G3M: A Generalized Multimedia Data Model Based on MPEG-7

Qiong Zuo, Zhongsheng Cao
Huazhong Univ of Sci & Tech, China
qiong_zuo@yahoo.com.cn, caozhongsheng@163.com

Abstract

In this paper, a generalized multimedia database data model based on MPEG-7 named G3M is proposed and formally defined. Different from those XML Database Management System (DBMS) solutions for MPEG-7 storage and retrieval, G3M analyses the MPEG-7 Multimedia Description Schemes (MDS) throughout to construct user-preferred database schemas, and the prototype system is built on extensible Object-Relational DBMSs. With the strong expressiveness of MPEG-7, G3M represents various aspects of multimedia data well. Domain knowledge for special application demand can be defined and used as references in G3M.

1. Introduction

Over the past decade, with the rapidly growing demand of multimedia applications, the research on multimedia databases has achieved a lot, and a great number of multimedia data models are developed.

In view of generalized data models, multimedia database management systems (MMDBMSs) can be distinguished between two directions [1]: multimedia extensions on traditional DBMSs (such as Oracle interMedia and IBM Informix DataBlades), and native MMDBMSs especially tuned for multimedia data (e.g., DISIMA and MARS). Since the latter can not support traditional data as well as multimedia data, we discard them. In the first direction, MMDDBMSs generally stand on three existed data models: Object-Relational data model, Object-Oriented data model and spatio-temporal data model. Among them, ORDBMS is most commercially used for multimedia extension. Because there are various formats of one media, syntax analysis and operations of each format should be added to the system. This work is very time-consuming. In most cases, such solutions are limited to special media data types and special applications.

Since Multimedia Content Description Interface (MPEG-7) is currently the most complete description standard of multimedia data, and MPEG-7 MDS is a XML Schema variant MPEG-7 DDL, XML DBMS is employed to manage MPEG-7 media descriptions [2]. In extended XML DBMS, various multimedia data are supported through one form, that is, MPEG-7 media descriptions. Retrieval on multimedia data is implemented by keyword searches on simple paths and elements of MPEG-7 descriptions. The focus of our criticism to such solutions is that MPEG-7 descriptions are not equal to XML. The valuable metadata (schema and type information available with MDS) for the management of MPEG-7 descriptions are neglected.

Attracted by the strong expressiveness, good extensibility and broad applicability of MPEG-7, we designed and implemented a generalized multimedia data model based on MPEG-7 named G3M. Different from those XML DBMS solutions mentioned above, we decompose the MPEG-7 MDS into ORDB schemas, where various aspects of media data are stored as specific tables and references for further retrieval.
G3M inherits lots of merits of MPEG-7 Description Tools and key contributions of existing multimedia data modeling methods, summarized as follows.
- Multi-level abstraction [3], which is a basic computer science tool for complex problem solving in general. Multimedia segmentation [4, 5] is a kind of effective abstraction on multimedia.
- Metadata interpretation. Metadata is essential for understanding and exploring unstructured data by representing multimedia documents in a structured way [6]. It can be used to solve the problem of information standardization.
- Salient object-based approaches. In most studies, salient physical objects extracted from media data and the spatio-temporal relationships among them are used for semantic expressing [7]. Recently, objects, events and their relationships are used to express "who is doing what" as nature languages do [8]. This way is still under construction to express high-level semantics.

The remainder of the paper is organized as follows. Section 2 presents the characteristics, architecture and formalized definition of G3M. Section 3 gives out the implementation of such model. Section 4 proposes a domain knowledge importing method. Section 5 draws the conclusion and discusses future works.

