COSMOS-7: Video-Oriented MPEG-7 Scheme for Modelling and Filtering of Semantic Content

HARRY AGIUS and MARIOS C. ANGELIDES*

Brunel University, School of Information Systems, Computing and Mathematics,

St John's, Uxbridge, Middlesex UB8 3PH, UK

Tel: {+44 1895 265993, +44 1895 265990}

Fax: {+44 870 125 0540, +44 1895 265990}

E-mail: {harryagius@acm.org, angelidesm@acm.org}

Short running title:

COSMOS-7: Video-Oriented MPEG-7 Scheme

* Corresponding author.

COSMOS-7: Video-Oriented MPEG-7 Scheme for Modelling and Filtering of Semantic Content

Abstract:

MPEG-7 prescribes a format for semantic content models for multimedia to ensure interoperability across a multitude of platforms and application domains. However, the standard leaves open how the models should be used and how their content should be filtered. Filtering is a technique used to retrieve only content relevant to user requirements, thereby reducing the necessary content-sifting effort of the user. This paper proposes an MPEG-7 scheme which can be deployed for semantic content modelling and filtering of digital video. The proposed scheme, COSMOS-7, produces rich and multi-faceted semantic content models and supports a content-based filtering approach that only analyses content relating directly to the preferred content requirements of the user.

1. INTRODUCTION

Digital video can be considered from two perspectives [1]. The *bit-centric perspective* focuses on the data of the digital video stream. Relevant concerns include encoding, bandwidth, synchronisation, error detection and correction, digital rights management, and so forth. The *content-centric perspective* focuses on a higher, conceptual level and is concerned with the content imparted by the digital video stream, which is the perspective of this paper. This content is rich in semantics which evolve predominantly over time and are subject to interpretation. The key concern here is that of access to digital video resources by users. For all application domains, it is simply impractical for users to spend inordinate amounts of time sifting through an entire digital video resource to find the few items of content that are relevant to them [2-6]. For example, consider that the user wishes to locate video segments within a digital video resource that feature a certain object or depict a certain event. Manually sifting through the digital video resource to obtain this information is a lengthy and time-consuming task, which rises drastically as the size of the video resource itself increases. If we further consider that the user may also wish to search for relationships between various aspects of content, such as video segments where certain objects are located next to each other or where one event occurs before another event, then the precise urgency for tools that assist the user is further appreciated.

Assisting the user in locating relevant video content requires that the contents of the video streams within a video resource have been modelled according to their semantics, such that the semantic video content models can function as ‘surrogates’ for the original video streams. That is, all enquiries and responses that are necessary for the user and system to interact are undertaken on the semantic video content model rather than the video streams. Effective

referencing strategies are utilised to tightly integrate the semantic video content model with the video streams. Since semantic video content models serve as surrogates, the richer the representation within the model, the more useful it is to the application and thus to the user [7-12]. Consequently, semantic video content models require extensive and flexible vocabularies and structures to enable them to express the diverse, subjective, and multiple-perspective meanings apparent within the video content. It has been advocated for some time now [3, 12-21] that the content should be represented at multiple structural levels so that the semantic video content model can express both the static content of frames and also the dynamic, temporal content of higher-level segment structures, such as shots and/or scenes. MPEG-7 [22-26] meets the above requirements and is the leading standard for specifying the format of semantic video content models. MPEG-7 standardises a set of descriptors, a set of description schemes, a language that specifies the description schemes and description encoding. The overarching aims of the standard are that it is scalable, extensible and flexible.

Filtering works in conjunction with semantic content models to present only relevant content to the user and to eliminate irrelevant content from consideration [11, 27-29]. The semantic video content model enables the filtering processes to determine the meaning conveyed by the video streams within the resource so that a comparison can be made against the specified filtering criteria and the matching video content returned. The return content may be the various segments from the video streams themselves or subsets of the semantic video content model itself. Filtering is therefore a crucial enabler for improving user access to digital video resources. MPEG-7 does not, however, prescribe how the semantic content model is to be produced or consumed, which includes how it is filtered. MPEG-7 has a restricted scope which is that of specifying the content model [24, 26, 30], as illustrated in Fig. 1.

(Fig. 1 near here)

Consequently, this paper examines how an MPEG-7 implementation can be deployed for semantic video content modelling and filtering and proposes the COSMOS-7 scheme. The models produced with COSMOS-7 are rich in content detail and multi-faceted in content semantics and the content-based filtering technique developed with COSMOS-7 only analyses content that relates directly to preferred and explicitly-stated user content requirements. The rest of this paper is organised as follows. Section 2 discusses collectively the key semantic aspects of video content that the research community argues should be modelled by any semantic content modelling scheme. Sections 3 and 4 discuss how COSMOS-7 models these semantic aspects in order to carry out content-based modelling and filtering respectively. Section 5 provides a review of related work. Section 6 concludes the paper.

2. SEMANTIC VIDEO CONTENT

Semantic content modellers argue that a number of key semantic content aspects should be represented in a semantic content model in order for that model to serve as a suitable surrogate for the original video streams. Previously [31-33] we have identified these key aspects as follows: events, temporal relationships among events, objects and object properties, and spatial relationships among objects. Repeated surveys [21, 34] have found that users are interested in ‘who’, ‘what’, ‘where’ and ‘when’ type filters. These semantic aspects reflect the importance of these filters.

Events within the semantic content model represent the context for objects that are present within the video stream at various structural levels. Events are therefore occurrences within the media that divide the media into shorter content segments involving specific objects. That is, more often than not, they serve to represent object behaviour. In this way, the user and the system are able to filter with regards to "What is happening here?" on both a general level, that is, *without* reference to specific objects, and a specific level, that is, *with* reference to specific objects (objects are discussed further below). For example, consider a TV talk show, where one event would be an interview taking place. At the general level, this is an interview event. At the specific level, the talk show host is interviewing the guest, e.g. Jonathan Ross interviewing Eddie Izzard, the key participant objects therefore being Jonathan Ross and Eddie Izzard.

Temporal relationships among events enable the semantic content model to express the dynamism in content that is apparent at these higher levels, thereby enabling filtering of non-static semantic content which is more in line with "What will or did happen here?" Again, this may occur on both a general and a specific level. For example, continuing the previous talk show example, a musical performance event may have taken place before the interview; thus 'before' serving as the temporal relationship in this case.

The expression of objects and object properties within the semantic content model provides for filtering with regards to objects that are readily identifiable within the video content stream. The term 'object' refers to any visible or hidden object depicted within a video frame at any necessary level of detail, from entire objects and groups of objects to the bottom-most component objects. For example, the Jonathan Ross object could be considered to be composed of sub-objects such as arms, legs, hands, etc. Similarly, if Jonathan Ross is facing

the camera side on, then one arm and leg may be obscured; thus, even though they are present (semantically), they are hidden rather than visible. Object properties serve to describe the object and represent additional information about the object. For example, properties of the Jonathan Ross object may include his hair colour, age, height, key career history, and so on. Because properties are also represented, other features of the objects may be used as filtering criteria, whether these are visible or not. Objects may themselves exist within a structural hierarchy thereby enabling inheritance of properties from higher level objects.

Representations of spatial relationships among objects within the semantic content model enable filtering concerning the relative location of objects, rather than the absolute location that comes from a co-ordinate based representation. This enables reference to be made to the relationships between objects within the content and can provide for 3D spatial representations and hidden object representations, which are difficult to derive from 'flat' two-dimensional co-ordinate representations. For example, if Jonathan Ross is shown as sitting to the right of Eddie Izzard, then the spatial relationship between the two objects would be 'right' or 'East', or we could represent the inverse using inverse spatial relationships such as 'left' or 'West'. Spatial relationships have a duration due to the object motion. Thus, the spatial relationships between two objects may differ over time within the same segment.

Fig. 2 depicts the semantic content aspects described above and their inter-relationships. When these semantic content aspects are integrated within a scheme, the resulting model can be used to model semantic content for most mainstream domains and user groups and, consequently, facilitate filtering for those domains and user groups. Potentially suitable domains would include all applications with added-value content; for instance, entertainment-on-demand, multimedia news, and organisational content.

(Fig. 2 near here)

3. MODELLING SEMANTIC VIDEO CONTENT WITH COSMOS-7

COSMOS-7 is compliant with Part 5 of the MPEG-7 standard [23] which contains the Multimedia Description Scheme (MDS) tools. It defines the semantic aspects described in the previous section in MPEG-7 format and, as a result, it allows interoperability of its semantic content models across a multitude of platforms and applications. Hence, it is an MPEG-7 Multimedia Content Description Interface [24].

MPEG-7 specifies a range of *top-level types*, whose purpose is to encapsulate complete descriptions of multimedia content and metadata. Thus, each top-level type descends description tools relevant for a particular media type, such as image or video, or additional metadata, such as describing usage or the scheme itself. The former top-level types are descended from the top-level abstract type *ContentDescriptionType* whereas the latter are descended from the *ContentManagementType* abstract type. Both are further descended from a *CompleteDescriptionType* which can be used to encapsulate a complete scheme. Fig. 3 illustrates the complete MPEG-7 type hierarchy.

(Fig. 3 near here.)

COSMOS-7 uses two of the MPEG-7 top-level types both of which are descended from the *ContentAbstractionType* abstract type, which itself is descended from the *ContentDescriptionType* abstract type:

- The *SemanticDescriptionType* is used to group the content modelled information regarding objects and object properties (including object inter-relationships), events, and temporal relationships among events.
- The *ContentEntityType* is used to group the content modelling information regarding spatial relationships among objects and provides identifiers for media files and segmentation.

While conforming to the MPEG-7 format, the arrangement of content modelled information contained by the above two top-level types and within a tight but rich content structure that is prescribed by COSMOS-7 (as detailed in the next sub-sections) negates the need for generic parsing and validation. Nevertheless, COSMOS-7-produced models may still be validated for correctness using any MPEG-7 validation service. Self containment of MPEG-7 data is beginning to emerge as a popular means for defining re-usable MPEG-7 structures and extensions [35].

