2005년 8월
석사학위논문

도메인 지식 구축에 의한 의미적 비디오 이벤트 표현

조선대학교 대학원

전 자 계 산 학 과

송 단

도메인 지식 구축에 의한 의미적 비디오 이벤트 표현

Semantic Video Event Description Assisted by Building Domain Knowledge

2005년 8월 일

조선대학교 대학원

전 자 계 산 학 과

송 단

도메인 지식 구축에 의한 의미적 비디오 이벤트 표현

지도교수 김 판 구

이 논문을 이학석사학위신청 논문으로 제출함.

2005년 4월 일

조선대학교 대학원

전자계산학과

송 단

송단의 석사학위논문을 인준함

위원장 조선대학교 교수 _____

위　원 조선대학교 교수 _____

위　원 조선대학교 교수 _____

2005년　5월　일

조선대학교 대학원

Contents

Figure Contents

Table Contents

ABSTRACT

Semantic Video Event Description Assisted by Building Domain Knowledge

Dan Song
Adviser : Prof. Kim Pan-Koo Ph.D.
Department of Computer Science,
Graduate School of Chosun University

The MPEG-7 visual standard under development specifies content-based descriptors that allow users or agents (or search engines) to measure similarity in images or video based on visual criteria, and can be used to efficiently identify, filter, or browse images or video based on visual content. More specifically, MPEG-7 specifies color, texture, object shape, global motion, or object motion features for this purpose. This paper outlines the aim, methodologies, and broad details of the MPEG-7 standard development for video event description. Except for assistant by the MPEG-7 tools, we also put forward a novel method for video event analysis and description based on the Domain Knowledge in this paper. Semantic concepts in the context of the video event are described in one specific domain enriched with qualitative attributes of the semantic objects, multimedia processing approaches and domain independent factors: low level features (pixel color, motion vectors and spatio-temporal relationship). In order to apply large-scale semantic knowledge in vision problems effectively, catering the naive user's retrieval and index processing with semantic (human) language, a few major issues are

resolved in this paper. Firstly, how can we get the semantic shot for the specific Domain Knowledge? The former existing algorithm has been adopted to solve the problem. Secondly, what visual observables should be collected? This is usually dependent on the problem domain. Here, we consider one shot of the billiard game clip as the specific Domain Knowledge. Thirdly, how can these observables be translated into the semantic representation, we are from two aspects to expose that issue: Firstly, video event representation using MPEG-7 high level descriptors which was defined in the MPEG-7 XML files. Secondly, video object motion analysis with the help of the MPEG-7 low level descriptors(video object motion detection and moving trajectory analysis). In addition, the most important contribution in this work is exploiting the video object ontology to map the MPEG-7's high-level descriptors to low level features descriptors which have been defined in the MPEG's logical structure.

1. Introduction

Nowadays, the rapid increase of the available amount of multimedia information has revealed an urgent need for developing intelligent methods for understanding and managing the conveyed information. To face such challenges developing faster hardware or more sophisticated algorithms has become insufficient. Rather, a deeper understanding of the information at the semantic level is required [1]. This results in a growing demand for efficient methods for extracting semantic information from such content. Although new multimedia standards, such as MPEG-4 and MPEG-7 [2], provide the needed functionalities in order to manipulate and transmit objects and metadata, their extraction, and that most importantly at a semantic level, is out of the scope of the standards and is left to the content developer. Extraction of low-level features and object recognition are important phases in developing multimedia database management systems [3].

We have got some significant results in the literature recently, with successful implementation of several prototypes [4]. However, the lack of precise models and formats for object and system representation and the high complexity of multimedia processing algorithms make the development of fully automatic semantic multimedia analysis and management systems a challenging task. Correspondingly, referring to the low level, the features of the moving object have become more and more concerned. Because moving objects refer to semantic real-world entity definitions that are used to denote a coherent spatial region and be automatically computed by the continuity of spatial low-level features. This is due to the difficulty that often mentioned as the semantic gap, in capturing concepts mapped into a set of image and low-level features that can be automatically extracted from the raw video data. The use of Domain Knowledge is probably the only way by which higher level semantics can be incorporated into techniques that capture the semantic concepts. Although there were some works about the domain application, they are restricted by the gap [5,6]. So, in this paper, a novel method for video event analysis and description based on the

specific Domain Knowledge was proposed.

The remainder of the paper is organized as follows: Section 2－The related work introduces the MPEG－7 fundamental knowledge. Section 3－Video analysis based on Domain Knowledge. Also, this section will introduce the video event representation using MPEG－7 high level descriptors and show the video object description with low level features. Domain Knowledge ontology infrastructure building, experiment results and query explanatory will be demonstrated in Section 4. After these comprehensive explanations, we will conclude in Section 5.

2. Related Work

MPEG-7, formally named "Multimedia Content Description Interface," is the standard that describes multimedia content so users can search, browse, and retrieve that content more efficiently and effectively than they could using today's mainly text-based search engines. It's a standard for describing the features of multimedia content. It provides the world's most comprehensive set of audiovisual descriptions. These descriptions are based on the catalog(title, creator, rights), semantic (the who, what, when, where information), and structural (the color histogram -measurement of the amount of color-associated with an image or the timbre of a recorded instrument) features of the AV content, and are leveraged on the AV data representation defined by MPEG-1, -2, and -4. MPEG-7 also uses XML Schema as the language of choice for content description.

However, MPEG-7 won't standardize the (automatic) extraction of AV descriptions/features. Nor will it specify the search engine (or any other program) that can make use of the description. It's left up to the creativity and innovation of search engine companies to manipulate and massage the MPEG-7-described content into search indexes that can be used by their browser and retrieval tools. Typical query examples enabled by MPEG-7 include:

- Audio: I want to search for songs by humming or whistling a tune or, using an excerpt of Pavarotti's voice, get a list of Pavarotti's records and video clips in which Pavarotti sings or simply makes an appearance.
- Graphics: Sketch a few lines on a screen and get a set of images containing similar graphics, logos, and ideograms.
- Image: Define objects, including color patches or textures, and get examples from which you select items to compose your image. Or check if your company logo was advertised on a TV channel as contracted.
- Video: Allow mobile phone access to video clips of goals scored in a soccer game, or automatically search and retrieve any unusual

movements from surveillance videos.

In addition, it's important to note that MPEG-7 addresses many different applications in many different environments. This means that it needs to provide a flexible and extensible framework for describing audiovisual data. MPEG-7 defines a library of methods and tools for many types of multimedia applications. It standardizes:

- A set of descriptors: A descriptor(D) is a representation of a feature that defines the syntax and semantics of the feature representation.
- A set of description schemes: A description scheme(DS) specifies the structure and semantics of the relationships between its components, which may be both descriptors and description schemes.
- A language that specifies description schemes (and possibly descriptors), the Description Definition Language(DDL): It also allows for the extension and modification of existing description schemes. MPEG-7 adopted XML Schema Language as the MPEG-7 DDL. However, the DDL requires some specific extensions to XML Schema Language to satisfy all the requirements of MPEG-7. These extensions are currently being discussed through liaison activities between MPEG and W3C, the group standardizing XML.
- One or more ways(textual, binary) to encode descriptions: A coded description is one that's been encoded to fulfill relevant requirements such as compression efficiency, error resilience, and random access.

