

A Fuzzy Ontology Based Automatic Video Content Retrieval

P.Anlet pamila suhi,S.Deepika

Department of CSE,Er.Perumal College of Engineering,India

deepikashanmugampt@gmail.com

Abstract

Recent advances in digital video analysis and extraction have made video more accessible than ever. The representation and recognition of events in a video is important for a number of tasks such as video surveillance, video browsing and content based video indexing. Rawdata and low-level features alone are not sufficient to fulfill the user's needs; that is, a deeper understanding of the content at the semantic level is required. Currently, manual techniques, which are inefficient, subjective and costly in time and limit the querying capabilities. Here, we propose a semantic content extraction system that allows the user to query and retrieve objects, events, and concepts that are extracted automatically. We introduce an ontology-based fuzzy video semantic content model that uses spatial/temporal relations in event and concept definitions. This metaontology definition provides a wide-domain applicable rule construction standard that allows the user to construct an ontology for a given domain. In addition to domain ontologies, we use additional rule definitions (without using ontology) to define some complex situations more effectively. The proposed framework has been fully implemented and tested on three different domains and it provides satisfactory results.

Index Terms-- Semantic content extraction, video content modeling, fuzziness, ontology.

1. INTRODUCTION

There is an increasing need to design efficient methods to semantically annotate video to store, retrieve and manage the information captured in them. Such extractions would not only help human users to easily query and manage their digital libraries, but also enable automated applications performing complicated tasks like video surveillance to create, store, exchange and reason with the data. The ultimate goal is to enable users to retrieve some desired content from massive amounts of video data in an efficient and semantically meaningful manner.

There are basically three levels of video content which are raw video data, low-level features and semantic content. First, raw video data consist of elementary physical video units together with some general video attributes such as format, length, and frame rate. Second, low-level features are characterized by audio, text, and visual features such as texture, color distribution, shape, motion, etc. Third, semantic content contains high-level concepts such as objects and events. These are the three levels of video content.

The first two levels on which content modeling and extraction approaches are based use automatically extracted data, which represent the low-level content of a video, but they hardly provide semantics which is much more appropriate for users. Users are mostly interested in querying and retrieving the video in terms of what the video contains. Therefore, raw video data and low-level features alone are not sufficient to fulfill the user's need; that is, a deeper understanding of the information at the semantic level is required in many video-based applications.

It is very difficult to extract semantic content directly from raw video data. This is because video is a temporal sequence of frames without a direct relation to its semantic content. Therefore, many different representations using different sets of data such as audio, visual features, objects, events, time, motion, and spatial relations are partially or fully used to model and extract the semantic content. No matter which type of data set is used, the process of extracting semantic content is complex and requires domain knowledge or user interaction.

A simple representation could relate the events with their low-level features using shots from videos, without any spatial or temporal relations. However, an effective use of spatiotemporal relations is crucial to achieve reliable recognition of events. Employing domain ontologies facilitate use of applicable relations on a domain. There are no studies using both spatial relations between objects, and temporal relations between events together in an ontology-based model to support automatic semantic content extraction.

A Video Event Recognition Language (VERL) that allows users to define the events without interacting with the lowlevel processing is defined. VERL is intended to be a language for representing events for the purpose of designing an ontology of the domain, and, Video Event Markup Language (VEML) is used to manually annotate VERL events in videos. The lack of low-level processing and using manual annotation are the drawbacks of this study.

In this study, a new Automatic Semantic Content Extraction Framework (ASCEF) for videos is proposed for bridging the gap between low-level representative features and high-level semantic content in terms of object, event, concept, spatial and temporal relation extraction. In order to address the modeling need for objects, events and concepts during the extraction process, an wide-domain applicable ontology-based fuzzy Video Semantic Content Model (VISCOM) that uses objects

and spatial/temporal relations in event and concept definitions is developed. VISCOM is a metaontology for domain ontologies and provides a domain-independent rule construction standard.

