VideoQ: A Fully Automated Video Retrieval System Using Motion Sketches

Shih-Fu Chang William Chen Hari Sundaram

Dept. of Electrical Engineering and New Media Technology Center
Columbia University
New York New York 10027.
E-mail: {sfchang,bchen,sundaram}@ctr.columbia.edu

Abstract

The rapidity with which digital information, particularly video, is being generated, has necessitated the development of tools for efficient search of these media. Content based visual queries have been primarily focused on still image retrieval. In this paper, we propose a novel, interactive system on the Web, based on the visual paradigm, with spatio-temporal attributes playing a key role in video retrieval. The resulting system VideoQ, is the first on-line video search engine supporting automatic object based indexing and spatio-temporal queries.

1. Introduction

Content-based visual queries (CBVQ) has emerged as a challenging research area in the past few years. While there has been substantial progress with the presence of systems such as QBIC [Flickner 1995], PhotoBook [Pentland 1996], Virage [Hamrapur 1997] and VisualSEEk [Smith 1996] most systems only support retrieval of still images. CBVQ research on video databases has not been fully explored yet. VideoQ is an advanced content-based video search system, with the following unique features:

- Automatic video object segmentation and tracking (including global motion stabilization).
- A rich visual feature library including color, texture, shape and motion.
- Query with multiple objects.
- Spatio-temporal constraints on the query.

Specifically, we present a novel video search system which allows users to search video based on a rich set of visual features and spatio-temporal relationships. What makes VideoQ powerful is the novel idea of an animated sketch to formulate the query. In an animated sketch, motion and temporal duration are the key attributes assigned to each object in the sketch in addition to the usual attributes such as shape, color and texture. Our objective is to investigate the full potential of visual cues in object-oriented content-based video search.

We also support a keyword based search, where the keywords have been manually generated. While the search on video databases ought to necessarily incorporate the diversity of the media (video, audio, text captions) our present work will integrate well into any such effort. An example query is shown in Figure 1.

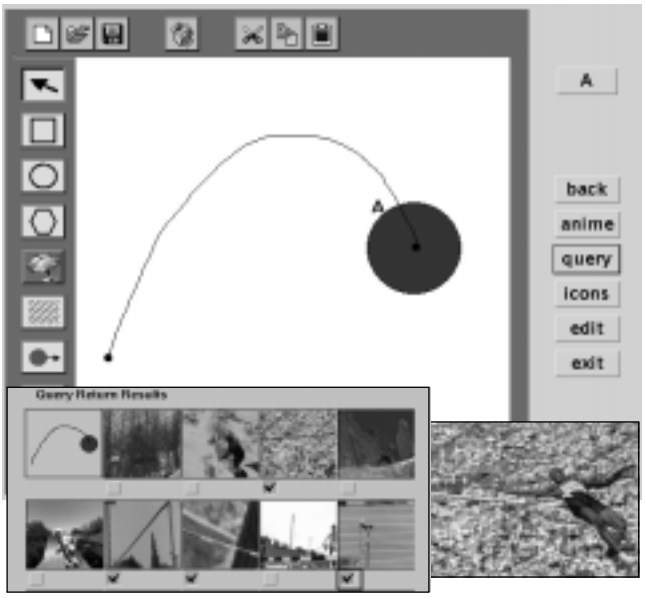

Figure 1. The figure shows an example query to retrieve high-jumpers. The marks on the second panel show the successful results. The third panel shows the highest ranked successful result.

2. The VideoQ System Overview

VideoQ is a Web based video search system, where the user queries the system using animated sketches. The system, which resides on the Web, incorporates a client-server architecture. The client (a java applet) is loaded up into a web browser where the user formulates (sketches) a query

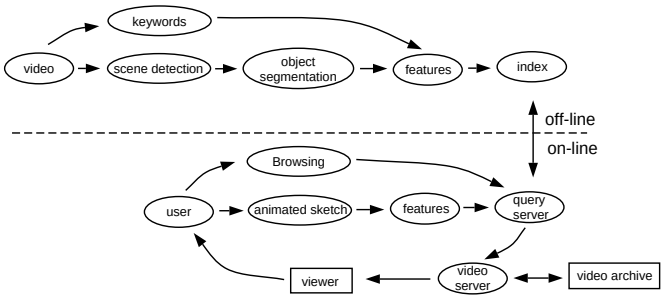

Figure 2. The VideoQ system where the queries are in the form of animated sketches. Both the animated sketch and the browsing modes support search with conjunction with keywords.

scene as a collection of objects with different attributes. Attributes include motion, spatio-temporal ordering, shape and the the more familiar attributes of color and texture. For the purposes of this paper, we define video objects to be regions homogeneous in the features that we use to index.