Description tools descended from the top-level types may be either a *Description Scheme (DS)*, a *Descriptor (D)* or a *Datatype*. DSs describe entities or relationships pertaining to multimedia content and specify the structure and semantics of their components, which may be other DSs, Ds, or datatypes. Examples of DSs specified in the standard include the *Event DS*, which enables content-modelling of events, and the *Object DS*, which enables content-modelling of objects. Ds describes a feature, attribute, or group of attributes of multimedia

content. Examples of Ds specified in the standard include the *TemporalMask D*, which enables the description of media segments, and the *MediaFormat D*, which can be used to describe the file format and the coding parameters of the media. Datatypes are basic reusable types employed by DSs and Ds. Examples of datatypes specified in the standard include the *Reference datatype*, which is used to refer to a part of a description (i.e. an instance of a D or DS), and the *mediaTimePoint datatype*, which is used to specify a point in time, e.g. within a video segment. While MPEG-7 provides a plethora of description tools, their implementation in any modelling and filtering scheme should be dependent upon requirements. Hence, COSMOS-7 implements only those tools from MPEG-7 Part 5 which are necessary for modelling the semantic aspects discussed in the previous section, and enabling their filtering thereafter. However, since COSMOS-7 is, as MPEG-7 is, an open modelling scheme, further MPEG-7 tools may be implemented if and when necessary without the need to re-implement the entire scheme from scratch. Fig. 4 shows the key MPEG-7 description tools that are used within COSMOS-7 and illustrates how they are inter-related. The figure also shows the use of a number of *Classification Schemes (CSs)*, such as the *TemporalRelation CS*. CSs are a type of DS that define a set of standard terms describing some domain. The following sections discuss the use of MPEG-7 tools in COSMOS-7 in more detail.

(Fig. 4 near here)

3.1 Events

Events are modelled in COSMOS-7 using the *Event DS* but one or more events are grouped using the *Semantic DS* and given a suitable *id* and *Label*. This enables related events, such as those related to specific objects, to be proximate and utilised efficiently. Each event uses several elements within the *Event DS* as follows. Each event is given a *Label* that describes

the event. The *MediaOccurrence* element uses the *MediaInformationRef* element to refer to the appropriate video segment and *TemporalMaskType*s within the *MediaInformationRef* element to define appropriate masks on the referred to video segment. Events are related to objects through the use of the *SemanticRelation CS* and the *agent* relation. The *agent* relation is defined as follows [23]: *A agent B* if and only if B is an agent of or performs or initiates A. The inverse, *agentOf*, is used when modelling objects (see later). An example of COSMOS-7 events is given in the Appendix.

3.2 Temporal relationships among events

Temporal relationships are grouped using the *Semantic DS* where "Temporal-Relationships" is the *id* and "Temporal Relationships" is the *Label*. The relationships themselves are represented as a graph via the *Graph DS*, with each temporal relationship defined using the *TemporalRelation CS*. The use of a graph in this way enables straightforward traversal of the relationships. COSMOS-7 supports the 14 binary and n-ary temporal relations specified in the MDS. These are given in Table 1, which also depicts the sub-set that map onto those originally specified by Allen [36]. The use of these relationships enables events to be logically ordered according to the requirements of a particular domain, and independently of their physical order within the video resource. This is particularly necessary where a single event can be apparent within multiple, disparate segments. Consequently, users may filter according to temporal locality in relation to known events, e.g. "Show me everything that happened before this event occurred". An example of COSMOS-7 temporal relationships is given in the Appendix.

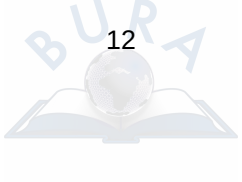
(Table 1 near here)

3.3 Objects and object properties

As is the case with events and temporal relationships among events, each object and its properties are grouped together using the *Semantic DS* with an appropriate *id* and *Label*. There are three parts to the COSMOS-7 representation of objects and object properties within the *Semantic DS*:

- Objects are related to events through the use of the *SemanticRelation CS* and the *agentOf* relation. The *agentOf* relation is defined as follows [23]: *B agentOf A* if and only if B is an agent of or performs or initiates A.
- The *Object DS* is used to content model the objects. It includes elements describing the composition of an object from sub-objects. Thus, each sub-object is incorporated into this representation within COSMOS-7. The *MediaOccurrence* elements (together with temporal masks) are used as previously described to relate specific video segments that reflect the occurrences of objects.
- The *SemanticState DS* is used to content model object properties. To enable *MediaOccurrences* to be related to specific object properties, each property is modelled as a separate *SemanticState*. The *AttributeValuePair* element is used to specify the properties themselves. The scheme makes no restrictions on which properties may be modelled, so long as they conform to the given structure. It can therefore be tailored for particular domains and user groups.

Objects are related to each other through the use of the *SemanticRelation CS*. The *Semantic DS* is used to group all object relationships together with a graph representing the relationships through the *specializes* relation (the inverse relation is *generalizes*). An example of objects and object properties in COSMOS-7 is provided in the Appendix.



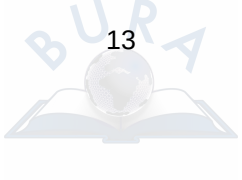
3.4 Spatial relationships among objects

Spatial relationships are specified on a per video stream basis. The video is segmented so that spatial-relationship inheritance may be deployed, whereby segments that are sub-segments of a segment *x*, inherit segment *x*'s spatial relationships as well as specifying their own spatial relationships. COSMOS-7 therefore uses the *VideoSegment DS* and creates segments using the *VideoSegmentTemporalDecompositionType* and the *VideoSegmentSpatioTemporalDecompositionType*. *MediaTime* elements are used to delineate the segments. The *MovingRegion DS* is used to identify objects as regions within the video which are related to the content-modelled objects specified in the previous section through the use of the *SemanticRef* element (a *Reference* data type) that the *VideoSegment DS* inherits from the *Segment DS* in the MDS. The spatial relationships themselves are specified using the *SpatialRelation CS* within a graph. Table 2 lists the spatial relations that are specified in MPEG-7 and which are employed within COSMOS-7. An example of spatial relationships among objects in COSMOS-7 is given in the Appendix.

(Table 2 near here)

3.5 COSMOSIS

There is currently a dearth of authoring tools to work with MPEG-7 data, though some are slowly emerging, e.g. [37, 38]. COSMOSIS is a front-end application, developed using version 1.4.2 of the Java 2 Platform Standard Edition (J2SE) and version 2.1.1 of the Java Media Framework (JMF). It supports the creation and editing of COSMOS-7 files in MPEG-7 format. On-line validation is supported through the service provided by NIST [39]. Four tabbed panes cater for each of the semantic video content aspects identified in Fig. 2, namely: events, temporal relationships among events, objects and object properties, and spatial



relationships among objects. The separation of semantic video content aspects via tabbed panes enables relevant information from each of the main sections of the file to be displayed together while also permitting comfortable cross-referencing between sections. All references and labels may be freely decided upon by the user, according to their requirements and the conventions of the organisation and domain of application. Prior references, such as those to segments and objects, may be chosen from a list of automatically-generated identifiers, which updates as new references are created. This maintains consistency and reduces the risk of errors during data creation and modification. COSMOSIS also supports the direct viewing and editing of the source file for users that wish to work in this manner, and changes made in this mode are also automatically reflected in the tabbed panes (with suitable error warnings issued when potential inconsistencies arise).

A screenshot of COSMOSIS is shown in Fig. 5, which depicts editing of the events given in the example in the previous section. A particular event grouping identifier is selectable from a drop down list which causes the corresponding label to be automatically displayed. The events contained within the selected event grouping may then be chosen from the second drop down list. Again, the corresponding label is displayed, together with the event's specified agent (if applicable) and the media occurrence masks. All displayed information is editable via the New, Modify and Delete buttons.

(Fig. 5 near here)

A Java Media Framework player (shown to the right) runs simultaneously with COSMOSIS to enable the location of segment start and duration parameters. These may be entered

directly by the user or retrieved in real-time from the media player. The player also serves to help the content creator visually verify and validate the semantic content information.

4. FILTERING SEMANTIC VIDEO CONTENT WITH COSMOS-7

A user will very often only be interested in certain video content, e.g. when watching a soccer game the user may only be interested in goals and free kicks. Identifying and retrieving subsets of video content in this way requires user preferences for content to be stated, such that content within a digital video resource may then be filtered against those preferences. While new approaches are emerging, such as those based on agents [40-42], filtering in video streams usually employs content-based filtering methods, which analyse the features of the material so that these may then be filtered against the user's content requirements [28, 29, 40, 43-48]. During the analysis, a set of key attributes is identified for each feature and the attribute values are then populated. The attribute values can either be simple preference ratings, such as those based on the Likert scale, or they can be weighted values, to indicate the relative success of the match of the attribute value to those preferences expressed by the user. The former type of values are commonly used in recommender systems, where content is recommended to the user based on their prior history, while the latter type of values tend to be employed in content-based retrieval domains, where the user is 'hunting' or 'exploring' content more specifically.

Content-based filtering is consequently most suitable when a computer can easily analyse the entities and where entity suitability is not subjective. Content-based filtering is also most effective when the content being analysed has features homogenous with the user's content

preferences, since heterogeneous features across the two domains would give rise to incompatibilities, i.e. not comparing like with like. This problem is particularly apparent when new content is introduced into the resource, since this may contain new features not previously contemplated. The use of standards which are comprehensive in their consideration of semantic aspects, such as MPEG-7, overcomes this problem.

Using a content-based filtering approach, COSMOS-7 matches the various semantic content of the digital video resource to the specified filter criteria of the user. This is achieved by building the content filter from the part of the COSMOS-7 content model that the user content requirements directly map on to. In this way, only content that relates directly to the preferred content requirements of the user is analysed and chosen. The filter is revisited every time new content is added or the user requirements change. Content filters are specified using the **FILTER** keyword together with one or more of the following: **EVENT**, **TMPREL**, **OBJ**, **SPLREL**, and **VIDEO**. These may be joined together using the logical operator AND. This specifies what the filter is to return. Consequently, the output of the filtering process may be either semantic content information and or video segments containing the semantic content information. The criteria are specified using the **WHERE** keyword together with one or more of the following clauses: **EVENT**, **TMPREL**, **OBJ**, and **SPLREL** clauses. Clauses may be joined together using the logical operators AND, NOT and OR. Clauses are terminated with semi-colons and grouped with braces. The clauses will be examined in further detail shortly, after a discussion of the Filtering Manager for COSMOS-7 which embodies these clauses to enable filtering of both the COSMOS-7 content model and related video segments.

