Fast and Robust Video Clip Search Using Index Structure

Ling-Yu Duan, Jun-Song Yuan, Qi Tian, Chang-Sheng Xu
Institute for Infocomm Research
21 Heng Mui Keng Terrace, Singapore 119613
{lingyu, jyuan, tian, xucs}@i2r.a-star.edu.sg

ABSTRACT
Content based retrieval of similar multimedia objects (e.g. images, text, and videos) is an important research issue in the field of multimedia database. In this demo, we present a fast and robust video clip searching system. This system consists of two major modules, namely, robust video representation and fast searching. Different from traditional key frame-based histogram methods, we employ the cumulative histogram to represent the ordinal features and color range features for a video segment. This representation provides a spatio-temporal description of the whole segment. Our experiment has shown it is effective for capturing the patterns of short video clips such as commercial, program lead in/out, flying logo in sports video, etc. In order to improve the performance in searching large video database, we introduce the index structure to deal with video search from the viewpoint of query processing (e.g. *K-NN query*, *Range query*, etc.) in high-dimensional spaces. Different query processing support different search tasks. In this demo, we employ the mrkd-tree index structure and the proposed video representation to fulfill fast and robust search of short video clips (i.e. news video lead-in/out, replay logo, commercial) in large video collections with the total length of 15 hours.

Categories and Subject Descriptors
H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval– *Search process, retrieval models.* I.4.7 [Image Processing and Computer Vision]: Feature Measurement– *Feature representation.*

General Terms
Algorithms, Experimentation.

Keywords
Video clip search, index structure, mrkd-tree, query processing.

1. INTRODUCTION
An important research issue in the field of multimedia databases is the content based retrieval of similar multimedia objects. Retrieval based on a video clip has attracted many research interests [1, 2, and 3]. A clip can be a sub-shot, a shot or a series of shots associated with a kind of coarse or clear semantic meaning. Applications include copy detection, video content monitoring, video structure analysis, etc.

Aiming at video clip search in a large video collection, a query clip is usually represented as a feature vector, which is expected to be a unique signature in a high-dimensional feature space. The obtained signature will be sequentially compared with those of a series of sliding windows in the target video stream, or used to do fast searching based on the previously built index structure. Then instance detection is completed based on the similarity values.

MM'04, October 10-16, 2004, New York, New York, USA.
ACM 1-58113-893-8/04/0010

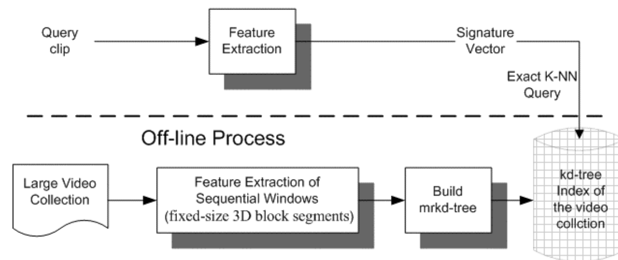

Figure 1. System Diagram

Video clip search concerns two challenging problems: *video representation* and *searching*, namely how to select appropriate features uniquely and robustly describing the video content, and how to accelerate the search process with the extracted features. Our goal in this demo is to present a set of robust spatio-temporal features and an index structure-based fast searching mechanism that has proved to be successful for short video clip (e.g. program lead in/out, commercials, and flying logo signifying replay scenes in sports video) searching in a large video collection with the total length of 15 hours.

From the database point of view, video search can be treated as a querying problem. That is, an essential purpose of video search is to find the contents that are the most similar to the query, namely the "neighbors" of the query point in the feature spaces. Different query strategies can be applied to different search purposes and applications [4]. For example, *K-NN Query* can retrieve nearby points efficiently and then is suitable for a content based retrieval. In contrast, *Range Query* is able to find all of individual points within a certain distance from a query and then it can be used for copy detection or content identification tasks. Such fast query strategies have ever been successfully applied to fast content-based image retrieval. Different from image database, however, video database contains large amounts of image sequences exhibiting a temporal order and a high redundancy. Thus indexing each frame like a still image is ineffective and practically impossible.

In this demo, we choose a series of fixed-length and overlapped video segments as the basic elements and build an index structure for those segments. Compared with previous key frames-based video representation, our approach avoids shot segmentation and key frame extraction. This is particularly useful for identifying those video clips containing ambiguous shot boundaries such as dynamic commercials and program lead-in/out clips. In order to robustly represent video segments, we have developed a set of spatio-temporal features incorporating both spatial and temporal information in addition to color range information. Our designed signature tends to capture the patterns of video segment globally rather than focusing on its sequential details like key frames-based representation.

In consideration of the successful story of mrkd-tree in terms of searching large and high-dimensional feature space, we employ it to construct an index structure for overlapped video segments. As mrkd-tree [5, 6] supports different queries (such as *point query*, *range query*, and *K-NN query*), it is able to support video copy detection, content identification and video retrieval respectively. There have been some index-based video searching schemes, e.g. VQ index, hash index. However, their purpose is an approximate *K-NN* query, where the effect of missed detection is probabilistically unavoidable. In contrast, our mrkd-tree based index emphasizes an exact *K-NN* query that is able to search out all the closest points of the given query, namely the *k* most similar video segments without any missed detection or false alarms.