In MPEG-7 the main tools for describing the structure of AV content are segment entities, segment features, and segment relations. In Figure1 you can see an example of a structural description of a video sequence that's described using such tools. The description is made up of three video segments, two moving regions, one segment relation, and several two-segment decompositions.

The root video segment corresponds to the entire video sequence. The remaining two video segments are formed from a segment decomposition of the root segment. The last two video segments are then related by the segment relation "Before". With a further decomposition of one of these segments, you obtain two moving regions that correspond to the

two objects: "player", "cue ball","object ball". How do them interact with each others? What does MPEG-7 DDL syntax look like? Let's look at this paper that was taking advantage of MPEG-7 syntax extended with XML files for Scheme Description.

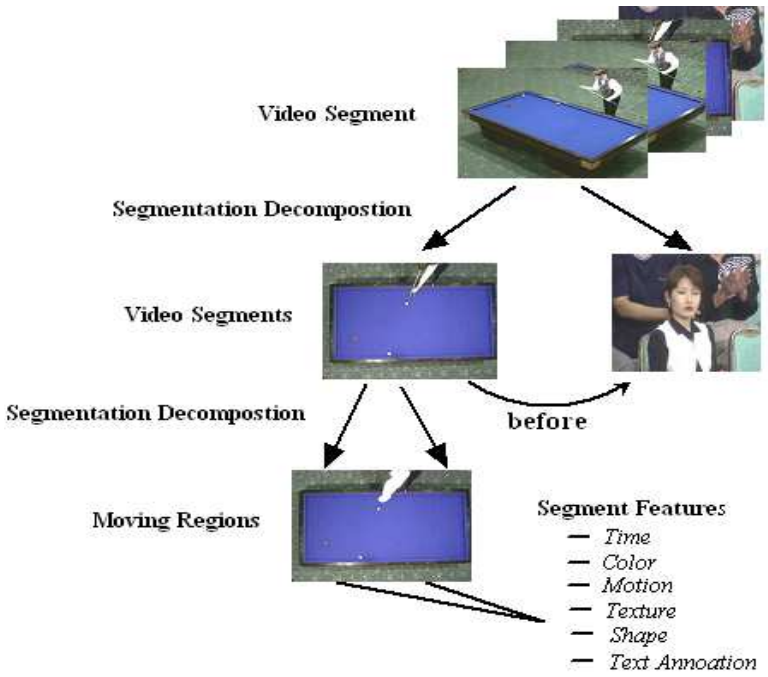

Figure 1. MPEG-7 Tools Application for A Video Description

3. Video Analysis Based on Domain Knowledge

3.1 Overview

Currently, object and event recognition algorithms are not mature enough to scale up to such a large problem domain, as this would require huge libraries of object and event representations and associated algorithms[7, 8].

Furthermore, the semantic meaning of many events can only be recognized in context. So, our approach is to couple video analysis with a structured semantic knowledge base. The use of the Domain Knowledge is probably the only way by which higher-level semantics can be incorporated into techniques that capture the semantic concepts.

Moreover, the building Domain Knowledge for the video object detection is the preferred method for mapping the domain dependent concepts into the domain independent low-level features.

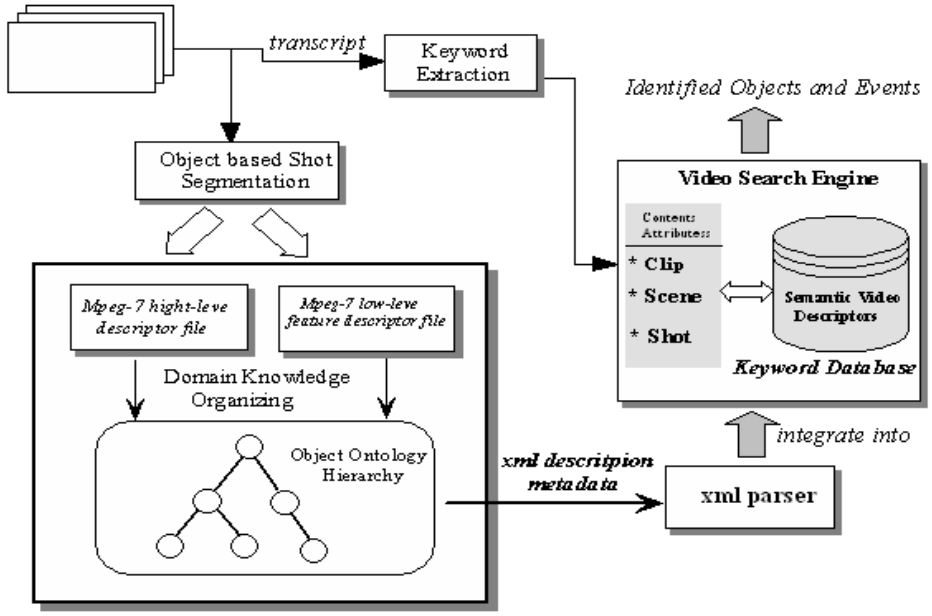

Figure 2. Framework for Video Event Understanding

Content-based analysis of multimedia (event as such) requires method which will automatically segment video sequences and key frames into image areas corresponding to salient objects (e.g. ball, player, spectator, table, etc), track these objects in time, and provide a flexible framework for object recognition, indexing, retrieval and for further analysis of their relative motion and interactions. Semantic concepts within the context of the examined domain are defined in ontology, enriched with qualitative attributes of the semantic objects(spatial relations, temporal relations, etc), multimedia processing methods(motion clustering, etc), and numerical data or low-level features generated via training.

The proposed approach, by exploiting the Domain Knowledge modeled in ontology, enables the recognition of the underlying semantics of the examined video. The retrieval system overall framework, shown in Figure 2, addresses some important issues mentioned in this paper. One is the domain obtainment, and its knowledge representation by the MPEG-7 low-level descriptors combining with high(intermediate-level) descriptors.

And, the most important issue is the video object ontology building. Its significant contribution is to realize the knowledge combination and reconstruct, in the other word is that it can realize mapping the low-level features to the semantic level. The domain-independent, primitive classes comprising the analysis ontology serve as attachment points allowing the integration of the two ontologies. Practically, each domain ontology comprises a specific instantiation of the multimedia analysis ontology providing the corresponding motion model, restrictions etc as will be demonstrated in more details in the next section.