4.1 Filtering Manager for COSMOS-7

Like COSMOSIS, the Filtering Manager is a Java-based front-end application. Its purpose is to construct and execute filters for COSMOS-7. A UML use case diagram is shown in Fig. 6 and serves as an overview of all activities. The two main use cases are as follows:

- *Specify Filter*, where the user creates the filter and specifies the filter criteria. This use case is extended by: (a) one or more *Specify … Filter* use cases, according to which combination of semantic aspects are being filtered on, (b) a *Specify Video* use case, when the user wishes to include video segments in the filter results, (c) a *Save Filter* use case, when the user wishes to save the created filter, and (d) an *Open Filter* use case, when the user wishes to load a previously specified filter.
- *Viewed Filtered Results*, where the results of the executed filter are made available to the user. This use case is extended by: (a) *Save Filtered Results*, where the user can save the filtering results for later display, (b) *Open Filtered Results*, where the user can view previously saved filtering results, and (c) *View Results by Video Segment*, where the user may view a list of video segments matching the filtering criteria, which in turn is extended by *Display Video Segment*, where the user may playback the video segments within the displayed results. The use case includes *View Results by Semantic Content*, where the user views the semantic content matching the filtering criteria, which in turn is extended by *Display COSMOS-7 Semantic Content*, where the user may view the portion of the COSMOS-7 semantic content model matching the filtering criteria.

(Fig. 6 near here)

The main dialog of the COSMOS-7 Filtering Manager is divided into two areas: the left pane enables the user to specify the filtering conditions, while the right pane displays the results. This is illustrated in Fig. 7(a), which shows the results of the example filter given above. The filtering pane requires the user to select one or more aspects that they wish to filter on, which maps on to the filter keywords presented earlier. Various condition clauses may then be constructed in the 'WHERE' area. To ensure full flexibility, the clauses are stated in text, but the various clause keywords may be selected and inserted from a drop-down list box. Four tabbed panes enable the specification of the event, temporal relationship, object and object property, and spatial relationship clauses. Fig. 7(b) shows a temporal relationship type being inserted. Once specified, filters may be saved for later retrieval. Once a filter is executed, the results are displayed in the right-hand area, within two tabbed panes. The first pane contains the semantic content results, displayed as *label*s within a tree structure, grouped by semantic aspect. These may be selected for viewing in COSMOS-7 format, if required, as is illustrated in the top right half of Fig. 7(a). The second pane contains the video segment results (if these were chosen by the user as part of the filter condition). Again, these are shown within a tree structure, grouped by the *idref* of the video segments. This is shown in Fig. 7(c). Each video segment may selected and viewed by the user and the complete set of results may be saved for later retrieval, if required.

(Fig. 7 near here)

4.2 Event clauses

The **EVENT** clauses enable event-related filtering criteria to be specified on the COSMOS-7 representation. The following clauses are supported:

- *eventGroupingId = <eventGroupingId>*, which enables filtering according to a particular event grouping, as specified by the *id* parameter of the <Semantics> tag of the COSMOS-7 representation.
- *eventGroupingLabel = <eventGroupingLabel>*, which enables filtering according to a particular event grouping label, as specified by the corresponding <Label> tag of the COSMOS-7 representation.
- *eventId = <eventId>*, which enables filtering according to a particular event, as specified by the id parameter of the <SemanticBase> tag of the COSMOS-7 representation.
- *eventLabel = <eventLabel>*, which enables filtering according to a particular event label, as specified by the corresponding <Label> tag of the COSMOS-7 representation
- *mediaOccurrence = <mediaId> | <mediaTimePoint> | <mediaRange>*, which enables filtering according to a particular media segment. The <mediaId>, <mediaTimePoint> and <mediaRange> enable correlation to the information specified within the <MediaOccurrence> section of the COSMOS-7 event representation.
- *agent = <objectId>*, which enables filtering according to event agents and correlates to the <Relation> tag within the COSMOS-7 event representation.

4.3 Temporal relationship clauses

The **TMPREL** clauses enable temporal-relationship-related filtering criteria to be specified on the COSMOS-7 representation. The following clauses are supported:

- *temporalRelationshipType = <temporalRelationshipType>*
- *temporalRelationshipSource = <objectId>*
- *temporalRelationshipTarget = <objectId>*

These all correspond with the parameters of the <Relation> element within the temporal relationship graph of the COSMOS-7 representation. However, in the case of <temporalRelationshipType>, an abbreviated version is used that omits the "urn:mpeg:mpeg7:cs:TemporalRelationCS:2001:" prefix and just states the actual temporal relationship, e.g. "precedes", since the prefix is redundant in this context.

4.4 Object and object property clauses

The **OBJ** clauses enable filtering criteria to be specified on the COSMOS-7 representation that are related to objects and their properties, respectively. The following clauses are supported:

- *objectGrouping = <objectGroupingId>*
- *objectGroupingLabel = <objectGroupingLabel>*
- *objectId = <objectId>*
- *objectLabel = <objectLabel>*

These four clauses are used to specify the same criteria as those in the **EVENT** clauses but for object groupings and objects, rather than event groupings and events.

- *subObjectId = <subObjectId>*
- *subObjectLabel = <subObjectLabel>*

These clauses are the same as the objectId and objectLabel clauses but enabling filtering of objects that are sub-objects of another object specifically. This enables the filter to match

only those objects that are sub-objects of another object and will reject objects that have the same label but are main objects.

- *mediaOccurrence = <mediaId> | <mediaTimePoint> | <mediaRange>*, which has the same functionality as that of the **EVENT** clause of the same name.
- *agentOf = <objectId>*, which has the the same functionality as that of the agent **EVENT** clause, but with refers meaning.

The next five clauses enable filtering according to object property identifiers, labels, attributes, units and values, respectively:

- *objectProperty = <objectPropertyId>*
- *objectPropertyLabel = <objectPropertyLabel>*
- *objectPropertyAttribute = <objectPropertyAttribute>*
- *objectPropertyUnit = <objectPropertyUnit>*
- *objectPropertyValue = <objectPropertyValue>*

These final three clauses have the same functionality as the **TMPREL** clauses but deal with the object hierarchy relationships:

- *objectHierarchyType = <objectHierarchyType>*
- *objectHierarchySource = <objectId>*
- *objectHierarchyTarget = <objectId>*

Again, the “urn:mpeg:mpeg7:cs:SemanticRelationCS:2001:” prefix is omitted due to its redundancy in this context.

4.5 Spatial relationship clauses

The **SPLREL** clauses deal with spatial-relationship-related filtering criteria, via the following clauses:

- *video* = <*videoId*> | <*mediaTimePoint*> | <*mediaRange*>
- *videoSegment* = <*videoSegmentId*> | <*mediaTimePoint*> | <*mediaRange*>

The above two clauses have the same functionality as that of the mediaOccurrence **EVENT** and **OBJ** clause, however here they relate to the <Video> and <VideoSegment> sections of the COSMOS-7 representation.

- *spatialRelationshipType* = <*spatialRelationshipType*>
- *spatialRelationshipSource* = <*objectId*>
- *spatialRelationshipTarget* = <*objectId*>

The above three clauses have the same functionality as the **TMPREL** clauses but deal with spatial relationships. The “urn:mpeg:mpeg7:cs:SpatialRelationCS:2001:” prefix is omitted for the same reasons as previously given.

The final clause enables filtering according to moving region identifiers or the objects related to them in COSMOS-7:

- *movingRegion = <movingRegionId> | <objectId>*

4.6 Example filter

Consider that the user wishes to filter both events and video where:

- events occur during the first 30 seconds or during the 59th minute of the video footage
- the agent of the events is the squirrel object
- the events precede the squirrel sleeping event
- the video footage that reflects the events features trees and lakes

The resultant filter would look like so:

```
FILTER EVENT AND VIDEO WHERE {
      EVENT {
            mediaOccurrence  =  T00:00:00-T00:00:30  OR  mediaOccurrence  =
T00:59:00;
            agent = Squirrel-O;
      }
      TMPREL {
            temporalRelationshipType = precedes;
            temporalRelationshipTarget = Squirrel-Sleeping-EV;
      }
      SPLREL {
            movingRegion = Tree-MR AND movingRegion = Lake-MR;
      }
}
```

This would return all events and all references to video that match the above criteria.

5. RELATED WORK

Although MPEG-7 is quite a recent standard, research literature discussing the application of the standard is beginning to emerge within a number of areas. One area is concerned with

MPEG-7 authoring tools. The MPEG-7 Metadata Authoring Tool [38] uses a structured dialog-driven interface to create and edit the MPEG-7 descriptions. The descriptions may also be viewed in an hierarchic tree-like structure. Mdefi [49] is an authoring tool for multimedia presentations that also enables the creation and editing of MPEG-7 descriptions. These descriptions are then integrated into a synchronisation model which is similar to SMIL (Synchronised Multimedia Integration Language).

A related area concerns the generation of the semantic content model itself. VideoAL [50] is an IBM-developed system that generates MPEG-7 content descriptions from MPEG-1 video sequences via a combination of shot segmentation, region segmentation, annotation, feature extraction, model learning, classification and XML rendering. The system utilises IBM's VideoAnnEx annotation tool [37]. Similarly, iFinder [6] automatically extracts MPEG-7 content descriptions from audio and video content and uses a client/server-based retrieval engine to enable search and retrieval of short video segments in multimedia databases.

Converse to generating the semantic content model from the video, a number of research projects have looked into using the semantic content model to assemble short video sequences. Fonseca and Pereira [51] address automatic video summarisation from MPEG-7 descriptions. Echigo et al. [52] use MPEG-7 content descriptions to generate 'video digests' for each video stream, which summarise that stream according to the user's preferences. The Automatic Video Summarizing Tool [53] uses MPEG-7 visual descriptors to generate video previews that permit non-linear access via thumbnails and searching based on shot similarities.

