行政院國家科學委員會專題研究計畫　成果報告

視訊資料庫中物件時空關係推理與相似性查詢之研究

計畫類別：個別型計畫
計畫編號：NSC91-2213-E-002-092-
執行期間：91年08月01日至92年07月31日
執行單位：國立臺灣大學資訊管理學系暨研究所

計畫主持人：李瑞庭

報告類型：精簡報告

處理方式：本計畫可公開查詢

中　華　民　國 92年12月2日

視訊資料庫中物件時空關係推理與相似性查詢之研究
A study on spatio-temporal reasoning and similarity retrieval in video databases

計畫編號：NSC 91-2213-E-002-092
執行期間：91/08/01-92/07/31

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、時空關係、2D 字串、3D 字串)

我們已提出一個新的視訊時空關係知識結構，3D C-string，並提出相關的字串產生與視訊重建演算法。本計畫提出一個新的視訊時空關係推理與相似性查詢的方法，我們擴充 2D C$^+$-string 中物件空間關係推理與相似性查詢方法到 3D C-string 中，以進行視訊資料庫中物件時空關係推理與相似性查詢。我們提出的方法主要包括兩個步驟，首先，我們推出視訊每一對物件的空間關係系列與時間關係，然後，使用所推出的關係來定義視訊的相似度並提出相似性查詢的演算法。我們所提出的方法可提供各種相似度的查詢。最後，我們進行一些實驗以評估我們所提出方法的效率。

英文摘要

(Video data, Video database, Spatio-temporal relationship, Knowledge structure, 2D string, 3D string)

We have proposed a new spatio-temporal knowledge structure 3D C-string to represent symbolic videos accompanying with the string generation and video reconstruction algorithms. In this paper, we extend the idea behind the similarity retrieval of images in 2D C$^+$-string to 3D C-string. Our extended approach consists of two phases. First, we infer the spatial relation sequence and temporal relations for each pair of objects in a video. Second, we use the inferred relations to define various types of similarity measures and propose the similarity retrieval algorithm. By providing various types of similarity between videos, our proposed similarity retrieval algorithm has discrimination power about different criteria. Finally, some experiments are performed to show the efficiency of the proposed approach.

二、計畫的緣由與目的

With the advances in information technology, videos have been promoted as a valuable information resource. Because of its expressive power, videos are an appropriate medium to show dynamic and compound concepts that cannot be easily explained or demonstrated using text or other media. Recently, there has been widespread interest in various kinds of database management systems for managing information from videos. The video retrieval problem is concerned with retrieving videos that are relevant to the users' requests from a large collection of videos, referred to as a video database.

In the previously proposed video retrieval methods, videos can be retrieved by textual descriptions of conventional database techniques, by attributes [1-2], or by visual features and browsing [3]. It is insufficient for video retrieval by attributes or textual descriptions of conventional database techniques, because there exist numerous interpretations of visual data and the current technologies of machine vision [4-5] can not provide automatic extraction of semantic information from general video programs. Retrieval by visual features can access video clips by colors, shapes, texture, sketch, object's motions and spatial constraints [6]. Retrieval by browsing allows users to find the desired video frames by navigation along links. Visual features or textual descriptions may be used to measure the similarity. Visual features can be extracted from a video automatically or semi-automatically and are independent of the semantic interpretation.

To retrieve desired videos from a video database, one of the most important methods for discriminating videos is the perception of the objects and the spatio-temporal relations that exist between the objects in a video. Much progress in the area of content-based video retrieval has been made in recent years, such as object searching based on localized color [7], moving object extraction and indexing [8], motion trail matching in MPEG [4], key-frame based video browsing [9], and abstraction of high level video units [10].

To represent the spatio-temporal relations between the objects in a video and keep track of objects' motions and size changes, we have proposed 3D C-string [11] to represent the spatio-temporal relations between the objects in a video. In the knowledge structure of 3D C-string, an object is approximated by a minimum bounding rectangle (MBR) and the information about its motions and size changes is recorded. 3D C-string is more accurate and efficient in representation and manipulation of videos and more spatio-temporal knowledge than 3D-list that was proposed by Liu *et al.* [12].

To represent the spatial relations between the objects in an image, many symbolic representations of images have been proposed such as 2D-string [13], 2D G-string [14], 2D C-string [15-18], 2D C^+-string [19], and 2D RS-string [20]. In image similarity retrieval, Chang *et al.* [13] defined three types of 2D subsequences to perform subpicture matching on 2D strings and transformed the image retrieval problem into a problem of 2D subsequence matching. Lee *et al.* [17] also proposed a similarity retrieval algorithm based on 2D string longest common subsequence (LCS). They showed that an LCS problem of 2D string can be transformed to the maximum complete subgraph problem [17]. Based on thirteen spatial relations and their categories, Lee and Hsu [15] proposed three types of similarity between images represented by 2D C-string. In the knowledge structure of 2D C^+-string [19], Huang and Jean used the same clique finding algorithm to find the most similar pictures.

In comparison with the spatial relation inference and similarity retrieval in image database systems, the counterparts in video database systems are a fuzzier concept. Therefore, we extend the idea behind the relation inference and similarity retrieval of images in 2D C^+-string to 3D C-string. We also define the similarity measures and propose a similarity retrieval algorithm.

