of occlusion

四、参考文献

- [1] G. Sudhir, John C. M. Lee and Anil K. Jain, “Automatic Classification of Tennis Video for High-Level Content-based Retrieval,” Proc. IEEE International Workshop Content-Based Access of Image and Video Database, 1998, pp. 81-90.
- [2] G. S. Pingali, Y. Jean and I. Carlbom,

- “Real Time Tracking for Enhanced Tennis Broadcasts,” Proc. IEEE Computer Society Conference Computer Vision and Pattern Recognition, 1998, pp. 260-265.
- [3] H. Miyamori and S. I. Iisaku, “Video Annotation for Content-based Retrieval using Human Behavior Analysis and Domain Knowledge,” Proc. Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000, pp. 320-325.
- [4] ISO/IEC JTC1/SC29/WG11/N3913, “Study of CD 15938-3 MPEG-7 Multimedia Content Description Interface – Part 3 Visual,” Pisa, January 2001.
- [5] M. Lee, and B. Ahn, “Robust Algorithm for Scene Analysis on Compressed Video,” Proc. IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, 1999, pp. 103-106.
- [6] H. L. Eng, and K. K. Ma, “Bidirectional Motion Tracking for Video Indexing,” Proc. 3rd IEEE Workshop on Multimedia Signal Processing, 1999, pp. 153-158.
- [7] J. Nang, S. Hong, and Y. Ihm, “An Efficient Video Segmentation Scheme for MPEG Video Stream using Macroblock Information,” Proc. ACM Multimedia Orlando, USA, 1999, pp. 23-26.
- [8] S. C. Pei, and Y. Z. Chou, “Efficient MPEG Compressed Video Analysis Using Macroblock Type Information,” IEEE Transactions on Multimedia, Vol. 1, No. 4, December 1999, pp. 321-333.
- [9] L. Favalli, A. Mecocci, and F. Moschetti, “Object Tracking for Retrieval Applications in MPEG-2,” IEEE Transactions on Circuits and Systems for Video Technology, Vol. 10, No. 3, April 2000, pp. 427-432.
- [10] R. Wang, and T. Huang, “Fast Camera Motion Analysis in MPEG domain,” Proc. IEEE International Conference on Image Processing, 1999, Vol. 3, pp. 691-694.
- [11] Y. P. Tan, D. D. Saur, S. R. Kulkarni, and P. J. Ramadge, “Rapid Estimation of Camera Motion from Compressed Video with Application to Video Annotation,” IEEE Transactions on Circuits and Systems for Video Technology, Vol. 10, No. 1, February 2000, pp. 133-146.
- [12] N. Haering, R. J. Qian, and M. I. Sezan, “A Semantic Event-Detection Approach and Its Application to Detecting Hunts in Wildlife Video,” IEEE Transactions on Circuits and Systems for Video Technology, Vol. 10, No. 6, September 2000, pp. 857-868.
- [13] N. Babaguchi, S. Sasamori, T. Kitahashi, and R. Jain, “Detecting Events from Continuous Media by Intermodal Collaboration and Knowledge Use,” Proc. IEEE International Conference on Multimedia Computing and Systems, 1999, Vol. 1, pp. 782-786.
- [14] S. Y. Lee, J. L. Lian and D. Y. Chen, “Video Summary and Browsing Based on Story-Unit for Video-on-Demand Service,” Proc. 3rd Intl. Conference on Information, Communications, and Signal Processing, Singapore, Oct. 2001.
- [15] D. Y. Chen and S. Y. Lee, “Motion-Based Semantic Event Detection for Video Content Descriptions in MPEG-7,” Proc. 2nd IEEE Pacific Rim Conference on Multimedia, Beijing, China, Oct. 2001, pp. 110-117.
- [16] D. Zhong and S. F. Chang, “Structure Analysis of Sports Video Using Domain Models,” Proc. IEEE Conference on Multimedia and Exhibition, Tokyo, Japan, Aug. 22-25, 2001. ISO/IEC JTC1/SC29 /WG11/N3964, “MPEG-7 Multimedia Description Schemes XM (v7.0),” Singapore, March 2001.
- [18] Coding of Moving Pictures and Associated Audio-for Digital Storage Media at up to about 1.5Mbit/s, Committee Draft of Standard ISO11172: ISO/MPEG 90/176, November 1991.
- [19] J. Meng, Y. Juan, S.F. Chang, “Scene Change Detection in a MPEG Compressed Video Sequence,” Proc. IS&T/SPIE, Vol. 2419, pp.14-25, 1995