Work addressing search and retrieval of MPEG-7 includes the following. SOLO [54] is an MPEG-7 search tool that is designed to search across multiple distributed databases rather than a single local database. Queries are structured within the scope of MPEG-7 descriptors. Graves and Lalmas [55] describe an inference-network-based model for video retrieval, where the document network is constructed using MPEG-7 content descriptions. The descriptions are grouped into structural aspects (video, shots and scenes), conceptual aspects (video, scene and shot content), and contextual aspects (context information about the position of conceptual content within the document). Martínez et al. [56] address the issue within the context of ubiquitous computing. Because some media characteristics will differ according to the device and network, they create a number of template-based variations of the MPEG-7 content descriptions. A searching system and a personalisation system provide examples of their approach. Walker [57] describes an MPEG-7-enabled prototype for content-based navigation of television programmes. Finally, Westermann and Klas [58] investigate current XML database solutions and critically assess their suitability with regards to database-style management of MPEG-7 content descriptions.

Last but not least, a wide range of research has looked at building additional functionality into the standard. Zhang et al. [59] propose a Video Stream Description Language for TV. The language specifies operations and processes that may be used in conjunction with MPEG-7 descriptions for annotation, browsing and retrieval. Rogge et al. [60] propose a modification to MPEG-7 so that code can be added to and executed from the semantic content model. Akrivas et al. [61] propose a method that applies fuzzy relational operations and fuzzy rules to an MPEG-7 content model in an attempt to reduce the need for human intervention during the model creation process. Charlesworth and Garner [62] describe the spoken content component of MPEG-7 and some experiments they have undertake as to its effectiveness.

Vakali et al. [63] argue the importance of reconsidering issues related to high-level multimedia modelling and representation in the light of MPEG-7. They propose a modelling database scheme based on multi-level video modelling and semantic video classification, which sits on top of MPEG-7. Smith [64] has looked at the harmonisation of all multimedia content modelling standards, such as SMPTE and DCMI, through an MPEG-7 harmonised model.

6. SUMMARY AND CONCLUSIONS

The accuracy and reliability of filtered content to user requirements relies on two factors: the richness and depth of detail used in the creation of the content model and the level of content-dependency of the filtering techniques employed. Furthermore, the interoperability of user-driven information retrieval, especially across platforms, is greatly enhanced if the underlying process is standardised. This is achieved through the various parts of the MPEG-7 standard and in our case through MPEG-7 Part 5. Part 5 allows the modelling of multimedia content using a rich set of description schemes that describe both low-level syntactic and high-level semantic concepts that can be used by any MPEG-7 compliant scheme. However, while MPEG-7 prescribes the format and structure of data, it does not specify how that data may be processed or used. COSMOS-7 is one approach to utilising MPEG-7 both for semantic-content-based modelling and semantic-content-based filtering, which has evolved from our pre-MPEG-7 work on content modelling.

COSMOS-7 is organised around a core set of semantic aspects which embody a subset of the full MPEG-7 specification. This not only enables COSMOS-7 modelling and filtering to be applicable to a broad range of applications, but also makes it easier for content modellers and

filterers without specific experience in MPEG-7 to get to grips with this important standard without intimidation. At the same time, because COSMOS-7 is open, further MPEG-7 functionality may be added as necessary, and as requirements and experience, grow. While modelling and filtering is undertaken manually in the implemented systems, the time taken tends to be reasonable because the semantic-aspects-based organisation of COSMOS-7 tends to be aligned with the common way in which semantic content is perceived by humans at a high level in terms of 'who, what, where and when'. While fully automated and reliable semantic content analysis and extraction is beyond the state-of-the-art at this time, the addition of some automated techniques to COSMOSIS which could work in conjunction with the human user is one possible further area of research and development with a view to further reducing the time taken. Filtering assumes some familiarity with the content model to which it is to be applied and also the COSMOS-7 format in general. This can be viewed positively since, in general, most effective filtering is achieved when there is understanding rather than ignorance of the underlying processes; typically because those processes may be better exploited.

We have experimented with modelling and filtering a range of footage using COSMOS-7 and have observed the resultant models and filters to generally be concise and for performance to scale reasonably, even when modelling greater than typical levels of detail. We have observed some variations according to the type of video content, however. For example, news footage tends to consist of a number of key events, which form the main headline stories, split into a number of small sub-events together with a number of key, 'newsworthy' objects. Events tend to exhibit a high degree of sequentiality within news footage which consequently simplifies the complexity of the temporal relationships. Similarly, focused non-studio footage tends to exhibit a lower degree of spatial variation between objects than some other types of video

content, which produces a consequent reduction in the complexity of spatial relationships. Conversely, for movie footage, many more events and objects become relevant and their interrelationships become more complex, leading to a rise in the required level of detail and thus the size of the derived content model and any applied filters. Generally, one may observe a broad correlation between the detail of the content model and the detail of created filters.

Currently, the model and filtering interfaces serve the needs of advanced users and our future work intends to address this by adding an additional layer to simplify these interfaces for novice users. We are also currently investigating how to improve the visualisation and exploration of the filtering results within the Filtering Manager. In the longer term, with the openness of COSMOS-7 and the evolution of another MPEG standard, MPEG-21, the considerations of ontologies [65] and content adaptation [66, 67] come to the fore.

7. APPENDIX: EXAMPLE COSMOS-7 IMPLEMENTATIONS

This appendix provides example implementations in COSMOS-7 for each of the four semantic aspects. The first example given below depicts events within COSMOS-7. It models a group of Squirrel events that consist of an Eating event and a Sleeping event. Both events are depicted by the WholeSquirrel-V-VS video segment within the subintervals given. The events are related to the Squirrel-O object through the *Relation* elements.

```
<Semantics id="Squirrel-Events">
```

```
      <Label>
            <Name>Squirrel events</Name>
      </Label>
      <SemanticBase xsi:type="EventType" id="Squirrel-Eating-EV">
            <Label>
                  <Name>Eating</Name>
            </Label>
            <MediaOccurrence>
                  <MediaInformationRef idref="WholeSquirrel-V-VS" />
                  <Mask xsi:type="TemporalMaskType">
                        <SubInterval>
                              <MediaTimePoint>T00:00:00</MediaTimePoint>
                              <MediaDuration>PT4M</MediaDuration>
                        </SubInterval>
                        <SubInterval>
                              <MediaTimePoint>T00:06:05</MediaTimePoint>
                              <MediaDuration>PT3M</MediaDuration>
                        </SubInterval>
                  </Mask>
            </MediaOccurrence>
            <Relation
type="urn:mpeg:mpeg7:cs:SemanticRelationCS:2001:agent"
   target="#Squirrel-O"/>
      </SemanticBase>
      <SemanticBase xsi:type="EventType" id="Squirrel-Sleeping-EV">
            <Label>
                  <Name>Sleeping</Name>
            </Label>
            <MediaOccurrence>
                  <MediaInformationRef idref="WholeSquirrel-V-VS" />
                  <Mask xsi:type="TemporalMaskType">
                        <SubInterval>
                              <MediaTimePoint>T00:12:00</MediaTimePoint>
                              <MediaDuration>PT2M</MediaDuration>
                        <SubInterval>
                              <MediaTimePoint>T01:00:00</MediaTimePoint>
                              <MediaDuration>PT2M</MediaDuration>
                        </SubInterval>
                        </SubInterval>
                  </Mask>
            </MediaOccurrence>
            <Relation
type="urn:mpeg:mpeg7:cs:SemanticRelationCS:2001:agent"
   target="#Squirrel-O"/>
      </SemanticBase>
</Semantics>
```

The second example shows temporal relationships in COSMOS-7. It depicts that the squirrel Eating event occurs before the squirrel Sleeping event. Note that there is no need to relate these events to the Squirrel-O object since this has already been done in the respective events.

```
<Semantics id="Temporal-Relationships">
      <Label>
            <Name>Temporal Relationships</Name>
      </Label>
      <Graph>
```

```
            <Relation
type="urn:mpeg:mpeg7:cs:TemporalRelationCS:2001:precedes"
                  source="#Squirrel-Eating-EV”
                  target="#Squirrel-Sleeping-EV"/>
      </Graph>
</Semantics>
```

The third example demonstrates the modelling of objects and object properties within COSMOS-7. The example depicts a squirrel object that consists of several sub-objects, one of which is its tail (please refer to the screenshot shown in Fig. 5). The height property is shown for the squirrel and specified with a value of 16cm. It is related to two media segments that illustrate the squirrel's height within the WholeSquirrel-V-VS segment. The object hierarchy is shown at the bottom of the example and is given the specific *id* "Object-Hierarchy" and the specific *Label* "Object Hierarchy". The relationship shows that the Squirrel-O object is a specialisation of the Rodent-O object, thus reflecting the fact that squirrels are rodents.

```
<Semantics id="Squirrel-SEM">
      <Label>
            <Name>Description of squirrel</Name>
      </Label>
      <SemanticBase xsi:type="ObjectType" id="Squirrel-O">
            <Label>
                  <Name>Squirrel</Name>
            </Label>
            <MediaOccurrence>
                  <MediaInformationRef idref="WholeSquirrel-V-VS" />
                  <Mask xsi:type="TemporalMaskType">
                        <SubInterval>
                              <MediaTimePoint>T00:07:00</MediaTimePoint>
                              <MediaDuration>PT3M</MediaDuration>
                        </SubInterval>
                        <SubInterval>
                              <MediaTimePoint>T00:15:00</MediaTimePoint>
                              <MediaDuration>PT9M</MediaDuration>
                        </SubInterval>
                  </Mask>
            </MediaOccurrence>
            <Relation
type="urn:mpeg:mpeg7:cs:SemanticRelationCS:2001:agentOf"
                  target="#Squirrel-Eating-EV" />
            <Relation
type="urn:mpeg:mpeg7:cs:SemanticRelationCS:2001:agentOf"
                  target="#Squirrel-Sleeping-EV" />
            <Object id="Squirrel-O-tail">
                  <Label>
                        <Name>Tail</Name>
                  </Label>
                  <MediaOccurrence>
                  <MediaInformationRef idref="WholeSquirrel-V-VS" />
```

```
                  <Mask xsi:type="TemporalMaskType">
                        <SubInterval>
                              <MediaTimePoint>T00:03:30</MediaTimePoint>
                              <MediaDuration>PT2M</MediaDuration>
                        </SubInterval>
                  </Mask>
                  </MediaOccurrence>
            </Object>
      </SemanticBase>
      <SemanticBase    xsi:type="SemanticStateType"    id="Squirrel-O-Props-
Height">
            <Label>
                  <Name>Height</Name>
            </Label>
            <MediaOccurrence>
                  <MediaInformationRef idref="WholeSquirrel-V-VS" />
                  <Mask xsi:type="TemporalMaskType">
                        <SubInterval>
                              <MediaTimePoint>T00:00:00</MediaTimePoint>
                              <MediaDuration>PT6M</MediaDuration>
                        </SubInterval>
                  </Mask>
            </MediaOccurrence>
            <AttributeValuePair>
                  <Attribute>
                        <Name>Height</Name>
                  </Attribute>
                  <Unit>
                        <Name>cm</Name>
                  </Unit>
                  <IntegerValue>16</IntegerValue>
            </AttributeValuePair>
      </SemanticBase>
</Semantics>

<Semantics id="Object-Hierarchy">
      <Label>
            <Name>Object Hierarchy</Name>
      </Label>
      <Graph>
            <Relation
type="urn:mpeg:mpeg7:cs:SemanticRelationCS:2001:specializes"
                  source="#Squirrel-O" target="#Rodent-O"/>
      </Graph>
</Semantics>
```

