DATA MODELLING AND HYBRID QUERY FOR VIDEO DATABASES

LILLY SURIANI AFFENDEY.

FSKTM 2006 7

DATA MODELLING AND HYBRID QUERY FOR VIDEO DATABASES

By

LILLY SURIANI AFFENDEY

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirement for the Degree of Doctor of Philosophy

October 2006

BISMILLAHIRAHMANIRAHIM

Alhamdulillah segala puji bagi Allah kerana dengan limpah rahmatNya dapat saya menyiapkan tesis ini.

Tesis ini didedikasi kepada suami, anak-anak dan keluarga tersayang.

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Doctor of Philosophy

DATA MODELLING AND HYBRID QUERY FOR VIDEO DATABASES

By

LILLY SURIANI AFFENDEY

October 2006

Chairman: Associate Professor Ali Mamat, PhD

Faculty: Computer Science and Information Technology

Video data management is important since the effective use of video in multimedia applications is often impeded by the difficulty in cataloging and managing video data. Major aspects of video data management include data modelling, indexing and querying. Modelling is concerned with representing the structural properties of video as well as its content. A video data model should be expressive enough to capture several characteristics inherent to video. Depending on the underlying data model, video can be indexed by text for describing semantics or by their low-level visual features such as colour. It is not reasonable to assume that all types of multimedia data can be described sufficiently with words alone. Although query by text annotations complements query by low-level features, query formulation in existing systems is still done separately. Existing systems do not support combination of these two types of queries since there are essential differences between querying multimedia data and traditional databases. These differences cause us to consider new types of queries.

The purpose of this research is to model video data that would allow users to formulate queries using hybrid query mechanism. In this research, we define a video data model that captures the hierarchical structure and contents of video. Based on this data model, we design and develop a Video Database System (VDBS). We compared query formulation using single types against a hybrid query type. Results of the hybrid query type are better than the single query types. We extend the Structured Query Language (SQL) to support video functions and design a visual query interface for supporting hybrid queries, which is a combination of exact and similarity-based queries.

Our research contributions include a video data model that captures the hierarchical structure of video (sequence, scene, shot and key frame), as well as high-level concepts (object, activity, event) and low-level visual features (colour, texture, shape and location). By introducing video functions, the extended SQL supports queries on video segments, semantic as well as low-level visual features. The hybrid query formulation has allowed the combination of query by text and query by example in a single query statement. We have designed a visual query interface that would facilitate the hybrid query formulation. In addition we have proposed a video database system architecture that includes shot detection, annotation and query formulation modules. Further works consider the implementation and integration of these modules with other attributes of video data such as spatio-temporal and object motion.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

PEMODELAN DATA DAN PERTANYAAN HIBRID UNTUK PANGKALAN DATA VIDEO

Oleh

LILLY SURIANI AFFENDEY

Oktober 2006

Pengerusi: Profesor Madya Ali Mamat, PhD

Fakulti: Sains Komputer dan Teknologi Maklumat

Pengurusan data video adalah penting kerana penggunaan video yang berkesan dalam aplikasi multimedia selalu terhalang oleh kesukaran mengkatalog dan mengurus data video. Aspek-aspek utama dalam pengurusan data video termasuk pemodelan data, pengindeksan dan pertanyaan. Pemodelan adalah berkenaan dengan mewakilkan sifat-sifat berstruktur dan juga kandungan video. Model data video mestilah mampu menunjukkan ciri-ciri khusus tentang video. Bergantung kepada model data yang menjadi dasar, video boleh diindeks secara teks untuk menerangkan semantik atau menggunakan ciri-ciri visual paras-rendah seperti warna. Sememangnya tidak munasabah mengandaikan bahawa semua jenis data multimedia boleh diterang secukupnya menggunakan perkataan semata-mata. Walaupun pertanyaan menggunakan anotasi teks melengkapkan pertanyaan melalui ciri-ciri paras-rendah, namun perumusan pertanyaan dalam sistem-sistem yang sedia ada masih dilakukan secara berasingan. Sistem-sistem yang sedia ada tidak menyokong gabungan kedua-dua jenis pertanyaan tersebut kerana terdapat perbezaan-perbezaan yang ketara di antara pertanyaan data

multimedia dan pangkalan data tradisional. Perbezaan-perbezaan ini menyebabkan kami mempertimbang jenis-jenis pertanyaan yang baru.

