Video Semantic Content Analysis Framework Based on Ontology Combined MPEG-7

Liang Bai[1,2], Songyang Lao[1], Weiming Zhang[1],
Gareth J.F. Jones[2], and Alan F. Smeaton[2]

[1] School of Information System & Management,
National University of Defense Technology,
ChangSha, China, 410073
lbai@computing.dcu.ie, laosongyang@vip.sina.com,
wmzhang@nudt.edu.cn
[2]Centre for Digital Video Processing, Dublin City University, Glasnevin, Dublin 9, Ireland
{gjones, asmeaton}@computing.dcu.ie

Abstract. The rapid increase in the available amount of video data is creating a growing demand for efficient methods for understanding and managing it at the semantic level. New multimedia standard, MPEG-7, provides the rich functionalities to enable the generation of audiovisual descriptions and is expressed solely in XML Schema which provides little support for expressing semantic knowledge. In this paper, a video semantic content analysis framework based on ontology combined MPEG-7 is presented. Domain ontology is used to define high level semantic concepts and their relations in the context of the examined domain. MPEG-7 metadata terms of audiovisual descriptions and video content analysis algorithms are expressed in this ontology to enrich video semantic analysis. OWL is used for the ontology description. Rules in Description Logic are defined to describe how low-level features and algorithms for video analysis should be applied according to different perception content. Temporal Description Logic is used to describe the semantic events, and a reasoning algorithm is proposed for events detection. The proposed framework is demonstrated in sports video domain and shows promising results.

Keywords: Video Semantic Content, MPEG-7, Ontology, OWL, Description Logic, Temporal Description Logic.

1 Introduction

Audiovisual resources in the form of image, video, audio play more and more pervasive role in our lives. Especially, the rapid increase of the available amount in video data has revealed an urgent need to develop intelligent methods for understanding, storing, indexing and retrieval of video data at the semantic level [1]. This means the need to enable uniform semantic description, computational interpretation and processing of such resources.

The main challenge, often referred to as the semantic gap, is mapping high-level semantic concepts into low-level spatiotemporal features that can be automatically

N. Boujemaa, M. Detyniecki, and A. Nürnberger (Eds.): AMR 2007, LNCS 4918, pp. 237–250, 2008.

extracted from video data. Feature extraction, shot detection and object recognition are important phases in developing general purpose video content analysis [2] [3]. Significant results have been reported in the literature for the last two decades, with several successful prototypes [4] [5]. However, the lack of precise models and formats for video semantic content representation and the high complexity of video processing algorithms make the development of fully automatic video semantic content analysis and management a challenging task. And, the mapping rules often are written into program code. This causes the existing approach and systems to be too inflexible and can't satisfy the need of video applications at the semantic level. So the use of domain knowledge is very necessary to enable higher level semantics to be integrated into the techniques that capture the semantics through automatic parsing.

Ontology is formal, explicit specifications of domain knowledge: it consists of concepts, concept properties, and relationships between concepts and is typically represented using linguistic terms, and has been used in many fields as a knowledge management and representation approach. At the same time, several standard description languages for the expression of concepts and relations in ontology have been defined. Among these the important are: Resource Description Framework (RDF) [6], Resource Description Framework Schema (RDFS), Web Ontology Language (OWL) [7] and, for multimedia, the XML Schema in MPEG-7.

Many automatic semantic content analysis systems have been presented recently in [8] [9] and [10]. In all these systems, low-level based semantic content analysis is not associated with any formal representation of the domain.

The formalization of ontology is based on linguistic terms. Domain specific linguistic ontology with multimedia lexicons and possibility of cross document merging has instead been presented in [11]. In [12], concepts are expressed in keywords and are mapped in object ontology, a shot ontology and a semantic ontology for the representation of the results of video segmentation. However, although linguistic terms are appropriate to distinguish event and object categories in a special domain, it is a challenge to use them for describing low-level features, video content analysis and the relationships between them.

An extending linguistic ontology with multimedia ontology was presented in [13] to support video understanding. Multimedia ontology is constructed manually in [14]. M.Bertini et al., in [15], present algorithms and techniques that employ an enriched ontology for video annotation and retrieval. In [16], perceptual knowledge is discovered grouping images into clusters based on their visual and text features and semantic knowledge is extracted by disambiguating the senses of words in annotations using WordNet. In [17], an approach for knowledge assisted semantic analysis and annotation of video content, based on an ontology infrastructure is presented. Semantic Web technologies are used for knowledge representation in RDF/RDFS. In [18], an object ontology, coupled with a relevance feedback mechanism, is introduced to facilitate the mapping of low-level to high-level features and allow the definition of relations between pieces of multimedia information.