In the automatic event and concept extraction process, objects, events, domain ontologies, and rule definitions are used. The extraction process starts with object extraction. Specifically, a semiautomatic Genetic Algorithm-based object extraction approach is used for the object extraction and classification needs of this study. Then, objects extracted from consecutive representative frames are processed to extract temporal relations, which is an important step in the semantic content extraction process. In these steps, spatial and temporal relations among objects and events are extracted automatically allowing and using the uncertainty in relation definitions.

The organization of the paper is as follows. In Section 2, the proposed video semantic content model is described in detail. The automatic semantic content extraction system is explained in Section 3. In Section 4, the performed experiments and the performance evaluation of the system are given. Finally, in Section 5, our conclusions and future research directions are discussed.

II. VIDEO SEMANTIC CONTENT MODEL

VISCOM is a well-defined metaontology for constructing domain ontologies. It is an alternative to the rule based and domain-dependent extraction methods. Constructing rules for extraction is a tedious task and is not scalable. Without any standard on rule construction, different domains can have different rules with different syntax. In addition to the complexity of handling such difference, each rule structure can have weaknesses. Besides, VISCOM provides a standardized rule construction ability with the help of its metaontology. It eases the rule construction process and makes its use on larger video data possible.

The rules that can be constructed via VISCOM ontology can cover most of the event definitions for a wide variety of domains. However, there can be some exceptional situations that the ontology definitions cannot cover. To handle such cases, VISCOM provides an additional rulebased modeling capability without using ontology. Hence, VISCOM provides a solution that is applicable on a wide variety of domain videos.

Objects, events, concepts, spatial and temporal relations are components of this generic ontology-based model. Similar generic models such as which use objects and spatial and temporal relations for semantic content modeling neither use ontology in content representation nor support automatic content extraction. To the best of our knowledge, there is no domain-independent video semantic content model which uses both spatial and temporal relations between objects and which also supports automatic semantic content extraction as our model does.

The starting point is identifying what video contains and which components can be used to model the video content. Keyframes are the elementary video units which are still images, extracted from original video data that best

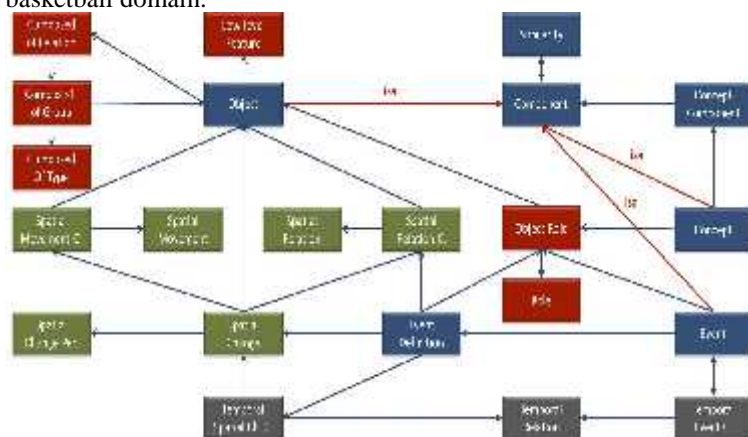
represent the content of shots in an abstract manner. Name, domain, frame rate, length, format are examples of general video attributes which form the metadata of video. Both the ontology model and the semantic content extraction process is developed considering uncertainty issues. For the semantic content representation, VISCOM ontology introduces fuzzy classes and properties. Spatial Relation Component, Event Definition, Similarity, Object Composed Of Relation and Concept Component classes are fuzzy classes as they aim to having fuzzy definitions.

III. ONTOLOGY-BASED MODELING

VISCOM is developed on an ontology-based structure where semantic content types and relations between these types are collected under VISCOM Classes, VISCOM Data Properties which associate classes with constants and VISCOM Object Properties which are used to define relations between classes. In addition, there are some domain independent class individuals. C-Logic is used for the formal representation of VISCOM classes and operations of the semantic content extraction framework. C-Logic includes a representation framework for entities, their attributes, and classes using identities, labels, and types.

