Development of Hypervideo Platform Using Object Databases

Qurban Memon* and Shakeel Khoja**
*Associate Professor, Karachi Institute of Information Technology
**Assistant Professor, Karachi Institute of Information Technology
ST-2, Sector-X/VII Gulshan-e-Maymar
Karachi-75340 Pakistan
qurban@kiit.edu.pk, shakeel@kiit.edu.pk

ABSTRACT

Web users are seizing on interactive capabilities that software suppliers have developed so far, and are eagerly awaiting new interactive capabilities now being demonstrated. This paper focuses the development of rich information environment based on a customized platform enabling hyperlinks on objects within a digital video. With the help of such environment, the viewer is able to view multiple videos concurrently and browse them temporally as well as spatially with the help of an object database. Specifically, the platform enables to traverse through that object by linking, and such links have been explored and created. The link may connect to within a video, or multiple running videos and/or World Wide Web object. The issues such as open hypermedia link base (static, generic or dynamic) and object database versus multiple videos are also investigated.

Keywords: Hypervideo, Video browser, Object database, Spatial browsing

1. Introduction

The exponential growth of the Internet has provided institutions a growing number of research people connected to Internet for information assessment and services development. The global nature of Internet and the growing number of software development community makes Internet a very attractive medium for publishing and educational development. Instant access to globally distributed information has popularized hypermedia for online publishing, and it is very likely that various forms of hypermedia will become dominant form of publishing. Hypervideo is the natural evolution of Hypertext. The related activity on Internet has provided boom to multimedia technologies on World Wide Web and have made it possible to browse video documents on the web. Traditional hypertext systems [Microcosm, Amsterdam System] allow textual information to be arbitrarily linked so that users can navigate between related parts of the information in the system. Many of them can embed images and videos as objects, but they are considered as binary large objects (blobs) only and usually have a general link attached to them. Systems like QBIC, Informix Visual information system (MAVIS) have additional system of understanding still images through content based processing, however fail to answer the big question of conceptual or high level navigation. The current trends and activities include classification of video documents to help users browse and search video documents according to their categories of interest. The authors in [1] propose classification based on contents while considering the hierarchical structure of video data. Thus it can only be applied to applications with video documents containing semantic descriptions and hierarchical contents structures. Many researchers have proposed image and video databases, which use a range of image properties for content-based retrieval[2-4]. Typically they are for use in specific application domains – like use of spectral information for content-based retrieval of satellite images, and specification of domain dependent image features and development of semantic features. A lot of work has also been reported in literature about developing education on demand using hypermedia. Users viewing a hypermedia video can get more information at any time, and the viewing software automatically resolves the hypermedia links to the relevant information – some thing like asking a question during a live presentation[5]. Specifically, the authors have proposed a conceptual Networked Hypermedia Quick Time (NHQT) system architecture – three parts being the users, network infrastructure and servers. The Internet acts as network infrastructure. Hypervideo necessitates a restructuring and rethinking of ideas about authoring and navigating links - the notion of links must be redefined to consider medium's spatial and temporal properties[6-8]. This complicates the framework for hypervideo. Structuring and organizing information is the main issue in hypermedia. True integration of video requires more powerful hypermedia model that takes into account its spatial and temporal dimensions as well as aesthetics and rhetorical aspects of integrating several media.

In our work, we present an object-based approach – by developing an object database that enables to navigate spatially as well temporally. Section 2 presents main issues of hypervideo currently being investigated around the research community. Here we examine the main aims and parameters of current developments in hypervideo. In section 3, we present object databases and discussion about how such databases will be linked through video. We discuss and analyze the proposed approach with world wide web (WWW) model in section 4. Section 5 presents concluding remarks and future work, followed by references in section 6.