Multimedia standards, MPEG-7 [19], provide a rich set of standardized tools to enable the generation of audiovisual descriptions which can be understood by machines as well as humans and to enable the fast efficient retrieval from digital archives as well as filtering of streamed audiovisual broadcasts on the Internet. But MPEG-7 is expressed solely in XML Schema and can not provide enough support for expressing

semantic knowledge, while most of video content is out of the scope of the standard at a semantic level. So a machine-understandable and uniform representation of the semantics associated with MPEG-7 metadata terms is needed to enable the interoperability and integration of MPGE-7 with metadata descriptions from different domain. Web ontology language (OWL) can be used to do this, which is an accepted language of the semantic web due to its ability to express semantics and semantic relationships through class a property hierarchies. Some new metadata initiatives such as TV-Anytime [20], MPEG-21 [21], NewsML [22] have tried to combine MPEG-7 multimedia descriptions with new and existing metadata standards for resource discovery, rights management, geospatial and educational.

In this paper, a framework for video semantic content analysis based on ontology combined MPEG-7 is presented. In the proposed video semantic content analysis framework, video analysis ontology is developed to formally describe the detection process of the video semantic content, in which the low-level visual and audio descriptions part of MPEG-7 is combined and expressed in OWL. This idea drives the work to investigate the feasibility of expressed MPEG-7 terms in OWL and how to express. Semantic concepts within the context of the examined domain area are defined in domain ontology. Rules in Description Logic are defined which describe how features and algorithms for video analysis should be applied according to different perception content and low-level features. Temporal Description Logic is used to describe the semantic events, and a reasoning algorithm is proposed for events detection. OWL language is used for ontology representation. By exploiting the domain knowledge modeled in the ontology, semantic content of the examined videos is analyzed to provide a semantic level annotation and event detection.

2 Framework of Video Semantic Content Analysis

The proposed video semantic content analysis framework is shown in Fig.1. According to the available knowledge for video analysis, a video analysis ontology is developed which describes the key elements in video content analysis and supports the detection process of the corresponding domain specific semantic content. The visual and aural descriptions of MPEG-7 are combined into this ontology expressed in OWL. Semantic concepts within the context of the examined domain are defined in domain ontology, enriched with qualitative attributes of the semantic content. OWL language is used for knowledge representation for video analysis ontology and domain ontology. DL is used to describe how video processing methods and low-level features should be applied according to different semantic content, aiming at the detection of special semantic objects and sequences corresponding to the high-level semantic concepts defined in the ontology. TDL can model temporal relationships and define semantically important events in the domain. Reasoning based DL and TDL can carry out object, sequence and event detection automatically.

Based on this framework, video semantic content analysis depends on the knowledge base of the system. This framework can easily be applied to different domains provided that the knowledge base is enriched with the respective domain ontology. OWL semantic definitions for MPEG-7 terms provide rich low-level visual and aural descriptions and importantly a common understanding of these descriptions for

different domains. Further, the ontology-based approach and the utilization of OWL language ensure that semantic web services and applications have a greater chance of discovering and exploiting the information and knowledge in the video data.