3.2 Video Shot Detection for Domain Knowledge

Video is a structured medium in which actions and events in time and space convey stories, so, a video program(raw video data) must be viewed as a document, not a non-structured sequence of frames. The Figure 3 has shown us one video clip about the Billiard Game's program modeling and

presentation. The Hierarchical levels of video stream abstraction, in decreasing degree of granularity:

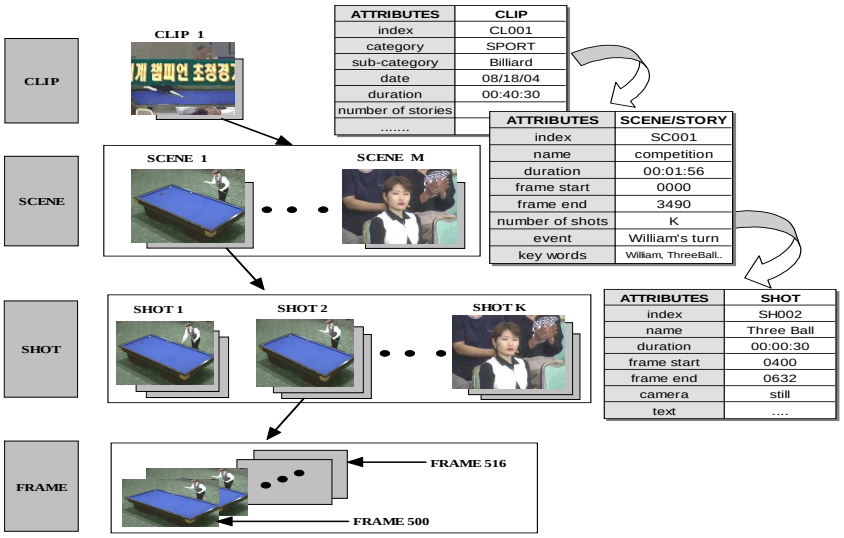

Figure 3. Video Modeling and Representation

From the Figure 3, we can see the fourth layer, video shots, which are directly related to video structures and contents, are the basic units used for accessing video and a sequence of frames recorded contiguously and re-presenting a continuous action in time or space. And we consider on shot that contains a series of actions that can be used to express one kind of meaningful event in the video as one Domain Knowledge. In order to get it to facilitate for our research, an automatic shot detection technique has been proposed for adaptive video coding applications [9]. We focus on video shot detection on compressed MPEG video of Billiard Game.

Since there are three frame types (I, P, and B) in a MPEG bit stream, we first propose a technique to detect the scene cuts occurring on I frames, and the shot boundaries obtained on the I frames are then refined by detecting the scene cuts occurring on P and B frames. For I frames, block-based DCT is used directly as:

$$F(u,v)=\frac{c_u c_v}{4}\sum_{x=0}^{7}\sum_{y=0}^{7} I(x,y)\times\cos\frac{(2x+1)u\pi}{16}\cos\frac{(2y+1)v\pi}{16}$$

where

$$C_{u,}C_{v}=\begin{cases}\frac{1}{\sqrt{2}}, & \text{for } u,v=0\\ 1, & \text{otherwise}\end{cases}$$

One finds that the dc image [consisting only of the dc coefficient (u=v=0) for each block] is a spatially reduced version of I frame. For a MPEG video bit stream, a sequence of dc images can be constructed by decoding only the dc coefficients of I frames, since dc images retain most of the essential global information of image components.

Yeo and Liu have proposed a novel technique for detecting shot cuts on the basis of dc images of a MPEG bit stream [10],in which the shot cut detection threshold is determined by analyzing the difference between the highest and second highest histogram difference in the sliding window. In this article, an automatic dc-based technique is proposed which adapts the threshold for shotcut detection to the activities of various videos. The color histogram differences (HD) among successive I frames of a MPEG bit stream can be calculated on the basis of their dc images as

$$HD(j,j-1)=\sum_{k=0}^{M}[H_{j-1}(k)-H_{j}(k)]^2$$

where Hj(k) denotes the dc-based color histogram of the jth I frame, Hj-1(k) indicates the dc-based color histogram of the (j-1)th I frame, and k is one of the M potential color components. The temporal relationships among successive I frames in a MPEG bit stream are then classified into two opposite classes according to their color histogram differences and an

optimal threshold $\overline{T_c}$.

The optimal threshold $\overline{T_c}$ can be determined automatically by using the fast searching technique given in Ref. [10]. The video frames ~including the I, P, and B frames. Between two successive scenes cuts are taken as one video shot. The following Figure4 has shown us the shot we have detected using the algorithm mentioned above. The shot has contains a series of I frames.

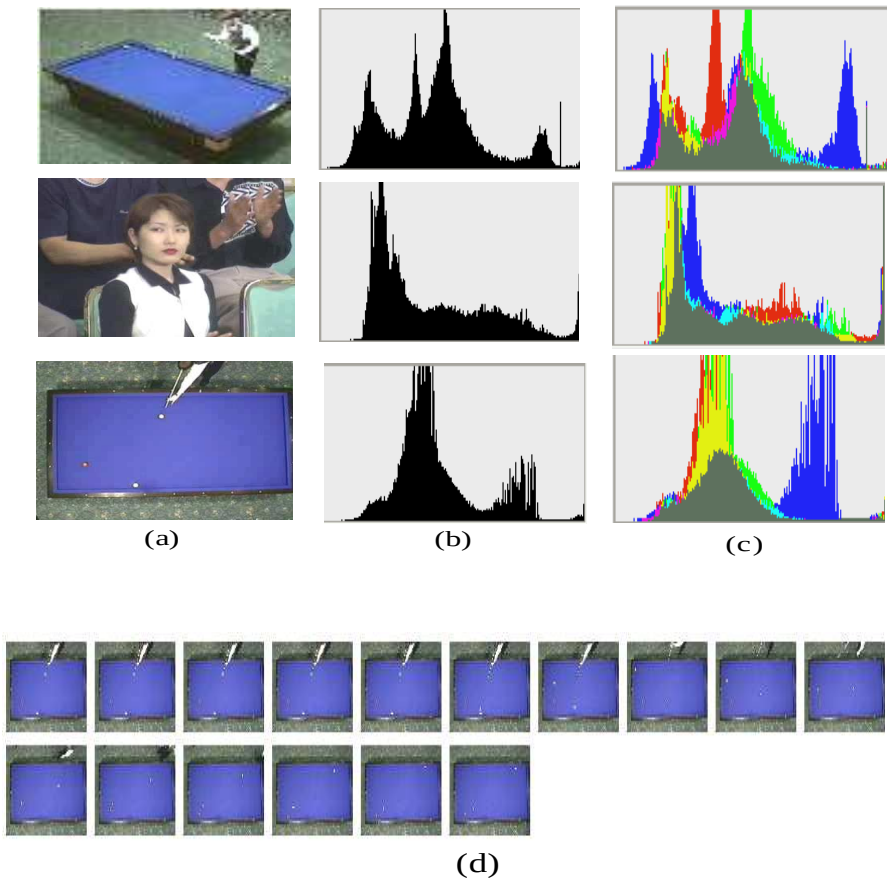

Figure 4. (b)and(c): the RGB histogram and color histogram difference of three snapshots(a). (d): Billiard Game's Shot Detection

3.3. Event Representation using MPEG-7 High Level Descriptors

Video event has the semantic meanings that can be expressed by the human language, it contains the feature elements, camera motion, time, key frame, annotation and object set elements, among others. Our task is how to analyze and describe the event taking advantage of those elements in the specific domain.

MPEG-7 is a new standard for describing the content of multimedia data [11]. MPEG-7 is a means of attaching meta-data to multimedia content. It offers a comprehensive set of audiovisual description tools including meta-data elements and their structures and relationships defined by the standard in the form of Descriptors and Description Schemes.