2. Generalized Multimedia Data Model: G3M

2.1. "Generalization" of G3M

The aim of our research is to design a generalized multimedia database data model. The keyword "generalized" of G3M means a lot.
1. G3M supports most media data types, e.g., audio, video, image, etc, that can be expressed by MPEG-7 descriptions, in a unified way. In other words, G3M only deals with one kind of data, the MPEG-7 descriptions. When a new media data type adds into MPEG-7, G3M can be easily modified to support it by converting the newly extended MPEG-7 MDS into a new database schema.
2. Every generally important aspect of multimedia data types can be represented well in G3M, such as flexible segmentations, low-level features, spatio-temporal relationships and high-level semantics etc.
3. G3M has strong extensibility for two points. One is that the new media type can be easily added into the system, as mentioned above. The other is due to the widely use of MPEG-7. All applications that use MPEG-7 as interface can be easily connected with G3M. Recently, several MPEG-7 ontology models in RDF or OWL have been proposed [9]. If these models come into being, the MPEG-7 will be more widely used, so will be G3M.
4. In most cases, more generalization means less specialization, which may limit the utilization of media properties in specific domains. G3M conquers this defect by supporting user-defined attributes, features, objects, events, concepts and relations. Furthermore, special domain knowledge can be imported into G3M, which will be presented in section 4 below in detail.

2.2 Architecture of G3M

To represent different aspects of multimedia, G3M divides the whole model into six layers, as figure 1 shows.
1. Raw data layer is the lowest layer, where MPEG-7 descriptions are fully held for further implied information to be dug out in near future.

2. Metadata layer stores the MPEG-7 MDS as metadata for unstructured multimedia data. Users can choose their interested MDS to build database schemas, and veil anything uninterested.

3. Segmentation layer contains the time-continued or content-related segments (noted as SegUnits) of multimedia data according to MPEG-7 MDS. MPEG-7 descriptions may first be divided by their audio-visual content, and then be cut into spatio-temporal or media source decomposition SegUnits.

4. Low-level feature layer stores features picked up from any SegUnit. A MPEG-7 Descriptor, such as color, texture, shape etc, is usually defined as a feature. G3M also supports user-defined features. Each feature is encapsulated as an object with its metadata information.

5. Spatio-temporal layer shows the spatio-temporal relationships of concrete objects. There are two kinds of concrete objects: simple object (noted as SObject) and combined object (noted as CObject). The spatial or temporal relations between SObjects will be recorded in G3M. CObject is a group of continuous or discontinuous SegUnits with same meanings. It is always used to describe abstract things. Concrete event (noted as CEvent) is also defined here with time and location attributes.

6. The highest layer is knowledge layer, where defines high-level semantics with salient objects, events and their relations which users are interested in. As natural languages do, G3M describes people, things and actions with background, using key factors “who, whom, what, when, where, why, how”.

1) There are two kinds of salient objects: salient concrete object (SCO) and salient abstract object (SAO), which are the abstraction forms of SObject and CObject separately. A SCO can be materialized by one or more SObjects, so does a SAO to CObjects. In other words, a SObject or a CObject is the occurrence of a SCO or a SAO separately in certain SegUnits.

2) Salient event noted as SE is an abstract notion without spatial or temporal location. A SE can be embodied by one or more CEvents in certain SegUnits.
2.3 Formalized Definition of G3M

The formalized definition of G3M is given below.

**Definition 1**: A generalized MPEG-7 multimedia database schema named G3MS is an 8-tuple

\[ G3MS = (\Sigma, Um, \Omega, Sem, F_{Um⧸\Omega}, F_{\Omega⧸Sem}, UAttr, T), \]

