View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by DCU Online Research Access Service

International Machine Vision and Image Processing Conference

# Video Semantic Content Analysis based on Ontology

Liang Bai[1, 2], Songyang Lao[1], Gareth J. F. Jones[2], 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*
[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 standards, such as MPEG-4 and MPEG-7, provide the basic functionalities in order to manipulate and transmit objects and metadata. But importantly, most of the content of video data at a semantic level is out of the scope of the standards. In this paper, a video semantic content analysis framework based on ontology is presented. Domain ontology is used to define high level semantic concepts and their relations in the context of the examined domain. And low-level features (e.g. visual and aural) and video content analysis algorithms are integrated into the ontology to enrich video semantic analysis. OWL is used for the ontology description. Rules in Description Logic are defined to 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. The proposed framework is demonstrated in a soccer video domain and shows promising results.*

## 1. Introduction

As a result of recent progress in high-speed broadband networks, digital video and hardware technologies, video has become a major source of content on the WWW, Digital TV and other multimedia application fields, such as: digital library and video on demand. The rapid increase in 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].

Although new multimedia standards, such as MPEG-4 and MPEG-7, provide the basic functionalities in order to manipulate and transmit objects and metadata, at a semantic level most video content is out of the scope of the standards. 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][6]. 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.

The main challenge, often referred to as the semantic gap, is mapping high-level semantic concepts into low-level spatio-temporal features that can be automatically extracted from video data. In many cases, the mapping rules must be 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.

An ontology is a formal, explicit specification 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) [7], Resource Description Framework Schema (RDFS), Web Ontology Language

0-7695-2887-2/07 $25.00 © 2007 IEEE
DOI 10.1109/IMVIP.2007.13



IEEE computer society

(OWL) [8] and, for multimedia, the XML Schema in MPEG-7.

Many automatic semantic content analysis systems have been presented recently [9] [10] [11] [12] and [18] 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 [13]. In [14], concepts are expressed as keywords and are mapped in an 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 any given domain, it is a challenge to use them for describing low-level features, video content analysis and the relationships between them.

An extended linguistic ontology with multimedia ontology was presented in [15] to support video understanding. A multimedia ontology is constructed manually in [16]. Marco Bertini et al., in [17], present algorithms and techniques that employ an enriched ontology for video annotation and retrieval. In [19], 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 [20], a visual descriptor ontology and a multimedia structure ontology, based on MPEG-7 visual descriptors and MPEG-7 MDS respectively, are used together with a domain ontology in order to support content annotation. In [21], 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.

In this paper, a framework for video semantic content analysis based on ontology is presented. Content-based analysis of video requires methods which will automatically segment video sequences and select key frames corresponding to semantic content that share similar spatio-temporal behaviors, and provide a flexible framework for indexing, and retrieval, and for further analysis of their relationships. In the proposed video semantic content analysis framework, a video analysis ontology is developed to formally describe the detection process of the video semantic content. Semantic concepts within the context of the examined domain area are defined in a 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. The 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.

The paper is organized as follows: the overall framework is outlined in section 2, a detailed definition of the ontology is given in section 3, the rules in DL based on the defined ontology for video processing and event detection are constructed in section 4, while in section 5 an application to the soccer domain is described, experimental results are presented in section 6, and in section 7 we provide conclusions and details of future works.

## 2. Framework for Video Semantic Content Analysis based on Ontology

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. Semantic concepts within the context of the examined domain area are defined in a domain ontology, enriched with qualitative attributes of the semantic content, low-level features and video processing algorithms which determined by the semantic content of video to be detected and its low-level features. The 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. Further, the ontology-based approach and the utilization of the 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.

