BilVideo Video Database Management System

Özgür Ulusoy, Uğur Güdükbay, Mehmet Emin Dönderler, Ediz Şaykol and Cemil Alper

Department of Computer Engineering
Bilkent University
06800, Ankara, Turkey

Abstract

A prototype video database management system, which we call BilVideo, is presented. BilVideo provides an integrated support for queries on spatio-temporal, semantic and low-level features (color, shape, and texture) on video data. BilVideo does not target a specific application, and thus, it can be used to support any application with video data. An example application, news archives search system, is presented with some sample queries.

1. Introduction

We present a video database management system (BilVideo), which provides an integrated support for queries on spatio-temporal, semantic and low-level features (color, shape, and texture) on video data. The aspects of BilVideo distinguishing it from the others can be listed as follows:

- BilVideo handles spatio-temporal queries using a knowledge-base, which consists of a fact-base and a comprehensive set of rules implemented in Prolog, while queries on semantic and low-level features are handled by an object-relational database. A spatio-temporal query may contain any combination of directional, topological, 3D-relation, object-appearance, trajectory-projection and similarity-based object-trajectory conditions. The rules in the knowledge-base significantly reduce the number of facts representing the relations that need to be stored for spatio-temporal querying of video data whilst keeping the query response time reasonably short.
- A novel approach is proposed for the segmentation of video clips based on the spatial relationships between salient objects in video data. Video clips are segmented into shots whenever the current set of relations between video salient objects changes.
- The systems proposed so far in the literature associate video features with scenes that are defined to be the smallest logical units of video clips. However, our data model supports a finer granularity for query processing, which is independent of semantic segmentation of video clips: it allows users to retrieve any segment of a video clip, in addition to semantic video units, as a result of a query.
- BilVideo supports any application with spatio-temporal and semantic query requirements on video data; therefore, it is application-independent. However, it can easily be tailored according to specific requirements of such applications through the definition of external predicates supported by its query language without any loss in performance.

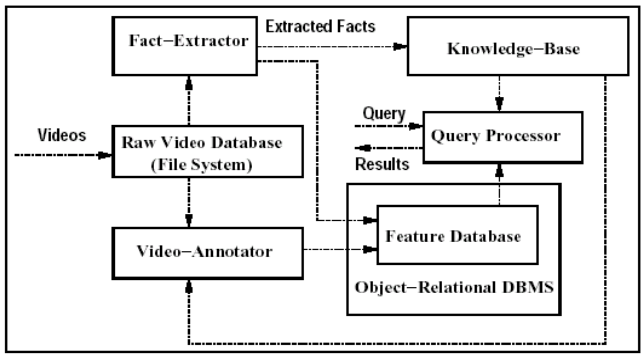

Figure 1: BilVideo System Architecture.

2. BilVideo System Architecture

In BilVideo, a visual query is formed by a collection of objects with some conditions, such as object trajectories with similarity measure, spatio-temporal ordering of objects, annotations and events. Object motion is specified as an arbitrary trajectory and annotations are used for keyword-based search. A text-based SQL-like query language is also available for the users. In the heart of the system lies the query processor (see Fig. 1). The feature database contains semantic properties of videos used for keyword, activity/event and category-based queries. These features are generated by *Video-Annotator*. The knowledge-base is used to support spatio-temporal queries and the facts-base is populated by *Fact-Extractor.*