2. Current Issues in Hypervideo Environment

The aim of our current work has been to explore and highlight the multimedia-matching problem areas, which were earlier tried and addressed by the systems mentioned above. The main problem areas include:

(a) Properties: For each media object there are different media properties (image, audio and video) that can be measured in different ways in which two or more objects can be considered similar. For example an image can be matched in terms of color, texture; audio object can be considered in terms of frequency, rhythm, tempo; and video objects can be matched in terms of motion vectors, segment matching etc.

(b) Signatures/Meta data: Each audiovisual property of a media can be given different signatures, which are manually edited or auto-generated, and there are different algorithms for calculating and comparing them in different situations.

(c) Similarity: Certain heuristics can be applied that certain perception properties, which are unable to be calculated by machine, should be implemented. Humans use usually world knowledge to recognize certain objects in scenes and sounds that require a level of media understanding, which is not available in computers.

The above-mentioned problems grow tremendously while working on video objects. The first and foremost problem is the different dimensions of a digital video, which are time domain and spatial domain. Automatic Object tracking is another issue, which is highlighted by many researchers[9-10]. By this we mean that users set anchor position in the start frame and the end frame. The system automatically tracks the objects in the specified segment, by using techniques like pattern matching or normalized correlation. Once the objects are traced, the next issue is to create semantics between them. Kanda and Tanaka in [11] discuss about object tracking in terms of similarity matching and developed certain threshold levels to match a particular link. Lewis and Tansley[12] in their MAVIS2 project came up with certain ideas about developing a layer model to develop a conceptual model for a particular media object.

The above mentioned projects are still in their infancy stages and no concrete results have been achieved, so we can term them as attempts to address these issues, however much work in these areas is needed to over come them. In the next session, we develop object database and explore such linking that enables traversing through that object.

3. Object-Database Design for Hypervideo

In this section, we describe an object database model that can be used to present various characteristics of the objects present in the image and/or video. Generally, object database model has to represent concepts well understood by users within an application domain. The application under study considers presence of objects or entities characterized by their attributes. These objects within the model are dictated by the requirement of the application as well as inherent structure of image and video.

The proposed model in Figure 1 shows the development structure of object database. The issues discussed in section two are the main constraints addressed in this model. The model shows spatio-temporal representation of object database. The database grows in space and time in respect of various object descriptors for multimedia matching thus enabling tracking of video objects followed by generation of semantics and further linking in space and time. In order to devise a mechanism for support of linking, conceptual layer is added at the top to facilitate mapping. In the rest of the section below, we highlight specific methodology to be used for each object description/property shown in Figure 1.

Methodologies:

Physical or conceptual objects in a video can be extracted through temporal or spatial features. These objects can re-occur in multiple frames or segments within a video, or occur once only. The following methodologies are preferably adopted in spatial domain while performing any action on video objects:

a. *Similarity Matching:* Through content-based retrieval techniques, similarities of objects can be measured. Typical dimensions are shape, color, shade and texture matching.

b. *Spectral components*: Similarly, spectral components can also be measured to compare different objects. They can be gathered by calculating the RGB values of a particular region, encompassing an object.

c. *Signatures/Annotations*: Once the above features are extracted, the unique automatic signatures and annotations are provided to different objects in developing indexes for video objects in databases. These signatures can be generated by concatenating different spectral components. High-level annotations can also be provided by the user at the later stage.

d. *Relations*: Depending on signatures and annotations, higher-level relations are developed between objects. This can be done by comparing low level annotations developed through spectral components and similarity matching.

e. *Motion Descriptors:* Motion descriptors are tilt up/down, pan left/right, roll, boom up/down, track left/right and dolly forward/backward. Further object motion is categorized by parametric object motion.