The fourth and final example depicts spatial relationships among objects. The example below shows a video, with *id* Squirrel-V, that has a segment defined on its entirety, *id* WholeSquirrel-V-VS. Two moving regions are then defined with *id*s Squirrel-MR and Tree-MR which are related to the Squirrel-O and Tree-O objects through the *SemanticRef* elements. The entire segment is then split into two segments. Both of these segments inherit the two moving regions and specify spatial relationships. The spatial relationship shown depicts that the squirrel is above the tree.

```
<Video id="Squirrel-V">
      <MediaLocator>
            <MediaUri>
                  squirrel003.mpg
            </MediaUri>
      </MediaLocator>
      <MediaTime>
            <MediaTimePoint>T00:00:00</MediaTimePoint>
            <MediaDuration>PT1M30S</MediaDuration>
      </MediaTime>
      <TemporalDecomposition gap="false" overlap="false">
            <VideoSegment id="WholeSquirrel-V-VS">
                  <Relation
type="urn:mpeg:mpeg7:cs:SpatialRelationCS:2001:below"
                        source="#Squirrel-MR" target="#Tree-MR"/>
                  <SpatioTemporalDecomposition gap="true" overlap="false">
                        <MovingRegion id="Squirrel-MR">
                              <SemanticRef idref="Squirrel-O"/>
                        </MovingRegion>
                        <MovingRegion id="Tree-MR">
                              <SemanticRef idref="Tree-O"/>
                        </MovingRegion>
                        </SpatioTemporalDecomposition>
                  <TemporalDecomposition>
                        <VideoSegment id="WholeSquirrel-V-VS-1">
                              <Semantic>
                                    <Label>
                                          <Name>Spatial
relationships</Name>
                                    </Label>
                                    <Graph>
                                          <Relation

            type="urn:mpeg:mpeg7:cs:SpatialRelationCS:2001:above"
                                    source="#Squirrel-MR" target="#Tree-MR"/>
                                          </Graph>
                                    </Semantic>
                                    <MediaTime>

      <MediaTimePoint>T00:00:00</MediaTimePoint>
                                          <MediaDuration>PT0M15S</MediaDuration>
                                    </MediaTime>
                        </VideoSegment>
                        <VideoSegment id="WholeSquirrel-V-VS-2">
                              <MediaTime>

      <MediaTimePoint>T00:00:15</MediaTimePoint>
                                          <MediaDuration>PT0M30S</MediaDuration>
                                    </MediaTime>
                        </VideoSegment>
                  </TemporalDecomposition>
            </VideoSegment>
      </TemporalDecomposition>
</Video>
```

8. REFERENCES

[1]. Huang, Q., Puri, A., and Liu, Z. (2000) Multimedia Search and Retrieval: New Concepts, System Implementation, and Application. *IEEE Transactions on Circuits and Systems for Video Technology*, **10**(5), 679-692.

[2]. Al-Safadi, L. and Getta, J. (2001) Semantic content-based retrieval for video documents. In S.M. Rahman (ed.), *Design and Management of Multimedia Information Systems: Opportunities & Challenges*. IDEA Group Publishing, Hershey, PA. 165-200.

[3]. Bolle, R.M., Yeo, B.-L., and Yeung, M.M. (1998) Video query: research directions. *IBM Journal of Research & Development*, **42**(2), 233-252.

[4]. Day, Y.F., Khokhar, A., Dagtas, S., and Ghafoor, A. (1999) A multi-level abstraction and modeling in video databases. *Multimedia Systems*, **7**(5), 409-423.

[5]. Jaimes, A., Echigo, T., Teraguchi, M., and Satoh, F. (2002) Learning personalized video highlights from detailed MPEG-7 metadata. *Proceedings of the 2002 IEEE International Conference on Image Processing (ICIP-02)*, Vol. 1, Rochester, NY, 22-25 September, pp. 133-136. IEEE Press, Piscataway, NJ.

[6]. Löffler, J., Biatov, K., Eckes, C., and Köhler, J. (2002) iFinder: an MPEG-7-based retrieval system for distributed multimedia content. *Proceedings of the 10th ACM International Conference on Multimedia (MM02)*, Juan-les-Pins, France, December 1-6, pp. 431-435. ACM Press, New York, NY.

[7]. Leung, C.H.C. (2001) Semantic-based retrieval of visual data. In M.S. Lew (ed.), *Principles of Visual Information Retrieval*. Springer-Verlag, New York, NY. 297-318.

[8]. Tusch, R., Kosch, H., and Böszörmenyi, L. (2000) VIDEX: an integrated generic video indexing approach. *Proceedings of the 8th ACM International Conference on*

Multimedia (MM00), Los Angeles, CA, 30 October-3 November, pp. 448-451. ACM Press, New York, NY.

[9]. Wang, Y., Liu, Z., and Huang, J.-C. (2000) Multimedia content analysis: using both audio and visual clues. *IEEE Signal Processing*, **17**(6), 12-36.

[10]. Vendrig, J.W., M. (2002) Interactive adaptive movie annotation. *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME 2002)*, Vol. 1, Lausanne, Switzerland, August 26-29, pp. 93-96 vol.1. IEEE Press, Piscataway, NJ.

[11]. van Beek, P., Smith, J.R., Ebrahim, T., Suzuki, T., and Askelof, J. (2003) Metadata-driven multimedia access. *IEEE Signal Processing*, **20**(2), 40-52.

[12]. Zhao, R. and Grosky, W.I. (2002) Bridging the semantic gap in image retrieval. In T.K. Shih (ed.), *Distributed Multimedia Databases: Techniques and Applications*. IDEA Group Publishing, Hershey, PA. 14-36.

[13]. Adami, N., Bugatti, A., Leonardi, R., Migliorati, P., and Rossi, L.A. (2001) The ToCAI description scheme for indexing and retrieval of multimedia documents. *Multimedia Tools and Applications*, **14**(2), 153-173.

[14]. Bryan-Kinns, N. (2000) VCMF: A Framework for Video Content Modelling. *Multimedia Tools and Applications*, **10**(1), 23-45.

[15]. Golshani, F. and Dimitrova, N. (1998) A language for content-based video retrieval. *Multimedia Tools and Applications*, **6**(3), 289-312.

[16]. Grosky, W.I. (1997) Managing multimedia information in database systems. *Communications of the ACM*, **40**(12), 72-80.

[17]. Hartley, E., Parkes, A.P., and Hutchison, A.D. (2000) A conceptual framework to support content-based multimedia applications. In H. Leopold and N. Garcia (eds.), *Lecture Notes in Computer Science*. Vol. 1629, Springer, New York, NY. 297-315.

[18]. Hunter, J. and Armstrong, L. (1999) A comparison of schemas for video metadata representation. *Computer Networks*, **31**, 1431-1451.

[19]. Leung, C.H.C. and Sutanto, D. (1999) Multimedia data modeling and management for semantic content retrieval. In B. Furht (ed.), *Handbook of Multimedia Computing*. CRC Press, Boca Raton, FL. 43-54.

[20]. Marcus, S. and Subrahmanian, V.S. (1996) Foundations of multimedia database systems. *Journal of the ACM*, **43**(3), 474-523.

[21]. Rowe, L.A., Boreczky, J.S., and Eads, C.A. (1994) Indices for user access to large video databases. *Proceedings of the Storage and Retrieval for Image and Video Database II, Proceedings of SPIE*, Vol. 2185, San Jose, CA, February 6-10, pp. 150-161. SPIE Press, Bellingham, WA.

[22]. Hunter, J. (2001) An Overview of the MPEG-7 Description Definition Language (DDL). *IEEE Transactions on Circuits and Systems for Video Technology*, **11**(6), 765-772.

[23]. ISO/IEC 15938-5 (2002) Information Technology – Multimedia Content Description Interface – Part 5: Multimedia Description Schemes. International Standard. International Organisation for Standardisation, Geneva, Switzerland.

[24]. Koenen, R. and Pereira, F. (2000) MPEG-7: a standardised description of audiovisual content. *Signal Processing: Image Communication*, **16**, 5-13.

[25]. Martínez, J.M. (2002) MPEG-7 Overview, ISO/IEC JTC1/SC29/WG11-N4980. International Organisation for Standardisation, Geneva, Switzerland.

[26]. Sikora, T. (2001) The MPEG-7 Visual Standard for Content Description - An Overview. *IEEE Transactions on Circuits and Systems for Video Technology*, **11**(6), 696-702.

[27]. Belkin, N.J. and Croft, W.B. (1992) Information filtering and information retrieval: two sides of the same coin? *Communications of the ACM*, **35**(12), 29-38.

[28]. Ferman, A.M., Errico, J.H., van Beek, P., and Sezan, M.I. (2002) Content-based filtering and personalization using structured metadata. *Proceedings of the 2nd ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL '02)*, Portland, Oregon, July 14-18, p. 393. ACM Press, New York, NY.

[29]. Naphande, M.R. and Huang, T.S. (2001) A probabilistic framework for semantic video indexing, filtering, and retrieval. *IEEE Transactions on Multimedia*, **3**(1), 141-151.

