A Semantic Content Retrieval and Browsing System Based on Associative Relation in Video Databases

Kyoung-Soo Bok
Department of Computer Science
Korea Advanced Institute of Science and Technology, Daejeon, Korea

Jae-Soo Yoo*
Department of Computer and Communication Engineering
Chungbuk National University, Cheongju, Korea

ABSTRACT

In this paper, we propose new semantic contents modeling using individual features, associative relations and visual features for efficiently supporting browsing and retrieval of video semantic contents. And we implement and design a browsing and retrieval system based on the semantic contents modeling. The browsing system supports annotation based information, keyframe based visual information, associative relations, and text based semantic information using a tree based browsing technique. The retrieval system supports text based retrieval, visual feature and associative relations according to the retrieval types of semantic contents.

***Keywords**: Video Data, Semantic Content, Associative Relation, Browsing System, Retrieval System.*

1. INTRODUCTION

In the past couple of decades, it has been increased to store and retrieve digital multimedia data such as audio, graphic, animation and video with the development of computer hardware system and computer software system[2]. Recently, it has been studying video databases that efficiently store and retrieve a large amount of video data. And various systems have been developed for retrieving and browsing video data such as OVID[1], JACOB[5], VideoStar[4], VideoQ[6] and so on.

M. H. Lee suggested the hybrid video information system called HVIS, which provides feature based queries and annotation based queries based on THOMM(Three layered Object-oriented Metadata Model)[16]. The THOMM consists of a raw-data layer, a metadata layer, and a semantic layer. H. Lee proposed a structural approach for the development of user interfaces for the video browsing system, a web based system for recording, browsing and playback of TV programmes. J. Chen proposed a paradigm for video database called ViBE, which introduces a variety of algorithms and techniques for processing, presenting, and managing video. M. Guillemot suggested an interactive content based browser allowing fast, non linear and hierarchical navigation for video over Internet[14].

To efficiently process content-based retrieval of video data, the new modeling of video data is required[7, 8]. A video data modeling should represent structural contents and semantic contents. And these contents need to be represented by visual features. There have been a number of the existing modeling schemes which can be grouped into two categories[11]. The first approach is a structural modeling to represent unstructured video data into a logical unit. An early proposal was a structural modeling to divide a video data into segments and then to describe every segment. In this approach, the raw video data is segmented into a sequence of shots by using automatic shot boundary detection. The structural units are created by combining complex units. Most of existing structural modelings have group-scene-shot-keyframe. But the structural modeling efficiently does not represent semantic contents such as objects and events. Therefore, a content modeling has been proposed to represent semantic contents.

The second approach is a content modeling to represent visual and semantic contents. The visual content is characterized by visual features such as colors, shapes, textures and so on. The semantic content contains high-level features such as objects and events. These contents can be represented

**Corresponding author. E-mail: yjs@chungbuk.ac.kr*
Manuscript received Oct. 14, 2005 ; accepted Mar. 01, 2006

This work was supported the Program for the Training of Graduate Students in Regional Innovation which was conducted by the Ministry of Commerce, Industry and Energy of the Korean Government and grant No.R01-2003-000-10627-0 from the Basic Research Program of the Korea Science & Engineering Foundation.

by many different visual presentations such as color, shape and so on.

Most of content modelings represent these semantic contents within structural units and do not represent associative relations between semantic contents as time goes by. And the existing systems efficiently don't provide visual abstraction and relations of semantic contents because of the absence of semantic modeling to represent them.

In this paper, we propose the semantic contents modeling called ARSM(Associative Relation based Semantic Modeling) to efficiently browse and retrieve semantic contents. The ARSM represents semantic contents such as objects, events and background, and associative relations among them. And the proposed ARSM represents semantic contents independent of structural contents in order to solve the problem of the existing modeling for browsing and retrieval of semantic contents. We implement the browsing and retrieval system of semantic video contents based on ARSM. The proposed browsing displays the properties and associative relations of semantic contents according to their types. The retrieval system supports searches based on properties of semantic contents, similarity searches based on visual feature extracted from keyframe and searches based on associative relations.

The rest of this paper is organized as follows. In section 2, we discuss the related works. In section 3, we propose a new semantic modeling called ARSM. In section 4, we discuss the procedure of semantic content construction. We explain browsing and retrieval systems in section 5 and section 6 concludes the paper.