VISCOM collects all of the semantic content under the class of Component. A component can have synonym names and similarity relations with other components. Component class has three subclasses as Objects, Events, and Concepts. Objects correspond to existential entities. An object is the starting point of the composition. An object has a name, low-level features, and composed-of relations. Basketball player, referee, ball and hoop are examples of objects for the basketball domain.

Events are long-term temporal objects and object relation changes. They are described by using objects and spatial/temporal relations between objects. Relations between events and objects and/or their attributes indicate how events are inferred from objects and/or object attributes. Jump ball, rebound, and free throw are examples of events for the basketball domain. Concepts are general definitions that contains related events and objects in it. Each concept has a relation with its components that are used for its definition. Attack and defense are examples of concepts for the basketball domain.



Besides, nearly every domain has a number of irregular situations that cannot be represented with the relation sets defined in the ontology. VISCOM is enriched with additional rule definitions where it is hard to define situations as a natural part of ontology. The second purpose of additional rules is to define such complex situations.

Rules can contain any class/property individual defined in the ontology. In fact, VISCOM is adequate to represent any kind of event definition in terms of spatial or/and temporal relations and similarity definitions. Rules give the opportunity to make the event definitions which contain a set of events or other class individuals defined in the domain ontology.

Spatial Change class is utilized to express spatial relation changes between objects or spatial movements of objects in order to model events. Spatial regions representing objects have spatial relations between each other. These relations change in time. This information is utilized in event definitions. Temporal relations between spatial changes are also used when more than one spatial change is needed for definition. This concept is explained under Temporal Relations and Event Definition classes. Spatial changes have an interval that is designated by the spatial relation individuals used in their definitions.

Spatial relations are momentary situations but periods of spatial relations can be extracted from consecutive frames. Whenever the temporal situation between Spatial Relation Component individuals defined in a Spatial Change individual is satisfied, the Spatial Change individual is extracted and Spatial Relation Component individuals' periods are utilized to calculate the Spatial Change individual's interval. According to the meaning of the spatial change, periods of spatial relations should be included or discarded in the calculation of spatial change intervals.

Second alternative to define a spatial change is using spatial movements. Spatial movements represent spatial changes of single objects. This class is used to define movement types. It has five individuals as; moving to left, moving to right, moving up, moving down, and stationary. Spatial Movement Component class is used to declare object movement individuals. "Ball moves left" is an example of an individual of this class. Temporal relations are used to order Spatial Changes or Events in Event Definitions. Allen's temporal relationships are used to express parallelism and mutual exclusion between components.

Temporal Event Component class is used to define temporal relations between Event individuals. Temporal Spatial Change Component class is used to define temporal relations between spatial changes in Event definitions. For instance, the temporal relation after is used between Ball hits Hoop and Player jumps. Spatial Change individual in the definition of Rebound event. An event can have several definitions where each definition describes the event with a certainty degree. In other words, each event definition has a membership value for the event it defines that denotes the clarity of description. Event definitions contain individuals of Spatial Change, Spatial Relation Component or Temporal Spatial Change Component classes.

IV. AUTOMATIC SEMANTIC CONTENT EXTRACTION FRAMEWORK

The Automatic Semantic Content Extraction Framework is illustrated in Fig. The ultimate goal of ASCEF is to extract all of the semantic content existing in video instances. In order to achieve this goal, the automatic semantic content extraction framework. There are two main steps followed in the automatic semantic content extraction process.

