

視訊分割、合併與索引之研究

A Study on Video Parsing and Indexing

計畫編號：NSC 88-2213-E-002-060

執行期間：87/08/01-88/07/31

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、景、索引、索引樹)

本研究提出一個新的視訊索引方法：Cone-tree。目前所有的索引方法都使用特徵向量的距離做為編索引的基礎。但是以距離為基礎的索引方法，不適合處理向量的平移(scalar shift)的問題。所以我們提出以向量夾角為基礎的索引方法，如此對於向量的平移(scalar shift)才有辦法加以處理。目前大多數的索引方法都使用距離來分割向量空間，將空間分割成一個個的矩形或球形區域然後，建立一個樹狀的索引樹。而我們的索引方法主要根據特徵向量的夾角，將空間分割成一個個錐形的區域，然後，建立一個樹狀的索引樹。索引樹本身是一個平衡樹(balanced tree)，它的高度為 $O(\log_c(N))$ ，其中 N 為特徵向量的個數， c 為一個節點可儲存的特徵向量數。根據和 SR-Tree 的比較結果，對於用特徵向量的夾角為基礎的 nearest neighbor query，我們發現我們的方法在大部分的情況下較 SR-Tree 為佳。

英文摘要(Keywords：Video data, Video

database, Video frame, Shot, Scene, Index, Index tree)

Many indexing methods have been proposed. However, most of them are built on the basis of distance between feature vectors. The indexing methods based on the distance can not handle the problem of scalar shifts. So, we propose a new index method, called ConeTree, which is suitable for nearest neighbor query based on the angle between two features vectors. Thus, our ConeTree can deal with the problem of scalar shifts. Instead of partitioning the data space into rectangular or spherical regions, the Cone-tree partitions the data space into cone-shape regions. The smaller the distances among points are, the closer the points are clustered in the SR-tree. Similarly, the smaller the angles among feature vectors, the closer the feature vectors are clustered in the Cone-tree. Our Cone-tree is a balanced tree whose height is bounded by $O(\log_c(N))$, where N is the number of feature vectors and c is the maximum number of feature vectors stored in a node. In the case of searching based on the angle between two features vectors, the results of our experiments demonstrate that a Cone-

tree outperforms an SR-tree in most of cases.

二、計畫的緣由與目的

近幾年來，隨著資訊科技的日益精進，如高效能的電腦、高容量的儲存體，以及處理多媒體資料的技術提昇，使得結合文字(Text)、聲音(Audio)、影像(Image)、圖形(Picture)、甚至視訊(Video)等多種媒體的應用逐漸成為目前的趨勢。由於視訊內容的特性較不同於一般的資料型態（如文字），其具有動態性的時間-空間相依的特性、龐大的數位視訊資料量（對於每秒播放 24 個視訊格，每個視訊格為具有 640X480 的全彩影像，其所需的資料量約為 21MB），以及非結構化的內容。這些特性，使得視訊資料變得難以管理與查詢。

因為視訊資料本身為一原始資料檔，我們難以從原始視訊資料中，直接判讀其內容。加上視訊的來源管道眾多，遂造成無法由眾多視訊資料中，迅速而有效的找到所需要的視訊片段。這對於視訊的後續處理工作，形成極不方便的現象。所以，如何發展一套有效的視訊處理與查詢系統，讓使用者可以不需要瀏覽過所有的視訊資料，而直接透過查詢的能力，迅速並正確的決定相關的視訊片段，就成為一個重要的研究議題。

目前最通用的視訊處理與查詢方法是：先將視訊資料分割成一個一個鏡(shot)或景(scene)，然後針對這些鏡或景來做視訊內容查尋。然後對這些視訊鏡、鏡編索引，以方便日後的搜尋的工作。然而，目前大多數的索引方法都是以特

徵向量間的距離來編索引。這些索引方法並不適合解決對於色差所造成的向量平移(scalar shift)的問題。因此，我們提出以特徵向量的夾角來編索引，如此，才能解決向量平移(scalar shift)的問題。