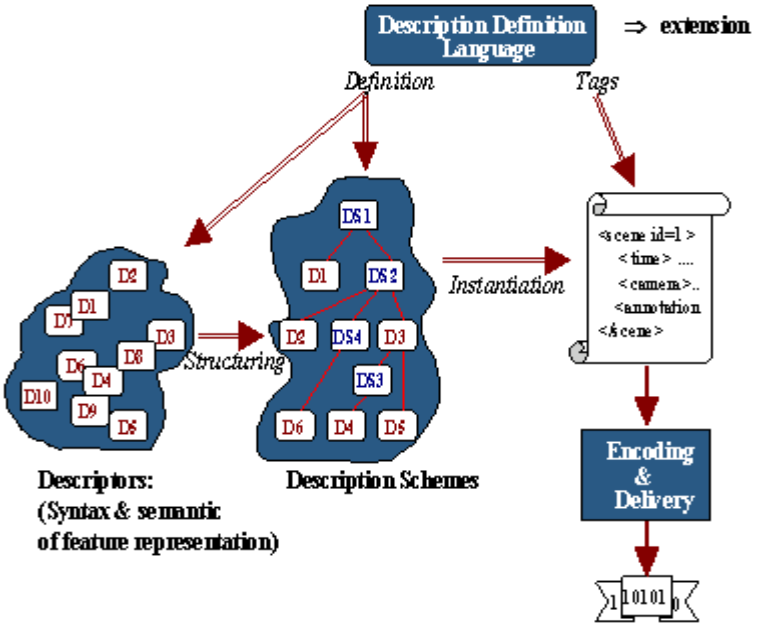

Figure 5. MPEG-7 Framework

Figure 5 shows the relationship among the different MPEG-7 elements

introduced above. The DDL allows the definition of the MPEG-7 description tools, both Descriptors and Description Schemes, providing the means for structuring the Ds into DSs [12]. The DDL also allows the extension for specific applications of particular DSs. The description tools are instantiated as descriptions in textual format (XML) thanks to the DDL (based on XML Schema). Binary format of descriptions is obtained by means of the BiM defined in the Systems part. And this kind of organizing mechanism facilitates our analysis about our instance "Billiard Game". We use the semantic (high-level) descriptors to define the components in the video shot of our example.

Figure 6 illustrates possible conceptual aspects and abstractions of a specific instance (image: "**billiard_Shot_I01.jpg**") of a video shot content.

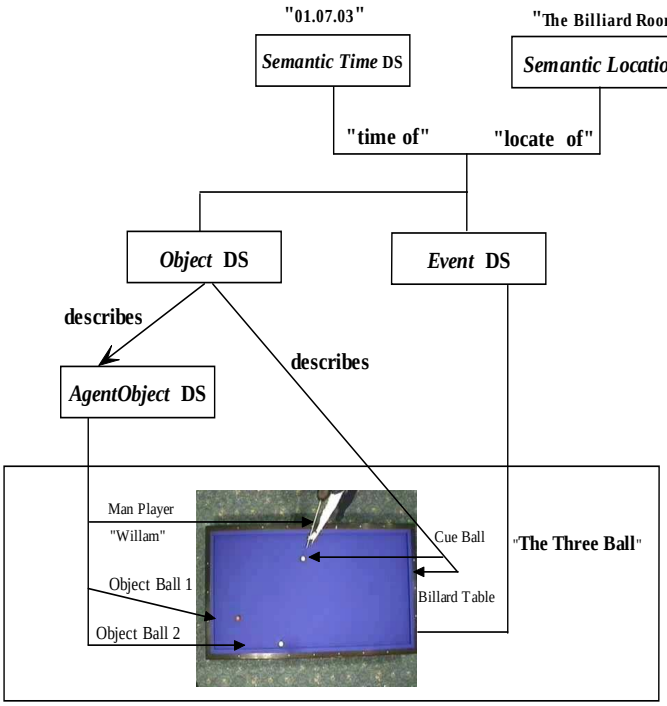

Figure 6. Conceptual Abstractions of a Video Shot Content

Table 1. Lexicon MPEG-7 XML file

```
<?xml version="1.0" encoding="iso-8859-1"?>
<Mpeg7 xmlns="urn:mpeg:mpeg7:schema:2001"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:mpeg7="urn:mpeg:mpeg7:schema:2001"
xsi:schemaLocation="urn:mpeg:mpeg7:schema:2001 Mpeg7-2001.xsd">
<Description xsi:type="ClassificationSchemeDescriptionType">
<ClassificationScheme uri="urn:example:MPEG7">
<Header xsi:type="DescriptionMetadataType">
</Header>
<Term termID="Events">
<Name xml:lang="en">Events</Name>
<Term termID="Semantic Time">
<Name xml:lang="en">01:07:02</Name>
<Term termID="Semantic Location">
<Name xml:lang="en">the billiard room</Name>
<Term termID="hit_the_ball">
<Name xml:lang="en">hit the ball</Name>
<Term termID="the_three_ball">
<Name xml:lang="en">the three ball</Name>
<Term termID="man_player_s_three_ball">
<Name xml:lang="en">man player's three ball</Name>
</Term>
</Term>
</Term>
</Term>
</Term>
</Term>
```

```
<Term termID="Key_Objects">
<Name xml:lang="en">Key Objects</Name>
<Term termID="Human">
<Name xml:lang="en">Human</Name>
<Term termID="Person">
<Name xml:lang="en">Person</Name>
<Term termID="Man_player">
<Name xml:lang="en">Man player</Name>
<Term termID="William">
<Name xml:lang="en">William</Name>
</Term>
</Term>
<Term termID="Woman_player">
<Name xml:lang="en">Woman player</Name>
</Term></Term></Term>
<Term termID="Man-Made_Object">
<Name xml:lang="en">Man-Made_Object</Name>
<Term termID="Billiard_Ball">
<Name xml:lang="en">Billiard Ball</Name>
<Term termID="cue_ball_1">
<Name xml:lang="en">object ball 1</Name>
</Term>
<Term termID="object_ball_1">
<Name xml:lang="en">object ball 2</Name>
</Term>
<Term termID="object_ball">
<Name xml:lang="en">object ball</Name>
</Term></Term></Term>
</ClassificationScheme>
</Description>
</Mpeg7>
```

The Structure DSs and Semantic DSs can be related by a set of links allowing the shot content to be described on the basis of both content structure and semantic structures together. The links relate different semantic concepts to the instances within the shot content described by the segments.

Furthermore, most of the MPEG-7 Content Description and Content Management DSs are linked together and in practice, also often included within each other in the MPEG-7 descriptions. As our instance, our Structure DS is the event -" The Three Ball", and it related to three kinds of DSs: "Semantic time", "Semantic Location" ,"Object" and with their corresponding semantic concept definitions.

We can make use of the MPEG-7 descriptions to do the annotation work both from semantic level and low level features respectively. MPEG-7 instances are XML documents that conform to a particular MPEG-7 schema, as expressed in the DDL and that describe visual content.