Tujuan penyelidikan ini adalah untuk memodelkan data video yang membolehkan pengguna merumus pertanyaan menggunakan mekanisma pertanyaan hibrid. Dalam penyelidikan ini, kami mentakrifkan model data video yang melambangkan struktur berhirarki dan kandungan video. Berdasarkan model data ini, kami mereka bentuk dan membangunkan Sistem Pangkalan Data Video. Kami membuat perbandingan di antara perumusan pertanyaan menggunakan jenis tunggal dengan jenis pertanyaan hibrid. Kami membuat lanjutan kepada *Structured Query Language (SQL)* untuk menyokong fungsi-fungsi video dan mereka bentuk antara muka pertanyaan visual bagi menyokong pertanyaan-pertanyaan hibrid, iaitu gabungan pertanyaan-pertanyaan tepat dan berdasarkan-persamaan.

Sumbangan penyelidikan kami termasuk model data video yang menyimpan struktur berhirarki video *(sequence, scene, shot* dan *key frame)*, di samping semantik (objek, aktiviti dan peristiwa) dan ciri-ciri paras-rendah (warna, tekstur, bentuk dan lokasi). Dengan memperkenalkan fungsi-fungsi video, lanjutan kepada *SQL* boleh menyokong pertanyaan ke atas segmen, semantik dan juga ciri-ciri paras rendah sesuatu video. Perumusan pertanyaan hibrid telah membolehkan pertanyaan menggunakan teks dan pertanyaan menggunakan contoh digabung dalam satu pernyataan pertanyaan. Kami telah mereka bentuk antara muka pertanyaan visual yang dapat membantu dalam perumusan pertanyaan hibrid. Di samping itu kami telah mencadangkan seni bina pangkalan data video yang mengandungi modul-modul pengesanan tangkapan gambar, anotasi dan perumusan pertanyaan. Kerja-kerja lanjutan mengkaji implementasi dan integrasi modul-modul tersebut dengan atribut-atribut video yang lain seperti *spatio-temporal* dan pergerakan objek.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor Associate Professor Dr. Hj. Ali bin Mamat for his constructive comments, suggestions, support and encouragement during this thesis work. I am also very much thankful to my co-supervisors, Associate Professor Dr. Hjh. Fatimah binti Ahmad and Associate Professor Dr. Hamidah binti Ibrahim for their guidance during my study.

I would like to take this opportunity to convey my sincere gratitude to members of the Faculty of Computer Science and Information Technology for supporting me to accomplish my research.

Finally, I am grateful to my family for their love, support and encouragement throughout my stressful journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Background

Multimedia data is a combination of video, audio, text, graphics, still images, and animation data. They are widely used for many applications such as computer-aided training, computer-aided learning, product demonstration, document presentation, electronic encyclopedias, advertisements, and broadcasting (Zhang, et.al., 1995, Lee, et.al., 1997, Lee, et.al., 1999, Donderler, et.al., 2003). Hence, there is a need for organizing and accessing them.

Recently, there has been much interest in databases that store multimedia data (Petkovic and Jonker, 2000, Donderler, et.al., 2003). Initially, multimedia data objects were treated as a single data item. In terms of data management, these data objects would be queried based on their associated attributes. The deficiencies of this approach for multimedia data objects quickly become apparent and researchers are now developing ways of retrieving multimedia data objects based on their content, mainly descriptive textual data (such as object, activity, event, etc.) and low-level features (such as colour, shape, etc.). (Decleir and Hacid, 1998, Jiang and Elmagarmid, 1998, Lee, et.al., 1999, Petkovic, 2000, Donderler, 2003).