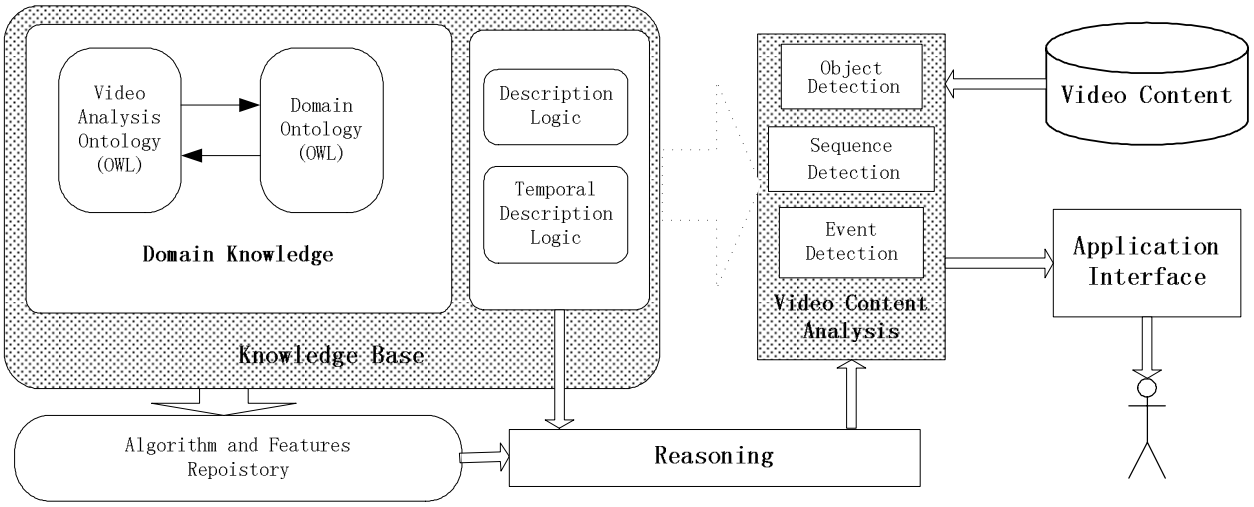


**Figure.1 Framework**

## 3. Ontology Development for Video Semantic Content Analysis

In order to realize the knowledge-based and automatic video semantic content analysis introduced in the previous section, the knowledge for video analysis is abstracted and a video analysis ontology is constructed. Many features and algorithms have been developed in the video content analysis field. 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 the 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.

- Class **Event**: the subclass and instance of the superclass "Event". Each event instance is a composition of special object instance and sequence instance and their temporal relationships.
- Class **Sequence**: the subclass and instance of the superclass "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 subclassed to **VisualSequence** 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 superclass "Object", all video objects can be detected through the analysis process at frame level. It is subclassed 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 superclass of video low-level features associated with each sequence and object, including audio track low-level features.
- Class **FeatureParameter**: denotes the actual qualitative descriptions of each corresponding feature. It is subclassed according to the defined features.
- Class **pRange**: is subclassed to Minimum and Maximun and allows the definition of value restriction to the different feature parameters.
- Class **Algorithm**: the superclass of the available processing algorithms to be used during the analysis procedure. It is linked to the instances of the **FeatureParameter** class through the **useFeatureParameter** property.

The classes defined above are expressed in the OWL language in our work.

## 4. Rules in Description Logic Construction

As mentioned in section 3, many features and algorithms for video content analysis have been proposed. 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 the algorithm's input feature parameters. The rules are represented in DL as follows:

- A sequence 'S' has features $F_1$, $F_2$, …, $F_n$:

$$\exists hasFeature(S, F_1, F_2, ..., F_n)$$

- A 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 $S \cap (\exists hasFeature.F \cap \exists has\text{Algorithm}.A)$

  Then $\exists useFeatureParameter(A, FP)$

  *(FP* is the parameter values of *F.)*

- If $O \cap (\exists hasFeature.F \cap \exists has\text{Algorithm}.A)$

  Then $\exists useFeatureParameter(A, FP)$

  *(FP* is the parameter values of *F.)*

In the next section, a domain ontology is constructed which provides the vocabulary and background knowledge of the domain. 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 constructors which can express temporal semantics. So in this paper, temporal description logic 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 $\mathcal{F}$ [22]. 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).

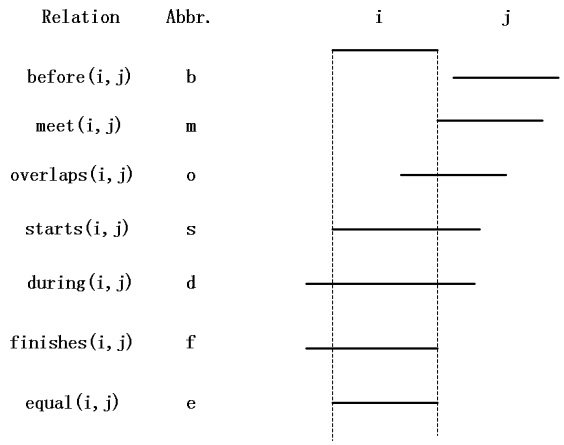


**Figure.2 Temporal Interval Relations**

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

## 5. Soccer Domain Ontology

As previously mentioned, for the demonstration of our framework an application in the soccer 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 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 soccer 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 soccer videos. Visual objects include: ball, player, referee (and assistant referees), coach, goalposts, sideline, corner arc, and on screen captions. According to the requirement of semantic content analysis in this paper we only select three objects as individuals of visual object class: caption, goal and referee.

