Content Based Editing of Semantic Video Metadata

Chitra L. Madhwacharyula, Mohan S. Kankanhalli*, Philippe Mulhem***
*School of Computing/IPAL-CNRS**,National University of Singapore, Singapore 117543*
E-mail: iscp1059@nus.edu.sg,mohan@comp.nus.edu.sg

Abstract

Bridging the 'signal-symbol gap' existing between multimedia signals generated through audio, video or other multimedia streams and the high level symbols(metadata) which describe them is presently one of the most vital areas of multimedia research. This paper attempts to bridge this significant gap by proposing a novel automatic mechanism for XML based video metadata editing, in tandem with video editing operations. An implementation framework for editing metadata in accordance with the video editing operations is demonstrated. Conflict resolution and regularization operations are defined and implemented with respect to video metadata editing operations.

1. Introduction

Multimedia computing has emerged in the past decade as a major area of research. Multimedia data consists of pixel values or compressed pixel values which by themselves do not convey much interesting information. Hence it is more relevant to index multimedia information on what the pixels represent, i.e. *semantic content* of the data, which is the 'metadata'. Each image, audio, or video clip contains potentially useful metadata. This metadata, if effectively utilized, can play a pivotal role in management and retrieval of digital image, video and other multimedia files.

One major hurdle encountered by current media management systems is the '*signal-symbol gap*' existing between multimedia signals generated through the audio, video or other multimedia streams and the high level symbols (metadata) which describe them. Bridging this gap is a very important area of study. Research on multimedia editing has extensively addressed the issue of editing the multimedia signal objects. However the associated symbol editing has not received the deserved attention so far. But since multimedia objects are increasingly being represented as integrated signal-symbol units, hence it is vital to manage and edit multimedia data, from the perspective of both the signals and their associated symbols.

Digital Video is an important component of multimedia consisting of raw video data and semantic information contained within it. With the rising popularity of personal camcorders, users are shifting from being passive spectators to active creators of video. Forrester Research predicts that by 2005, 92% of online consumers will create personal multimedia content at least once a month [8]. Therefore research on 'Home Video production techniques' as exemplified by Garage Cinema [3] is becoming inevitable. Video editing tools edit home video data by building, assembling and trimming video clips. Unfortunately all these operations are carried out independent of associated video metadata thus further widening the breach between multimedia signals and symbols.

This paper attempts to bridge this significant gap by addressing the issue of 'automatic content based editing of semantic video metadata' in tandem with video editing operations. Video metadata editing can be 'format based' which involves converting metadata from one document format to another (XML to HTML,SGML etc) and is typically handled by XSLT [11], 'structure based' or 'content-based' by editing structure or content or both. We propose a method for representing and implementing content based metadata editing.

2. Related Works

Content representation for video and multimedia applications has been addressed by [1, 2]. Standardization of multimedia metadata annotation format has been attempted in [5]. Multimedia metadata based querying has been handled by [4]. As XML [10] is becoming a standard multimedia annotation tool, it is useful to study editing mechanisms from the perspective of XML documents. Available XML document editing techniques [6, 7, 9] perform simple merge operations. Many of them do not stress on conformance to schema. Conflict resolution during

editing operations has also not received much attention. One of the most difficult problems in the current XML based metadata editing scenario is to develop standard generic methods for metadata editing whose annotation structure, in most of the cases, is not standardized.

3. Metadata transformation representation

The scope of our proposed novel metadata editing representation is restricted to one underlying schema for all editing operations. The two main editing operations handled are:

- Merge of two metadata documents.
- Projection of a portion of the metadata document.

3.1 Definitions

Definitions for metadata editing representation are:

Definition 1: Model of Conformant document

Every metadata document will have the structure *(S_i, D_j)* where *S_i denotes a Schema, D_j denotes a XML metadata Document and Dj conforms to the schema S_i.* Therefore if we have 2 documents conforming to same schema they will be represented as ***(S_1, D_1)*** and ***(S_1, D_2).***

Definition 2: Structure of a metadata document

Every metadata document D is represented as:

D = (R, E, <comp, order)

