International Journal of Engineering & Technology, 8 (1.10) (2019) 151-154

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Review on Semantic Content Extraction in Videos

[1]Anuranji R, [2]Dr. Srimathi H

[12]*S.R.M Institute of Science & Technology, Chennai*

Abstract

In recent years the need for extracting contents from video in applications such as medical, forensics has increased. Raw data representing the elementary physical video attributes and low-level features representing audio, video, text alone are not sufficient to fulfill the user's needs resulting in a deeper understanding of the content at the semantic level. Hence digital video databases have come to be more pervasive and finding video clips quickly in video databases have become a major challenge. Manual techniques were used which were inefficient, subjective and costly in time and limit the querying capabilities. Here Video Semantic Content Model is presented which helps in the automatic extraction of objects, events and concepts which is based on domain ontology. In proposed framework, viewing angle of camera to improve extraction capabilities is been considered.

Keywords: *ontology, fuzziness, VISCOM*

1. Introduction

Digital video plays a vital role in education, entertainment and various other multimedia applications. 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. The main aim is to enable users to retrieve objects, events and concepts more efficiently. The three basic levels of video contents are raw video data, low-level features and semantic content.

Raw video data consist of elementary physical video units together with some general video attributes such as format, length, and frame rate.

Low-level features are characterized by audio, text, and visual features such as texture, color distribution, shape, motion, etc.

Semantic content contains high-level concepts such as objects and events.

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.

Video content based searching is very different from the normal searching concept and much more difficult than the query mechanism from a database. In Manual extraction, several rules are been applied by the user which represent semantic information. Manual techniques are not suitable for large amount of video data. Manual extraction approaches are tedious, subjective and time consuming thereby limiting querying capabilities. Hence it is necessary to fix the semantic gap between the low level feature and the high level feature [3]. For this purpose, multimedia retrieval system with ontology concept is attempted.

2. Related Work

A video is a temporal sequence of pixel regions at the physical level; it is very difficult to explore its semantic content. In order to solve this problem, several domain-dependent research approaches were introduced and these approaches take an advantage of using domain knowledge to facilitate extraction of high-level concepts directly from features.

In [2], new video data model is proposed which uses rule-based approach and supports spatial-temporal formalization of high-level concepts with a stochastic approach. This model helps to extract semantics automatically from raw video data by integrating two different approaches to map low-level features to high-level concepts.

In [3], a video semantic content analysis framework is presented which is based on ontology. In domain ontology, high level semantic concepts and their relations are defined. Integration of low-level features and video content analysis algorithms are made into the ontology. For ontology description, OWL is used. Description Logic is used to describe how features and algorithms for video analysis should be applied according to different perception content and low-level features. Semantic events are described by Temporal Description Logic.

In [4], extraction techniques that separate objects from background of video and their limitations are discussed.

3. Video Semantic Content Meta Model

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. Ontology provides many advantages to content modeling. There are many studies for ontology based video content modeling but they are for specific domain only. So, it is necessary to have wide domain applicable video content model. VSCMM is used as meta-ontology for constructing domain ontology. It is also used for rule construction process. Rule construction process becomes easier by using VSCMM. VSCMM is presented with Class logic formulation:
VSCMM:

$$\left\{\begin{array}{l}\left[\begin{array}{c} class \rightarrow \{Ck\}, \\ TemporalRelation \rightarrow \left\{\begin{array}{c} start, finishes, before, \\ after, meets \end{array}\right\}, \\ SpatialRelation \rightarrow \left\{\begin{array}{c} far, near, above, below, \\ left, right \end{array}\right\}, \\ Movement \rightarrow \{up, down\} \end{array}\right] \\ where\,(Ck, VSCMM\ Class) \end{array}\right. \quad (1)$$

Most of the event definitions for a wide variety of domain are covered by rules that are constructed by VSCMM. However there can be complex situations that ontology definitions cannot cover. The model VSCMM provides rule based modeling capability without using ontology. Objects, events, concepts, spatial and temporal relations are components of VSCMM.
For the semantic content extraction, VSCMM ontology provides fuzzy classes for every component and they have fuzzy definitions. Object instances have membership value as an attribute which represent relevance of the given Object Bound Rectangle (OBR) to the object type.