where:
- \( \Sigma \) is a 4-tuple \((Mid, Mr, Mm, Am)\), where \( Mid \) is the unique id of the media data, such as an image or video segment, written in MPEG-7 DDL; \( Mr \) is its raw data; \( Mm \) is its full MPEG-7 media description scheme; and \( Am \) is the privilege definition of the elements in \( Mm \), which is defined by database administrator (DBA) for safety management;
- \( Um \) is a 4-tuple \((Uid, Up, Au, Mu)\), where \( Uid \) is the unique id of a user; \( Au \) is the user’s database access privileges. If a user’s read privilege is higher than the privilege of the element in \( Mm \) which he/she is interested in, then the scheme of such element can be picked up from \( Mm \) to \( Mu \). \( Up \) represents those elements in \( Mm \) that the user are interested in; \( Mu \) is the cut media description scheme of \( \Sigma \), with the limitation of the user’s preference and access privileges. In G3M, each user has one corresponding media description scheme over one media data;
- \( \Omega \) is a collection of SegUnit according to \( Mu \);
- \( Sem \) is a group of high-level semantic information collections appear in the SegUnits in \( \Omega \);
- \( F_{Um⧸\Omega} \) is a mapping between the \( Mu \) in \( Um \) and each SegUnit in \( \Omega \);
- \( F_{\Omega⧸Sem} \) represents the high-level semantics of each SegUnit in \( \Omega \);
- \( UAttr \) is the common attributes of the media data, such as the CreationInfo from MPEG-7 descriptions. Each attribute is characterized by its value domain and value;
- \( T \) is the text annotation according to MPEG-7, recording the media semantics in text format.

**Definition 2**: A SegUnit can be a simple segment unit or a collection of segments. The information of concrete objects and events that appear in this SegUnit is also recorded here. It is a 7-tuple

\[ \text{SegUnit} = (\text{SegId}, \text{Type}_{seg}, \Gamma_{seg}, M_{seg}, \Sigma_{seg}, \Delta_{seg}, \text{Sem}_{seg}), \]

where:
- \( \text{SegId} \) is the unique id of a SegUnit;
- \( \text{Type}_{seg} \) is the type of a SegUnit. There are three kinds of SegUnit: simple segment, time-continued or content-related combined segments;
- \( \Gamma_{seg} \) represents the temporal information of a SegUnit. It is defined as a time interval \([Ts, Te]\), where \( Ts \) is the beginning time, and \( Te \) is the end time of the SegUnit; When \( Ts \) is equal to \( Te \), it means such SegUnit happens at a time point;
- \( M_{seg} \) is the temporal, spatial or spatio-temporal connectivity of the sub segments in this SegUnit, which is defined according to the definitions of those Mask Descriptors of MPEG-7 Segment DS;
- \( \Sigma_{seg} \) is a collection of sub segments of this SegUnit;
- \( \Delta_{seg} \) is a collection of low-level features which appear in the SegUnit;
- \( \text{Sem}_{seg} \) is a collection of spatio-temporal semantics which appear in the SegUnit.

**Definition 3**: A low-level feature named LowF is a 5-tuple

\[ \text{LowF} = (Fid, Fm, Fd, U_{seg}, U_{el}), \]

where:
- Fid is the unique id of a LowF;
- Fm is the metadata of such feature;
- Fd is the value of a feature written in MPEG-7 DDL, which is stored and processed as a whole;
- Useg is a collection of SegUnits where this feature appears in.
- Uel is a set of elements or attributes of such LowF that users are interested in.

Definition 4: A spatio-temporal semantic collection named SemST includes a set of SObjects, CObjects, CEvents and the spatio-temporal relationships of SObjects that happen at the same time interval in some region. There may be no spatio-temporal relationship between a SObject and a CObject, or between two CObjects. A SemST is a 9-tuple
\[ \text{SemST} = (\text{Sid}, \text{Attrst}, \Gamma_{st}, \text{Lst}, \text{Uo}, \text{Ue}, \text{SpaRoo}, \text{TempRoo}, \text{Tst}) \]
where:
- Sid is the unique id of a SemST;
- Attrst is the attribute set of this SemST;
- \( \Gamma_{st} \) is the time interval \([Ts,Te]\) of the SemST;
- Lst is the location information of the SemST;
- Uo is a collection of SObjects and CObjects in this SemST;
- Ue is a collection of CEvents in this SemST;
- SpaRoo represents the relative spatial relations among SObjects. The relations among m SObjects are represented by a m*m 2-dimension array, whose node type is SRoo, the relative spatial relations between two SObjects;
- TempRoo is the relative temporal relations among SObjects. It also uses a 2-dimension array, whose node type is TRoo, the relative temporal relations between two SObjects;
- Tst is the text annotation of such SemST.