三、研究方法與成果

The capability of similarity retrieval is important in video database management systems. Therefore, in this paper, we extend the idea behind the similarity retrieval of images in 2D C^+-string to 3D C-string. Our extended approach consists of two phases. First, we infer the spatial relation sequence and temporal relations for each pair of objects in a video. Second, we use the inferred relations and sequences to define various types of similarity measures and propose the similarity retrieval algorithm. By providing various types of similarity between videos, our proposed similarity retrieval algorithm has discrimination power about different criteria. Our proposed approach can be easily applied to an intelligent video database management system to infer spatial and temporal relations between the objects in a video and to retrieve the videos similar to a given query video from a video database.

四、結論與討論

We have proposed a new spatio-temporal knowledge structure called 3D C-string to represent symbolic videos accompanying with the string generation and video reconstruction algorithms. In this paper, we extend the idea behind the similarity retrieval of images in 2D C^+-string [19] to 3D C-string. Our extended approach consists of two phases. First, we infer the spatial relation sequence and temporal relations for each pair of objects in a video. Second, we use the inferred relations and sequences to define various types of similarity measures and propose the similarity retrieval algorithm. Three criteria are used to define similarity measures. The concept of processing spatial relation sequences and temporal relations can also be easily extended to process other criteria such as velocities, rates of size changes, distances, and so on. We also show that different types of similarity have different multi-granularity to meet users' need by examples. By providing various types of similarity between videos, our proposed similarity retrieval algorithm has discrimination power about different criteria. Our proposed approach can be easily applied to an intelligent video database management system to infer spatial and temporal relations between the objects in a video and to retrieve the videos similar to a query video from a video database.

A video contains rich visual and audio (or sound) information. In the 3D C-string representation and the similarity retrieval algorithm based on the 3D C-string, we focused on utilizing the visual information to process videos. How to integrate the audio information with the visual information to represent a video and perform similarity retrieval is worth further study.

五、參考文獻

[1] T. D. C. Little, G. Ahanger, R. J. Folz, and J. F. Gibbon, A digital on-demand video service supporting content-based queries, In *Proc. ACM Intl. Conf. On Multimedia*, Ahaheim, CA, 427-436, (1993).

[2] L. A. Rowe, *et al.*, Indexes for user access to large video databases, In *Proc. Of Storage and Retrieval for Image and Video Database II, IS&T/SPIE Symposium on Electronic Imaging Science & Technology*, San Jose, CA, 150-161, (1994).

[3] J. K. Wu, *et al.*, CORE: A content-based retrieval engine for multimedia information systems, *Multimedia Systems*, Vol.3, No.1, 25-41, (1995).

[4] N. Dimitrova and F. Golshani, Rx for semantic video database retrieval, In *Proc. ACM Intl. Conf. On Multimedia*, San Francisco, CA 219-226, (1994).

[5] N. Dimitrova and F. Golshani,

Motion recovery for video content classification, *ACM Trans. On Information Systems*, Vol.13, No.4, 408-439, (1995).

[6] A. Gupta and R. Jain, Visual information retrieval, *Communications of ACM* , Vol.40, No.5, 71-79, (1997).

[7] A. Nagasaka and Y. Tanaka, Automated video indexing and full-video search for object appearance, *Visual Database Systems II*, (1992).

[8] S. Y. Lee and H. M. Kao, Video indexing-an approach based on moving object and track, In *Proc. IS&T/SPIE*, Vol. 1908, 25-36, (1993).

[9] D. Zhong, H. J. Zhang and S.-F. Chang, Clustering methods for video browsing and annotation, In *Proc. Of Storage and Retrieval for Image and Video Database IV, IS&T/SPIE's Electronic Imaging*, (1996).

[10] M. M. Yeung, B. L. Yeo and B. Liu, Extracting story units from long programs for video browsing and navigation, In *Proc. Multimedia Computing and Systems*, (1996).

[11] Anthony J.T. Lee, Han-Pang Chiu and Ping Yu, 3D C-string: a new spatio-temporal knowledge structure for video database systems, *Pattern Recognition*, Vol. 35, No. 11, 2521-2537, (2002).

[12] C.C. Liu and A.L. P. Chen, 3D-list: a data structure for efficient video query processing, *IEEE Trans. on Knowledge and Data Engineering*, Vol. 14, No. 1, 106-122, (2002).

[13] S.K. Chang, Q.Y. Shi and C.W. Yan, Iconic indexing by 2D strings, *IEEE Trans. Pattern Analysis and Machine Intelligence*, PAMI-9, 413-429, (1987).

[14] S.K. Chang, E. Jungert and Y. Li, Representation and retrieval of symbolic pictures using generalized 2D strings, Technical Report, University of Pittsburg, (1988).

[15] S.Y. Lee and F.J. Hsu, 2D C-string: a new spatial knowledge representation for image database system, *Pattern Recognition*, Vol. 23, 1077-1087, (1990).

[16] S.Y. Lee and F.J. Hsu, Picture algebra for spatial reasoning of iconic images represented in 2D C-string, *Pattern Recognition Letter*, Vol. 12, 425-435, (1991).

[17] S.Y. Lee and F.J. Hsu, Spatial reasoning and similarity retrieval of images using 2D C-string knowledge representation, *Pattern Recognition*, Vol. 25, 305-318, (1992).

[18] S. Y. Lee, M. K. Shan and W. P. Yang, Similarity retrieval of iconic image database, *Pattern Recognition*, Vol. 22, 675-682, (1989).

[19] P.W. Huang and Y.R. Jean, Using 2D C^+-string as spatial knowledge representation for image database systems, *Pattern Recognition*, Vol. 27, 1249-1257, (1994).

[20] P.W. Huang and Y.R. Jean, Spatial reasoning and similarity retrieval for image database systems based on RS-strings, *Pattern Recognition*, Vol.29, No.12, 2103-2114, (1996).