

# 行政院國家科學委員會補助專題研究計畫成果報告

## 使用動態規劃法之內涵式視訊搜尋

計畫類別：×個別型計畫      整合型計畫

計畫編號：NSC 89-2213-E-002-015

執行期間： 88年8月1日至 89年 7月 31日

計畫主持人：李瑞庭

共同主持人：

本成果報告包括以下應繳交之附件：

赴國外出差或研習心得報告一份

赴大陸地區出差或研習心得報告一份

出席國際學術會議心得報告及發表之論文各一份

國際合作研究計畫國外研究報告書一份

執行單位：臺灣大學資訊管理學系

中 華 民 國 89年 10月 30日

## 使用動態規劃法之內涵式視訊搜尋

### A Dynamic Programming Approach to Content-based Video Searching

計畫編號：NSC 89-2213-E-002-015

執行期間：88/08/01-89/07/31

主持人：李瑞庭 臺灣大學資訊管理學系 副教授

一、中文摘要(關鍵詞:視訊資料、視訊資料庫、視訊格、鏡、幕、索引、動態規劃法)

本計畫提出一個新的視訊搜尋的演算法，以動態規劃(dynamic programming)的精神來進行視訊的比對。首先，我們將一個視訊切割成一個一個鏡，並選取一個關鍵視訊格來代表每個鏡。再選取適當的特徵向量來代表視訊中每一個關鍵視訊格的特徵，例如：顏色、紋理、形狀等特徵。然後，使用者可以用一個詢問視訊(query video)來找尋和詢問視訊最相似的視訊，或與詢問視訊的相似度(similarity)大於某個門檻值(threshold)的視訊。在此，我們使用動態規劃法來找尋符合詢問的視訊。所提出的內涵式視訊搜尋方法具有下列的特性：

1. 使用視訊之內涵來搜尋視訊，並支援以範例詢問(query by example)。
2. 使用動態規劃法，可以求得最佳的解。
3. 本方法之時間複雜度為  $O(mn)$ ，其中  $m$  為詢問視訊(query video)所包括之關鍵視訊格數目， $n$  為儲存於資料庫中所有視訊所含之關鍵視訊格數的總和。
4. 時間資訊能夠充分被利用，保持詢問視訊中視訊格的時間順序。
5. 並提供相似度的查詢。
6. 所使用的詢問視訊包括愈多鏡，則所得的結果精確度(precision)與檢出率(recall)愈高。

英文摘要(Keywords : Video data, Video frame, Shot, Scene, Index, Dynamic Programming)

In this project, we propose a dynamic programming approach to content-based video searching. First, each video is segmented into several shots, each of which is represented by a key frame. Each key frame is characterized by a feature vector, including color, texture, shape, etc. Second, a user may query the video database by a query video to search the best matched video or a set of videos with the similarity greater than a threshold. The proposed content-based video searching algorithm has the following properties:

1. Query a video database by the content of an example video.
2. Use a dynamic programming approach to find the best matched video.
3. The complexity of the algorithm is bounded by  $O(mn)$ , where  $m$  is the number of key frames in the query video and  $n$  is the total number of key frames of videos in the database.
4. Keep the temporal constraints on the query.
5. Provide similarity search.
6. The more shots are contained in the query video, the higher precision and recall are.

## 二、計畫的緣由與目的

多媒體的應用與研究在近年來日漸興盛，主要的動力在於硬體速度的提昇、儲存空間的增大、網路使用者的急速增加、以及日漸成長的頻寬、對於高品質的資訊內涵的需求、多媒體的資訊系統越來越容易使用。這些動力都促使多媒體相關的研究日漸重要。在可預見的未來，這些因素並不會消失，反而會更形重要。因此多媒體系統與其應用可預期有非常高的重要性與發展空間。

傳統的視訊查詢(video retrieval)都是以快轉、倒轉，然後播放的方式來瀏覽並搜尋所要的片段。這種方式在效率上來說是線性的(也就是 sequential 的方式)，在實用上非常浪費時間。因此為了克服此種限制，加速視訊(video)的查詢，就必須將視訊切割成有意義的片段，並以索引(index)或摘要視訊(summary video)等方式來加速存取與瀏覽的速度，或是讓使用者可以在不同的階層來瀏覽視訊，然後放大某一段，選取所要的片段。

就詢問(query)的階層來看，目前使用查詢視訊資訊的做法有以鏡為主來尋找對應的鏡的，也有以影像為主來直接比對影像的，也有以整個視訊的特徵為主來查詢的。內涵式的查詢在最近幾年就變成一個相當富挑戰性的研究議題。最近幾年在影像(image)的內涵式查詢方面較有實質性的進展，如 QBIC[FSN95]，Photobook[PPS96]，Virage[HGH97]，與 VisualSEEK[SC96]。在視訊的內涵式查詢方面，Chang 等人[CCM97]首先提出一個內涵式視訊查詢系統，叫做 VideoQ。