- where *R* is the Root Element of the document.
- *E* is the set of elements of *D*. *E* represents a set of elements consisting of n elements represented as E_1……E_n. Each Element E_i is associated with a *start tag* $<E_i>$ and an *end tag* $</E_i>$
  Structure of *E* (E_1{ text, $attribute_1$:value, $attribute_2$:value…..attribute $_n$: value } ,E_2 {text, $attribute_1$:value……},……..E_n{ }) .
- ***<comp*** denotes composition of elements or relationships among the elements. It consists of a set of ordered pairs of distinct elements of *E*.
- *order d*enotes order of occurrence of the elements in a metadata document. It denotes document flow in the following manner. If E_1, E_2 and E_3 are elements present in the metadata document then E_1→ E_2→E_3 denotes order of occurrence of elements as E_1 followed by E_2 followed by E_3 in the actual document instance.

Information Content of metadata document ***D*** is defined as ***I(D) = (R,E, <comp)***. Difference between metadata document *D* and its Information Content *I(D)* is that order of occurrence of elements is not taken into consideration in *I(D)*. All metadata editing operations are carried out with respect to the Information Content.

Definition 3: Merging of Video Metadata

Two metadata documents having the same schema are defined as merged if their content is interwoven into a new metadata document according to the specific merge criterion while still conforming to the same original schema. If we have two conformant documents (S_1,D_1) and (S_1,D_2) Then merged document D_3 is defined as $D_3 = D_1 \oplus^{R}_{mc} D_2$ where [1] $\oplus^{R}_{mc}$ is the merge operator consisting of merge criteria **mc** and regularization operator **R.** The corresponding conformant document is represented as ***(S_1, D_3).*** Structure of D_3 can vary according to the merge condition **mc** *(refer definition 6)*.

Definition 4: Projection of Video metadata

*Projection of metadata Document **D** conforming to a schema **S** is defined as extracting a copy of the sub parts of original document **D** as D_1 based on a projection criterion such that D_1 conforms to original schema **S**.* The document resulting from projection should be complete in describing the component or frame attached to it. Considering a conformant document *(S_1, D_1)*, suppose we want to project document D_1 as D_2 based on a criterion. This can be represented as $D_2 = \prod^{R}_{pc} (D_1)$ where [2] $\prod^{R}_{pc}$ is projection operator and pc is the projection condition.

Definition 5: Regularization

Regularization with respect to metadata editing is defined as process of conflict resolution and introduction of minimal redundancy to maintain validity, consistency, accuracy of metadata document with respect to conformance to original schema.

The functions of the regularization operator **R** are:

- Conflict resolution: Resolving conflicts arising during metadata editing.
- Introduction of minimal redundancy to ensure conformance to original schema.

Possible conflicts occurring during metadata editing:

- **Element related Conflicts:** Primary key conflicts or range based conflicts with respect to elements of an XML metadata document.

[1] $\oplus^{R}_{mc}$: Merge Operator operates on two operands which are metadata documents based on given constraints ***mc*** *(Refer Definition 6)* and returns merged result in the form of an edited metadata document. Also acts as regularization operator (*Refer Definition 5*) to handle inconsistencies and ensure conformance to original schema

[2] $\prod^{R}_{pc}$: Projection Operator operates on a single operand consisting of metadata document and results in projection of original document based on given projection constraint p***c*** *(Refer Definition 6)*. Also acts as regularization operator to handle the ID inconsistencies and ensure conformance to original schema

- **Attribute related Conflicts:** When we have exactly the same element with different or additional attribute values.
- **Content related Conflicts:** Same element referring to same object has different content in two different documents.
- **Structure related Conflicts:** Arise due to restrictions on metadata documents by the schema with respect to number of elements that should be present and other similar constraints.

Definition 6: Merge/Project Criteria

Metadata editing is governed by merge/project condition. Conditional merge/project criteria are specified in syntax similar to that of XQL [12]. The schema designer knows the precise semantics of the application and thus has to explicitly specify the criterion. Some of the editing criteria can be as follows:

- **Merge Criteria mc**
  - Concatenation(c): Entire content of both documents added to the merged document.
  - Union(u): Same as concatenation but duplicate elements not repeated.
  - Intersection(i): Common elements between both metadata documents are merged.
  - Conditional(Element, Attribute or Content based): Merge operation carried out based on merge condition specified i.e. only selected elements, attributes or content included in merged document.
- **Projection Criteria pc**
  - Conditional(Element, Attribute or Content based): Project operation carried out based on projection condition specified. Only selected elements, attributes or content are included in the projected document.