Some sounds are useful for semantic analysis of games. In general, in a soccer match there are two kinds of important audio: whistle and cheers. So the individuals of the aural object class are: whistle and cheers.

## 5.2. Sequences

In soccer videos we observe just three distinct visual sequence classes: Loose View, Medium View and Tight View. In soccer videos 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). Further based on the different areas of the playing area, Normal View can be divided into eight subclasses: Left Goal Area, Right Goal Area, Midfield Upper, Midfield Lower, Left Upper Corner, Left Lower Corner, Right Upper Corner and Right Lower Corner.

## 5.3. Features and Algorithms

According to definitions of the objects and sequences in the soccer domain and much observation of soccer video data, we found that the visual objects and sequences in soccer videos can be characterized by similar color or similar shape. So color features and shape features are used in soccer domain ontology. MPEG-7 visual descriptors are selected for our work [27]. Dominant color and color layout are two individual features of color features. Dominant color can represent local features where a small number of colors are enough to characterize the color information in the region of interest. Color layout is effective to describe the spatial distribution of the color of visual signals in a very compact form. The color features are effective for distinguishing different visual sequences and detecting the visual objects that are characterized by similar colors, such as "Referee". The region shape feature makes use of all pixels constituting the shape within a frame, it can describe complex shapes and be robust to minor deformation along the boundary of the object. This feature is useful for detection of the objects in soccer videos that have fixed shape [23], such as: "caption". In previous work [23], an HMM was used for distinguishing different visual sequences, Sobel edge detection and Hough transform are used for detecting "Goalposts" object, and image cluster algorithm based on color features have been proved to be effective in the soccer video 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 [24]. For detection of aural objects, frequency energy can be used in an SVM model for detection of "Cheers" [25], and "Whistles" can be detected according to peak frequencies which fall within a threshold range [26].

## 5.4. Event Description and Detection

It is possible to detect events in soccer videos by means of reasoning in TDL once all the sequence and objects defined above are detected using 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 the audience occur 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 the soccer domain which characterize the semantic events in soccer videos and can be used to formally describe the events and detect them automatically. These are also the same patterns as used by Sadiler and O'Conner in [18]. The events described in this paper are goal scored and foul. 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 to a concept if it is declared at the nearest temporal quantifier in the body of which it occurs.

- Goal scored

In soccer videos, a goal scored event often includes goal object, whistle object, cheers object, caption object, and, Goal Area (GA), TV sequence, OOF

sequence, SMR sequence. The goal scored event is described in TDL as follows:

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

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

- Foul

A foul event in soccer videos often happens in a NV sequence with a whistle object, followed by a TV sequence with a referee object, and a MSR sequence in the end. The foul event is described in TDL as follows:

$$Foul = \Diamond\left(d_{whistle}, d_{referee}, d_{NV}, d_{TV}, d_{SMR}\right)$$
$$\left(d_{whistle}\ d\ d_{NV}\right)\left(d_{referee}\ d\ d_{TV}\right)\left(d_{NV}\ m\ d_{TV}\right)\left(d_{TV}\ m\ d_{MSR}\right).$$
$$(whistle@d_{whistle} \cap referee@d_{referee}$$
$$NV@d_{NV} \cap TV@d_{TV} \cap SMR@d_{SMR})$$

$d_{whistle}, d_{referee}, d_{NV}, d_{TV}, d_{SMR}$ represent the temporal intervals of responding objects and sequences.

If the foul causes a yellow card or a red card, a caption object will occur. The description of a yellow or red card event is as follows:

$$Foul = \Diamond\left(d_{whistle}, d_{referee}, d_{caption}, d_{NV}, d_{TV}, d_{SMR}\right)$$
$$\left(d_{whistle}\ f\ d_{NV}\right)\left(d_{referee}\ d\ d_{TV}\right)\left(d_{caption}\ d\ d_{TV}\right)$$
$$\left(d_{NV}\ m\ d_{TV}\right)\left(d_{TV}\ m\ d_{MSR}\right).$$
$$(whistle@d_{whistle} \cap referee@d_{referee} \cap caption@d_{caption}$$
$$NV@d_{NV} \cap TV@d_{TV} \cap SMR@d_{SMR})$$

$d_{caption}$ is the temporal interval of caption object.