Furthermore, physical shape detection can also be categorized into grid layout structure and histogram

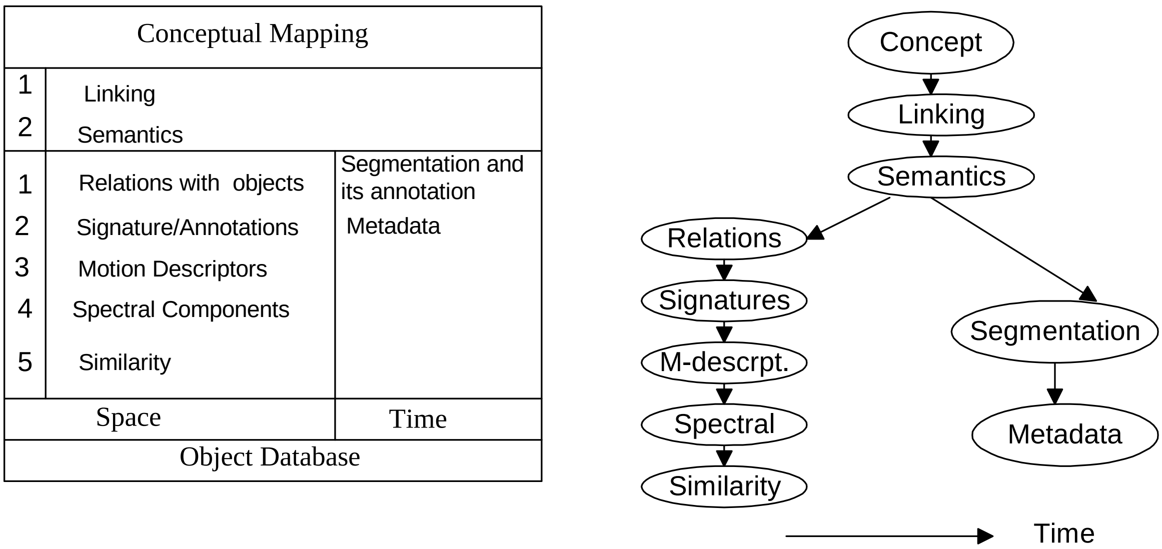

Figure 1: Development Structure of Object Database (Abstraction View)

structure. Further, they are divided by object bounding box properties, region based shape descriptors and contour bases shape descriptors. Color descriptors are models (RGB Scattering, YUH, HSV or HMMD), using dominant color description and histogram. Texture is tagged with luminance edge, histogram description, and/or homogeneous/ heterogeneous textures.

The temporal properties are important from the perspective that a viewer is navigating through a video, and are used in formulating queries that contain object movement constraint among the video frames. Following temporal properties are considered for our hypervideo data model.

a. *Segmentation:* The first step of any video processing system is to segment long chunks of video data into small segments (or streams). These segments, temporal in nature, are usually generated by template matching or histogram matching. Once segments are obtained, they are usually assigned some related information, known as temporal annotations. These annotations can be manually edited or automatically generated. Manual annotations are more specific and depict higher level of intellect, as compared to automatically generated annotations. A segment can also have multiple descriptions, depending upon the semantics used in the segment.

b. *Metadata:* Metadata for temporal aspect of a video will be stored in order to describe the relationship between objects. This metadata is generated by taking into view the temporal relations of different video objects. For example an object is 'overlapping' or is 'in front of' or 'in left direction' etc. Metadata of different frames and segments will also enable non-linear navigation into a video through hyperlinks.

At one layer above these spatial and temporal descriptions, we develop semantics of the objects to enable linking within/across video(s). Following approaches are preferred at this layer for our model:

a. *Semantics:* Once the spatial and temporal features of a video are described, semantics are developed at a higher level. These semantics are developed by assessing all of the spatial and temporal properties of a video object. Further latent Semantics define how closely two objects are related to each other[13]. So by using semantics, we develop relations between similar objects.

b. *Linking:* Hyperlinks allow users to brows the information and access it according to the user's particular subject of interest. Hyperlink systems as compared to traditional database systems provide an ad-hoc jump or non-linear access to information. For video objects, anchors are placed within the video and the link can traverse within or outside the video document.