The SRoo and TRoo relations between two SObjects are defined below.

\[ \text{SRoo} ::= \text{south} | \text{north} | \text{west} | \text{east} | \text{northwest} | \text{northeast} | \text{southwest} | \text{southeast} | \text{left} | \text{right} | \text{below} | \text{above} | \text{onTopOf} | \text{under} | \text{equal} | \text{inside} | \text{contains} | \text{coveredBy} | \text{covers} | \text{overlaps} | \text{.touches} | \text{disjoint} | \text{separated} | \text{UserDefined} \]

\[ \text{TRoo} ::= \text{before} | \text{after} | \text{meets} | \text{metBy} | \text{overlaps} | \text{overlapsBy} | \text{strictDuring} | \text{during} | \text{contains} | \text{strictContains} | \text{starts} | \text{startedBy} | \text{finishes} | \text{finishedBy} | \text{equal} | \text{UserDefined} \]

Definition 5: A high-level semantic collection named Sem\( \Sigma \), which contains a group of salient objects, events, concepts and their relationships to represent the same meaning, is an 8-tuple
\[ \text{Sem\( \Sigma \)} = (\text{SemId}, \text{Attrs}_{\Sigma}, \text{Uo}, \text{Ue}, \text{URoo}, \text{UREe}, \text{Uc}, \text{Tst}_{\Sigma}) \]
where:
- SemId is the unique id of a Sem\( \Sigma \);
- Attrs\( \Sigma \) is the attribute set of the Sem\( \Sigma \);
- Uo is a collection of salient objects (SCO and SAO) appear in Sem\( \Sigma \);
- Ue is a collection of salient events appear in Sem\( \Sigma \);
- URoo is the relationship set among objects in Uo. It uses a 2-dimension array, whose node type is Roo;
- URee is the relationship set among events in Ue. It uses a 2-dimension array, whose node type is Ree;
- Uc is a collection of concepts appear in Sem\( \Sigma \);
- Tst\( \Sigma \) is the text annotation of this Sem\( \Sigma \).

The relationship between two objects named Roo defines as follows.
R_{oo} ::= \text{componentOf} \mid \text{hasComponentOf} \mid \text{memberOf} \mid \text{hasMemberOf} \mid \text{substanceOf} \mid \text{hasSubstanceOf} \mid \text{knowledgeRef} \mid \text{knowledgeReferredBy} \mid \text{UserDefined} \mid \text{R}_{\text{com}}

The relationship between two events named R_{ee} defines as follows.
R_{ee} ::= \text{resultOf} \mid \text{causeOf} \mid \text{summaryOf} \mid \text{elaborationOf} \mid \text{entails} \mid \text{entailmentOf} \mid \text{mannerOf} \mid \text{knowledgeReferredBy} \mid \text{KnowledgeReferredBy} \mid \text{UserDefined} \mid \text{R}_{\text{com}}

The R_{com} appeared above, is the common relationship between two objects, two events, or two concepts, which is defined below.
R_{com} ::= \text{specializationOf} \mid \text{generalizationOf} \mid \text{similarTo} \mid \text{oppositeTo} \mid \text{exampleOf} \mid \text{partOf} \mid \text{exemplifiedBy} \mid \text{equivalentTo} \mid \text{hasPartOf} \mid \text{IdentifiedWith} \mid \text{contains} \mid \text{isContainedBy} \mid \text{refines} \mid \text{refinedBy} \mid \text{providesContrastfor}