We are trying to create a standard description is an appropriate way to address semantic level concepts based the detection video shot. .An example of lexicon MPEG-7 XML file is shown in Table 1. In this file, it defines the event, objects with their attributes and describes the semantic logical structure. That the set of lexicon is dependent on the summarization application, and can be easily modified and imported into the annotation tool.

3.4. Video Object Description with Low Levels Features

Previous sections have been analyzed the video event's logical structure and schema by using the MPEG-7 high level descriptors from the point of view of the human meanings. However, we know that, what the Video Database stored is the "Raw Video Data", such as the video object's dominant colors, the motion parameters which have described the spatio-temporal relationships between moving objects, and the trajectories of them. So, analyzing the video object motions become our challenge task in

this section.

According to the characteristics of the "Billiard Game", the trajectory is preferred to considers our research object. We try to use the object's motion trajectory to express the event meanings whose definitions are described in the MPEG-7 high-level descriptors.First issue is that we apply the DCT algorithm to pick up the key-frames in the shot that was detected before. Just like the below Figure 7 showed us, key-frames are obtained.

These four key-frames have present the whole process of the event of "The Three Ball": the "cue ball" touches the "object ball 1"(the gray ball) then bounce back the Northwest direction, then come back to hit another "object ball 2"(the red ball). And the 3-D Figure 8 defined as a series of spatial-temporal localization of one of its representative points. Therefore, the core information of trajectory is a list of points, which has both spatial and temporal locations.

In order to describe the spatial localization of the points, spatial coordinate system must be specified. In fact, this shot records about 60 points from the 20 frames. Each 20 points present one Ball's trajectory. Here, we also use the MPEG-7 XML file to describe the video object moving state.

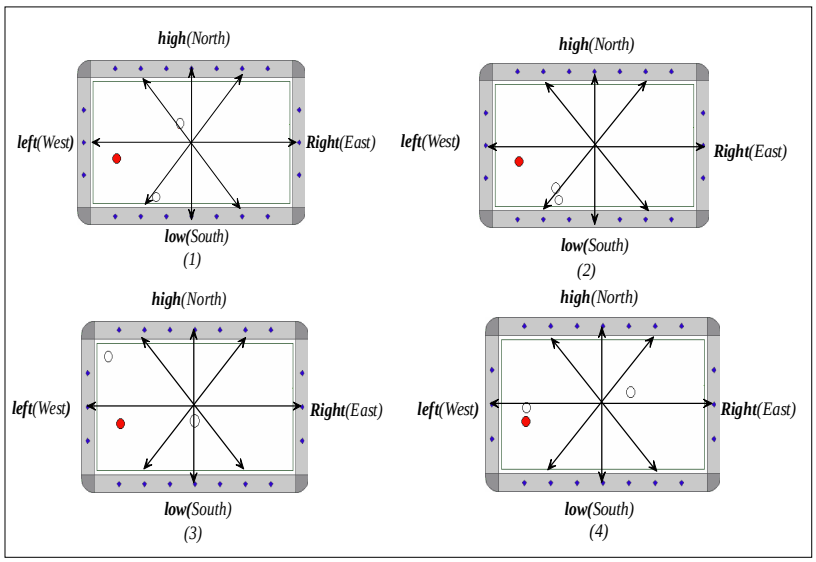

Figure 7. Key-frame's Moving Object Trajectory

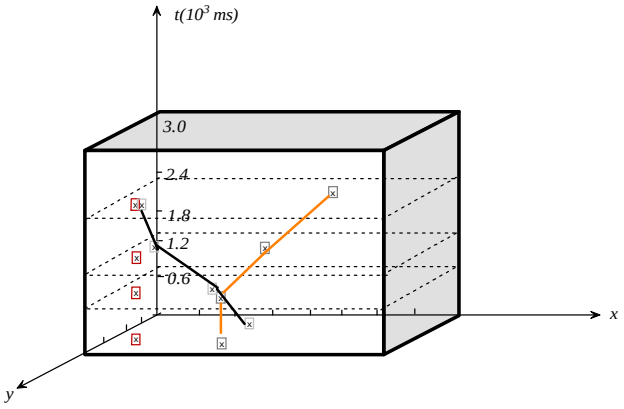

Figure 8. Moving Object trajectory Recording

The motion trajectory records the moving path of a specific object. It is a high-level feature associated with a moving region. As mentioned before, in the Table2, we annotate the shot, which includes16 frames, of a video as "the three ball" and annotate a rectangular region in the key frame-514. In MPEG-7, each video shot is defined as a Video Segment, where the shot start-time (shown in the Thh:mm:ss:nnF500 format) and duration are given and the annotations are described. Furthermore, the embedded <SpatioTemporalDecomposition> tag allows us to specify the region location and the corresponding text annotation in a key frame. In this table, the key frame is the 514th frame of the video sequence.

The annotated region is specified by the <SpatialLocator> tag. It is identified by a polygon whose n vertex coordinates are recorded in the order of <x0, y0>, <x1, y1>, ···, <xn-1, yn-1> after the <CoordsI> tag. For multiple regions in a key frame, the system needs to repeat the section between <StillRegion> and its closing tag inside the <SpatialDecomposition> section.

Table 2. Low-level Features in MPEG-7 XML file

```
<?xml version="1.0" encoding="iso-8859-1"?>
<Mpeg7 xmlns="urn:mpeg:mpeg7:schema:2001"  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:mpeg7="urn:mpeg:mpeg7:schema:2001"  xsi:schemaLocation="urn:mpeg:mpeg7:schema:2001  Mpeg7-
2001.xsd">
<Description xsi:type="ContentEntityType">
<MultimediaContent xsi:type="Video Type">
<Video>
<TemporalDecomposition>
<MediaTime>
<MediaTimePoint>T00:00:00:00F500</MediaTimePoint>
<MediaIncrDuration timeUnit="PTIN500">20</MediaIncrDuration>
</MediaTime>
<TextAnnotation type="shot" relevance="1" confidence="1">
<FreeTextAnnotation>the three ball</FreeTextAnnotation>
</TextAnnotation>
<SpatioTemporalDecomposition>
<StillRegion>
<MediaIncrTimePoint
```

```
timeUnit="PTIN500F">514</MediaIncrTimePoint>
<FreeTextAnnotation>completingtheshooting</FreeTextAnnotatin>
<SpatialDecomposition>
<StillRegion>
<SpatialLocator>
<Poly>
<CoordsI>41 135 290 135 290 230 41 230 </CoordsI>
</Poly>
</SpatialLocator>
</StillRegion>
</SpatialDecomposition>
</StillRegion>
</SpatioTemporalDecomposition>
</TemporalDecomposition>
</Video>
</MultimediaContent>
</Description>
</Mpeg7>
```

4. Domain Knowledge Ontology Infrastructure Building

We already described the video content both from the semantic level and the low level features. However, the MPEG-7 formalism does not allow reasoning mechanism because it is defined solely in XML schema. Although XML schema is suitable for expressing the syntax, structural, cardinality and typing constraints required by MPEG-7, it displays some limits when we need to extract implicit information. This is due to the fact that there is neither formal semantics nor inference capabilities associated with the elements of an XML document. In this setting, the MPEG-7 standard is extended by an ontology for the Domain Knowledge representation allowing some reasoning mechanism over the MPEG-7 description.