**Proceedings of the 30th VLDB Conference,
Toronto, Canada, 2004**

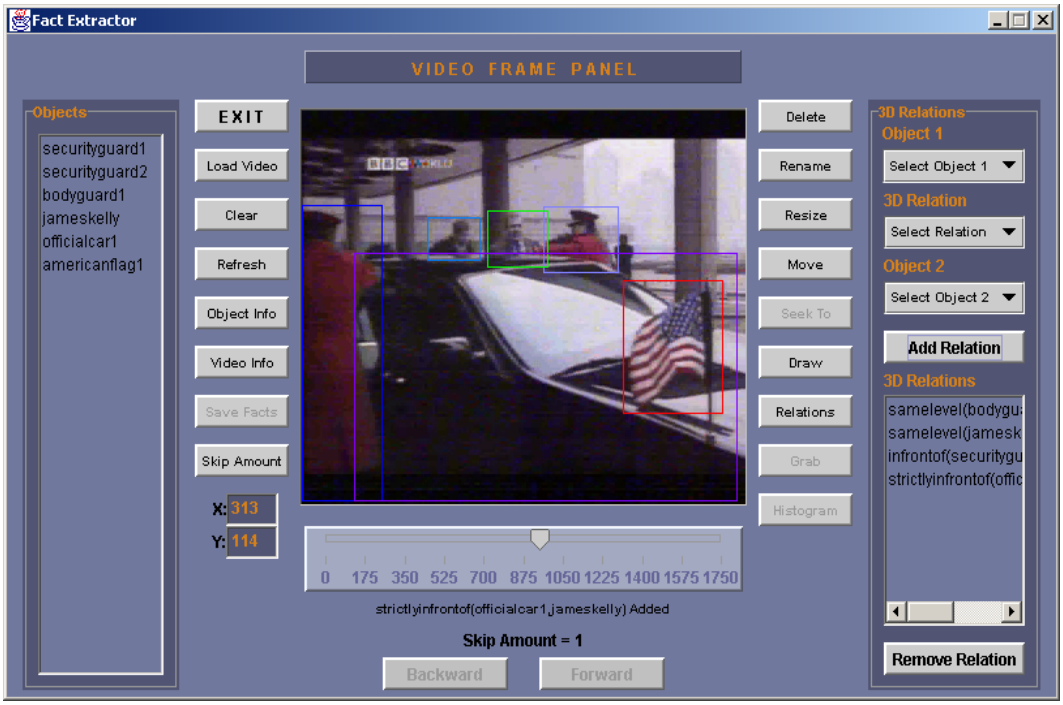

Figure 2: Fact-Extractor

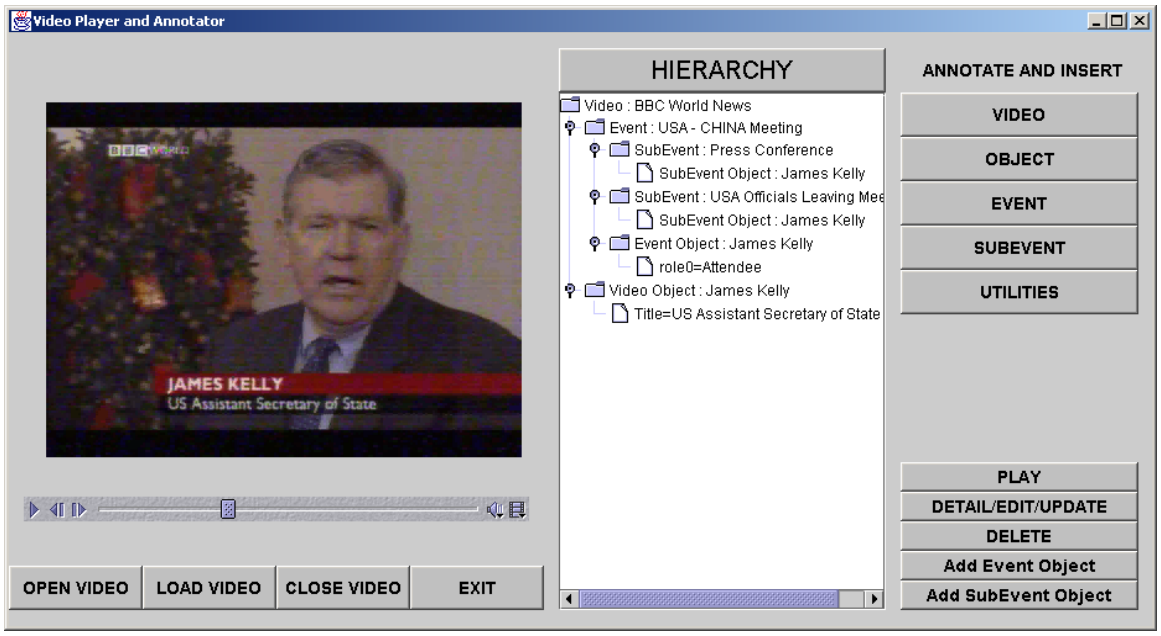

Figure 3: Video Annotator Tool.

2.1 Fact-Extractor

Fact Extractor computes a set of spatio-temporal relations between the salient objects in frames using the objects' minimum bounding rectangles (MBRs). The rules in the knowledge-base are used to eliminate the relations that can be derived using existing ones. Object trajectories and object-appearance relations are also extracted (Fig. 2).

2.2 Video Annotator

The *Video-Annotator* tool is used to extract semantic data from video clips (Fig. 3). Our semantic video hierarchy contains three levels: *video*, *sequence* and *scene*. *Videos* consist of *sequences* and *sequences* contain *scenes* that need not be consecutive in time. With this semantic data model, we plan to answer *video*, *event/activity* and *object*

queries. *Video* queries can be based on title, length, producer, production year, category etc. *Event/activity* queries can be used to retrieve videos by specifying events that occur at semantic layer *sequence* because events are associated with sequences. *Object* queries are used to retrieve videos by specifying semantic object features. As videos are annotated, video salient objects are also associated with annotations.

2.3 Web-Based User Interface

BilVideo can handle multiple requests over the Internet through a graphical query interface.