\textbf{Definition 6:} An object named Obj, represents the people or things that users are interested in, is an 8-tuple
\begin{align*}
\text{Obj} = (\text{Oid}, \text{Type}_o, \text{Attr}_o, \text{T}_o, \text{L}_o, \text{U}_{\text{Seg}}, \text{U}_f, \text{U}_o),
\end{align*}
where:
\begin{itemize}
  \item \text{Oid} is the unique id of an Obj;
  \item \text{Type}_o represents the object type: SObject, SCO, CObject or SAO;
  \item \text{Attr}_o is the attribute set of the Obj;
  \item \text{T}_o is the exact time interval of Obj. SCO and SAO may have no time information;
  \item \text{L}_o is the exact location information of SObject. Other objects may have no location information;
  \item \text{U}_{\text{Seg}} is a collection of SegUnits where the Obj appears in;
  \item \text{U}_f is a collection of features that the Obj owns;
  \item \text{U}_o is a collection of objects that this Obj contains.
\end{itemize}

\textbf{Definition 7:} An event named Event, represents the event that users are interested in, is an 8-tuple
\begin{align*}
\text{Event} = (\text{Eid}, \text{Attr}_e, \text{T}_e, \text{L}_e, \text{U}_{\text{Seg}}, \text{U}_o, \text{U}_e, \text{UR}_{\text{ee}}),
\end{align*}
where:
\begin{itemize}
  \item \text{Eid} is the unique id of an Event;
  \item \text{Attr}_e is the attribute set of the Event;
  \item \text{T}_e is the exact time interval of an Event. SE may have no time information;
  \item \text{L}_e is a 2-tuple (\text{Loc}, \text{F}_e), where \text{Loc} is the exact location information of an Event; and \text{F}_e is the relationship between the location and the Event. SE may have no location. \text{F}_e is defined below.
\end{itemize}
\begin{align*}
\text{F}_e ::= \text{locationOf} \mid \text{hasLocationOf} \mid \text{sourceOf} \mid \text{hasSourceOf} \mid \text{destinationOf} \mid \text{pathOf} \mid \text{hasDestinationOf} \mid \text{hasPathOf} \mid \text{UserDefined}
\end{align*}
\begin{itemize}
  \item \text{U}_{\text{Seg}} is a collection of SegUnits where this Event appears in;
  \item \text{U}_o is a collection of objects appeared in this Event;
  \item \text{U}_e is a collection of events that this Event is comprised of;
  \item \text{UR}_{\text{ee}} is a collection of relationships between Objs and this Event. It uses a 2-dimension array, whose node type is R_{ee}. Its definition is below.
\end{itemize}
\begin{align*}
R_{ee} ::= \text{agentOf} \mid \text{hasAgentOf} \mid \text{hasPatientOf} \mid \text{PatientOf} \mid \text{experiencedOf} \mid \text{causerOf} \mid \text{hasExperiencedOf} \mid \text{stimulusOf} \mid \text{sourceOf} \mid \text{hasStimulusOf} \mid \text{hasCauserOf} \mid \text{locationOf} \mid \text{hasLocationOf} \mid \text{hasSourceOf} \mid \text{themeOf} \mid \text{destinationOf} \mid \text{hasDestinationOf} \mid \text{resultOf} \mid \text{beneficiaryOf} \mid \text{hasBeneficiaryOf} \mid \text{pathOf} \mid \text{hasThemeOf} \mid \text{hasResultOf} \mid \text{instrumentOf} \mid \text{hasInstrumentOf} \mid \text{hasPathOf} \mid \text{accompanierOf} \mid \text{hasAccompanierOf} \mid \text{UserDefined}
\end{align*}
Definition 8: A Concept, which is an abstract semantic entity as defined in MPEG-7, represents anything that can not be described by object, event, time, location or state. It is a 3-tuple

\[ \text{Concept} = (\text{CCId}, \text{CCName}, (\text{SimilarCC}, R_{cc})), \]

where:
- \text{CCid} is the unique id of the Concept;
- \text{CCName} is the name of the Concept;
- \text{(SimilarCC, R_{cc})} is a couple to represent the similar concepts and their relations with this concept.