4.1 Building Object Ontology in Domain Knowledge

In order to make the annotation can range from high-level semantic concepts to low-level feature descriptions. The low-level features automatically extracted from the resulting moving objects are mapped to high-level concepts using ontology in a specific Domain Knowledge, combined with a relevance feedback mechanism is the main contribution in this part. In this study, ontologies [13, 14] are employed to facilitate the annotation work using semantically meaningful concepts (semantic objects).

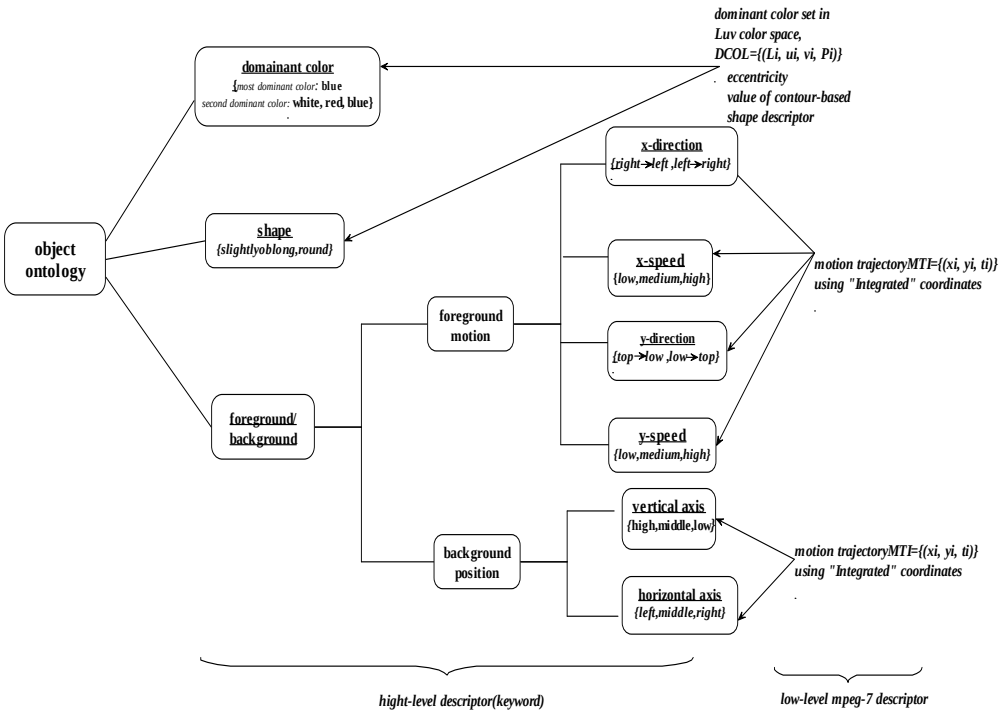

Figure 9. Object Ontology Structure.
The Correspondence Between Low-level MPEG-7
Descriptors and High-level Descriptors

The object ontology is presented in Figure 9, where the possible intermediate level descriptors and descriptor values are shown. Each intermediate level descriptor value is mapped to an appropriate range of values of the corresponding low-level, arithmetic descriptor. With the exception of color (e.g. "black") and direction (e.g. "low?high") descriptor values, the value ranges for every low-level descriptor are chosen so that the resulting intervals are equally populated.

This is pursued so as to prevent an intermediate-level descriptor value from being associated with a plurality of spatio-temporal objects in the database, since this would render it useless in restricting a query to the potentially most relevant ones. Overlapping, up to a point, of adjacent value ranges, is used to introduce a degree of fuzziness to the descriptor values;

for example, both "slightly oblong" and "moderately oblong"values may be used to describe a single object. Regarding color, a correspondence between the 11 basic colors used as color descriptor values and the values of the HSV color space is heuristically defined.

More accurate correspondences based on psycho visual findings are possible; this is however beyond the scope of this work. Regarding the direction of motion, the mapping between values for the descriptors "x direction", "y direction" and the MPEG-7 Motion Trajectory descriptor is based on the sign of the cumulative displacement of the foreground spatio-temporal objects.

4.2 Concepts Definitions in Domain Knowledge

Through this domain ontology, we map the low-level features with high-level concepts using spatio-temporal relations between objects. Moreover, in order to give specific annotation to frames, we defined some semantic concepts with clear and typical characteristics to describe shot events.

- Class Object: **the subclass and instance of the superclass "Object", all video objects that can be detected through the analysis process. Each object instance is related to appropriate feature instances by hasFeature property. Each object can be related to one or more other objects through a set of predefined spatial properties.**

- Class feature: **the subclass and instance of the superclass "feature", which is the low-level features of multimedia associated with each object. In our video shot domain, the features covered low-level background features like *color, moving object direction, each of which has a closely relation with the object that will be detected andthroughout the whole process of shot analysis; for* the high-level semantics associated with foreground features**

such as *motionactivity, status… it emphasize on their instinct meaningful description of certain video events.*

- Class: **featureParameter: denotes the actualqualitativedescriptions of each corresponding feature. It is subclassed according to the definedsuperclass,"featureParemeter",including:ColorfeatureParameter,MotionfeatureParameter,PositionfeatureParameter,DirectionfeatureParameter.**

- Spatial Relations: **approach, touch, disjoint. These three spatial descriptors are used to describe the object relations between each individual frame; and describe the differences and relations between two adjacent frames.**

- Temporal Relations: **before, meet, after, starts, completes. Plus the three spatial descriptors, a whole process of describing the moving video objects, emphasize the semantics and relations between shots, come up with a meaningful metadata in the backend.**

- Actions: **beforeShooting, cue ball hitting, object ball hitting, finishShooting, change player. We defined five actions to describe the semantic event in one shot, and could use it as akeyword annotation of free-text, as well as index the key frame instead of getting all the I-frames an annotation.**

The developed domain ontology mainly focused on the representation of semantics in each detected shot and its frames. As a consequence, we could simply pick up the key-frames from the shot depository, and annotate them sequentially and semantically according to the inner presentation of moving objects. To facilitate the explanation, we consider one shot in the Billiard Game put forward above as our Domain Knowledge. Because one shot usually has the semantic meaning that contains a sequence of frames, we call one

shot an Event. The event we are going to describe and analysis is named as "The Three Ball". So, the following section shows the application to "the three ball" billiard game video shots

4.3 Video Event Representation

The proposed approach in the section 5 was tested in one specific Domain Knowledge: The Billiard Game Shots In the domain, appropriate domain ontology was defined. In this case, Figure10 is the result of the video object's motion trajectory detection through object ontology's knowledge restructure for mapping the MPEG-7 high level descriptors to the low level features.

For each video object created by applying the segmentation algorithm to a collection of video shots, MPEG-7 low-level descriptors were calculated and the mapping between them and the intermediate-level descriptors defined by the object ontology was performed. Subsequently, the object ontology was used to define, using the available intermediate-level descriptors, semantic objects. Querying using these definitions resulted in initial results produced by excluding the majority of spatio-temporal objects in the database.