Spatial Query Specification: Spatial content of a keyframe is the relative positioning of its salient objects with respect to each other. This relative positioning consists of *directional*, *topological* and *3D-relations*. In the spatial query specification window (Fig. 4), objects are sketched by rectangles representing MBRs of the objects. Similar to the database population phase, the directional and topological relations between objects are extracted automatically in the query specification phase. Since it is not possible to extract 3D-relations from 2D-data, users are guided to select appropriate 3D-relations.

Trajectory Query Specification: Trajectory of a salient object is described as a path of vertices corresponding to the locations of the object in different video frames. Displacement values and directions between consecutive frames (vertices) are used in defining the *trajectory* fact of an object. In the trajectory query specification window, users can draw trajectories of objects (Fig. 5). Object-trajectory queries are similarity-based; therefore, users specify a similarity value between 0 and 100.

Semantic Query Specification: The semantic query specification interface (Fig. 6) currently supports three types of semantic queries, which are *event*, *object* and *metadata* queries. In an event query, user can specify three item types: the event itself, a sub-event that occurs in this event, and the objects appearing in this event or in a sub-event of it. Object queries are used for specifying the attributes of the objects. A semantic query for the BBC news video that we used is as follows:

Query 1: *"Retrieve all news videos produced in year 2003 that has a `USA - CHINA Meeting' event in which one of the attendees `James Kelly', the US assist. Secr. of state, is giving a press conference behind microphones."*

```
select video
from all
where meta(vtype:news and pyear:2003) and
    etype:`USA - CHINA Meeting' with
        (`James Kelly':role = attendee and
        setype:`Press Conference') and
    odata(`James Kelly' (Title:`US Assistant
    Secretary of State')) and
    behind(`James Kelly',Microphones);
```

Some of the data used for specifying the query are retrieved from the database to guide the user for entering a valid query. Such data include video name and attributes, event and subevent types, and object attributes.

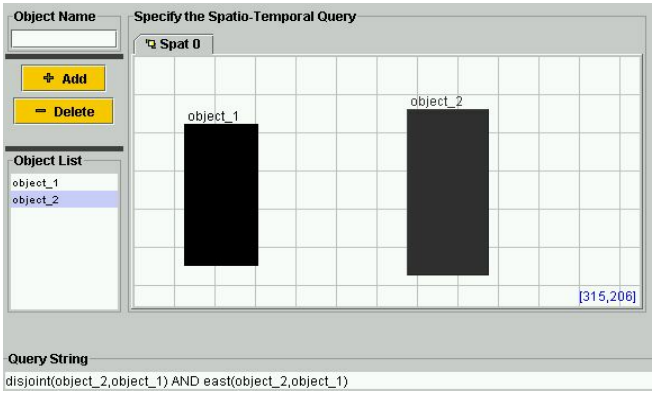

Figure 4: Spatial Query Formulation.

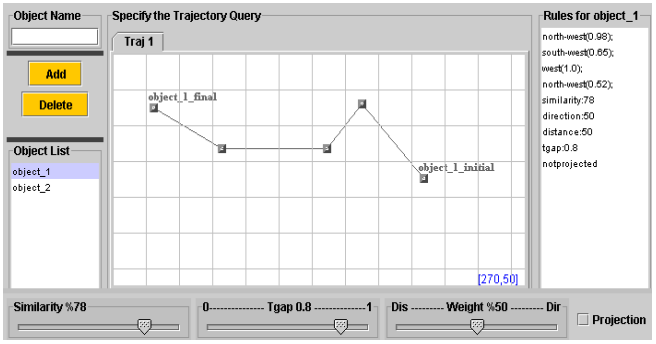

Figure 5: Trajectory Query Specification.

3. Example: News Archives Search System

In this section, we present an application, *news archives search system*, for BilVideo. A news archives search system contains video clips of news broadcasts and is used to retrieve specific news fragments based on some descriptions given as query conditions. The traditional approach for accomplishing this task is to provide some keywords that would describe the semantic content of the news fragments for retrieval. For this, a traditional database system would suffice since news fragments are indexed by some textual data. However, spatio-temporal relations between objects and object trajectories are not considered. BilVideo fills up this gap by providing support for spatio-temporal and semantic video queries. Users may also query news archives by some specific application-dependent external predicates that they define (see Query 2). For Query 2 and 3, we have used two sample news video clip fragments captured from BBC News and Kanal D (a Turkish TV channel).