Based on the descriptions of event in TDL, reasoning on event detection can be designed. After detection of sequences and objects in a soccer video, every sequence and object can be described formally in TDL as follows:

$$\Diamond x(\ ).C@x$$

*C* is the individual of sequence or object; *x* is the temporal interval of *C*. *()* denotes C do not have 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 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:**

**Step1.** 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 $\{S_0, S_1, ..., S_{n-1}, S_n\}$.

**Step2.** 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 $y_{goal}\ f\ GA_k$, $y_{whistle}\ d\ \mathrm{GA}_k$, $GA_k\ o\ y_{cheers}$, $y_{caption}\ e\ \mathrm{TV}_{k+1}$, $y_{cheers}\ e\ \mathrm{TV}_{k+1}$. If all of such objects exist, $E_{Ck}$ is a goal scored event.

Other events can be detected using a similar 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 event description and detection in TDF based on domain ontology is that the user can define and describe different events, and use different descriptions in TDL for the same event based on their domain knowledge. For example, the user can define a different TDL description from the one used here for the goal scored event.

## 6. Experiment and Results

The proposed framework was tested in the soccer domain. In the domain, an appropriate domain ontology was constructed which describes knowledge of the associated rules for the application of the

appropriate video processing algorithms using suitable features and parameter values, and the definition of objects, sequences and events. For ontology creation the Protégé ontology engineering environment was used, OWL DL is used as the output language.

The detection of objects and sequences has previously been introduced in [23]. 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 soccer videos.

Experiments were carried out using five soccer game 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.

Table1 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.

**Table.1. Precision and recall for three soccer semantics**

| *semantic* | *Actual Num* | True | False | Precision | Recall |
|---|---|---|---|---|---|
| *Goal* | *10* | *8* | *0* | *100%* | *80%* |
| *Foul* | *193* | *141* | *11* | *92.8%* | *73.1%* |
| *YR Card* | *26* | *22* | *2* | *91.7%* | *84.6%* |

From Table 1, it can be seen that the precision results of event detection are higher than 91%, 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. A shortcoming of using different descriptions for same event is that it increases the complexity of the knowledge base. The wrong results for yellow (or red) card event occur because when a player is injured, a MSR and a Caption object occur. In this case, a yellow card event is detected wrongly.

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, as well as using automatically-determined low-level annotations.

## 7. Conclusion and Discussions

In this paper, a video semantic content analysis framework based on ontology is presented. A domain ontology is used to define high level semantic concepts and their relations in the context of the examined domain. Low-level features (e.g. visual and aural) and video content analysis algorithms are integrated into the ontology to enrich video semantic analysis.

In order to create a 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. An ontology in the soccer 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 soccer videos. The proposed framework supports flexible and managed execution of various application and domain independent video low-level analysis tasks.

Experiments have shown the proposed framework is effective for video content analysis at the semantic level. Results for the reasoning algorithm for event detection have been presented in terms of precision and recall. High precision but relatively low recall are shown and analysis of results is given in detail.

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, as well as the use of automatically determined low level features. Further exploration of low-level multimedia features is expected to lead to more accurate and thus efficient representations of semantic content.