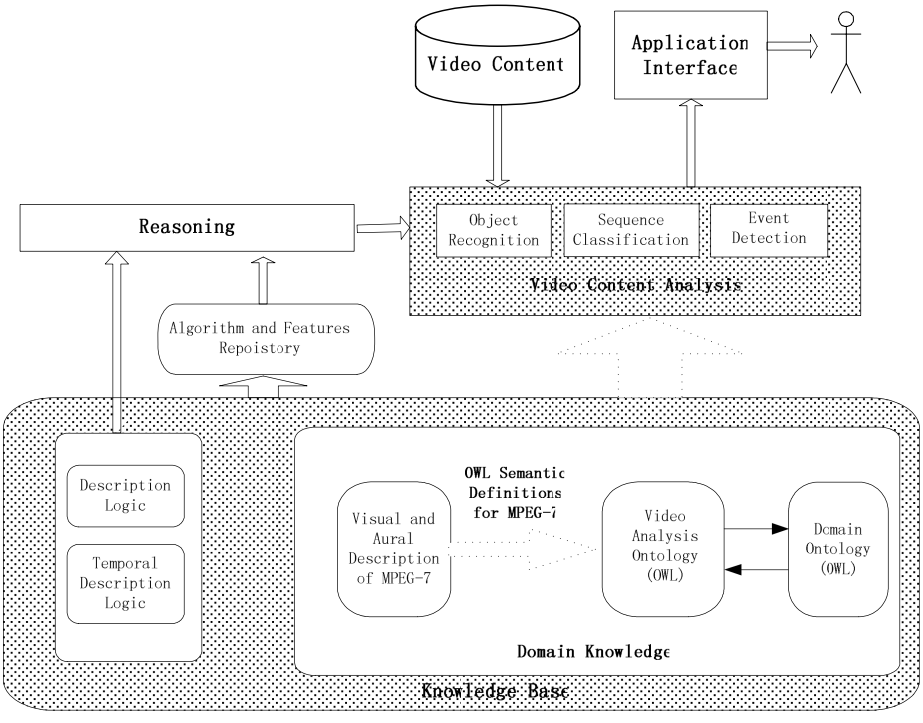

Fig. 1. Framework for Video Semantic Content Analysis based on Ontology

3 Video Analysis Ontology Development

3.1 The Definition for Video Analysis Ontology

In order to realize the knowledge-based and automatic video semantic content analysis explained in section 2, the knowledge for video analysis is abstracted and a video analysis ontology is constructed. In general, video content detection, such as objects, considers the utilization of content characteristic features in order to apply the appropriate detection algorithms for the analysis process in form of algorithms and features. So all elements for the video content analysis, including content, features, algorithms and necessary restrictions, must be described clearly in a video analysis ontology. The audio track in video data, including aural sequences and objects, is important information for video semantic content analysis. The development of the proposed video analysis ontology deals with the following concepts (OWL classes) and their corresponding properties, as illustrated in Fig. 2. The classes defined above are expressed in OWL language in our work.

- Class **Sequence**: the subclass and instance of the super-class "Sequence", all video sequences can be classified through the analysis process at shot level, such as: long-view shot or tight-view shot in sports video. It is sub-classed to **Visual-Sequence** and **AuralSequence**. Each sequence instance is related to appropriate feature instances by the **hasFeature** property and to appropriate detection algorithm instances by the **useAlgorithm** property.
- Class **Object**: the subclass and instance of the super-class "Object", all video objects can be detected through the analysis process at frame level. It is sub-classed to **VisualObject** and **AuralObject**. Each object instance is related to appropriate feature instances by the **hasFeature** property and to appropriate detection algorithm instances by the **useAlgorithm** property.
- Class **Feature**: the super-class of video low-level features associated with each sequence and object. It is linked to the instances of **FeatureParameter** class through the **hasFeatureParameter** property.
- Class **FeatureParameter**: denotes the actual qualitative descriptions of each corresponding feature. It is sub-classed according to the defined features. It is linked to the instances of **pRange** class through **hasRange** property
- Class **pRange**: is sub-classed to Minimum and Maximum and allows the definition of value restriction to the different feature parameters.
- Class **Algorithm**: the super-class of the available processing algorithms to be used during the analysis procedure. It is linked to the instances of **Feature-Parameter** class through the **useFeatureParameter** property.

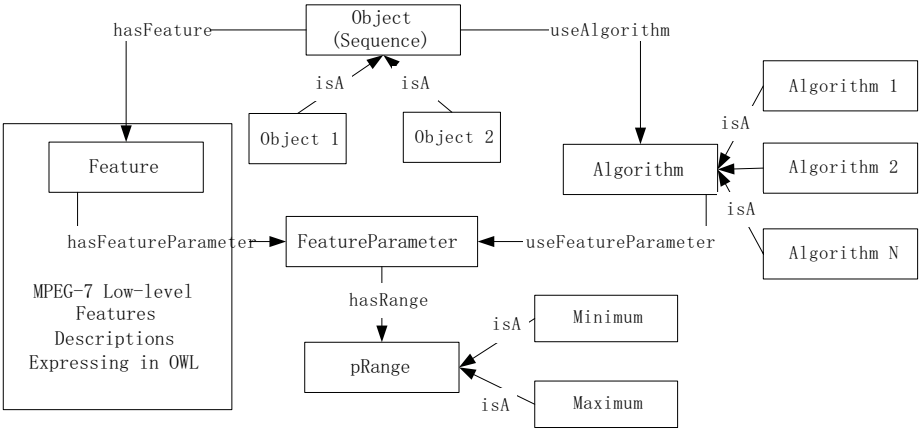

Fig. 2. Classes and Properties in Video Analysis Ontology

3.2 Expressing MPEG-7 in OWL

In this paper, we try to combine the low-level visual and aural descriptions of MPEG-7 into video analysis ontology for constructing a common understanding low-level features description for different video content. In the same way, we can combine other parts of MPEG-7 into an OWL ontology.

The set of features or properties which is specific to the visual entities defined in MPEG-7 include: Color, Texture, Motion and Shape. Each of these features can be represented by a choice of descriptors. Similarly there is a set of audio features which is applicable to MPEG-7 entities containing audio: Silence, Timbre, Speech and Melody.

Taking the visual feature descriptor "Color" as an example, we demonstrate in Figure 3, how MPEG-7 descriptions are combined into video analysis ontology with OWL definitions.