[30]. Tseng, B.L., Lin, C.-Y., and Smith, J.R. (2004) Using MPEG-7 and MPEG-21 for personalizing video. *IEEE Multimedia*, **11**(1), 42-53.

[31]. Agius, H.W. and Angelides, M.C. (1999) COSMOS - Content Oriented Semantic Modelling Overlay Scheme. *The Computer Journal*, **42**(3), 153-176.

[32]. Agius, H.W. and Angelides, M.C. (2000) A method for developing interactive multimedia from their semantic content. *Data & Knowledge Engineering*, **34**(2), 165-187.

[33]. Agius, H.W. and Angelides, M.C. (2001) Modelling content for semantic-level querying of multimedia. *Multimedia Tools and Applications*, **15**(1), 5-37.

[34]. Ferman, A.M., Takalp, A.M., and Mehrotra, R. (1998) Effective content representation for video. *Proceedings of the 1998 International Conference on Image Processing (ICIP 98)*, Vol. 3, Chicago, IL, October 4-7, pp. 521-525. IEEE Press, Piscataway, NJ.

[35]. Döller, M., Kosch, H., Dörflinger, B., Bachlechner, A., and Blaschke, G. (2002) Demonstration of an MPEG-7 Multimedia Data Cartridge. *Proceedings of the 10th*

ACM International Conference on Multimedia (MM02), Juan-les-Pins, France, December 1-6, pp. 85-86. ACM Press, New York, NY.

[36]. Allen, J.F. (1983) Maintaining knowledge about temporal intervals. *Communications of the ACM*, **26**(11), 832-843.

[37]. IBM VideoAnnEx website. *http://www.research.ibm.com/VideoAnnEx/.*

[38]. Ryu, J., Sohn, Y., and Kim, M. (2002) MPEG-7 metadata authoring tool. *Proceedings of the 10th ACM International Conference on Multimedia (MM02)*, Juan-les-Pins, France, December 1-6, pp. 267-270. ACM Press, New York, NY.

[39]. NIST MPEG-7 Validation Service. *http://m7itb.nist.gov/M7Validation.html.*

[40]. Good, N., Schafer, J., Konstan, J., Borchers, A., Sarwar, B., Herlocker, J., and Riedl, J. (1999) Combining collaborative filtering with personal agents for better recommendations. *Proceedings of the 16th National Conference on Artificial Intelligence (AAAI-99)*, Orlando, FL, 18-22 July, pp. 439-446. AAAI Press, Menlo Park, CA.

[41]. Wenyin, L., Chen, Z., Lin, F., Zhang, H., and Ma, W.-Y. (2003) Ubiquitous media agents: a framework for managing personally accumulated multimedia files. *Multimedia Systems*, **9**(2), 144 - 156.

[42]. Zhou, W., Vellaikal, A., and Dao, S. (2001) Cooperative content analysis agents for online multimedia indexing and filtering. *Proceedings of the Third International Symposium on Cooperative Database Systems for Advanced Applications (CODAS 2001)*, Beijing, China, 23-24 April, pp. 118-122. IEEE Press, Piscataway, NJ.

[43]. Angelides, M.C. (2003) Multimedia content modelling and personalization. *IEEE Multimedia*, **10**(4), 12-15.

[44]. Eirinaki, M. and Vazirgiannis, M. (2003) Web mining for web personalization. *ACM Transactions on Internet Technology*, **3**(1), 1-27.

[45]. Kuflik, T. and Shoval, P. (2000) Generation of user profiles for information filtering - research agenda. *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Athens, Greece, July 24-28, pp. 313-315. ACM Press, New York, NY.

[46]. Specht, T.K.G. and Kahabka, T. (2000) Information filtering and personalisation in databases using Gaussian curves. *Proceedings of the IEEE International Databases Engineering and Applications Symposium (IDEAS'00)*, Yokohama, Japan, 18-20 September. IEEE Press, Piscataway, NJ.

[47]. van Meteren, R. and van Someren, M. (2000) Using content-based filtering for recommendation. *Proceedings of the Machine Learning in the New Information Age: MLnet / ECML2000 Workshop*, Barcelona, Spain, 30 May.

[48]. Wallace, M.S., G. (2002) Towards a context aware mining of user interests for consumption of multimedia documents. *Proceedings of the 2002 IEEE International Conference on Multimedia and Expo (ICME 2002)*, Vol. 1, Lausanne, Switzerland, pp. 733-736. IEEE Press, Piscataway, NJ.

[49]. Tran-Thuong, T. and Roisin, C. (2003) Multimedia modeling using MPEG-7 for authoring multimedia integration. *Proceedings of the 5th ACM SIGMM International Workshop on Multimedia Information Retrieval*, Berkeley, CA, November 7, pp. 171-178. ACM Press, New York, NY.

[50]. Lin, C.-Y., Tseng, B.L., Naphade, M., Natsev, A., and Smith, J.R. (2003) VideoAL: a novel end-to-end MPEG-7 video automatic labeling system. *Proceedings of the IEEE International Conference on Image Processing 2003*, Vol. 3, Barcelona, Spain, September 14-17, pp. III-53-56. IEEE Press, Piscataway, NJ.

[51]. Fonseca, P.M. and Pereira, F. (2004) Automatic video summarization based on MPEG-7 descriptions. *Signal Processing: Image Communication*, **19**, 685-699.

[52]. Echigo, T., Masumitsu, K., Teraguchi, M., Etoh, M., and Sekihuchi, S. (2001) Personalized delivery of digest video managed on MPEG-7. *Proceedings of the 2001 International Conference on Information Technology: Coding and Computing*, Las Vegas, NV, USA, 2-4 April, pp. 216-220. IEEE Press, Piscataway, NJ.

[53]. Lee, J.-H., Lee, G.-G., and Kim, W.-Y. (2003) Automatic Video Summarizing Tool using MPEG-7 Descriptors for Personal Video Recorder. *IEEE Transactions on Consumer Electronics*, **49**(3), 742-749.

[54]. Lay, J.A. and Ling, G. (2000) SOLO: an MPEG-7 optimum search tool. *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME 2000)*, Vol. 2, New York, NY, 30 July-2 August, pp. 777-780. IEEE Press, Piscataway, NJ.

[55]. Graves, A. and Lalmas, M. (2002) Video retrieval using an MPEG-7 based inference network. *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Tampere, Finland, August 11-15, pp. 339-346. ACM Press, New York, NY.

[56]. Martínez, J.M., González, C., Fernández, O., Garcia, C., and de Ramón, J. (2002) Towards universal access to content using MPEG-7. *Proceedings of the 10th ACM International Conference on Multimedia (MM02)*, Juan-les-Pins, France, December 1-6, pp. 199-202. ACM Press, New York, NY.

[57]. Walker, T. (2000) Content-based navigation of television programs using MPEG-7 description schemes. *Proceedings of the 2000 IEEE International Conference on Consumer Electronics (ICCE)*, Los Angeles, CA, 13-15 June, pp. 272-273. IEEE Press, Piscataway, NJ.

[58]. Westermann, U. and Klas, W. (2003) An analysis of XML database solutions for the management of MPEG-7 media descriptions. *ACM Computing Surveys*, **35**(4), 331-373.

[59]. Zhang, W., Yaginuma, Y., and Sakauchi, M. (2001) TV drama management system based on the video stream description language. *Proceedings of the 6th International Symposium on Signal Processing and its Applications (ISSPA 2001)*, Kuala Lumpur, Malaysia, 13-16 August, pp. 186-189. IEEE Press, Piscataway, NJ.

[60]. Rogge, B., Van de Walle, R., Lemahieu, I., and Philips, W. (2001) MPEG-7 based dynamic metadata. *Proceedings of the 2001 IEEE International Conference on Multimedia and Expo (ICME 2001)*, Tokyo, Japan, August 22-25, pp. 165-168. IEEE Press, Piscataway, NJ.

[61]. Akrivas, G., Stamou, G.B., and Kollias, S. (2004) Semantic Association of Multimedia Document Descriptions Through Fuzzy Relational Algebra and Fuzzy Reasoning. *IEEE Transactions on Systems, Man and Cybernetics - Part A: Systems and Humans*, **34**(2), 190-196.

[62]. Charlesworth, J.P.A. and Garner, P.N. (2000) Spoken Content Metadata and MPEG-7. *Proceedings of the 2000 ACM Workshops on Multimedia*, Los Angeles, CA, October 30-November 3, pp. 81-84. ACM Press, New York, NY.

[63]. Vakali, A., Hacid, M.-S., and Elmagarmid, A. (2004) MPEG-7 based description schemes for multi-level video content classification. *Image and Vision Computing*, **22**, 367-378.

[64]. Smith, J. (2003) MPEG-7 Multimedia Content Description Standard. In D. Feng, W.C. Siu, and H.J. Zhang (eds.), *Multimedia Information Retrieval and Management: Technological Fundamentals and Applications*. Springer, New York, NY.

[65]. Hunter, J. (2003) Enhancing the semantic interoperability of multimedia through a core ontology. *IEEE Transactions on Circuits and Systems for Video Technology*, **13**(1), 49-58.

[66]. Libsie, M. and Kosch, H. (2002) Content adaptation of multimedia delivery and indexing using MPEG-7. *Proceedings of the 10th ACM International Conference on Multimedia (MM02)*, Juan-les-Pins, France, December 1-6, pp. 644-646. ACM Press, New York, NY.

[67]. Vetro, A. (2004) MPEG-21 digital item adaptation: Enabling universal multimedia access. *IEEE Multimedia*, **11**(1), 84-87.