三、研究方法與成果

本研究的主要概念是利用分割特徵向量空間(簡稱空間)的方式，將空間分割成好幾個錐形子空間，直到每一錐形子空間所包含的特徵向量可以存入一個磁碟頁(Disk page)中，則停止分割。否則將每一錐形子空間用上述的方法再進行分割，直到每一錐形子空間所包含的特徵向量可以存入一個磁碟頁中為止。

首先，我們將視訊資料加以分割與結構化剖析，使用一個關鍵視訊格來代表每個分割後的視訊鏡(shot)，並萃取關鍵視訊格的特徵(如顏色、紋理等)當作它的特徵向量。再將所有關鍵視訊格的特徵向量輸入到 ConeTree 中，方便日後的查詢。並且提供類似性查詢(Similarity search)。

在本研究中，我們提出的索引方法主要包括下列演算法：insert, delete, nearest neighbor query。所建立的樹的高度為 $O(\lg(N))$ 。根據和 SR-Tree 的比較結果，對於用特徵向量的夾角為基礎的 nearest neighbor query，我們發現我們的方法在大部分的情況下較 SR-Tree 為佳。

四、結論與討論

我們提出的索引方法，能達到更有

效率與更有彈性的管理與查詢。所提出的視訊資料索引方法將具有下列的特性：

1. 將視訊資料切割成一個個具意義的鏡(shot)；並使用一個關鍵視訊格來代表此鏡。
2. 對於每一個關鍵視訊格，萃取出一個特徵向量，包括色彩、紋理等等。
3. 將這些特徵向量建成一個索引樹，此索引樹為一個平衡樹(Balanced Tree)。
4. 此索引樹的高度為 $\log(N)$ 。
5. 每一個非葉節點(non-leaf)所對應的空間為一個錐形區域，且它包括下層節點所對應的錐形區域或所包括的特徵向量。
6. 提供類似性查詢(Similarity search)。

五、參考文獻

- [1] F. Arman, R. Depommier, A. Hsu, and M-Y. Chiu, "Content-based Browsing of Video Sequences," *In Proceedings of ACM International Conference on Multimedia*, 1994.
- [2] M.G. Brown, J.T. Foote, G.J.F. Jones, K.S. Jones, S.J. Young, "Automatic Content-based Retrieval of Broadcast News," *In Proceedings of ACM International Conference on Multimedia*, 1995.
- [3] T.S. Chua, and L.Q. Ruan, "A Video Retrieval and Sequencing System," *ACM Transactions on Information Systems*, Vol. 13, No. 4, Oct 1995.
- [4] Akio Nagasaka and Yuzuru Tanaka,

"Automatic Video Indexing and Full-video Search for Object Appearances," *In Proceedings of 2nd Working Conference on Visual Database Systems*, Oct 1991.

- [5] K. Otsuji and Y. Tonomura, "Projection-detecting Filter for Video Detection," *Multimedia Systems*, 1994.
- [6] R. Weiss, A. Duda, and D.K. Gifford, "Content-Based Access to Algebraic Video," *IEEE Transactions on Knowledge and Data Engineering*, 1994.
- [7] R. Zabih, J. Miller, and K. Mai, "A Feature-Based Algorithm for Detecting and Classifying Scene Breaks," *In Proceedings of ACM International Conference on Multimedia*, 1995.
- [8] H. Zhang, Y. Gong, S. W. Smoliar, and S.Y. Tan, "Automatic Parsing of News Video," *IEEE Transactions on Knowledge and Data Engineering*, May 1994.
- [9] H.J. Zhang, C.Y. Low, S.W. Smoliar, and J.H. Wu, "Video Parsing, Retrieval and Browsing: An Integrated and Content-Based Solution," *In Proceedings of ACM International Conference on Multimedia*, 1995.