The syntax definition of \( R_{cc} \) is as follows.

\[ R_{cc} ::= \text{propertyOf} | \text{hasPropertyOf} | \text{isUsedBy} | \text{uses} | \text{knowledgeReferredBy} | \text{knowledgeRef} | R_{com} | \text{UserDefined} \]

3. Implementation of G3M

The G3M prototype system was implemented on ORDBMS, where MPEG-7 descriptions are mapped to corresponding database data types, objects and tables, and various relationships are expressed by relational keys and object references. The storage granularity of a feature, an object or even a segment is decided by user interests and access privileges permitted. Once a schema is created, it can be changed with the transition of the user’s attention, even if there is no variation of such media content. For example, when an object is first inserted into the system, it can be stored as a BLOB (coarse-grained). If a user with right access privileges wants to know more details of such object, it can be unfurled to database tables (fine-grained).

As defined above, the inner- and inter- relationships of segments, features, objects and events are so intricate, and what users want may vary frequently, so collection and classification operations are urgently needed. G3M supports aggregation and decomposition operations on content, segments, features, and objects. Spatial and temporal composition and decomposition operations on segments are supported for semantic clustering and multimedia presentation.

4. Introducing Domain Knowledge into G3M

In special domain applications (such as sports and broadcasting), the especial features can help to acquire high level semantics more efficiently. The problem of domain knowledge importing is a main task for G3M design. As we mentioned above, \text{SCO}, \text{SAO} and \text{SE} are abstract forms of SObject, CObject and CEvent separately. The abstract entities can be materialized by their concrete ones. Suppose the special domain knowledge can be represent as abstract objects, abstract events and their relations, then the special domain knowledge can be seen as examples or references for concrete objects and events in the database. So we especially designed two kinds of relations, \text{“KnowledgeRef”} and \text{“knowledgeReferredBy”}, with which, G3M system can easily recognize the domain knowledge and use it as examples or references for hidden information reasoning or digging. Besides, G3M supports lots of kinds of relationships among semantic entities, such as \text{R_{so}}, \text{R_{oc}} and \text{R_{cc}}, which are much richer than those defined in most MMDDBMSs. Furthermore, the relation \text{“UserDefined”} guarantees the free expression of the special relationships in special domain knowledge of real world.
The right use of domain knowledge for information inferences is still under construction.

5. Conclusions

Combining the well-known MPEG-7 with modern DBMS is a great idea for the development of MMDBMS. Guided by such idea, we defined a Generalized Multimedia data model based on MPEG-7(G3M) and implemented it on ORDBMS. Different from XML DBMS methods, G3M takes full advantage of MPEG-7 MDS. Six layers are abstracted from MPEG-7 media descriptions to represent every aspect of multimedia. Besides, G3M supported user-preferred safety management of database schemas and data. A distinguishing characteristic of G3M is that, special domain knowledge can be imported by defining it as a semantic collection of object-event-relations. For more general-purpose models, G3M needs to be enhanced with implied information reasoning, uncertainty expression and fuzzy retrieval in near future.

References

Authors

Qiong Zuo received her B.S. and M.S. from Huazhong University of Science and Technology (HUST), China. Now she is a lecturer, College of Computer Science and Technology, HUST, China. She is also a Ph.D. student at HUST. Her main research areas are database management systems, information retrieval and multimedia systems.

Zhongsheng Cao received his B.S. from Fudan University, China. And he had his M.E. and Ph.D. from HUST, China. Now he is an associate professor, College of Computer Science and Technology, HUST, China. His main research areas are information theories and technologies, including information modeling, multimedia systems, database management systems, GIS etc.