2. RELATED WORKS

To retrieve and browse the semantic contents of video data, the modeling representing semantic contents of video data is required. In [11], M. Petkovic proposed a layered video data modeling in order to overcome the problem of mapping from features to high level concepts. The video data modeling has the four layers such as the raw data, the feature, the object and the event. The object layer consists of entities which can be assigned to one or more regions. An object is defined as a collection of video regions. These objects should be semantically consistent, representing one real world objects and subject of interest to users or applications. The event layer is the highest layer that consists of events, which describe object interactions in spatio-temporal manner. Predefined event types can together be a part of the compound event type description. [9, 10, 15, 17] proposed a video modeling that supports associative relations of semantic contents. A. Ekin proposed an integrated semantic-syntactic video model to include all of elements within a single framework to enable structural video search and browsing[17]. The model includes semantic elements such as object sand events and the relation between them. And C. Yong proposed a semantic associative model based on semantic networks [18].

M. H. Yoon suggested the IHVRS which provides the semantic retrieval and similarity retrieval[16]. The IHVRS composed of raw-data layer for physical video stream of video, a metadata layer for a semantic retrieval and a semantic layer for a query reformulation. The semantic layer represents the relationship between the concept layer andobject feature layer and can reform the query provided that query result doesn't exist. A similarity search uses the metadata layer to retrieve the most similar scene and feature based retrieval is performed using object-feature layer. F. Li designed and implemented a prototype software video browsing application that provides a wide array of features enabled by digital technologies[12]. The prototype provided rich indices for navigation, speedup playback feature, the ability to make personal annotations that are anchored to the video timeline, and other advanced browsing controls. To allow the user to rapidly view a video data, F. Arman proposed content based browsing system which forms an abstraction to represent each shot of the sequence by using a representative frame or an Rframe[3]. This system includes management techniques to allow the user to easily navigate the Rframes.

3. SEMANTIC CONTENTS MODELING

In this section, we propose a new semantic modeling called ARSM which efficiently represents contents about objects, events and backgrounds for contents based browsing and retrievals. The semantic contents layer represents semantic contents and associative relations among each other. Video data contains the contents of various types. Content layer represents structural and semantic contents contained in the video data. The structural contents represent a logical hierarchy such as such as shots, scenes and sequences with partitioning raw data but the semantic contents represents semantic data such as objects, events and backgrounds and associative relations among each other.

When semantic contents are dependent on structural contents, they do not represent their durations and associative relations among them. Therefore, semantic contents are independent of structural contents to efficiently represent duration of semantic contents and associative relations among them. Figure 1 shows relations between structural contents and semantic contents. In Figure 1, ST_1, ST_2 and ST_3 are structural contents and SE_1, SE_2, SE_3 and SE_4 are semantic contents. SE_1 and SE_2 are fully included within ST_3 and ST_2. But SE_3 and SE_4 are partially included within ST_1 and ST_2 and within ST_1, ST_2 and ST_3 respectively. Therefore, when the semantic contents such as SE_3 and SE_4 are modeled, their durations and the relations among them not are exactly represented.

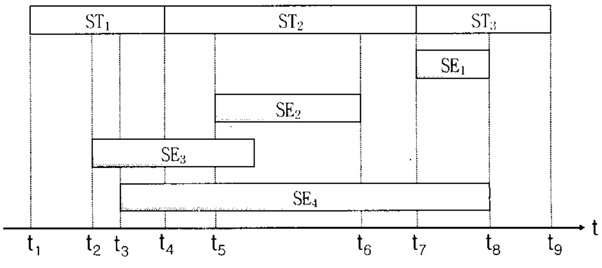


Fig. 1. Representing semantic contents

If some semantic content appears through a discontinuous time line and are allocated to the same identifier, it is determined that the semantic content exists while the semantic content does not exist. Therefore, the semantic contents that appear several times through a discontinuous time line are given to different identifiers of each other.

3.1 Raw Data Layer

Raw data layer represents a video data itself stored in the physical storage. Video data stored in the BLOB space consists of a sequence of frames. Raw data layer does not represent any semantic contents, because the raw data is an unstructured data. Therefore, raw data layer represents physical features and production features. The video data on raw data layer is described as follows :

<RID, NAME, BLOB, PHY, PRO>

, where RID is an identifier of video data, NAME is a name of video data, BLOB is a identifier of a BLOB space storing video data and PHY represents physical features of video data such as duration, frame rate per second, formant and so on. PRO represents the production features of video data such as actor, production, director and so on.