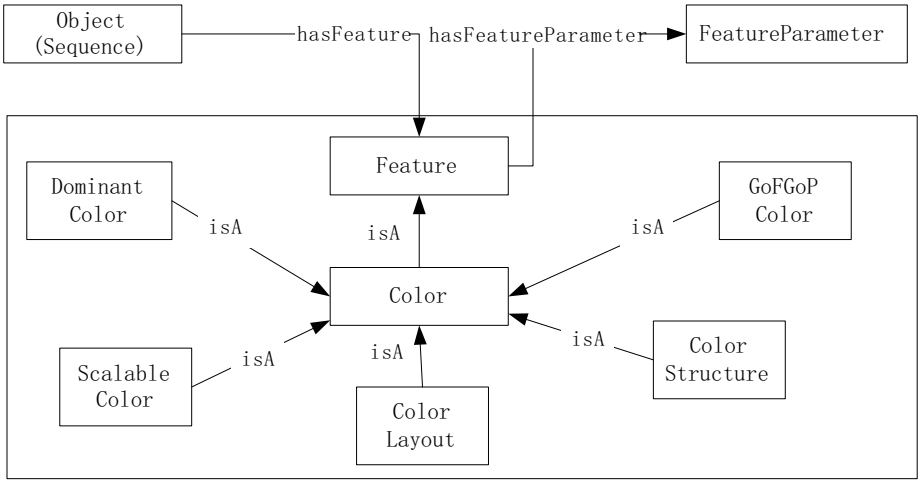

Fig. 3. Definitions of MPEG-7 Color Descriptor in Video Analysis Ontology

An example of color descriptor expressed in OWL is shown in List 1.

List 1. Example of Color Descriptor Expressing in OWL

```
…
<owl:Class rdf:ID = "Color"/ >
      <rdfs:label>Color</rdfs:label>
      <rdfs:subClassOf rdf:resource="#Feature"/>
</owl:class>
<owl:Class rdf:ID = "DominantColor">
       <rdfs:label>DominantColor</rdfs:label>
        <rdfs:subClassOf rdf:resource="#Color"/>
</owl:class>
<owl:Class rdf:ID = "ScalableColor">
       <rdfs:label> ScalableColor </rdfs:label>
        <rdfs:subClassOf rdf:resource="#Color"/>
</owl:class>
…
```

4 Rules in Description Logic Construction

The choice of algorithm employed for the detection of sequences and objects is directly dependent on its available characteristic features which directly depend on the domain that the sequences and objects involve. So this association should be considered based on video analysis knowledge and domain knowledge, and is useful for automatic and precise detection. In our work, the association is described by a set of properly defined rules represented in DL.

The rules for detection of sequences and objects are: rules to define the mapping between sequence (or object) and features, rules to define the mapping between sequence (or object) and algorithm, and rules to determine algorithms input feature parameters. The rules are represented in DL as follows:

- An sequence 'S' has features F_1, F_2, …, F_n: $\exists hasFeature(S, F_1, F_2, ..., F_n)$
- An sequence 'S' detection use algorithms A_1, A_2, …, A_n :
  $\exists useA\lg orithm(S, A_1, A_2, ..., A_n)$
- An object 'O' has features F_1, F_2, …, F_n: $\exists hasFeature(O, F_1, F_2, ..., F_n)$
- An object 'O' detection uses algorithms A_1, A_2, …, A_n :
  $\exists useA\lg orithm(O, A_1, A_2, ..., A_n)$
- An algorithm 'A' uses features parameters FP_1, FP_2, …, FP_n :
  $\exists useFeatureParameter(A, FP_1, FP_2, ..., FP_n)$
- If $S \cap (\exists hasFeature.F \cap \exists hasA\lg orithm.A)$
  Then $\exists useFeatureParameter(A, FP)$ *(FP* is the parameter values of *F.)*
- If $O \cap (\exists hasFeature.F \cap \exists hasA\lg orithm.A)$
  Then $\exists useFeatureParameter(A, FP)$ *(FP* is the parameter values of *F.)*

In the next section, a sports ontology is constructed which provides the vocabulary and domain knowledge. In the context of video content analysis the domain ontology maps to the important objects, their qualitative and quantitative attributes and their interrelation.

In videos events are very important semantic entities. Events are composed of special objects and sequences and their temporal relationships. A general domain ontology is appropriate to describe events using linguistic terms. It is inadequate when it must describe the temporal patterns of events. Basic DL lacks of constructors which can express temporal semantics. So in this paper, Temporal Description Logic (TDL) is used to describe the temporal patterns of semantic events based on detected sequences and objects. TDL is based on temporal extensions of DL, involving the combination of a rather expressive DL with the basis tense modal logic over a linear, unbounded, and discrete temporal structure. $\mathcal{TL}$-$\mathcal{F}$ is the basic logic considered in this paper. This language is composed of the temporal logic $\mathcal{TL}$, which is able to express interval temporal networks, and the non-temporal Feature Description Logic $\mathcal{F}$ [23].

The basic temporal interval relations in $\mathcal{TL}$-$\mathcal{F}$ are: before (b), meets (m), during (d), overlaps (o), starts (s), finishes (f), equal (e), after(a), met-by (mi), contains (di), overlapped-by (oi), started-by (si), finished-by (fi).