4. Implementation

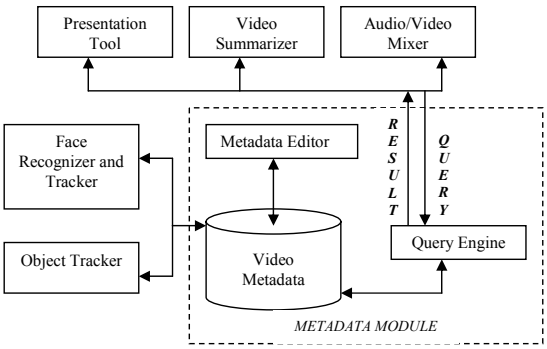

Figure 1: Modules of DVA system

The video metadata editing mechanism has been implemented in the context of Digital Video Album (DVA) system [13] for editing metadata in accordance with presentation/ summarization operations performed on video. Digital Video Album (DVA) focuses on techniques to index, retrieve, summarize and access digital video from home video and digital TV. Figure 1 shows the DVA System modules.

Object and Face Trackers automatically detect and track objects and faces in video and store their information in the form of XML metadata. Metadata Editor allows manual input of video metadata which is accessed by Summarization and Presentation Tools to create presentations and summaries. Face detector, Object Tracker and Metadata Modules serve as annotation tools to generate video metadata. Metadata is an integral part of the DVA system because it serves as the database and co-ordinates information of the whole system. Workflow of video metadata editing application is as shown in Figure 2.

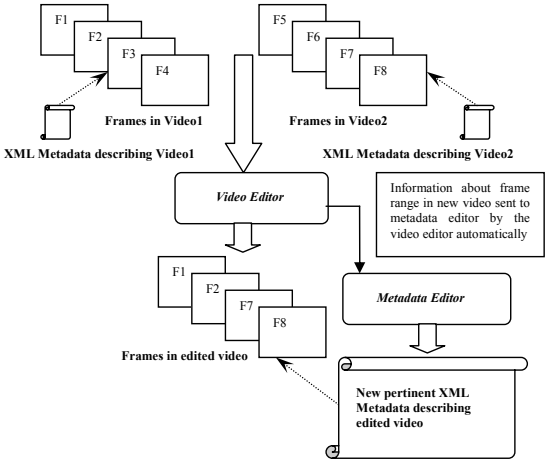

Figure 2: Workflow of metadata editing process

4.1. DVA Document Structure

DVA metadata structure consists of three main files:

- **Index File:** Root to record basic information such as Video ID and physical path for any video file loaded in the system.
- **Object File:** Stores metadata about available objects and people in the video. Every Object assigned unique Object ID and its details annotated. Reference is made in Video Description File to Object ID's in Object File for objects occurring in a particular video.
- **Video Description File:** Every video is associated with a Video Description File. Stores information about a particular video pertaining to sequences, objects and other relevant details. Metadata editing operations are performed on this file. The newly generated edited metadata file is associated with the edited video clip generated by the video editor.