3.2 Semantic Contents Layer

The semantic contents layer represents semantic contents on objects, events and backgrounds. An event is appeared in video with relation to many objects. And one or more objects and events are contained in a background. Therefore, semantic contents not only consist of semantic contents itself but also associative relations among them. An object represents a certain thing and person that come out for continuous frames. In video data, an object carries out a certain action and changes spatial positions as time goes by. An action carried out by an object is related to a certain event. In semantic contents layer, an object is described as follows :

<OID, NA, D, AL, AT>

, where OID is an identifier of an object, NA is the name of an object, D is the duration of an object, AL is an action list that an objectis executed and AT is additional attributes by users or administrators. If an object appears several times within discontinuous time line, other identifiers are allocated to objects of the same name. Therefore, an object of the same name may appear several times. An object executes a certain action as time goes by. AL represents the action of an object. If an object executes actions n times for its duration, AL is described as follows :

$< (AL_1, D_1), (AL_2, D_2), ..., (AL_n, D_n) >$

, where AL_i is the name of an action executed by an object for D_i and D_i is the duration presented by means of a start frame and an end frame of the object.

An event represents special conditions appeared by actions of several objects. In video data, objects execute several actions. And an event is related to these actions. In semantic contents layer, therefore an event not only represents itself but also represents associative relations of objects related to a certain action. An event is described as follows :

<EID, NA, D, OL, AT >

, where EID is an identifier of an event, NA is the name of an event. D is the duration of an event, OL is the object list related to an event and AT represents additional attributes by users or administrators. If an event appears several times within discontinuous time line, other identifiers are allocated to events of the same name. Therefore, an event of the same name may appear several times. The OL represents objects not only related to an event but also represents actions executedby objects. If an event is related to n objects, OL is described as follows :

$< (OID_1, AL_1, D_1), (OID_2, AL_2, D_2), ..., (OID_n, AL_n, D_n) >$

, where OID_i is an identifier of an object, $AL_i$is an action executed by the object whose identifier is OID_i and D_i is the duration of an action A_i.

A background represents background information in which objects and events come out. Because a background is related to n objects and m events, BG is described as follows :

< BID, NA, D, OL, EL, AT >

, where BID is an identifier of a background, NA is the name of a background, D is the duration of a background, OL is the object list related to a background, EL is the event list related to a background and AT represents additional attributes by users or administrators.

3.3 Keyframe Layer

Keyframe layer represents visual features and spatial features extracted from semantic contents. Visual features areextracted according to a certain time interval or changes of visual features. Spatial features are spatial positions of an object in the keyframe. We extract one or more keyframes from the semantic contents in order to browse visual summaries and to support similarity searches based on visual features to users. Keyframe layer is described as follows :

< KID, VF, OBL, SID, AT>

, where KID is an identifier of a keyframe extracted from shots and semantic contents, VF is a set of visual features of a keyframe such as color, texture and shape, OBL is the object list contained in the keyframe and SID is an identifier of an semantic contents. AT represents additional attributes by users or administrators. OBL that represents object features in the keyframe have to describe spatial positions and colors of objects. OBL is described as follows :

$< (OID_1, COLOR_1, MBR_1), (OID_2, COLOR_2, MBR_2), ..., (OID_n, COLOR_n, MBR_n) >$

, where OID_i is the identifier of an object, $COLOR_i$ is the color feature of an object and MBR_i is the spatial position of an object. MBR of objects in the keyframe is represented by (X_1, X_2, Y_1, Y_2), where X_1 and X_2 are the start point and end point of X-axis and Y_1 and Y_2 are the start point and end point of Y-axis. Figure 2 shows the spatial position of an object in keyframe.

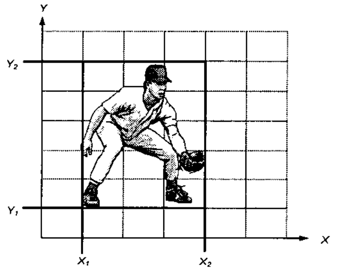


Fig. 2. Spatial position of an object in keyframe

4. DATABASE DESIGN AND CONSTRUCTION

4.1 Semantic Contents Construction

In this subsection, we describe how semantic contents are constructed based on the proposed Modeling. Figure 3 shows how semantic contents are constructed based on the proposed modeling. To construct semantic contents, objects, events and backgrounds are created by a semantic extraction method from raw data. When each semantic contents is constructed, one or more keyframes are extracted from them. Keyframes support visual summaries and similarity searches to user. And to represent spatial and visual features of objects in the keyframe, the object is extracted by object feature extraction from each keyframe.