2. OVERVIEW

Figure 1 illustrates our video similarity search based on mrkd-tree structure. An index structure of the video database can be built off-line before querying process. Compared with traditional video similarity search method, our approach has the following features:

- In contrast to sequential search scheme [1] (temporal pruning), *K-NN* query based on the mrkd-tree employs spatial pruning to accelerate searching. The prior query results are always the most similar contents. A user can decide whether to do further searching according to the available results. Exhaustive seeking on the whole database is thus avoided.
- Instead of using representative images to characterize video content, we employ spatio-temporal features and cumulative color histogram as a compact and robust signature to describe video content unlike a feature string. Complex training phase is unnecessary for extracting these features. Good performance has been proved in this demo.
- As the index of video database is built for those temporally overlapped video segments with a predefined length, this scheme is applicable to a sub-shot, a shot, and a series of shots. By using different query strategies (*point*, *K-NN*, and *range*), our method has provided a more general framework to fulfill different user search requirements.

More technical details can be found in [7] [8].

3. EXAMPLES OF SHORT VIDEO CLIPS

We illustrate several representative short video clips as follows.

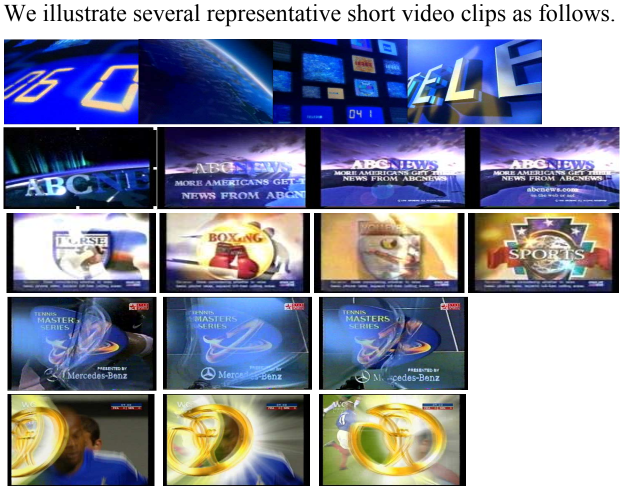

Figure 2. Representative frames of typical short video clips

4. PERFORMANCE

Below we briefly give the simulation results on 10.5 hours ABC TV news video. The 83 commercials appear in 209 instances in these 22 half-hour news programs; and the lead-out clip appears in total 11 instances. The instances searching performance of the given 84 clips in the 10.5 hour video collection is presented in Figure 3. It is obvious that better performance can be achieved by using the combined feature compared with considering only color feature or only ordinal feature [7] [8] respectively.

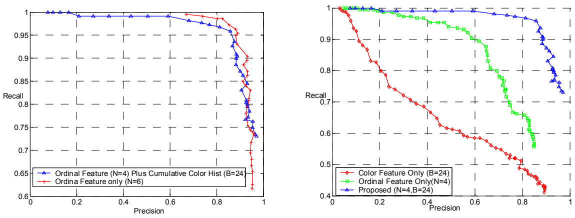

Figure 3. Performance comparisons using different features, the detection curves are generated by varying the searching range. More details can be found in [7] [8].

In this demonstration, we will show more results on the detection of flying logo inserted at replay scenes in broadcast sports video.

5. CONCLUSIONS

In this demo, we have presented a fast and robust video clip searching system using an index structure. The mrkd-tree index structure provides an efficient mechanism for examining only those feature points closest to the query, thereby greatly saving computation towards good matching. Unlike the key-frames based video matching, our matching method is based on a series of overlapped video segments with a predefined length. Thus it is applicable to a sub-shot, a shot or a sequence of shots. Since the research in high-dimensional index structure has been very active and fruitful over the past few years, there are many existing index structures (e.g. mrkd-tree, etc.) capable of supporting various types of querying. Such an index structure-based video searching scheme provides a more general framework to fulfill different user search requirements.

6. REFERENCES

[1] K. Kashino, *et al.*, "A Quick Search Method for Audio and Video Signals Based on Histogram Pruning," *IEEE Trans. on Multimedia* 5(3):348-357, 2003.

[2] A. K. Jain, *et al.*, "Query by video clip," *Multimedia System* 7:369-384, 1999.

[3] D. DeMenthon, *et al.*, "Video retrieval using spatio-temporal descriptors," In *Proc. of ACM MM'03*, pp. 508-517, 2003.

[4] Christian Bohm, *et al.*, "Searching in high-dimensional spaces – index structures for improving the performance of multimedia databases," *ACM Computing Surveys* 33(3): 322-373, 2001.

[5] A. Moore, "A tutorial on kd-trees," *University of Cambridge Computer Laboratory Technical Report No. 209*, 1991.

[6] K. Deng and A.Moore, "Multi-resolution instance based learning," In *Proc. of Int. Joint Conf. On Artificial Intelligence'95*, pp. 1233-1239, 1995.

[7] http://ma.i2r.a-star.edu.sg/Staff/~lingyu/FastVSearch.pdf

[8] http://ma.i2r.a-star.edu.sg/Staff/~lingyu/MrkdFastVSearch.pdf