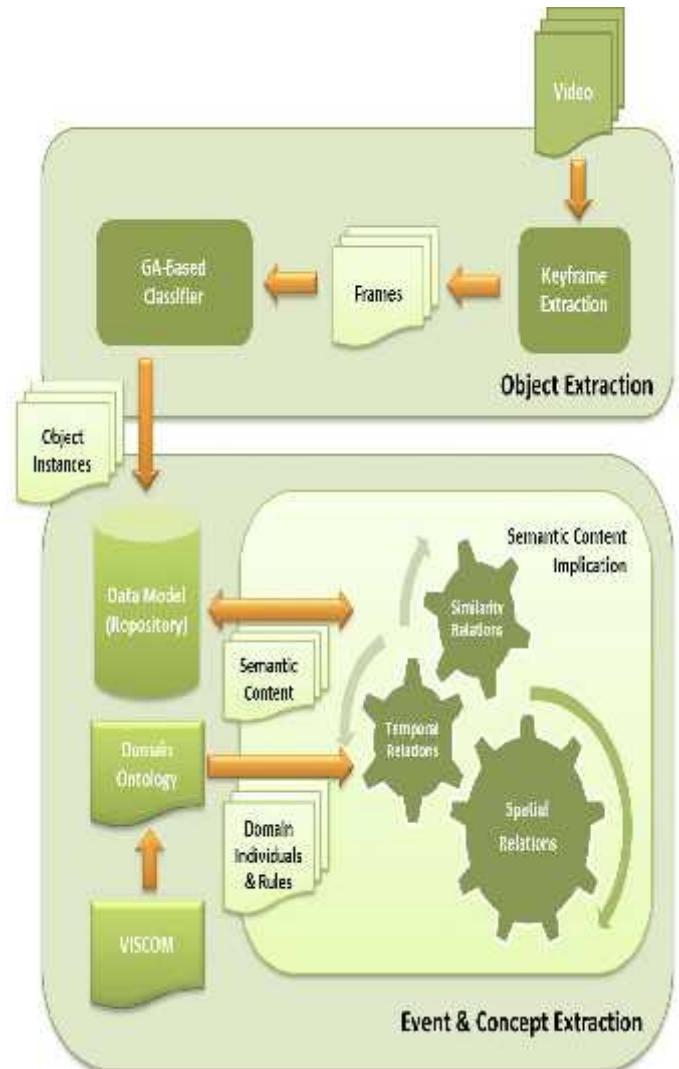


Fig. Automatic semantic content extraction framework.

The first step is to extract and classify object instances from representative frames of shots of the video instances. The second step is to extract events and concepts by using domain ontology and rule definitions. A set of procedures is executed to extract semantically meaningful components in the automatic event and concept extraction process. The first semantically meaningful components are spatial relation instances between object instances. Then, the temporal relations are extracted by using changes in spatial

relations. concepts are extracted by using the spatial and temporal relations.

a. OBJECT EXTRACTION

Object extraction is one of most crucial components in the framework, since the objects are used as the input for the extraction process. However, the details of object extraction process is not presented in detail, considering that the object extraction process is mostly in the scope of computer vision and image analysis techniques. It can be argued that having a computer vision-based object extraction component prevents the framework being domain independent. However, object extraction techniques use training data to learn object definitions, which are usually shape, color, and texture features. These definitions are mostly the same across different domains.

b. SPATIAL RELATION EXTRACTION

Object instances are represented with the MBR. There can be object instance (as regions) represented with R in a frame F. Every spatial relation extraction is stored as a Spatial Relation Component instance which contains the frame number, object instances, type of the spatial relation, and a fuzzy membership value of the relation.

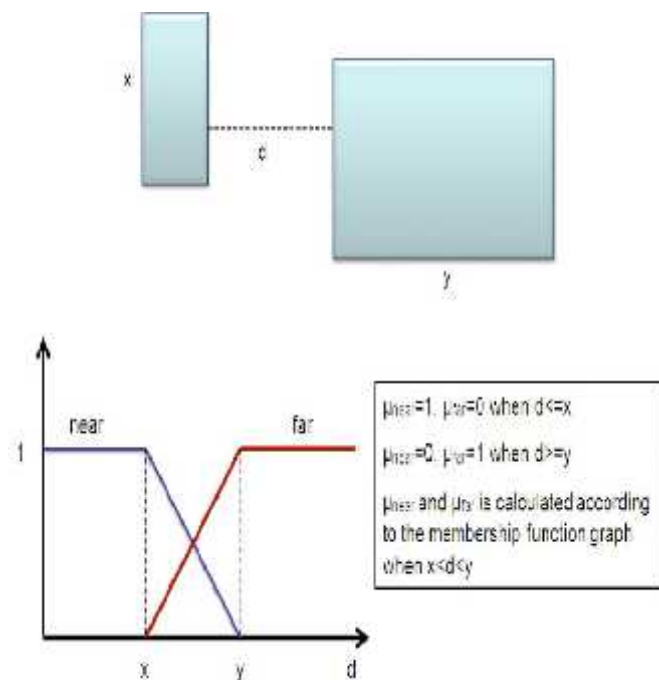


Fig. Graph for distance relation membership function.