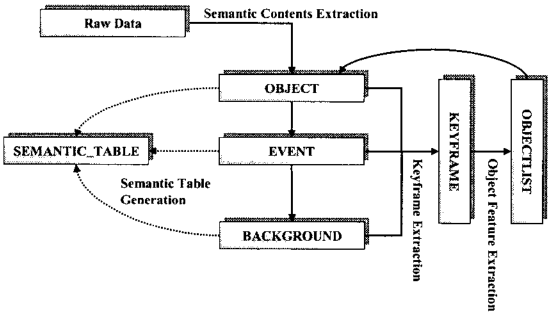


Fig. 3. Semantic contents construction

4.2 Database Design

To retrieve and browse the semantic contents of video data, we design the database that supports efficiently semantic contents modeling. Figure 4 shows database based on semantic content modeling. In figure 4, RAW_DATA table is video data represented in the raw data layer. EVENT, OBJECT and BG tables maintain semantic contents on event, object and background respectively. And EORLT, BORLT and BERLT tables maintain associative relations among semantic contents. EORLT table represents associative relations between events and objects, BORLT table represents associative relations between backgrounds and objects, and BERLT table represents associative relations between backgrounds and events. ACTION and TRA tables represent changes of objects in terms of their behaviors and trajectories as time goes. SEMANTIC_TABLE represents abstract features of semantic contents such as events, objects and backgrounds, which are name, occurrence number and so on. Through the abstract features, semantic table represents relations among raw data and semantic contents such as event, object and background. KEYFRAME table maintains information on the extracted keyframes. Because the keyframe extracted from semantic contents may contain several objects, OBL table represents features of objects contained in the keyframe.

5. BROWSING AND RETRIEVAL SYSTEM

In this section, we design and implement browsing and retrieval system based on semantic contents modeling. The system is implemented using Microsoft Visual C++ 6.0 based on Window 2000 Server and MS SQL Server 2000 using Directshow 8.1.

5. 1 Semantic Contents Retrieval

We implement a video retrieval system that supports various query types according to the proposed video modeling. Our retrieval system enables hybrid searches as well as text based searches and similarity searches based on visual features. The latter is classified into query-by-example and query-by-sketch.