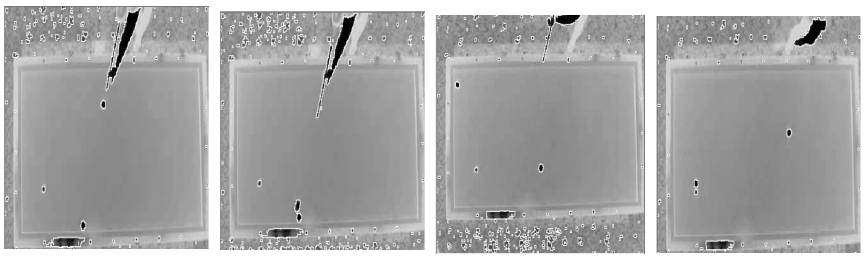

Figure 10. Motion Detection in Key-frames

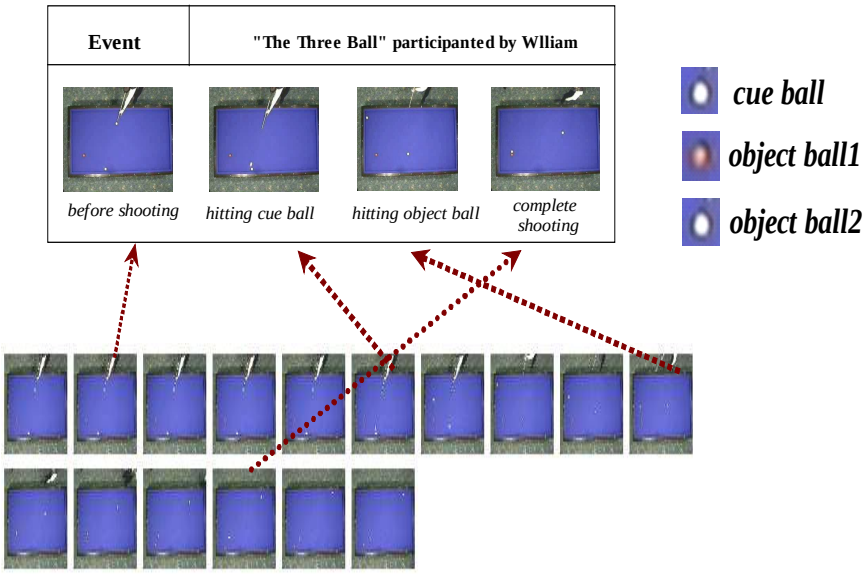

Figure 11. Key-Frames Annotation of Specific Event "The Three Ball"

Figure 11 has shown the results of the annotation of key-frame of the video event. These descriptors are presented by the XML description output from the ontology modeling described above, then, parsed by the XML parser. All of the keywords, not only extracted from the clip, scene from the video storage, but the annotators from the ontology modeling are integrated into the keywords database for catering the user's retrieval.

4.4 Retrieval

This retrieval model is designed based on the SVQL(semantic views query languages). SVQL is a high level query language that allows users to express their requirements following the Semantic Views Model in a concise, abstract and precise way.

Moreover, SVQL is designed particularly to make possible the retrieval of audiovisual data described by MPEG-7 standard. SVQL allows the

users to formulate high level queries on top of MPEG-7 descriptions without getting involved into the implementation details of MPEG-7. Compared to traditional database query languages such as SQL, OQL, and XQuery, SVQL has the advantage of being specially designed for the retrieval of audiovisual data based on a semantic model of user's requirements. Query languages such as SQL, OQL, and XQuery are designed to fit specific structures of data.

However these languages are not suited to express the abstract level of the Semantic Views Model. Using XQuery and XML Schema is an advantageous method for the "implementation" of the Semantic Views Model and Semantic Views Query Language on top of MPEG-7 instances. Nevertheless, from a conceptual point of view, XQuery does not allow the user to directly express his/her abstract requirements following the Semantic Views Model.

Table 3: SVQL Query

LET	$semanticViews := *semanticViews*("Video data Repository")
	$newsItem := *newsItem*($semanticViews),
	$fact := *fact*($semanticViews),
	$shot := *shot*($semanticViews),
	$videoSegment := videoSegment*($semanticViews),*
WHERE	match(getDescription($fact,Event),*event*("the three ball"))AND
	match(getDescription($shot,Person),*person*("man_player"))AND
	greaterThan(getDescription($videoSegment,shot),*shot*("3"))AND
	corresponds($videoSegment,$shot)
RETURN	$video Segment

As can be observed in the example, the query is based on a "LET-WHERE-RETURN" structure[16], This type of query syntax, refereed to as "keyword oriented syntax" [17], is used in the most well-known query languages, such as SQL, OQL, and XQuery, and is a familiar mode of query expression for specialized end users and

application programmers.

As can be seen in the above query, the LET clause contains two types of expressions: In the first expression, the SemanticViews() function is called with the name of the file containing the MPEG-7 instances to be retrieved. This function creates the Semantic Views Document corresponding to the MPEG-7 file. In the next series of expressions, a set of functions are called to get different required BasicViewEntities of the Semantic Views Document and to assign them into a set of variables. Each function name indicates the type of the BasicViewEntity, e.g. NewsItem,Shot, and VideoSegment.

The WHERE clause also contains two different types of expressions: The first four expressions represent a set of conditions that should be held on the BasicViewEntities of different Views. These conditions are expressed via a set of ViewOperators: here match(), and greaterThan() are used.

The last expression represents the condition concerning the InterViewRelation of BasicViewEntities via corresponds() operator. It determines which of the cited BasicViewEntites correspond to each other. This expression is used if more than one View is used in the query. Finally, the RETURN clause identifies a variable containing a BasicViewEntity to be returned to the user. In the above query the variable containing the VideoSegment is returned.

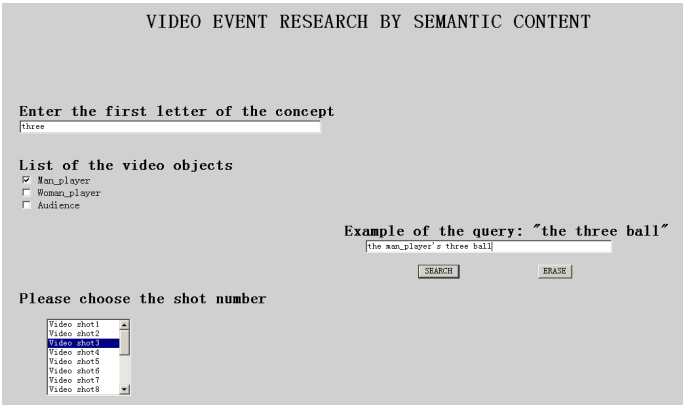

Figure 12. Video Research by Semantic Content

Querying interface based on SVQL is showed in Figure 12. We provide a friendly, visually rich inter-fact that allows the user to interactively query the database, browse the results, and view the selected video clips.

In order to make sure the user can get the exact results. we make some conceptual restrictions on the video retrieval interface, we provide the possible concepts to facilitate the user's choice and two text boxes for the user's own will. And this search engine is responsible for searching the database according to the parameters provided by the user.