Objects and sequences in soccer videos can be detected based on video analysis ontology. Events can be described by means of the occurrence of the objects and sequences, and the temporal relationships between them. The events description and reasoning algorithm for event detection are introduced in next section.

5 Sports Domain Ontology

As previously mentioned, for the demonstration of our framework an application in the sports domain is proposed. The detection of semantically significant sequences and objects, such as close-up shots, players and referees, is important for understanding and extracting video semantic content, and modeling and detecting the events in the sports video. The features associated with each sequence and object comprise their definitions in terms of low-level features as used in the context of video analysis. The category of sequences and objects and the selection of features are based on domain knowledge. A sports domain ontology is constructed and the definitions used for this ontology are described in this section.

5.1 Objects

Only a limited number of object types are observed in sports videos. Visual objects include: ball, player, referee, coach, captions, goalposts in soccer, basket in basketball and so on (see figure 4).

In general, in a sports match there are two kinds of important audio: whistle and cheers. So the individuals of aural object class are: whistle and cheers.

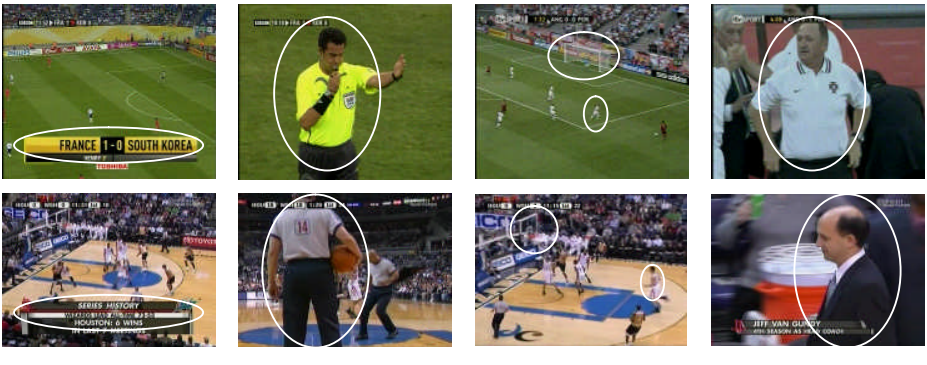

Fig. 4. Objects in Sports Videos

5.2 Sequences

In sports videos we observe just three distinct visual sequence classes: Loose View, Medium View and Tight View. The loose view and medium view share analogical

visual features and are often associated with one shot zooming action, so they can be defined as one visual sequence style named Normal View. When some highlights occur, the camera often captures something interesting in the arena, called Out-of-field. Important semantic events are often replayed in slow motion immediately after they occur. So individuals of visual sequence class are: Normal View (NV), Tight View (TV), Out-of-field (OOF) and Slow-motion-replay (SMR). (For example in soccer, see Figure 5).

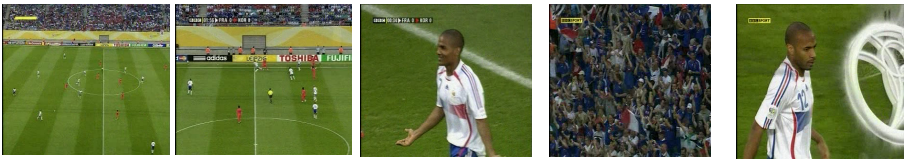

Fig. 5. Sequences in Soccer Game

5.3 Features and Algorithms

In section 3.2, we have combined MPEG-7 visual and aural descriptions into video analysis ontology expressed in OWL. The definitions of these visual and aural features are used for the detections of the sequences and objects defined in the sports domain ontology.

In our previous work [24], HMM was used for distinguishing different visual sequences, Sobel edge detection algorithm and Hough transform are used to detect "Goalposts" object, and image cluster algorithm based on color features have been proved to be effective in the soccer videos content analysis domain. The pixel-wise mean square difference of the intensity of every two subsequent frames and RGB color histogram of each frame can be used in a HMM model for slow-motion-replay detection [25]. For detection of aural objects, frequency energy can be used in SVM model for detection of "Cheers"[26], "Whistles" can be detected according to peak frequencies which fall within a threshold range [27].

5.4 Events Description and Detection

It is possible to detect events in sports videos by means of reasoning on TDL once all the sequence and objects defined above are detected with the video content analysis ontology. In order to do this we have observed some temporal patterns in soccer videos in terms of series of detected sequences and objects. For instance, if an attack leads to a scored goal, cheers from audience occurs immediately, then sequences are from "Goal Area" to "Player Tight View", "Out-of-Field", "Slow Motion Replay", and another player "Tight View", and finally returning to "Normal View", then a "Caption" is shown. Essentially these temporal patterns are the basic truth existing in sports domain which characterize the semantic events in sports videos and can be used to formally describe the events and detect them automatically. TDL is used for descriptions of the events. And the necessary syntaxes in TDL are listed as follows:

x, y denote the temporal intervals;

$\Diamond$ is the temporal existential quantifier for introducing the temporal intervals, for example: $\Diamond(x, y)$;

@ is called bindable, and appears in the left hand side of a temporal interval. A bindable variable is said to be bound in a concept if it is declared at the nearest temporal quantifier in the body of which it occurs.