Smoliar 與 Zhang[SZ94]發展出一套方法查詢視訊資料。首先將視訊資料加以切割後以框(frame)代替，這裡的框是指一種資料結構，主要是將特徵與節點相結合，以代表某種觀念或物件。這些

節點只儲存特定領域的知識，然後以標題式的階層架構來分類之。每一種分類則由一些特徵，像是顏色、紋理、形狀等來加以區別。查詢時則是看代表視訊資料的關鍵視訊格與詢問之間各種特徵的相似度來判斷。但是這個方法完全忽略視訊資料所具有的時間特性。

CONIVAS 系統[ADD96]則使用邊緣或顏色之一作為影像主要的特徵。在視訊的查詢方面則提供兩種方法：第一種是以關鍵視訊格(keyframe)為主，先將視訊切割成鏡，再自動算出對應每個鏡的關鍵視訊格，然後使用者以現成資料庫中的影像例子來查詢這個關鍵視訊格。查詢結果相似的關鍵視訊格會顯示出來讓使用者點選後播放視訊。第二種方法是使用 DC+M 的視訊查詢方法，根據兩段視訊的分佈特徵與移動特徵來計算相似程度。

然而，在視訊的內涵式查詢方面，則少有人探討以一個視訊來搜尋資料庫，找出與之最相似的視訊或與它的相似度大於某個門檻值的視訊。因此，本計畫將針對此議題做深入的探討。

## 三、研究方法與成果

本計畫提出一個視訊搜尋的演算法，以動態規劃(dynamic programming)的精神來進行視訊的比對。首先，我們將一個視訊切割成一個一個鏡，並選取一個關鍵視訊格來代表每個鏡。再選取適當的特徵向量來代表視訊中每一個關鍵視訊格的特徵，例如：顏色、紋理、形狀等特徵。然後，使用者可以用範例視訊(example video)或詢問視訊(query video)來找尋和詢問視訊最相似的視訊，或與詢問視訊的相似度(similarity)大於某個門檻值(threshold)的視訊。

我們提出的演算法，利用動態規劃(dynamic programming)的精神來計算詢問視訊 Q 與資料庫中某一個視訊 V 的相

似度(similarity), 以便搜尋出相似度最高的視訊。

我們先定義兩個視訊格的相似度, 兩個視訊格都可以以  $k$  維度的特徵向量  $q=(a_1, a_2, \dots, a_k)$  與  $f=(b_1, b_2, \dots, b_k)$ , 首先, 我們先計算兩個視訊格的特徵向量的距離, 而其距離之計算如下:

$$D(q, f) \equiv \sqrt{\sum_{i=1}^k (a_i - b_i)^2}$$

兩個視訊格的相似度, 則依據其特徵向量的距離來定義:

$$S(q, f) = 1 - \frac{D(q, f)}{MaxD}$$

其中  $MaxD$  表示特徵空間中, 兩特徵向量最大的可能距離。因此  $S(q, f)$  將會介在  $[0, 1]$  之間。當然也可以在每個特徵值上加上權重(weight), 再計算相似度。

假設  $V$  含有  $s$  個關鍵視訊格  $F_1, F_2, \dots, F_s$ , 其相對應的特徵向量為  $f_1, f_2, \dots, f_s$ ;  $Q$  含有  $m$  個關鍵視訊格  $QF_1, QF_2, \dots, QF_m$ , 其相對應的特徵向量為  $q_1, q_2, \dots, q_m$ ,  $s \geq m$ 。當  $QF_i$  比對到  $F_j$  的時候, 以  $M_{i,j}$  表示  $QF_i$  對應到  $F_j$  時的最大的相似度總和。則  $M_{i,j}$  可表示如下:

$$M_{i,j} = \begin{cases} S(q_i, f_j) & \forall i=1 \\ S(q_i, f_j) + \text{Max}(M_{i-1, i-1}, \dots, M_{i-1, j-1}) & \forall j = i, i+1, \dots, s-m+i \\ 0 & \forall j < i \\ 0 & \forall i < j-s+m \end{cases}$$

因此, 只要從最後往前遞迴算出  $M_{m,s}$  的最大值即為其相似度。

如此, 我們將資料庫中每一個視訊一一與  $Q$  做比對, 即可算出每個視訊與  $Q$  的相似度。所以, 我們就可以找出與  $Q$  相似度最高的視訊, 或找出與  $Q$  最相似的前幾個視訊。

此外, 我們共提供三種相似性的查詢, 三種查詢之定義如下:

} 第一種查詢: 查詢所得的視訊, 與詢問視訊所對應的鏡必須符合出現的先後順序。

} 第二種查詢: 除了先後順序之外, 鏡與鏡之間的時間間隔還必須符合使用者所輸入的限制。

} 第三種查詢: 沒有順序的限制。只要視訊包含使用者所輸入的詢問視訊的鏡即可。

#### 四、結論與討論

我們使用動態規劃法來找尋符合詢問的視訊。這樣的搜尋方式, 可以兼顧效率、準確、彈性。所提出的內涵式視訊搜尋方法具有下列的特性:

1. 使用視訊之內涵來搜尋視訊, 並支援以範例詢問(query by example)。
2. 使用動態規劃法, 可以求得最佳的解。
3. 本方法之時間複雜度為  $O(mn)$ , 其中  $m$  為詢問視訊(query video)所包括之關鍵視訊格數目,  $n$  為儲存於資料庫中所有視訊所含之關鍵視訊格數的總和。
4. 時間資訊能夠充分被利用, 保持詢問視訊中視訊格的時間順序。
5. 並提供相似度的查詢。
6. 所使用的詢問視訊包括愈多鏡, 則所得的結果精確度(precision)與檢出率(recall)愈高。

#### 五、參考文獻

- [ADD96] M. Abdel-Mottaleb, N. Dimitrova, R. Desai, and J. Martino, "CONIVAS: Content-based Image and Video Access System," *In Proceedings of ACM International Conference on Multimedia*, Boston USA, 1996.
- [AHK96] P. Alshuth, T. Hermes, J. Krey, and M. Röper, "Video Retrieval with IRIS," *In Proceedings of ACM International Conference on Multimedia*, Boston, USA, 1996.
- [BY98] R. Bolle and B. Yeo, "Computer

- Vision for Visual Computing: Techniques and Applications," *Computer Vision and Image Understanding*, Vol. 71, No. 2, August, 1998.
- [CCM97] S. Chang, W. Chen, H. Meng, H. Sundaram, and D. Zhong, "VideoQ: An Automated Content Based Video Search System Using Visual Cues," *In Proceedings of ACM International Conference on Multimedia*, Seattle USA, 1997, pp. 313-324.
- [Dan93] Daneels D. et al., "Interactive Outlining: An Improved Approach Using Active Geometry Features," *In Proceedings of IS&T/SPIE. Conf. On Storage and Retrieval for Image and Video Databases II*, San Jose, CA, 1993.
- [FSN95] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkanic, J. Hafner, D. Lee, D. Petkovic, D. Steele, P. Yanker, "Query by Image and Video Content: The QBIC System," *IEEE Computer Magazine*, Vol. 28, No. 9, 1995, pp. 23-32.
- [Gon94] Gong Y. et al., "An Image Database System with Content Capturing and Fast Image Indexing Abilities," *In Proceedings of International Conference on Data Engineering*, Boston, MA, 1994, pp.121-130.
- [HGH97] A. Hamrapur, A. Gupta, B. Horowitz, C.F. Shu, C. Fuller, J. Bach, M. Gorkani, R. Jain, "Virage Video Engine," *SPIE Proceedings on Storage and Retrieval for Image and Video Databases V*, 1997, pp. 188-197.
- [Kat92] Kato T. et al., "A Sketch Retrieval Method for Full Color Image Database: Query by Visual Example," *In Proceedings of 11<sup>th</sup> International Conference on Pattern Recognition*, Amsterdam, HOLLAND, 1992, pp.530-533.
- [MJ92] J. Mao, K. Jain, "Texture Classification and Segmentation Using Multiresolution Simultaneous Autoregressive Models," *Pattern Recognition*, Vol. 25, No.2, 1992, pp.93-98.
- [NT91] A. Nagasaka, and Y. Tanaka, "Automatic Video Indexing and Full-video Search for Object Appearances," *In Proceedings of Second Working Conference on Visual Database Systems*, October, 1991.
- [PPS96] A. Pentland, R.W. Picard, and S. Sclaroff, "Photobook: Content-based Manipulation of Image Databases," *International Journal of Computer Vision*, Vol. 18, No. 3, 1996, pp. 233-254.
- [SC96] J.R. Smith and S.F. Chang, "VisualSEEK: A Fully Automated Content-based Image Query System," *In Proceedings of ACM International Conference on Multimedia*, 1996, pp. 87-98.
- [SZ94] S. Smoliar and H. Zhang, "Content-based Video Indexing and Retrieval," *IEEE Multimedia*, Vol. 1, No. 2, 1994, pp.62-72.
- [TMY79] H. Tamuran, S. Mori, and T. Yamawaki. "Texture Features Corresponding to Visual Perception," *IEEE Trans. On Syst. Man, And Cybern.*, Vol.6, No.4, 1979, pp.460-473.