Spatial relations are fuzzy relations and membership values for each relation type can be calculated according to the positions of objects relative to each other. Below, we explain how membership values for each of the distance, topological, and positional relation categories are calculated.

c. TEMPORAL RELATION EXTRACTION

In the framework, temporal relations are utilized in order to add temporality to sequence Spatial Change or Events individuals in the definition of Event individuals. One of the well-known formalisms proposed for temporal reasoning is Allen's temporal interval algebra [24] which describes a temporal representation that takes the notion of a temporal interval as primitive. Allen's algebra is used to express parallelism and mutual exclusion between model components of VISCOM.

d. EVENT EXTRACTION

Event instances are extracted after a sequence of automatic extraction processes. Each extraction process outputs instances of a semantic content type defined as an individual in the domain ontology. Algorithm 2 describes the whole event extraction process. In addition, relations between the extraction processes are illustrated in Fig.

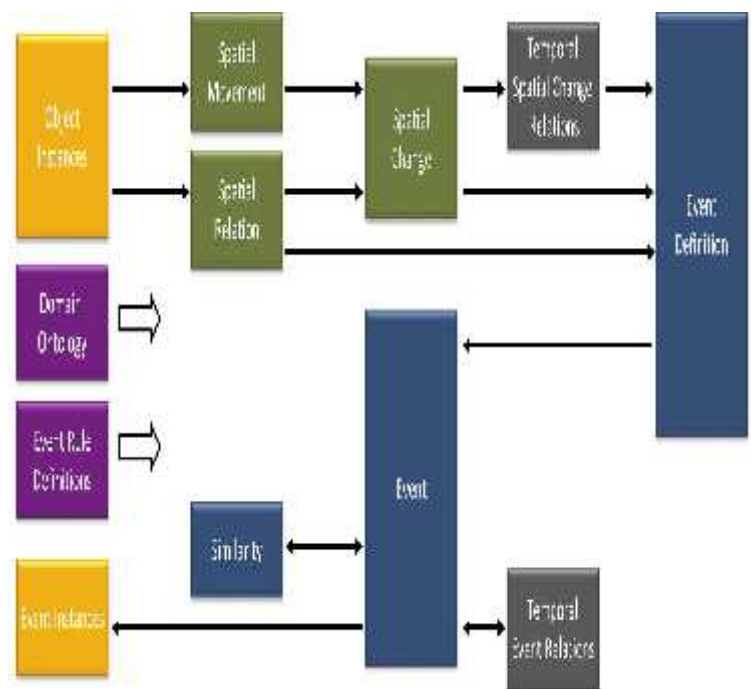


Fig. Event Extraction process

e. CONCEPT EXTRACTION

In the concept extraction process, Concept Component individuals and extracted object, event, and concept instances are used. Concept Component individuals relate objects, events, and concepts with concepts. When an object or event that is used in the definition of a concept is extracted, the related concept instance is automatically extracted with the relevance degree given in its definition. In addition, Similarity individuals are utilized in order to extract more concepts from the extracted components. The last step in the concept extraction process is executing concept rule definitions. Concept Extraction Algorithm given as Algorithm 3 simply describes the whole concept extraction process. In addition,

relations between the concept extraction processes are illustrated in Fig.