Ontology Based Modeling
VSCMM is developed on an ontology based structure. Semantic content types and relations between these types are collected under VSCMM Classes, VSCMM Data Properties associates classes with constants and VSCMM Object properties are used to define relations between classes.
The VSCMM classes are introduced with their formal representation, description and important relation descriptions.

a) Component: All the semantic content comes under the class of component. A component can have similarity relation with the other components. Component class has three subcategories: Objects, Events and Concepts [5].

b) Object
Objects are nothing but existential entity. An object has low level features, names and composed-of relation. Ball, referee, player are the examples of objects for the football domain.

Object:

$$\left\{\begin{array}{l}\left[\begin{array}{l} Name \rightarrow [string] \\ LowLevelFeature \rightarrow \{Lm\}, \\ Composed\ Of \rightarrow \{COMRn\}, \end{array}\right] \\ where\,(Lm, LowLevelFeature) \end{array}\right. \quad (2)$$

c) Event
Events are temporal objects and changes in object relation. They are given by objects and spatial/temporal relation between objects. Sidekick, goal, pass the ball are the examples of events in football domain.

Event:

$$\left\{\begin{array}{l}\left[\begin{array}{l} Name \rightarrow [string], EventDef \rightarrow \{EDk\} \\ ObjectRole \rightarrow \{ORm\}, \\ TemporalEventComp \rightarrow \{TECn\} \end{array}\right] \\ where\,(EDi, EventDefinition) \\ where\,(TEC, TemporalEventComponent) \end{array}\right. \quad (3)$$

d) Concept
Concepts contain related events and objects in it. Components which are used for concept definition has relation with that concept.

Concept:

$$\left\{\begin{array}{l}\left[\begin{array}{c} Name \rightarrow [string], \\ ConceptComponent \rightarrow \{CCk\} \end{array}\right] \\ where\,(CCk, ConceptComponent) \end{array}\right. \quad (4)$$

e) Spatial Relation
Spatial relation is a relative position between objects. Relative position is as above, near, inside. The spatial relation types are subcategorized as topological, positional and distance spatial relation.

f) Temporal Relation
Temporal Relations are used to sequence Spatial Changes or Events in Event Definitions.

g) Similarity Class
Similarity Class is used to give the relevance of a component to another component in fuzzy manner. Whenever a component is having similarity relation with another is extracted, semantically related components are automatically extracted [6].

Rule Based Modeling
Content modeling capabilities can be extended by using additional rules. There are two parts as body and head for each rule. Each body part has any number of domain classes and head part has only one individual with value μ, where μ represents certainty of the definition given in body part to represent the definition in head part and $0 \leq \mu \leq 1$. Rules have basic syntax having parenthesis and logical connectives in body and head part. Using rule definitions one can lower the computation cost for spatial relation. Rule definitions help to strengthen the framework for both semantic content representations and semantic content extraction.

4. Automatic Semantic Content Extraction Framework

The Automatic Semantic Content Extraction Framework is presented in Fig. 1

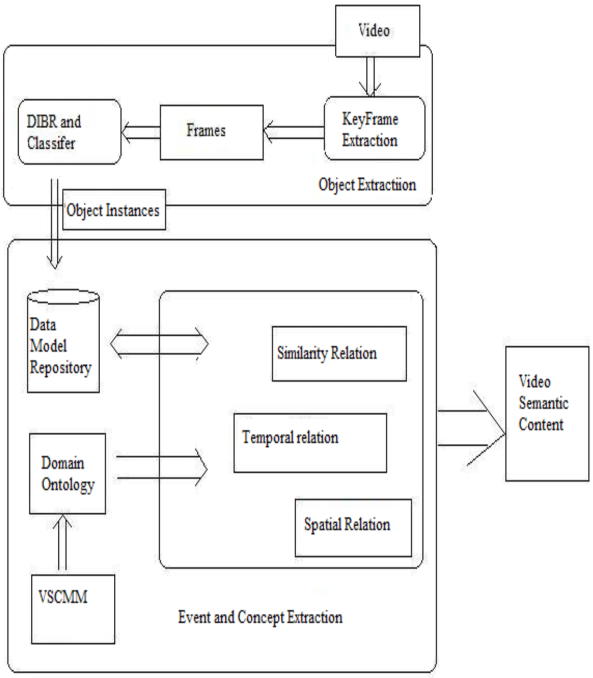

Fig. 1: Automatic Semantic Content Extraction Framework

To extract semantic content from video instance, this framework takes Vi, Ri and ONTi as input, where Vi is video instance, ONTi is domain ontology to which Vi belongs and Ri is set of rules for domain ontology. The output of the extraction process is semantic content i.e. events and concepts.