List of tables

List of figures

Table 1: Temporal relationships used in COSMOS-7

MPEG-7 relation	MPEG-7 inverse relation	Allen's relation	Conceptual example
Binary			
precedes	*follows*	before	AAA BBB
coOccurs	*coOccurs*	equal	AAA BBB
meets	*metBy*	meets	AAABBB
overlaps	*overlappedBy*	overlaps	AAAA BBBBBB
strictDuring	*strictContains*	during	AAA BBBBBB
starts	*startedBy*	starts	AAA BBBBBB
finishes	*finishedBy*	finishes	AAA BBBBBB
contains	*during*	-	*Any of the above 3*
N-ary			
contiguous	-	-	AAABBBCCC
sequential	-	-	AAA BBBCCC
coBeing	-	-	AAA BBBBBBB CC
coEnd	-	-	AAA BBBBBBB CC
parallel	-	-	AAAAAA BBBBB CCCCCCC
overlapping	-	-	AAAAA BBBBBBB CCCCCCC

Table 2: Spatial relationships used in COSMOS-7

MPEG-7 relation	MPEG-7 inverse relation
south	*north*
west	*east*
northwest	*southeast*
southwest	*northeast*
left	*right*
below	*above*
over	*under*

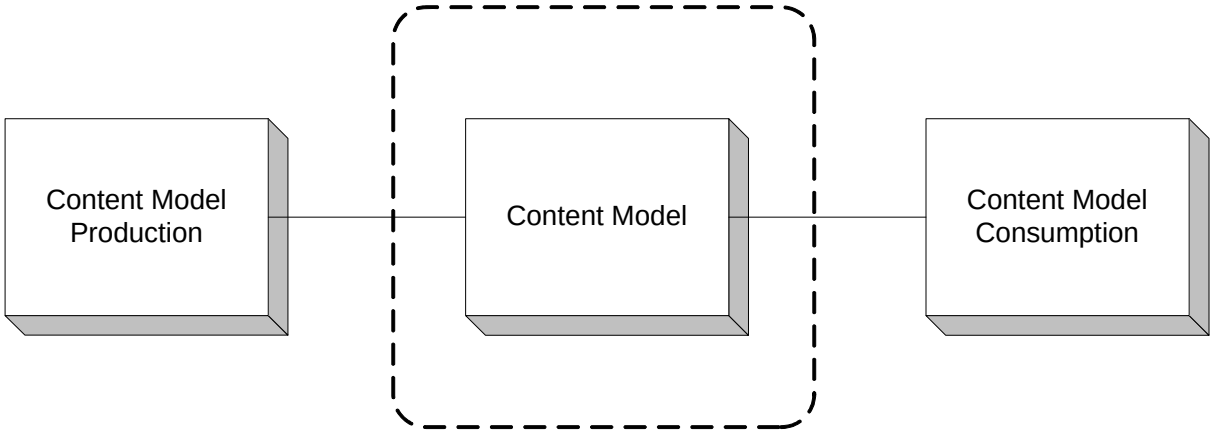


Fig. 1: MPEG-7 focus.

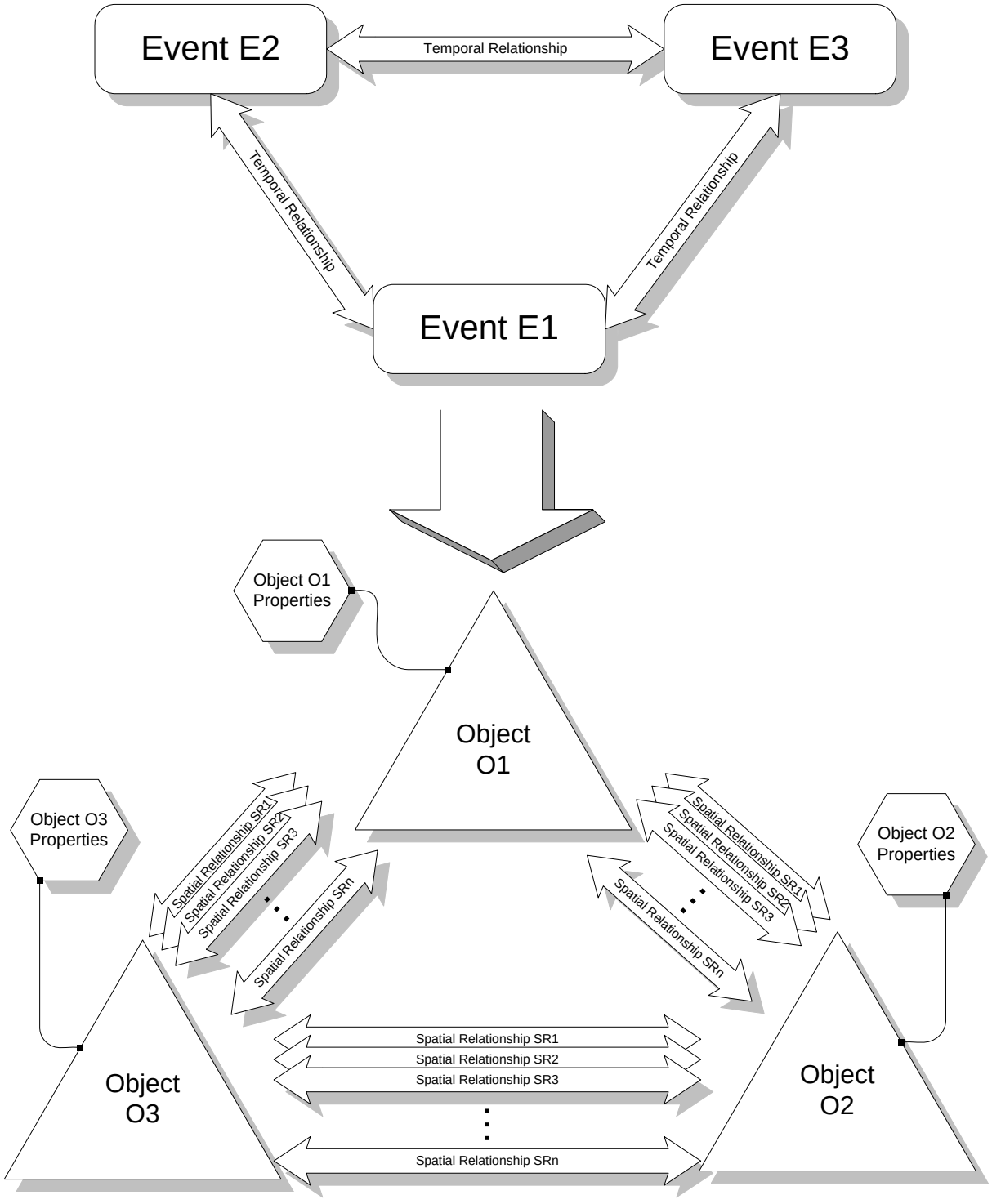


Fig. 2: Semantic content aspects for video and their inter-relationships.

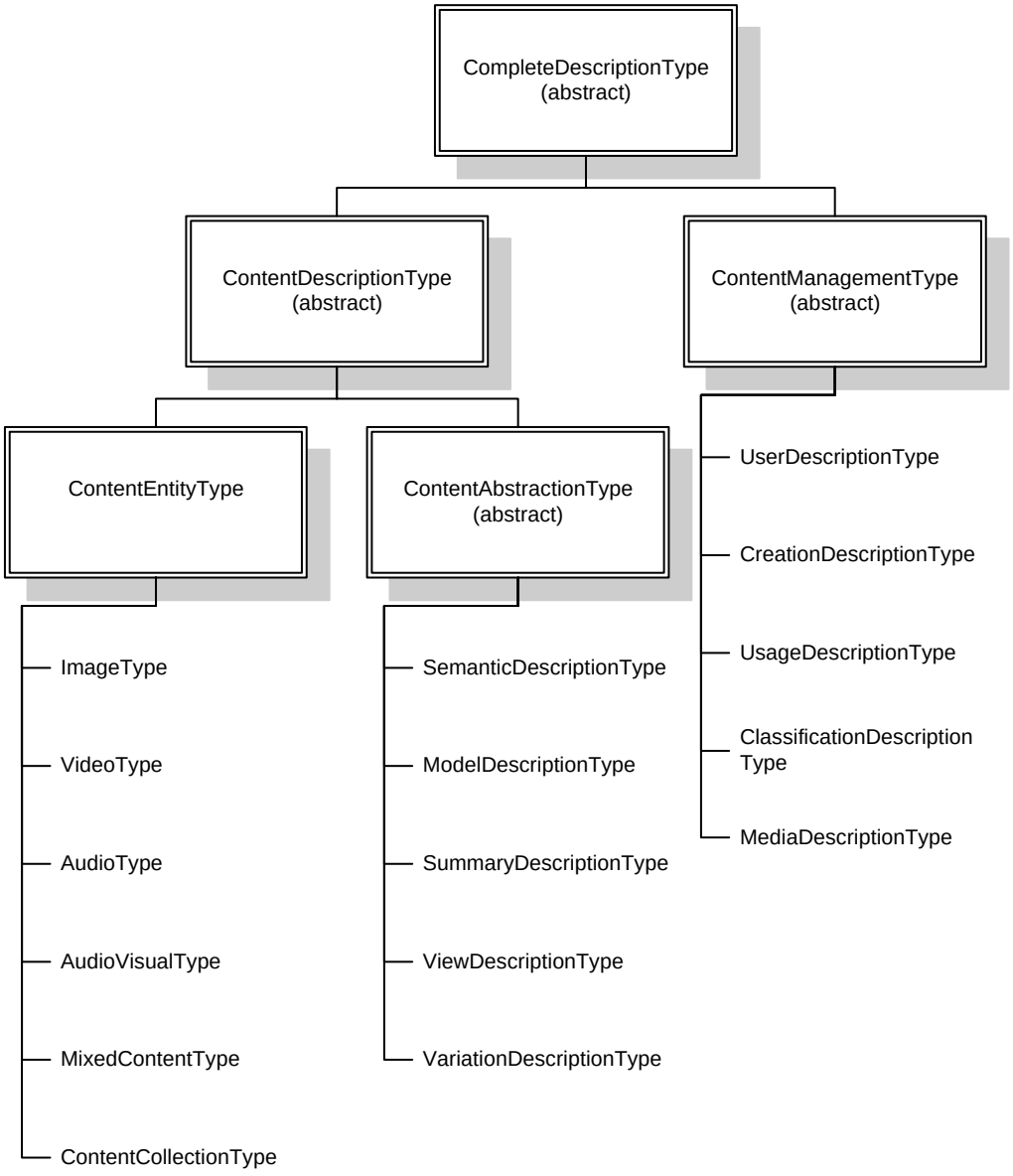


Fig. 3: MPEG-7 top-level types.