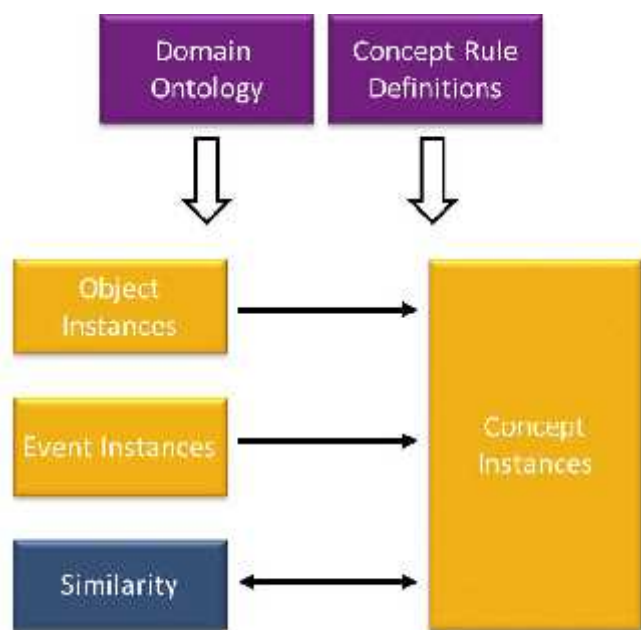


Fig. Concept Extraction Process.

V. CONCLUSION

The primary aim of this research is to develop a framework for an automatic semantic content extraction system for videos which can be utilized in various areas, such as surveillance, sport events, and news video applications. The novel idea here is to utilize domain ontologies generated with a domain-independent ontology-based semantic content metaontology model and a set of special rule definitions. Automatic semantic Content Extraction Framework contributes in several ways to semantic video modeling and semantic content extraction research areas. First of all, the semantic content extraction process is done automatically. In addition, a generic ontology-based semantic metaontology model for videos (VISCOM) is proposed. Moreover, the semantic content representation capability and extraction success are improved by adding fuzziness in class, relation, and rule definitions. An automatic Genetic Algorithm-based object extraction method is integrated to the propose system to capture semantic content. In every component of the framework, ontology-based modeling and extraction capabilities are used. The test results clearly show the success of the developed system.

VI. REFERENCES

- M. Petkovic and W. Jonker, "An Overview of Data Models and Query Languages for Content-Based Video Retrieval," *Proc. Int'l Conf. Advances in Infrastructure for E-Business, Science, and Education on the Internet*, Aug. 2000.
- M. Petkovic and W. Jonker, "Content-Based Video Retrieval by Integrating Spatio-Temporal and Stochastic Recognition of Events," *Proc. IEEE Int'l Workshop Detection and Recognition of Events in Video*, pp. 75-82, 2001.

L.S. Davis, S. Fejes, D. Harwood, Y. Yacoob, I. Haratoglu, and M.J. Black, "Visual Surveillance of Human Activity," *Proc. Third Asian Conf. Computer Vision (ACCV)*, vol. 2, pp. 267-274, 1998.

G.G. Medioni, I. Cohen, F. Bre´mond, S. Hongeng, and R. Nevatia, "Event Detection and Analysis from Video Streams," *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 23, no. 8, pp. 873-889, Aug. 2001.

S. Hongeng, R. Nevatia, and F. Bre´mond, "Video-Based Event Recognition: Activity Representation and Probabilistic Recognition Methods," *Computer Vision and Image Understanding*, vol. 96, no. 2, pp. 129-162, 2004.

A. Hakeem and M. Shah, "Multiple Agent Event Detection and Representation in Videos," *Proc. 20th Nat'l Conf. Artificial Intelligence (AAAI)*, pp. 89-94, 2005.

M.E. Do´nderler, E. Saykol, U. Arslan, O . . Ulusoy, and U. Gu'du' kbay, "Bilvideo: Design and Implementation of a Video Database Management System," *Multimedia Tools Applications*, vol. 27, no. 1, pp. 79-104, 2005.

T. Sevilmis, M. Bastan, U. Gu'du' kbay, and O . . Ulusoy, "Automatic Detection of Salient Objects and Spatial Relations in Videos for a Video Database System," *Image Vision Computing*, vol. 26, no. 10, pp. 1384-1396, 2008.

M. Ko'pru' lu', N.K. Cicekli, and A. Yazici, "Spatio-Temporal Querying in Video Databases," *Information Sciences*, vol. 160, nos. 1-4, pp. 131-152, 2004.