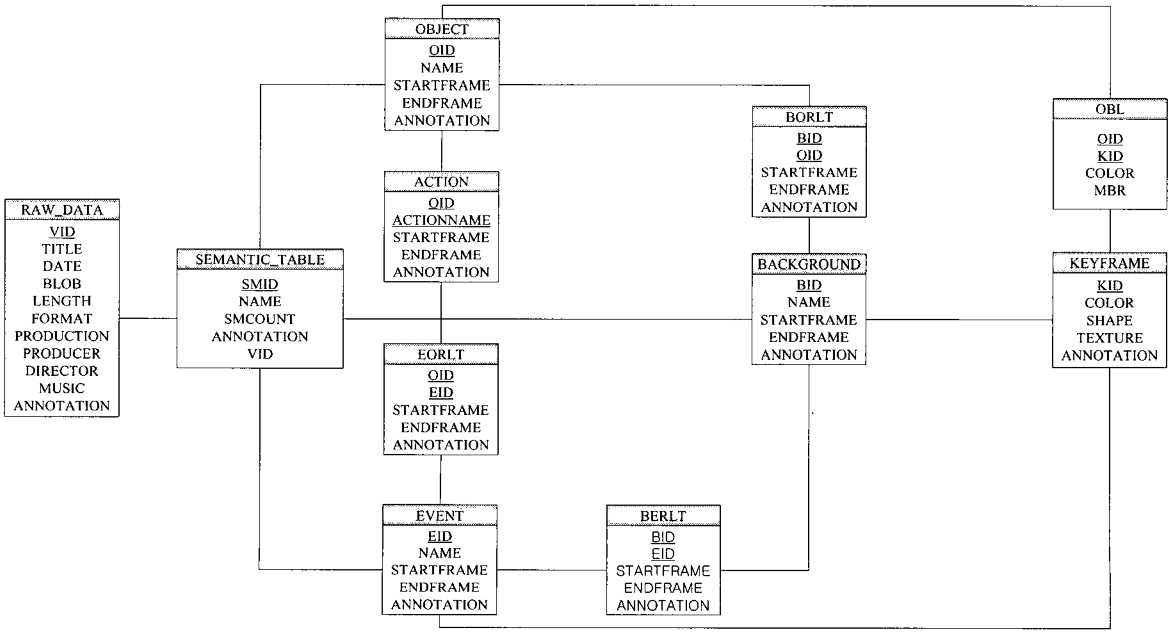


Fig. 4. Database design

Figure 5 shows the user interface of semantic contents retrieval. The user interface of semantic contents retrieval consists of two parts such as input for search conditions and output for search results. The user interface includes associative relation-based searches and text-based searches. In left side, input parts consist of color information and spatial relationships among semantic contents extracted in the keyframe. It enables to describe the weight and the degree of similarity for the similarity searches using color information of the object. The user interface includes the part to select an image in the sample images, which enables to perform the similarity searches based on the visual features, their weight and the degree of similarity for the semantic contents. The right part of the semantic search interface shows the search results that include the visual and text information identical to the structure search. The visual summary information is the keyframe representing the semantic contents.

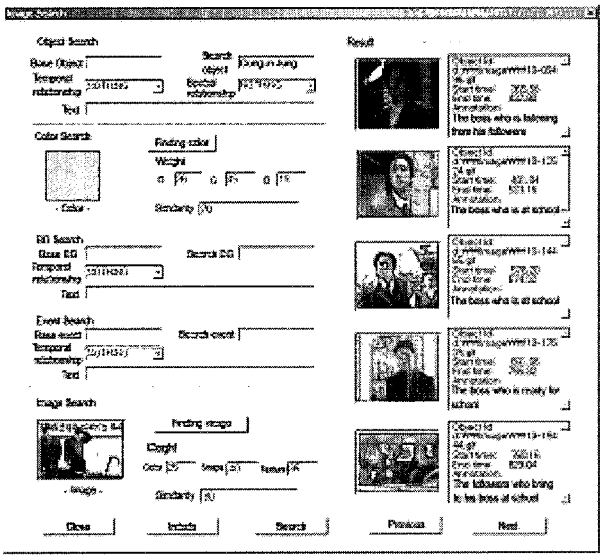
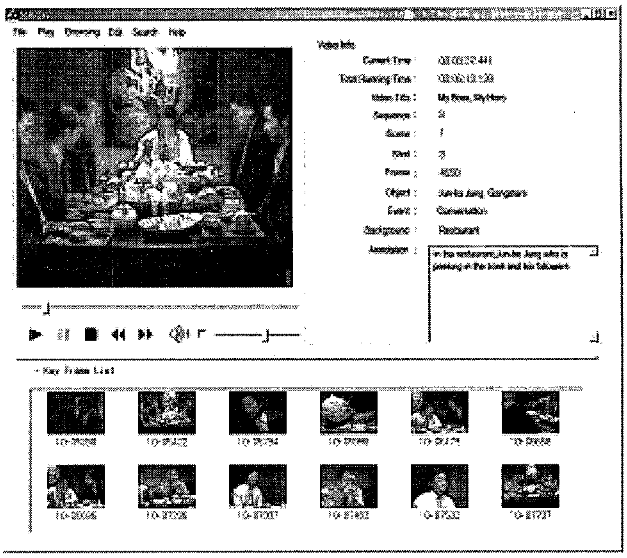


Fig. 5. Semantic contents retrieval

If we choose an image in the search results, the video related to the image is played. The search results are revealed on the right part of the user interface. Figure 6 shows the user interface that selects the keyframe image and then plays its video. The play interface consists of a screen, video properties, control buttons and keyframe lists. The screen shows the video that is being played. And the running time is determined by time information of search results. The video properties show the semantic details about the screen. The control buttons consist of the buttons to control the screen such as play, pause, fast-forward, stop and so on. Finally the keyframe lists show the images of the semantic content to be played.