5. Conclusions

Our paper was mainly designed for video processing, video analysis, and video event presentation. From the video shot and key-frame detection to internal moving objects description, from the MPEG-7 descriptors definitions for the low-level features and high-level semantic to the video object ontology, we all did the comprehensively and clearly exposition. We combined the traditional algorithms with our own methods put forward for applying to our specific examples.

In this paper, not only did we analysis the MPEG-7 in one specific domain comprehensively, such as have focused on a few specific visual dimensions such as color, motion and spatial information. but we make a lot of improvement and novelty works based on that. We chiefly take advantage the XML files to perform knowledge enrichment on top of the primary information extracted from the video which can be considered the extended format of MPEG-7 files, since the MPEG-7 formalism does not allow reasoning mechanism.

This enriched data representative helps users search for multimedia video content more efficiently. Although XML schema is suitable for expressing the syntax, structural, cardinality and typing constraints required by MPEG-7,it displays some limits when we need to extract implicit information. This is due to the fact that there is neither formal semantics nor inference capabilities associated with the elements of an XML document. So, we proposed a novel method for video event analysis and description on the fundamental of the Domain Knowledge object ontology, the MPEG-7 standard is extended by an ontology for the human knowledge representation allowing some reasoning mechanism over the MPEG-7 descriptions, the most contribution is that it helps us overcome gap between the low-level features and high level semantics, but combining these two aspects in the most efficient and flexible (expressive) manner.

The proposed approach aims at formulation a domain specific analysis

model facilitating for the video event retrieval and video object motion detection. And at the last part, we designed a friendly retrieval interface for the users based on the SVQL querying language.

Our future work includes the enhancement of the domain ontology with more complicated model representation and we try to use the more complex spatio-temporal relationship reference rules to analyze the moving features. We will also do more technical work to intensify our retrieval part function. Extending the modeling scheme with more concepts and conceptual relationships and test our system on other categories of video like documentaries, more complex sports(like football) or movies also are our further research.

References

1. S.-F. Chang. The holy grail of content-based media analysis. IEEE Multimedia,9(2):610, Apr-Jun. 2002.
2. S.-F. Chang, T. Sikora, and A. Puri. Overview of the MPEG-7 standard. IEEE Trans. on Circuits and Systems for Video Technology, 11(6):688-695, June 2001.
3. A. Yoshitaka and T. Ichikawa. A survey on content-based retrieval for multimedia databases. IEEE Transactions on Knowledge and Data Engineering, 11(1):81-93, Jan-Feb 1999.
4. P. Salembier and F. Marques. Region-Based Representations of Image and Video:Segmentation Tools for Multimedia Services. IEEE Trans. Circuits and Systems for Video Technology, 9(8):1147-1169, December 1999.
5. S.-F. Chang, T. Sikora, and A. Puri, "Overview of the MPEG-7 standard," IEEE Trans. Circuits Syst. Video Technol., vol. 11, pp. 688-695, June 2001.
6. W. Al-Khatib, Y. F. Day, A. Ghafoor, and P. B. Berra, "Semantic modeling and knowledge representation in multimedia databases," IEEE Trans. Knowledge Data Eng., vol. 11, pp. 64-80, Jan-Feb. 1999.
7. R. Nelson and A. Selinger. Learning 3D recognition models for general objects from unlabeled imagery: An experiment in intelligent brute force. In Proceedings of ICPR,volume 1,pages 18, 2000.
8. C. Town and D. Sinclair. A self-referential perceptual inference framework for video nterpretation. In Proceedings of the International Conference on Vision Systems, volume 26, pp. 54-67, 2003.
9. J. Fan, D. K. Y. Yau, W. G. Aref, and A. Rezgui, "Adaptive motioncompensated video coding scheme towards content-based bit rate allocation,"J. Electron. Imaging: 521-533, 2000
10. B. L. Yeo and B. Liu, "Rapid scene change detection on compressed video,"IEEE Trans. Circuits Syst. Video Technol. 5, 533-544, 1995.
11. Sikora, T.: The MPEG-7 Visual standard for content description - an overview.IEEE Trans. on Circuits and Systems for Video

Technology,special issue on MPEG-7 696-702, 2001
12.Multimedia content description interface-part 8: extraction and use of MPEG-7 descriptors. ISO/IEC, 2002.
13.P. Martin and P. W. Eklund, "Knowledge retrieval and the World Wide Web," IEEE Intell. Syst., vol. 15, pp. 18-25, May-June, 2000.
14.A. T. Schreiber, B. Dubbeldam, J. Wielemaker, and B. Wielinga, "Ontology-based photo annotation," IEEE Intell. Syst., vol. 16, pp. 66-74, May-June 2001.
15.M.K.Smith, C.Welty, and D.L.McGuinness, "OWLWeb Ontology Language Guide,"W3C Candidate Recommendation, Feburary 2004.
16.Chapitre du livre: N.Day, N. Fatemi, O. Abou Khaled. Search and Browsing. Introduction to MPEG-7: Multimedia Content Description Interface. Wiley. England: B.S. Manjunath ,Philippe Salembier, Thomas Sikora,ISBN 0-471-48678-7, pp. 335-352, 2002.
17.Group,Cotton, P. and Robie J. The W3C XML Query Working XML Europe 2001, Berlin, Germany, May 2001.

저작물 이용 허락서

학 과	전자계산학과	학 번	20037087	과 정	석사

성 명	한글: 송단 한문 : 宋 丹 영문 : Dan Song
주 소	조선대학교 여자기숙사 638호실
연락처	E-MAIL : star_sd821@hotmail.com
논문제목	한글 :**도메인 지식 구축에 의한 의미적 비디오 이벤트 표현** 영문 :**Semantic Video Event Description Assisted by Building Domain Knowledge**

본인이 저작한 위의 저작물에 대하여 다음과 같은 조건아래 -조선대학교가 저작물을 이용할 수 있도록 허락하고 동의합니다.

- 다 음 -

1. 저작물의 DB구축 및 인터넷을 포함한 정보통신망에의 공개를 위한 저작물의 복제, 기억장치에의 저장, 전송 등을 허락함
2. 위의 목적을 위하여 필요한 범위 내에서의 편집형식상의 변경을 허락함. 다만, 저작물의 내용변경은 금지함.
3. 배포전송된 저작물의 영리적 목적을 위한 복제, 저장, 전송 등은 금지함.
4. 저작물에 대한 이용기간은 5년으로 하고, 기간종료 3개월 이내에 별도의 의사표시가 없을 경우에는 저작물의 이용기간을 계속 연장함.
5. 해당 저작물의 저작권을 타인에게 양도하거나 또는 출판을 허락을 하였을 경우에는 1개월 이내에 대학에 이를 통보함.
6. 조선대학교는 저작물의 이용허락 이후 해당 저작물로 인하여 발생하는 타인에 의한 권리 침해에 대하여 일체의 법적 책임을 지지 않음
7. 소속대학의 협정기관에 저작물의 제공 및 인터넷 등 정보통신망을 이용한 저작물의 전송출력을 허락함.

2005 년 5 월 일

저작자: (서명 또는 인)

조선대학교 총장 귀하