Query 2: "*Retrieve the segments from the news clips, where James Kelly and his assistant appear together alone (no other object of interest is in the scene) and Kelly is to the right of his assistant.*"

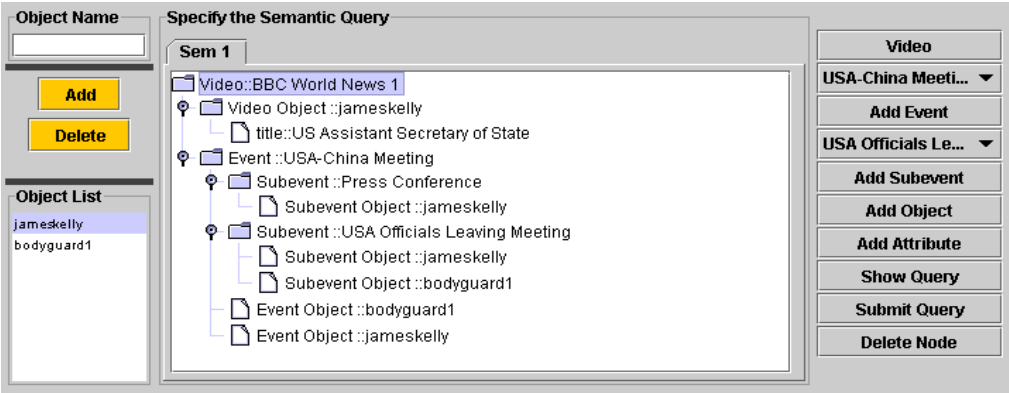

Figure 6: Semantic Query Specification.

```
select segment from vid
where appear_alone(james_kelly, assistant1)
       and right(james_kelly, assistant1);
```

In this query, *appear_alone/2* is an external (application-dependent) predicate. It is used to search for video keyframes, where the only objects appearing are those specified. The predicate *right*/2 is a directional predicate. *Vid* is a unique video identifier assigned to a video clip.

Query 3: "*Retrieve the segments from the sample news clip, where James Kelly moves west together with two security guards on his right and left, given a similarity threshold 0.8 and an allowed time gap 1 second.*"

```
select segment from vid
where (tr(james_kelly, [[west]]) sthreshold 0.8
       tgap 1) repeat
       and   right(james_kelly,   securityguard1)
       and left(james_kelly, securityguard2);
```

In this query, a similarity-based trajectory condition is given, along with directional and topological conditions. The keywords *tgap* and *repeat* are used to specify the trajectory conditions, which ensure that all the segments in the clip satisfying the given conditions are returned.

4. Conclusion

BilVideo does not target a specific application, and thus, it can be used to support any application, where vast amount of video data needs to be searched by spatio-temporal and semantic queries. Moreover, the query language of BilVideo provides a simple way to extend the system's query capabilities through *external predicates*, which makes the system application-independent but yet easily fine-tunable for specific needs of such applications without much effort and any loss in performance at all. This can be achieved by adding to the knowledge-base some application-dependent rules and/or facts that will be used for queries. In this article, we presented an example application of BilVideo; *news archives search system*. Among other applications that may be supported by BilVideo are sports event analysis systems (soccer, basketball, etc.), object movement tracking systems (medical, biological, astrophysical, etc.) and video archive search systems (movie retrieval, digital libraries, etc.).

5. Current Status Of BilVideo

We have already implemented the Query Processor and Web-based User Interface as well as the Fact-Extractor and Video-Annotator tools. We have also developed the knowledge-base of the system. Our work on semantic query execution and query-by-low-level-properties (color, shape, and texture) is ongoing. We are currently working on the integration of the Web-based query interface and the query processor. Tutorial video clips demonstrating the query interface and the tools of BilVideo are available at ***http://www.cs.bilkent.edu.tr/~oulusoy/BilVideo/***.

6. Acknowledgements

This work was supported by Scientific and Technical Research Council of Turkey (TUBITAK) under the grant number EEEAG-199E025, and is supported in part by Turkish State Planning Organization (DPT) under grant number 2004K120720, and European Commission 6th Framework Program MUSCLE Network of Excellence Project with grant number FP6-507752.

7. Related Publications

[1]. M.E. Dönderler, E. Şaykol, U. Arslan, Ö. Ulusoy, U. Güdükbay, BilVideo: Design and Implementation of a Video Database Management System, to appear in Multimedia Tools and Applications.

[2]. M.E. Dönderler, E. Şaykol, Ö. Ulusoy, U. Güdükbay, BilVideo: A Video Database Management System, IEEE Multimedia, Vol.10, No.1, pp. 66-70, 2003.

[3]. M.E. Dönderler, Ö. Ulusoy and U. Güdükbay, A rule-based video database system architecture, Info. Sci., Vol. 143, No. 1-4, pp. 13-45, 2002.

[4]. M.E. Dönderler, Ö. Ulusoy and U. Güdükbay, Rule-based spatio-temporal query processing for video databases, VLDB Journal, Vol.13, No.1, pp. 86-103, 2004.