For example, the description of goal scored event in soccer event is as follows:

$$Scoredgoal = \Diamond\left(d_{goal}, d_{whistle}, d_{cheers}, d_{caption}, d_{GA}, d_{TV}, d_{OOF}, d_{SMR}\right)$$
$$\left(d_{goal}\ f\ d_{GA}\right)\left(d_{whistle}\ d\ d_{GA}\right)\left(d_{GA}\ o\ d_{cheers}\right)\left(d_{caption}\ e\ d_{TV}\right)$$
$$\left(d_{cheers}\ e\ d_{TV}\right)\left(d_{GA}\ m\ d_{TV}\right)\left(d_{TV}\ m\ d_{OOF}\right)\left(d_{OOF}\ m\ d_{MSR}\right).$$
$$(goal\ @\ d_{goal} \cap whistle\ @\ d_{whistle} \cap cheers\ @\ d_{cheers} \cap caption\ @\ d_{caption} \cap$$
$$GA\ @\ d_{GA} \cap TV\ @\ d_{TV} \cap OOF\ @\ d_{OOF} \cap SMR\ @\ d_{SMR})$$

$d_{goal}, d_{whistle}, d_{cheers}, d_{caption}, d_{GA}, d_{TV}, d_{OOF}, d_{SMR}$ represent the temporal intervals of responding objects and sequences.

Based on the descriptions of event in TDL, reasoning on event detection can be designed. After detection of sequences and objects in a sports video, every sequence and object can be described as formal in TDL as: $\Diamond x(\ ).C\ @\ x$. C is the individual of sequence or object; x is the temporal interval of C. $()$ denotes C dose not any temporal relationship with itself. So the reasoning algorithm is described as follows:

Suppose: $\{S_0, S_1, ..., S_{n-1}, S_n\}$ is a sequence individuals set from detection results of a soccer video. Each element S_i in $\{S_0, S_1, ..., S_{n-1}, S_n\}$ can be represented as follows:

$$S_i = \Diamond x_i(\ ).S_i\ @\ x_i$$

The definition of $\{S_0, S_1, ..., S_{n-1}, S_n\}$ includes a latent temporal constraint: $x_i\ m\ x_{i+1}, i = 0, 1, ..., n-1$ which denotes two consecutive sequences in $\{S_0, S_1, ..., S_{n-1}, S_n\}$ are consecutive in the temporal axis of the video.

$\{O_0, O_1, ..., O_{m-1}, O_m\}$ is object individuals set from detection results of a soccer video. Each element O_i in $\{O_0, O_1, ..., O_{m-1}, O_m\}$ can be represented as follows:

$$O_i = \Diamond y_i(\ ).O_i\ @\ y_i$$

Reasoning algorithm for goal scored event in soccer video:

Step 1. Select the subsets in $\{S_0, S_1, ..., S_{n-1}, S_n\}$ which are composed of consecutive sequences individuals *GA*->TV->OOF->MSR. Each of the subsets is a candidate goal scored event E_{Ck}.

$$E_{Ck} = \{GA_k, TV_{k+1}, OOF_{k+2}, MSR_{k+3}\}$$

where k is the subscript mark of the current *NV* of the current candidate event in $\{S_0, S_1, ..., S_{n-1}, S_n\}$.

Step 2. For each candidate event E_{Ck}, Search goal objects O_{goal}, $O_{whistle}$, O_{cheers}, $O_{caption}$ in $\{O_0, O_1, ..., O_{m-1}, O_m\}$, they have corresponding temporal intervals y_{goal}, $y_{whistle}$, y_{cheers}, $y_{caption}$, and satisfy corresponding temporal constrains $y_{goal}\ f\ GA_k$, $y_{whistle}\ d\ GA_k$, $GA_k\ o\ y_{cheers}$, $y_{caption}\ e\ TV_{k+1}$, $y_{cheers}\ e\ TV_{k+1}$. If all of such objects exist, E_{Ck} is a goal scored event.

Other events can be detected using same reasoning algorithm. We just need to adjust the definition of candidate event subset and searched objects. A particular strength of the proposed reasoning algorithm for events description and detection in TDF based on domain ontology is that the user can define and describe different events, and use different description in TDL for the same event based on their domain knowledge.