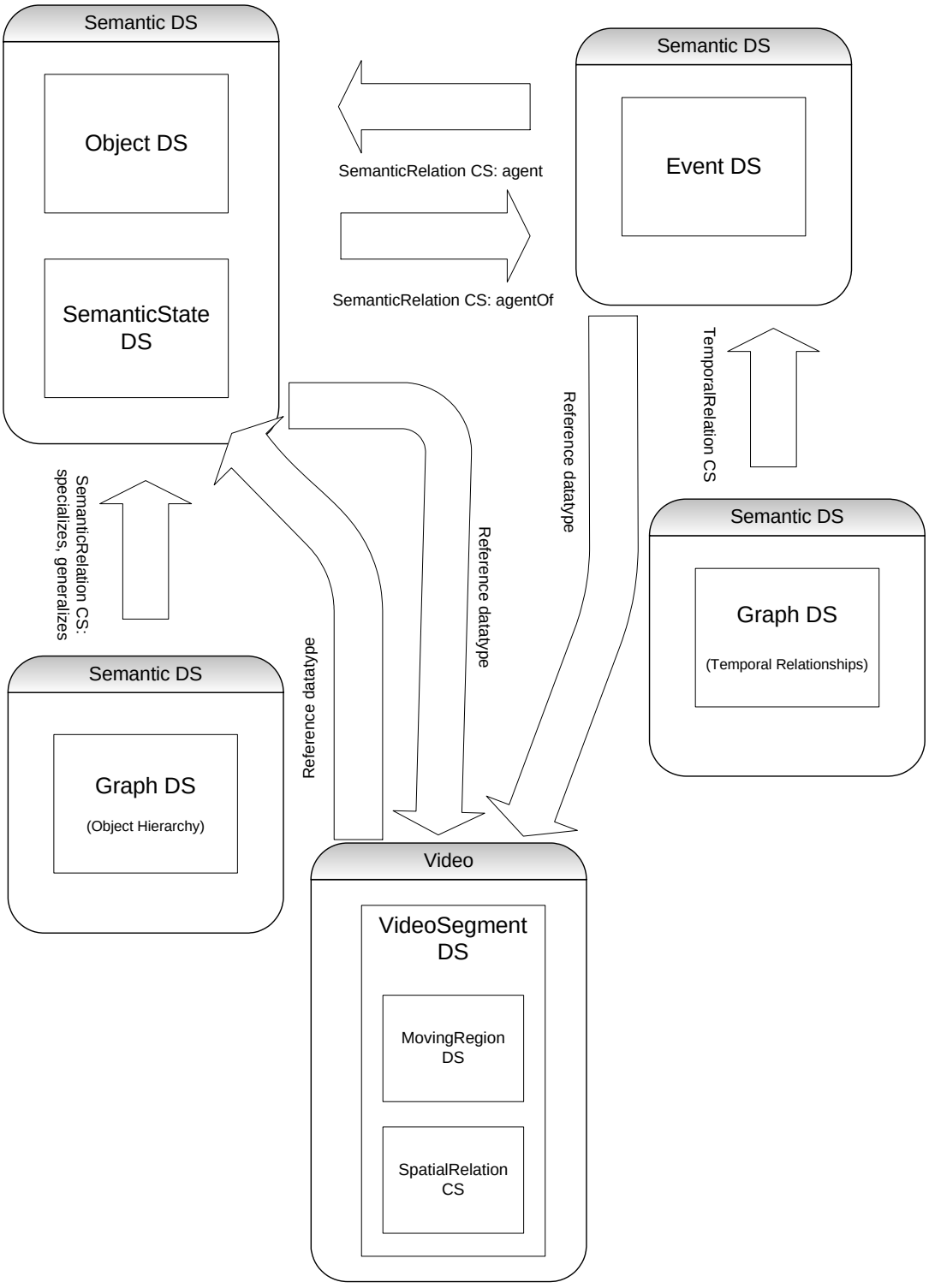


Fig. 4: Key MPEG-7 description tools used in COSMOS-7.

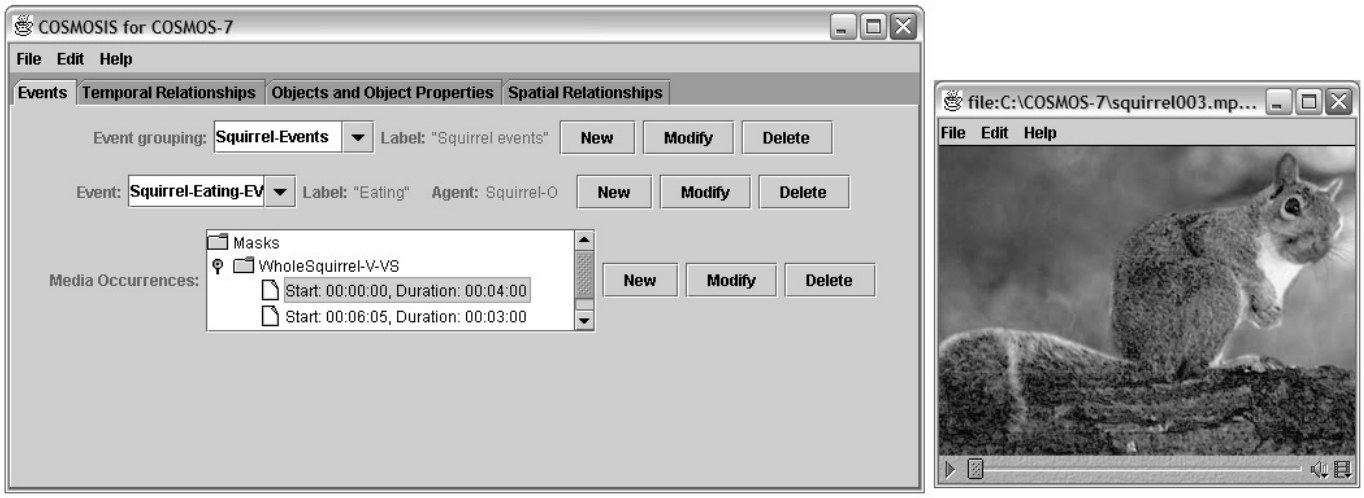


Fig. 5: COSMOSIS.

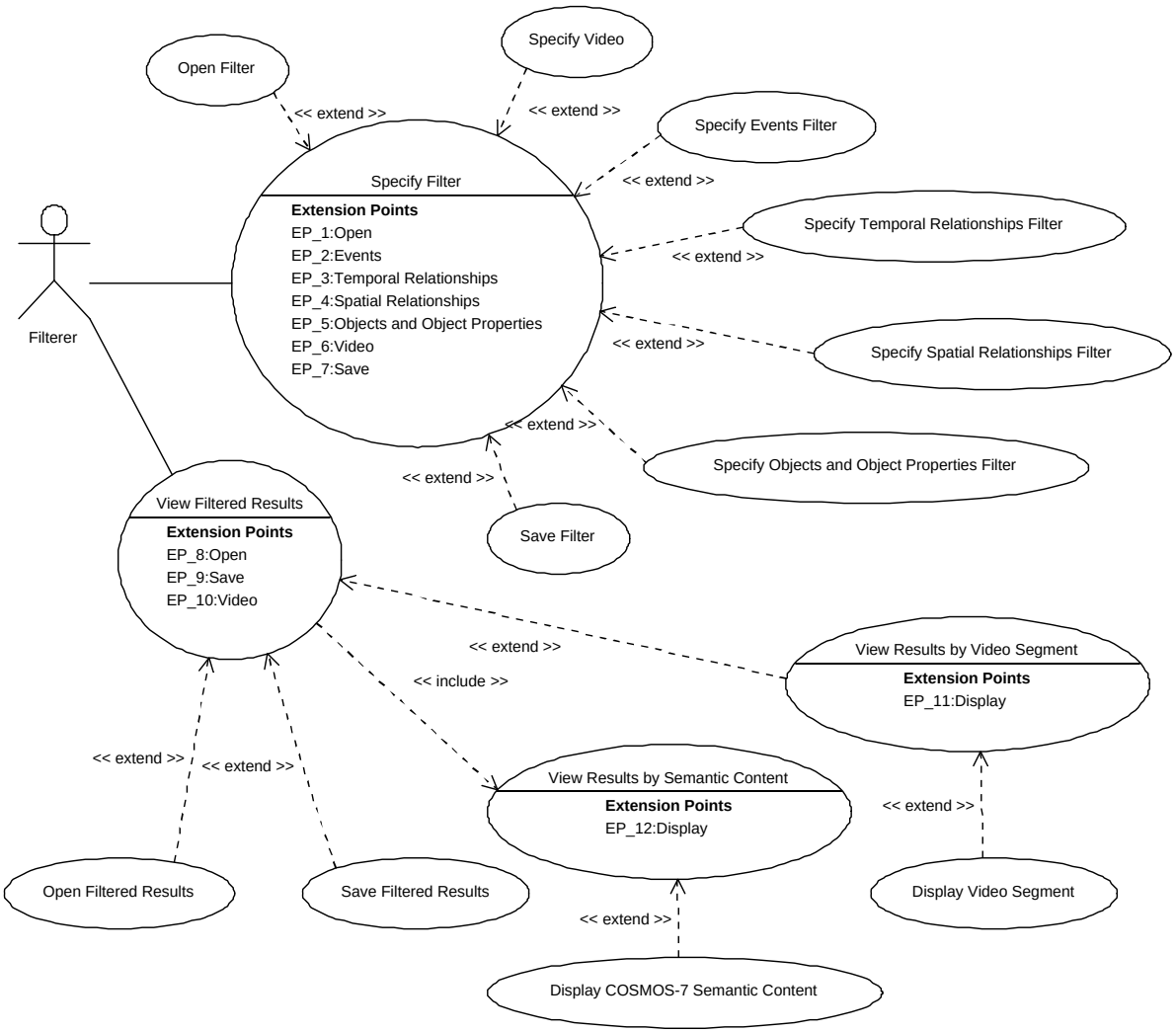


Fig. 6: COSMOS-7 Filtering Manager use cases.

Fig. 7: COSMOS-7 Filtering Manager: (a) returned COSMOS-7 semantic content results, (b) inserting a temporal relationship into the filter, and (c) returned video segments results.