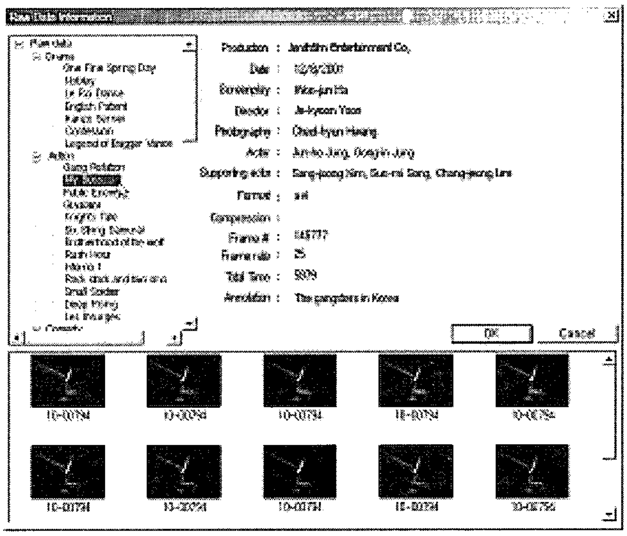


Current Time : 00:00:29:441
Total Running Time : 02:00:13:20
Video Title : My Boss, My Hero
Sequence : 3
Scene : 1
Shot : 3
Frame : 4633
Object : Jun-ho Jung, Gangsters
Event : Conversation
Background : Restaurant
Annotation : In the restaurant, Jun-ho Jung who is [illegible] the boss and his followers.

Fig. 6. playing video data

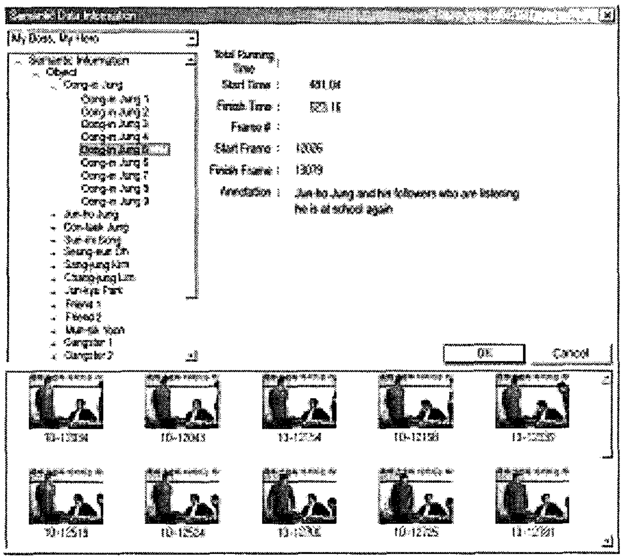
5.2 Semantic Contents Browsing

According to semantic content modeling, users can browse the information of semantic contents and raw data. The proposed browser supports text information based on annotation and a visual information based on keyframe. Figure 7 and figure 8 shows the user interface for raw data browser and semantic content browser. The browser consists of three parts such as a contents list, text information, and visual information. Users do not always want to browse whole contents. In the left side, the user can select semantic contents which user wants to browse in content list. In figure 7, content list is group by six categories such as drama, action, and so on. In figure 8, content list is group by semantic content type. Each of semantic contents is grouped by its name independent of structural contents to efficiently represent duration of semantic contents and associative relations among them. We provide users with visual summary information by keyframe list extracted by contents and a text based information by annotation and additional properties, when users select the content list.



Production : Jindafilm Entertainment Co,
Date : 12/8/2001
Screenplay : Won-jun Ha
Director : Je-hyeon Yoon
Photography : Chul-hyun Hwang
Actor : Jun-ho Jung, Gang-jin Jung
Supporting actor : Sang-joong Kim, Sue-mi Song, Chang-jeong Lim
Format : avi
Compression :
Frame # : 145777
Frame rate : 25
Total Time : 5839
Annotation : The gangsters in Korea

Fig. 7. Raw data browser



Total Running Time :
Start Time : 481.04
Finish Time : 525.18
Frame # :
Start Frame : 12026
Finish Frame : 13079
Annotation : Jun-ho Jung and his followers who are listening he is at school again

Fig. 8. Semantic contents browser

6. CONCLUSION

We have proposed a semantic video modeling that supports the efficient browsing and retrieval of the semantic contents. The proposed semantic modeling represents semantic contents and represents associative relations among them. We have also implemented browsing and retrieval system based on the proposed semantic contents modeling. Our browsing system supports text information and visual summarized information. And our retrieval system supports text based searches, visual feature based similarity searches, and associative relations based searches.