6 Experiment and Results

The proposed framework was tested in the sports domain. In this paper we focus on developing the framework for video content analysis based on ontology and demonstrating the validity of the proposed reasoning algorithm in TDL for event detection. So the experiments described here used a manually annotated data set of objects and sequences in sports videos. Experiments were carried out using five soccer games and three basketball games recordings captured from 4:2:2 YUV PAL tapes which were saved as MPEG-1 format. The soccer videos are from two broadcasters, ITV and BBC Sport, and are taken from the 2006 World Cup, taking a total of 7hs 53mins28s. The basketball videos are NBA games recorded from ESPN, FOX Sports and CCTV5 taking a total of 6hs 47mins 18s.

For soccer videos we defined Goal Scored, Foul in Soccer and Yellow (or Red) Card events. And Highlight Attack and Foul events are defined and detected in basketball videos. Table 1 shows “Precision” and “Recall” for detection of the semantic events. “Actual Num” is the actual number of events in entire matches, which are recognized manually; “True Num” is the number of detected correct matches, and “False Num” is the number of false matches.

From Table 1, it can be seen that the precision results of event detection are higher than 89%, but the recall results are relatively low. This is because the description in TDL is very strict in logic and do not allow any difference between the definition of events and the occurrence of events to be detected, thus the reasoning algorithm for event detection can ensure high precision, but it may lose some correct results. If we define different descriptions in TDL for the same event which has different composition of objects, sequences and temporal relationship, high recall can be obtained.

Table 1. Precision and recall for five soccer and basketball semantics

semantic	Actual Num	True Num	False Num	Precision (%)	Recall (%)
Goal Scored	10	8	0	100	80
Foul in Soccer	193	141	11	92.8	73.1
Yellow (or Red) Card	26	22	2	91.7	84.6
Highlight Attack	45	36	4	90.0	80.0
Foul in Basketball	131	106	12	89.8	80.9

We also compared the proposed approach with other approaches. In our previous work, a Petri-Net (PN) model is used for video semantic content description and detection [28]. HMM is a popular model for video event detection. In our experiments, we use the PN based approach and HMM based approach proposed in [24] to detect semantic content using same video data set. The results are shown in Table 2.

Table 2. Results based on PN and HMM Approach

semantic		Goal Scored	Foul in Soccer	Yellow(Red) Card	Highlight Attack	Foul in Basketball
PN	**Pre(%)**	85.2	86.6	91.7	85.8	84.5
	Rec(%)	100	84.1	97.5	91.6	90.3
HMM	**Pre(%)**	75.4	63.8	77.6	61.5	59.2
	Rec(%)	80.1	72.5	83.1	64.9	67.3

From Table 2, we can find the precision and recall of PN based approach is almost equivalent with the proposed approach. It is because both of these approaches detect high-level semantic events based on middle semantics, objects and sequences. Low precision and recall are shown in the experimental results of HMM based approach, in which low-level features are extracted to training different HMM models for different semantic content. This approach maps low-level features to high-level semantic directly, which can capture perception feature pattern well but not be effective to model and detect spatiotemporal relationship between different semantic content.

Based on the above experimental results, we believe that the proposed framework for video content analysis and event detection method based on TDL have considerable potential. We are currently conducting a more thorough experimental investigation using a larger set of independent videos and utilizing the framework in different domains.

7 Conclusions and Discussions

In this paper, a video semantic content analysis framework based on ontology combined MPEG-7 is presented. A domain ontology is used to define high level semantic concepts and their relations in context of the examined domain. MPEG-7 low-level feature descriptions expressing in OWL and video content analysis algorithms are integrated into the ontology to enrich video semantic analysis.

In order to create domain ontology for video content analysis, owl is used for ontology description language and Rules in DL are defined to describe how features and algorithms for video analysis should be applied according to different perception content and low-level features, and TDL is used to describe semantic events. A ontology in the sports domain is constructed using Protégé for demonstrating the validity of the proposed framework. A reasoning algorithm based on TDL is proposed for event detection in sports videos. The proposed framework supports flexible and managed execution of various application and domain independent video low-level analysis tasks.

Future work includes the enhancement of the domain ontology with more complex model representations and the definition of semantically more important and complex events in the domain of discourse.

Acknowledgement

This work is supported by the National High Technology Development 863 Program of China (2006AA01Z316), the National Natural Science Foundation of China (60572137) and China Scholarship Council of China Education Ministry.