Among the multimedia data, video is the most complex data object, since it incorporates image and audio in addition to its own attributes (Lee, 1997, Ponceleon, 1998). Other attributes of video data include temporal and object trajectory (Bimbo, et.al.,1995, Sawhney and Ayer, 1996, Liu, et.al., 1999). Video data management is important since the effective use of video in multimedia applications is often impeded by the difficulty of cataloging and managing video data (Chua and Ruan, 1995, Carrer, et.al., 1997, Donderler, et.al., 2002). Major challenges in designing a video database system includes data modeling, indexing, query formulation, query language and query processing (Aref, et.al., 2003).

The purpose of the data modelling process is to structure the data to reflect the relationships that exist between the various data items. The data modelling should facilitate the queries and operations that are to be performed on the data. The data model of a video should reflect the inherent hierarchical structure of sequences and frames within the object in order that functions such as retrieving the sequence can be performed. Recent works focused on modelling the video content (Decleir and Hacid, 1998, Petkovic & Jonker, 2000, Naphade, et.al., 2002, Donderler, 2002, Chen, et.al., 2003).

The indexing issue is directly related to the techniques for storing and retrieving video metadata. Metadata is any data description that "tell us something" about the video content. It can be in the form of textual or visual attributes and these can be used as index terms for video retrieval. Currently, there are two main approaches used in indexing and retrieving video data (Jiang, et.al., 1997, Dagtas, et.al., 1999, Tusch, et.al., 2000, Fan, et.al., 2004). The first approach is text annotations. It is often used to provide semantic content-based access. However, one of the major difficulties of this approach is the time consuming-effort required in manual image annotation. Another difficulty arises from perception subjectivity and imprecise annotation that may cause a mismatch during the retrieval process. However, automatic semantic interpretation of video data is not feasible given the state of the art of computer vision and machine intelligence.

To overcome the difficulties faced by the text-based approach, the second approach, the content-based image retrieval was proposed in the early 1990's (Rui, et.al., 1999, Aslandogan and Yu, 1999). It supports accesses based on the visual content of the image data such as colour, texture, and shape. These visual features are automatically extracted to form visual indices. This visual-based approach, mostly studied by the researchers in computer vision, supports accesses based on visual content of the image data (Bimbo, 1998, Natsev, et.al., 1999).

The final issues pertain to query formulation, query language and query processing. To formulate a database query the user must specify which data objects are to be retrieved, the database tables from which they are to be extracted and the predicate on which the retrieval is based. Traditional queries are expressed in a textual format using a query language, such as the industry standard query language SQL. Video database queries require additional functionality for content-based retrieval. Proposals for extensions to SQL (Amato, 1997), new text based query languages (Decleir, 1998, Donderler, 2003) and visual query languages (Hibino, 1996, Assfalg, et.al., 2000) have been put forward. Research by the Information Retrieval group has made used of partial or fuzzy textual matching (Jiang, 1998, Bimbo, 1998). Meanwhile the database community has used exact matching as in normal textual query (Carrer, 1997). Another research community, the Computer Vision has used similarity-based matching which is meant for content-based image retrieval (Natsev, 1999, Atnafu and Brunie, 2001). Since text queries complement visual queries, the necessity of using combined query system becomes apparent.

1.2 Problem Statement

From the database point of view, a powerful video model will enable a good basis for content-based search and retrieval of video data (Petkovic and Jonker, 2000). It is recognized by the database research community that video data requires a new data model that is different from the traditional data model. While the traditional data model deals only with data structure, the video data model has to include not only the representation of video structure but also elements that represent the content of video data. Thus, an expressive video data model is needed to capture several characteristics inherent to video. Given the importance of different video representations, which is not reflected in the state of the art video retrieval systems, our goal is to identify a video data model that combines low and high level representation of video content and support for content-based video retrieval. In other words, to enable the semantic content provided by manual annotation complement the query using visual features such as colour, texture and shape.