a) Object Extraction
In this framework, Object extraction is the most important component because if objects are detected correctly then events and concept are extracted correctly. The extracted objects are stored in database and later they are used as input for the event and concept extraction. Firstly, key frames are extracted from video and from each key frame object are extracted. For this purpose, Depth Image Based Rendering is used. DIBR considers views of images so that object is detected more correctly. These extracted objects are classified using Invariant Image classification. These extracted objects are stored with membership function, frame number and Object Bound Rectangle data.

Object Bound Rectangle:

$$\begin{bmatrix} x \rightarrow [float], y \rightarrow [float], \\ width \rightarrow [float], \\ length \rightarrow [float] \end{bmatrix} \quad (5)$$

b) Spatial Relation Extraction
Object Bound Rectangle represents object instances. It is possible to have n objects instances as region represented with R in a frame F i.e. F= {R0, R1, R2, ……. , Rn}. The area inside Ri is represented with RiI and edges of Ri are represented with RiE. Spatial relations are fuzzy relations, so for each relation type membership value is calculated. Membership value is calculated according to relative position of objects with each other.

c) Temporal Relation Extraction
Spatial change or events individuals in the definition of Event individuals should be sequenced. So, temporal relations are used to add temporality to spatial change or events individuals.

d) Event and Concept Extraction
All instances of a semantic content type are defined in as individuals in domain ontology are extracted. After extracting all the instances sequentially, event instances are extracted. In concept extraction process, extracted object, event and concept instances and Concept Component individuals are used. When an object and event used in concept definition is extracted, related concept instance is automatically extracted with relevance degree given in definition.

Algorithm : Concept Extraction algorithm

Input : Domain Ontology, Object Instances and Event Instances.
Output : Concept Instances.
1: for all Class Component individuals defined in the ontology do
2: Based on the individual def. check whether there are Objects or Events instances.
3: end for
4: for all Similarity individuals in the ontology do
5: Based on the individual def. extract Concept instances which satisfy it.
6: end for
7: execute all rules which are defined for Concept individuals.

5. Conclusion

The aim of this research is to develop a flexible framework for an automatic semantic content extraction system for video based applications. Semantic content extraction is done automatically using Automatic Semantic Content Extraction Framework. Also, generic Video Semantic Content Meta-Model is developed. Depth Image Based Rendering is used to consider views of objects and it is integrated to proposed system. In order to classify the objects, Invariant Image Classification will be used. It takes less time to extract semantic content from video. As further study, one can improve the extraction capabilities of framework for spatial relation by considering motions in depth.

References

[1] Y. Yildirim, A. Yazici, T. Yilmaz, "Automatic Semantic Content Extraction in Videos Using a Fuzzy Ontology and Rule-Based Model", IEEE Transactions on Knowledge & Data Engineering, vol.25, no. 1, pp. 47-61, Jan. 2013, doi:10.1109/TKDE.2011.189
[2] 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.
[3] L. Bai, S.Y. Lao, G. Jones, and A.F. Smeaton, "Video Semantic Content Analysis Based on Ontology," IMVIP '07: Proc. 11th Int'l Machine Vision and Image Processing Conf., pp. 117-124, 2007.
[4] Andrea Cavallarao and Touradj Ebrahimi, "Object Based video: Extraction tools, evaluation metrics and application" .
[5] 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 a Internet, Aug. 2000.
[6] 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.
[7] Hakeem and M. Shah, "Multiple Agent Event Detection and Representation in Videos", Proc. 20th Nat'l Conf. Artificial Intelligence (AAAI), pp. 89-94, 2005.
[8] 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.
[9] 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.
[10] 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.
[11] 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.
[12] J. Fan, W. Aref, A. Elmagarmid, M. Hacid, M. Marzouk, and X. Zhu, "Multiview: Multilevel Video Content Representation and Retrieval," J. Electronic Imaging, vol. 10, no. 4, pp. 895-908, 2001.