At a higher layer one above, the concepts are generated by evaluating semantics matching. Concept is an abstract entity corresponding to a real world `object`. Each concept is associated with one or more media representations, i.e. multimedia objects that represent the concept. These representations may be a text term or phrase, a portion of an image, a segment of video or any other medium.

Inter-video links and Open hypermedia linking:

The object database model presented above along with associated methodologies for each of its descriptors at various layers facilitates development of object database for easy access and traversal through hypervideo browser. Furthermore, in order to enable linking objects across multiple videos, the metadata of a video document can have such information, and can only be placed at a further higher layer in the model as it relates to different videos.

An Open Hypermedia System (OHS) is typically a middleware component in the computing environment offering hypermedia functionality to applications orthogonal to their stage and display functionality. An important theme in OHS is the distinction between structure and content i.e. hypermedia links or structural data (metadata in our case) has to be stored outside the document and are not embedded within the original document. This holds true for our model, as the metadata is stored as a wrapper to the document, but not embedded within the video and the actual format of the video is not modified. This feature provides the facility to generate dynamic hyperlinks at the run time. We propose the similar approach as given in [14], as that fits our model objectives. Dynamic links are generated, when a user searches for a particular keyword within a video document, and if any hit is found in metadata, unique URLs are provided to the metadata and the link-base (stored separately) is updated. When a generic link is authored, only the content of the source selection is stored in the source anchor. The original location of the source anchor is not retained. At the link (in the following step) only the content (and not the location) of the user selection is matched with the content section of generic link source anchors and the user may follow links wherever a match is found. This provision is kept in our hypervideo model, so that if the metadata is modified or altered by the user, generic and dynamic links will automatically be updated.

In the next section, we analyze hypervideo model in comparison with WWW model and highlight its critical aspects.

4. Analyses and Comparison with WWW

In this section, we highlight main functional differences between the two media types. Due to its time-based nature, hypervideo requires different aesthetic and rhetoric consideration than traditional static hypermedia. The main differentiating categories are:

(a) *Relationships:* Hypermedia applications allow these relationships to be instantiated as links, which connect the various information elements, and are classified as structural links, associative links and referential links. The hypervideo applications, on the other hand as proposed in section 3, develop such links based on spatio-temporal descriptors and hence are associated at higher level.

(b) *Synchronization:* The proposed hypervideo model can find the appropriate video events and coordinate its properties within the routine events. Even a dynamically segmented video can be developed to meet user specific needs. Such need has actually arisen due to presence of temporal objects, where as such complexity is not present within hypermedia.

(c) *Authoring:* This task in hypervideo is rather tedious than hypermedia, as it requires different rhetoric to accommodate annotations, metadata (provided by users) at different levels (i.e., spatial, temporal and conceptual), and even adding of hotspots links. Although, this part is not directly related to our work, but we thought to address it for the purpose of completeness. For the information of interested readers, it is important to state here that such tools have been developed by various researchers/organizations for authoring a hypervideo.

(d) *Management and Presentation:* This factor becomes more serious when multiple videos are considered such as traversing a link from source video to destination video. The concerns such as pause/play versus continue/end in source video, while destination video has been selected, become serious issues, if desired to be managed within hypervideo. Such a scenario is completely absent in hypermedia. As stated earlier, like authoring, this part is also out of the scope of our current work.

5. Concluding Remarks and Future Work

We highlighted main issues related to development of hypervideo currently under investigation by research community. For this, we also proposed a customized object-database model, which can enable hyperlinks on objects within a video or across videos. We demonstrated that such an object-based database could develop necessary metadata for traversing through objects by linking. We also discussed main differentiating factors between such a platform and existing hypermedia. Currently, we are working on a design structure of metadata and management of such a platform, and hence such development will be reported in due time.

6. References

[1]. D. Liu, C. Lin, and J. Hwang, "Classifying Video Documents by Hierarchical Structure of Video Contents", *The Computer Journal*, Vol. 43, No.5, 2000, pp. 396-410.