With the rapid growth of video data following the progression of the digital television technology and the Internet, problems are encountered with the respect to the retrieval of the audio-visual data. It is almost impossible to use free browsing due to the huge amount of data. For a user who wishes to find a specific segment of a particular video it would be a tedious and time-consuming process. Still, the retrieval process can rely on textual annotation of video data (Oomoto and Tanaka, 1993, Chua and Ruan, 1995, Hjelsvold, 1996, Jiang, 1998, Fan, 2004). Video data contains bibliographic information such as title, descriptive content such as events, as well as low-level features such as colour. Whilst bibliographic data is easily obtainable, time-consuming textual annotation is still required to provide semantic content that cannot be automatically extracted by visual analysis of video data. Furthermore, the text associated with the video segments is often vague and incomplete due to subjective human perception of the video content.

The limitation of the annotation-based approach has resulted in a demand for new techniques that can manipulate other attributes of video data such as the visual features. Much research has been done in the area of indexing and accessing video based on its visual features, such as colour, shape, motion, etc. (Ardizzone and Carsia, 1997, Ponceleon, et.al., 1998, Lim, 2000, Assfalg, et.al., 2000). However, applications under this category tend to be domain dependent, and do not cater for all types of video. Furthermore, querying by visual features alone is not sufficient to express semantic content.

When addressing the problem of video query, the query formulation is one difficult part of the problem. Typically querying systems should be organized so as to cater for all possible users' needs. Each type of querying should concentrate on representing all search characteristics. However, combining query types is not so trivial since it involves mixing parameters that may not be coherent with one another. One common approach has been to deal with each type of query separately (Kuo and Chen, 2000, Naphade, et.al., 2002). This however defeats the advantage of being able to use logical connectives such as AND, OR, NOT, on the desired characteristics. One way to combine the querying system is to normalize the influence of each component and to ask the user to provide weighs for each component of the query (Fagin and Wimmers, 2000). Therefore, query expression must be enhanced to allow the combination of query types.

Although much has been said about the possibility of integrating exact and similarity-based queries (Bimbo, 1998, Donderler, 2002), to the best of our knowledge none of the literature has perform a comparison between the these two types of queries. We anticipate that hybrid query formulation could present a better result as opposed to queries using a single type. Furthermore, in addition to the basic query formulation, users of a content-based video database system should be allowed to further interact with the search results, for example to play a particular shot, scene, sequence or the whole video.

1.3 Objectives of the Study

The objectives of this research are to model video data and to provide a query mechanism for video databases, which allows a query to be expressed in combination of text and visual attributes (content-based) in a single mode. Furthermore, it is to show that hybrid query mechanism can give better results than query formulation using a single type.

1.4 Research Methodology

To address the issues regarding video modelling we survey existing video modelling approaches, and content-based video retrieval systems and analyse their advantages and drawbacks. Next, we propose an approach that overcome the identified shortcomings and develop a modelling framework. The framework is developed to facilitate validation of our ideas regarding video modelling and to support the integration of low-level and high-level representation of video content. An additional goal is to provide the basis for the system that can be used to validate the use of different attributes for querying video content.

To support the proposed video data model and hybrid query mechanism, we designed and developed a prototype Video Database System (VDBS). Our data consists of more than 30 minutes of video clips that had to be preprocessed. To populate our database, we performed video shot detection and then video annotations. Some video processing and feature extraction techniques are integrated within the prototype to support the content-based retrieval of video data. We use VDBS to experimentally compare the accuracy of the retrieval when it uses a single type of attribute to formulate the query, with its performance when hybrid query type is used. Furthermore, we extended the Structured Query Language (SQL) with video functions to support query result presentation. This is to facilitate video play back in the media player.