REFERENCE

[1] E. Oomoto and K. Tanaka, "OVID: Design and Implementation of a Video-Object Database System", IEEE Transactions on Knowledge and Data Engineering, Vol. 5, No. 4, pp.629-643, 1993.

[2] W. I. Grosky, "Multimedia Information Systems", IEEE Multimedia , Vol. 1, No. 1, pp.12-24, 1994.

[3] F. Arman, R. Depommier, A. Hsu, and M. Y. Chiu, "Content-Based Browsing of Video Sequences", Proc. the Second ACM International Conference on Multimedia, pp.97-103, 1994

[4] R. Hjelsvold, "VideoSTAR - A Database for Video Information Sharing", Ph.D. Thesis, Norwegian Institute of Technology, 1995.

[5] M. La Cascia, and E. Ardizzone, "JACOB: Just a content-based query system for video databases", Pro. IEEE International Conference on Acoustics, Speech, and Signal Processing, 1996.

[6] S. F. Chang, W. Chen, H. J. Meng, H. Sundaram, and D. Zhong, "VideoQ: An Automatic Content-Based Video Search System Using Visual Cues", Pro. ACM Multimedia 97 Conference, Seattle, WA, 1997.

[7] M. Petkovic and W. Jonker, "An Overview of Data Models and Query Languages for Content-based Video Retrieval", Pro. International Conference on Advances in Infrastructure for Electronic Business, Science, and Education on the Internet, 2000.

[8] U. Srinivasan and G. Riessen, "A Video Data Model for Content-Based Search", Pro. the Eighth International Workshop on Database and Expert Systems, pp.178-185, 1997.

[9] J. L. Koh, C. S. Lee, and A. L. P. Chen, "Semantic video model for content-based retrieval", Pro. the International Conference on Multimedia Computing and Systems, pp.472-478, 1999.

[10] A. Safadi, L.A.E and J. R. Getta, "Semantic modeling for video content-based retrieval systems", 23rd Australasian on Computer Science Conference, pp.2-9, 2000.

[11] M. Petkovic and W. Jonker, "A Framework for Video Modeling", Pro. Eighteenth IASTED International Conference on Applied Informatics, 2000.

[12] F. Li, A. Gupta, E. Sanocki, L. He, and Y. Rui, "Browsing Digital Video", Proc. ACM conference on Computer Human Interaction, pp. 169-176, 2000.

[13] J. Chen, C. M. Taskiran, A. Albiol, C. A. Bouman and E. J. Delp, "ViBE : A Video Indexing and Browsing Environment", Pro. SPIE Conference on Multimedia Storage and Archiving Systems IV vol. 3846, pp.148-164. 1999.

[14] M. Guillemot, P. Wellner, D. Gatica-Perez, and J. M.Odobez, "A Hierarchical Keyframe User Interface for Browsing Video over the Internet", Pro. the 9th International Conference on Human-Computer Interaction, 2003.

[15] A. Ekin, A. M. Tekalp, and R. Methrotra, "Integrated semantic-syntactic video event modeling for search and browsing", IEEE Transaction on Multimedia, vol. 6, no. 6, pp.839-851, 2004.

[16] M. H. Yoon, Y. I. Yoon, and Kio Chung Kim, "Hybrid video information system supporting content-based retrieval, similarity retrieval and query reformulation", Proc. IEEE International on Fuzzy Systems, pp.1159-1164, 1999.

[17] A. Ekin, A. M. Tekalp, and R. Methrotra, "Integrated semantic-syntactic video event modeling for search and browsing", IEEE Transaction on Multimedia, Vol. 6, No. 6, pp.839-851, 2004.

[18] C. Yong and X. De, "Content-based semantic associative video model", Proc. 6th International Conference on Signal Processing, pp.26-30, 2002.

Kyoung-Soo Bok

He received the B.S. in Mathematics from Chungbuk National University, Korea in 1998 and also received M.S. and Ph.D. in Computer and Communication Engineering from Chungbuk National University, Korea in 2000 and 2005. He is now Postdoc in KAIST, Korea. His main research interests include location based services, spatio-temporal database, storage management system and content-based retrieval system.

Jae-Soo Yoo

He received the B.S. in Computer Engineering from Chunbuk National University, Korea in 1989 and also received M.S. and Ph.D. in Computer Science from KAIST, Korea in 1991 and 1995. He is now a professor in Computer and Communication Engineering, Chungbuk National University, Korea. His main research interests include database system, multimedia database, location based services, distributed computing and storage management system.