4.2. Steps in video metadata editing process

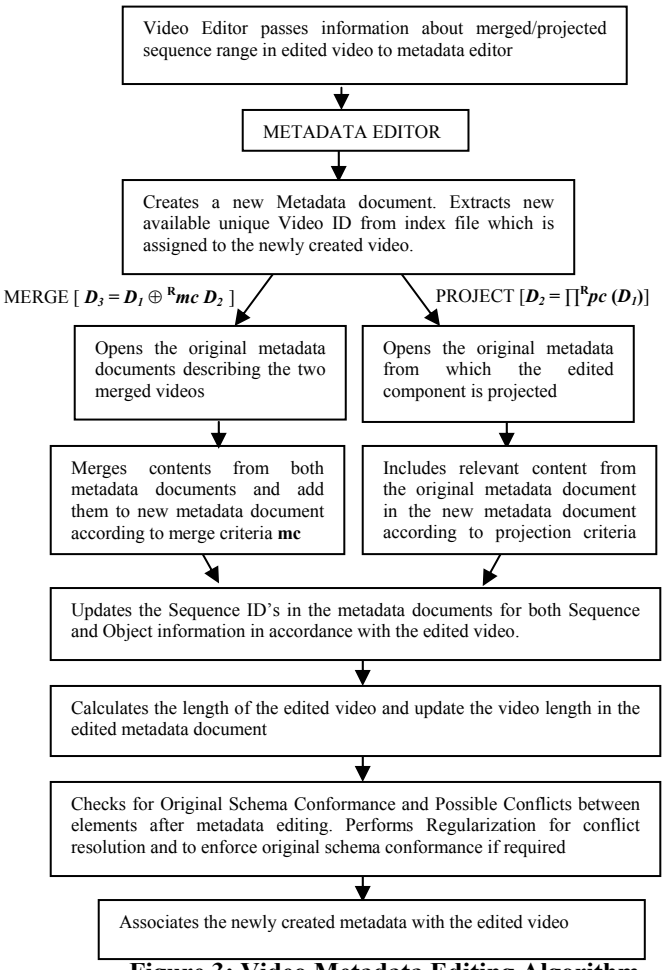

Figure 3: Video Metadata Editing Algorithm

5. Experiments and Results

We conducted a number of video metadata editing experiments on Video Description Files conforming to one schema. Our metadata editor consistently reacted accurately to information provided by the video editor to automatically produce valid metadata documents which describe the edited video. Since space constraints prevent us from presenting results, detailed examples are provided at [13].

6. Conclusion

The paper attempts to bridge the signal-symbol gap existing between low level multimedia signals and high level symbols that describe them by presenting an automatic XML based video metadata editing mechanism which functions in tandem with video editing. The framework is implemented in context of DVA metadata. A similar methodology can be adapted to any video metadata editing scenario. The possible conflicts that may arise with regards to the conformance to original schema have been handled to ensure that the newly formed metadata documents are valid and well formed.

For future work, an approach can be developed to handle metadata editing and conflict resolution at the basic schema design level by developing a standardized schema for metadata editing. The advantage of doing this is that as the schema designer designs the structure of metadata document hence he/she will know best about the types of editing that can be allowed on it. Thus potential conflicts during editing process can be identified and appropriate conflict resolution mechanisms can be enforced at design level itself. In order to do this available XML editors need to be modified or new ones need to be developed to recognize the new schema with the additional components.

7. References

[1] Chang, S.F., Alexandros, E., Mcclintock, R.**:** "Next-generation content representation, creation, and searching for new-media; applications on education." *Proc.of IEEE* 86(5), pp 884-904, 1998.

[2] Christel, M:"Addressing the contents of video in a digital library." *Elec. Proc. of ACM Workshop on Effective Abstractions in Multimedia,* 1995. http://www.informedia.cs.cmu.edu/documents/acmmm95_christel.htm

[3] Davis M.: "Garage Cinema and the Future of Media Technology." *Communications of the ACM (50th Anniversary Edition Invited Article)* 40 , pp 42-48,1997.

[4] Grosky W., Farshad F., Zhaowei J.: "Using metadata for the intelligent browsing of structured media objects." *Multimedia Data Management: Using Metadata to Integrate and Apply Digital Media*, pp 123-148, 1998.

[5] Hunter J, Renato I**.:** "The Application of Metadata Standards to Video Indexing", *In Second European Conference on Research and Advanced Technology for Digital Libraries***,** 1998.

[6] La Fontaine,R.: *"*Merging XML files : a new approach providing intelligent merge of XML data sets*"* , *XML Europe,*2002,www.idealliance.org/papers/xmle02/dx_xmle02/papers/03-03-04/03-03-04.pdf.

[7] Manger.G.W: "A Generic Algorithm for Merging SGML/XML"*, XML Europe*, pp 21-25, 2001.

[8] Schwartz, J., Rhinelander, T., Dorsey, M.: "Personal Rich Media Takes Off", *The Forrester Report, Forrester Research Inc*., 2000.

[9] Tardif.L,Euzenat.J: "XML transformation flow processing", *2nd Conference on Extreme markup languages*, pp 61-72, 2001.

[10] XML, http://www.w3.org/TR/REC-xml

[11] XSLT: http://www.w3.org/TR/xslt

[12]XQL:http://www.ibiblio.org/xql/xql-proposal.html

[13] http://origin.bic.nus.edu.sg/~chitra/MetadataEditing