Scene cut detection and global motion stabilization are incorporated into the object segmentation process. Automatic object segmentation and tracking involves fusing multiple features including color, texture and motion.

After each object in the video shot has been segmented and tracked [Chang 1998a], we then compute the different features of the object and store them in our feature library. For each object we store the following features: Color, texture, shape, size and motion. Hence, the query server contains several feature databases, one for each of the individual features that the system indexes on. The source video shot database is stored as a compressed MPEG stream.

Once the user is done formulating the query, the client sends it over the network to the query server. There, the features of each object specified in the query are matched against the features of the objects in the database. Then, lists of candidate video shots are generated for each object specified in the query. The candidate lists for each object are then merged to form a single video shot list.

3. Feature Metrics and Querying

The nature of the metric, plays a key role in any image or video retrieval system. For it is not enough to be able to locate images or videos that are close under a metric, they must be perceptually close to the query. While we employ well accepted metrics for color, texture and shape, we have designed new metrics to exploit the spatio-temporal information in the video [Chang 1998a].

Using these feature space metrics and the composite distance function, we compute the composite distance of each object in the database with the each object in the query. The the query process for a single object is summarized in Figure 3. The detailed experimental results can be found in [Chang 1998a].

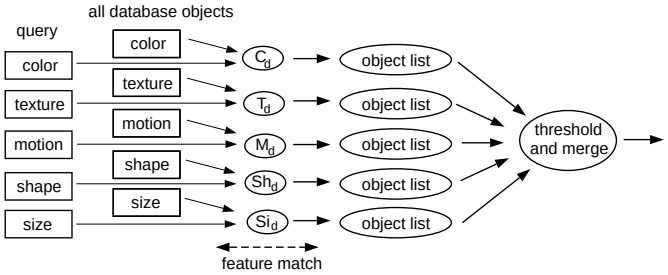

Figure 3. Generating the candidate video shot list for a single object query.

4. Conclusions

Video search in large archives is an emerging research area. Although integration of the diverse multimedia components is essential in achieving a fully functional system, we focus on exploiting visual cues in this paper. Using the visual paradigm, our experiments with VideoQ show considerable success in retrieving diverse video clips such as soccer players, high jumpers and skiers. Indexing video objects with motion attributes and developing good spatio-temporal metrics have been the key issues in this paradigm.

The other interesting contributions include developing a fully automated video analysis algorithm for object segmentation and feature extraction and the content-based visual matching of spatio-temporal attributes. Our current work includes region grouping, object classification and more accurate shape representation. Spatio-temporal serach tools are being extended to operate on higher level objects (e.g MPEG-4 objects). We are also attempting to bridge the gap between the semantics that the user wishes to query on and the low level features present in the database. This is the idea behind Semantic Visual Templates (SVT) [Chang 1998b].

References

[Chang 1998a] S.F. Chang, W. Chen, J. Meng, H. Sundaram, D. Zhong, *A Fully Automated Content Based Video Search Engine Supporting Spatio-Temporal Queries*, to appear in IEEE Trans. on Circuits and Systems for Video Technology, Sep. 1998.

[Chang 1998b] S.F. Chang, W. Chen, H. Sundaram, *Semantic Visual Templates: Linking Visual Features to Semantics*, to appear in ICIP '98, Workshop on Content Based Video Search and Retrieval, Chicago IL, Oct. 1998.

[Flickner 1995] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, 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, pp. 23-32, Sep. 1995.

[Hamrapur 1997] 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, 188-97, San Jose, Feb. 97.

[Pentland 1996] A. Pentland, R.W. Picard, and S. Sclaroff, *Photobook: Content-Based Manipulation of Image Databases*, International Journal of Computer Vision, Vol. 18, No. 3, pp. 233-254, 1996.

[Smith 1996] J.R. Smith, S.F. Chang, *VisualSEEk: A Fully Automated Content-Based Image Query System*, ACM Multimedia Conference, Boston, 87-98, Nov. 1996. http://www.ctr.columbia.edu/VisualSEEk