[2]. J. Murthy, S. Chen, and S. Chatterjee, " Similarity Measures for Image Databases", *Proceedings of SPIE Conference on Storage and*

Retrieval for Image and Video Databases III, 1995, pp.58-65.
[3]. N. Ramesh, Sethi, "Feature Identification as an Aid to Content Based Retrieval", *Proceedings of SPIE Conference on Storage and Retrieval for Image and Video Databases III*, 1995, pp.2-11.
[4]. S. Smollar, H. Zhang, "Content based video indexing and retrieval", *IEEE Multimedia Magazine* , 1, 2, 1994, pp. 62-72.
[5]. W. Ma, Y. Lee, and D. Du, "Video-Based Hypermedia for Education-on-Demand", *IEEE Multimedia Magazine*, January-March 1998, pp. 72-83.
[6]. M. Dobie, P. Lewis, "Object Tracking in Multimedia Systems", *4th IEE International Conference on Image Processing and its Applications*", April 1994, pp. 41-44.
[7]. N. Sawhney, D. Balcom and I. Smith, "Authoring and Navigating Video in Space and Time", *IEEE Multimedia Magazine*, October-December 1997, pp.30-39.
[8]. A. Salem, J. Potter, " Multimedia Browser", *Proceedings of SPIE Conference on Internet Multimedia Management Systems II*, Vol. 4519, July 2001, pp.203-214.
[9]. W. Chen, and S. Chang, "VISMAP: An Interactive Image/Video Retrieval
System Using Visualization and Concept Maps", *IEEE International Conference on Image Processing*, Greece, Oct. 2001.
[10]. E. Oomato, and K. Tanaka, "OVID, Design and Implementation of a video object database system", *IEEE Transactions on Knowledge and Data Engineering*", Vol. 5, No. 4, August 1993, pp 629 - 643.
[11]. J. Kanda, K. Wakimoto, H. Abe, S. Tanaka, "Video Hypermedia Authoring using Automatic Object Tracking", *Proceedings of SPIE Conference on Storage and Retrieval for Still Image and Video Database VI*, Vol. 3312, January 1998, pp. 108-115.
[12]. M. Dobie, R. Tansley, D. Joyce, M. Weal, P. Lewis, W. Hall, "MAVIS 2: A New Approach to Content and Concept Based Navigation", *Proceedings of the IEE Colloquim on Multimedia Databases and MPEG-7*, 99(056), 1-1999, pp. 9/1-9/5.
[13]. T. Landauer, P. Foltz, and D. Laham, *An Introduction to Latent Semantic Analysis, Discourse Processes*, University of Colorado, 25:259-284, 1998.
[14] Steve Griffiths, Rob J, et. al., "Media-based Navigation with Generic Links", *7th ACM Hypertext 96*, Washington DC, USA, March 16-20, 1996.

Biography of First Author:

Dr. Qurban A. Memon has received MS and Ph.D. degrees from Florida in 1993 and 1996 respectively. He has served as Assistant Professor at GIK Institute of Engineering and Technology from January 1997 to July 1998, and later as Associate Professor at Hamdard Institute of Information Technology. Currently, he is Associate Professor at Karachi Institute of Information Technology. He carries an overall professional experience of over ten years. He has served as Project Director, Hamdard Education Net and Administrator, Regional Cisco Network Academy, Hamdard University. He has twenty-six research publications to his credit in International Journals and Conferences. His research interests include Internet Video, Network Security and Mobile Computing. His biography has been included in millennium edition of *Whos' Who in the World*.

Biography of Co-Author:

Dr. Shakeel A. Khoja received his PhD from University of Southampton, UK, in the field of Multimedia Databases in 2002. He has worked on various projects related to digital video, MPEG-4 and MPEG-7, and has 11 publications related to these topics. Currently he is working as an Assistant Professor at Karachi Institute of Information Technology (KIIT), Pakistan. He is an active researcher in areas of video databases and multimedia frameworks for E-learning.