## 8. 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 Science Foundation Ireland through grant 03/IN.3/I361.

## 9. References

[1] S.-F. Chang. The holy grail of content-based media analysis. IEEE Multimedia, 9(2):6-10, Apr.-Jun. 2002
[2] A.Yoshitaka, 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.
[3] A.Hanjalic, L.Q.Xu. Affective video content representation and modeling. IEEE Transactions on Multimedia, 7(1):143-154, Feb.2005.
[4] S.Muller-Schneiders, T.Jager, H.S.Loos, W.Niem. Performance evaluation of a real time video surveillance system. 2nd Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance. 137-143, 15-16 Oct. 2005.
[5] X.S.Hua, L.Lu, H.J.Zhang. Automatic music video generation based on the temporal pattern analysis. 12th annual ACM international conference on Multimedia, October 2004.
[6] Informedia-II: Auto-Summarization and Visualization Over Multiple Video Documents and Libraries. Technical Report, September 2001, http://www.informedia.cs.cmu.edu
[7] Resource description framework. Technical report, W3C, http://www.w3.org/RDF/, Feb 2004.
[8] Web ontology language (OWL). Technical report, W3C, http://www.w3.org/2004/OWL/, 2004.
[9] R.Leonardi and P.Migliorati. Semantic index of multimedia documents. IEEE Multimedia, 9(2):44-51, April-June 2002.
[10] A.Ekin, A.M.Tekalp, and R.Mehrotra. Automatic soccer video analysis and summarization. IEEE Transactions on Image Processing, 12(7):796-807, July 2003.
[11] X.Yu, C.Xu, H.Leung, Q.Tian, Q.Tang, and K.W.Wan. Trajectory-based ball detection and tracking with applications to semantic analysis of broadcast soccer video. In ACM Multimedia 2003, volume 3, pages 11-20, Berkeley, CA(USA), 4-6 Nov.2003.
[12] H.X.Xu, T-S Chua. Fusion of AV features and external information sources for event detection in team sports video. ACM Transactions on Multimedia Computing, Communications and Applications, Vol. 2, No.1, Pages 44-67, February 2006.
[13] D.Reidsma, J.Kuper, T.Declerck, H.Saggion, and H.Cunningham. Cross document ontology based information extraction for multimedia retrieval. In Supplementary proceedings of the ICCS03, Dresden, July 2003.
[14] V.Mezaris, I.Kompatsiaris, N.Boulgouris, and M.Strintzis. 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.
[15] A.Jaimes, B.Tseng, and J.Smith. Modal keywords, ontologies, and reasoning for video understanding. In International Conference on Image and Video Retrieval (CIVR 2003), July 2003.
[16] A.Jaimes and J.Smith. Semi-automatic, data-driven construction of multimedia ontologies. In Proc.of IEEE Int'l Conference on Multimedia & Expo, 2003.
[17] M.Bertini, A.DelBimbo, C.Torniai. Enhanced ontoloies for video annotation and retrieval. In ACM MIR'2005, November 10-11, 2005, Singapore.
[18] D.A.Sadlier and N.E. O'Connor, "Event Detection in Field Sports Video Using Audio-visual Features and A SVM", IEEE Transactions on Circuits and Systems for Video Technology, 15(10), 1225-1233, 2005.
[19] S.Dasiopoulou, V.K.Papastathis, V.Mezaris, I.Kompatsiaris and M.G.Strintzis. An Ontology Framework for Knowledge-Assisted Semantic Video Analysis and Annotation. Proc. 4th International Workshop on Knowledge Markup and Semantic Annotation (SemAnnot 2004) at the 3rd International Semantic Web Conference (ISWC 2004), November 2004.
[20] J.Strintzis, S.Bloehdorn, S.Handschuh, S.Staab, N.Simou, V.Tzouvaras, K.Petridis, I.Kompatsiaris, and Y.Avrithis. Knowledge representation for semantic multimedia content analysis and reasoning. In European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology, Nov.2004.
[21] I.Kompatsiaris, V.Mezaris, and M.G.Strintzis. Multimedia content indexing and retrieval using an object ontology. Multimedia Content and Semantic Web Methods, Standards and Tools, Editor G.Stamou, Wiley, New York, NY, 2004.
[22] A.Artale, E.Franconi. A temporal description logic for reasoning about actions and plans. Journal of Artificial Intelligence Research 9, 463-506, 1998.
[23] J.Y. Chen, Y.H. Li, S.Y. Lao, et al, Detection of Scoring Event in Soccer Video for Highlight Generation. Technical Report, National University of Defense Technology, 2004.
[24] Hao Pan, P. van Beek, M. I. Sezan. 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'01), Salt Lake City, UT, USA, May 2001.
[25] Bai Liang, Hu Yanli, Lao Songyang, Chen Jianyun, Wu Lingda. "Feature Analysis and Extraction for Audio Automatic Classification", IEEE SMC 2005, Hawaii USA, October 10-12, 2005
[26] Zhou, W., S. Dao, and C.-C. Jay Kuo, On-line knowledge and rule-based video classification system for video indexing and dissemination. Information Systems, 2002. 27(8): p. 559-586.
[27] MPEG-7 Overview, http://www.chiariglione.org/mpeg, October 2004.