References

1. Chang, S.-F.: The holy grail of content-based media analysis. IEEE Multimedia 9(2), 6–10 (2002)
2. Yoshitaka, A., Ichikawa, T.: A survey on content-based retrieval for multimedia databases. IEEE Transactions on Knowledge and Data Engineering 11(1), 81–93 (1999)
3. Hanjalic, A., Xu, L.Q.: Affective video content representation and modeling. IEEE Transactions on Multimedia 7(1), 143–154 (2005)
4. Muller-Schneiders, S., Jager, T., Loos, H.S., Niem, W.: Performance evaluation of a real time video surveillance system. In: 2nd Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, October 15-16, 2005, pp. 137–143 (2005)
5. Hua, X.S., Lu, L., Zhang, H.J.: Automatic music video generation based on the temporal pattern analysis. In: 12th annual ACM international conference on Multimedia (October 2004)
6. Resource description framework. Technical report, W3C (February 2004), http://www.w3.org/RDF/
7. Web ontology language (OWL). Technical report, W3C (2004), http://www.w3.org/2004/OWL/
8. Ekin, A., Tekalp, A.M., Mehrotra, R.: Automatic soccer video analysis and summarization. IEEE Transactions on Image Processing 12(7), 796–807 (2003)

9. Yu, X., Xu, C., Leung, H., Tian, Q., Tang, Q., Wan, K.W.: Trajectory-based ball detection and tracking with applications to semantic analysis of broadcast soccer video. In: ACM Multimedia 2003, Berkeley, CA(USA), November 4-6, 2003, vol. 3, pp. 11–20 (2003)
10. Xu, H.X., Chua, T.-S.: Fusion of AV features and external information sources for event detection in team sports video. ACM transactions on Multimedia Computing, Communications and Applications 2(1), 44–67 (2006)
11. Reidsma, D., Kuper, J., Declerck, T., Saggion, H., Cunningham, H.: Cross document ontology based information extraction for multimedia retrieval. In: Supplementary proceedings of the ICCS 2003, Dresden (July 2003)
12. Mezaris, V., Kompatsiaris, I., Boulgouris, N., Strintzis, M.: Real-time compressed-domain spatiotemporal segmentation and ontologies for video indexing and retrieval. IEEE Transactions on Circuits and Systems for Video Technology 14(5), 606–621 (2004)
13. Jaimes, A., Tseng, B., Smith, J.: Modal keywords, ontologies, and reasoning for video understanding. In: Bakker, E.M., Lew, M., Huang, T.S., Sebe, N., Zhou, X.S. (eds.) CIVR 2003. LNCS, vol. 2728, Springer, Heidelberg (2003)
14. Jaimes, A., Smith, J.: Semi-automatic, data-driven construction of multimedia ontologies. In: Proc. of IEEE Int'l Conference on Multimedia & Expo (2003)
15. Bertini, M., Bimbo, A.D., Torniai, C.: Enhanced ontoloies for video annotation and retrieval. In: ACM MIR'2005, Singapore, November 10-11 (2005)
16. Bentitez, A., Chang, S.-F.: Automatic multimedia knowledge discovery, summarization and evaluation. IEEE Transactions on Multimedia (submitted, 2003)
17. Dasiopoulou, S., Papastathis, V.K., Mezaris, V., Kompatsiaris, I., Strintzis, M.G.: An Ontology Framework for Knowledge-Assisted Semantic Video Analysis and Annotation. In: McIlraith, S.A., Plexousakis, D., van Harmelen, F. (eds.) ISWC 2004. LNCS, vol. 3298, Springer, Heidelberg (2004)
18. Kompatsiaris, I., Mezaris, V., Strintzis, M.G.: Multimedia content indexing and retrieval using an object ontology. In: Stamou, G. (ed.) Multimedia Content and Semantic Web Methods, Standards and Tools. Wiley, New York (2004)
19. MPEG-7 Overview (October 2004), `http://www.chiariglione.org/mpeg`
20. TV-Anytime Forum, `http://www.tv-anytime.org/`
21. MPEG-21 Multimedia Framework, http://www.cselt.it/mpeg-21_pdtr.zip
22. NewsML, http://www.newsml.org
23. Artale, A., Franconi, E.: A temporal description logic for reasoning about actions and plans. Journal of Artificial Intelligence Research 9, 463–506 (1998)
24. Chen, J.Y., Li, Y.H., Lao, S.Y., et al.: Detection of Scoring Event in Soccer Video for Highlight Generation. Technical Report, National University of Defense Technology (2004)
25. Pan, H., van Beek, P., Sezan, M.I.: Detection of Slow-motion Replay Segments in Sports Video for Highlights Generation. In: Proceedings of IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP 2001), Salt Lake City, UT, USA (May 2001)
26. Liang, B., Yanli, H., Songyang, L., Jianyun, C., Lingda, W.: Feature Analysis and Extraction for Audio Automatic Classification. In: IEEE SMC 2005, Hawaii USA, October 10-12 (2005)
27. Zhou, W., Dao, S., Jay Kuo, C.-C.: On-line knowledge and rule-based video classification system for video indexing and dissemination. Information Systems 27(8), 559–586 (2002)
28. Lao, S.Y., Smeaton, A.F., Jones, G.J.F., Lee, H.: A Query Description Model Based on Basic Semantic Unit Composite Petri-Nets for Soccer Video Analysis. In: Proceedings of ACM MIR 2004, New York, USA, October 15-16 (2004)