International Journal of Modern Trends in Engineering and Research
www.ijmter.com
e-ISSN No.:2349-9745, Date: 2-4 July, 2015

VIDEO SEMANTIC CONTENT EXTRACTION USING k-MEANS and FUZZY ONTOLOGY

Priyanka Nikam[1], Mr. B. R. Nandwalkar[2]

[1]*Computer Department, Late G.N. Sapkal COE, Anjaneri, Nashik, priyankanikam28@gmail.com*
[2]*Computer Department, Late G.N. Sapkal COE, Anjaneri, Nashik , nandwalkar.bhushan@gmail.com*

Abstract— Increased use of video based applications revealed need of efficiently extracting the video contents. The video contains various features which can be classified into normal data, relative features and logic content. For core content of video semantic level understanding is required. So to get video content automatic semantic content framework is proposed. In proposed system a semantic content extraction system which allows the user to query and regain objects, events, and concepts that are extracted automatically. VISCOM is a video semantic content model which contains classes and relations between classes. Objects and events are represented by some VISCOM classes and other classes are used in the automatic semantic content extraction framework. VISCOM classes collect the semantic content types and relations.

Ontology based fuzzy video data semantic model which uses spatial and temporal relations in event and concept definition is proposed. Extracted objects from consecutive representative frames are processed to extract temporal relations. Additional rules to lower spatial relation computation cost and to define some difficult situations more successfully are used. To extract objects from video we apply k-means clustering algorithm. By, which we get the more relevant objects related to user query.

Keywords-Fuzziness, Ontology, Semantic Content Extraction, Spatial Relations, Video Content Modeling.

I. INTRODUCTION

As a large amount of video data is available it is necessary to model them so, that it enables users to retrieve some desired contents from video in semantically meaningful manner. Video content modeling and extraction is required in the applications surveillance, crime investigation, sports, border monitoring etc. It helps to give desired content from video.

Video content has 3 levels which are normal data, relative features and logic content. Normal data is elementary physical units with attributes frame rate, format, length. While audio, text, and visual features like texture, color distribution, shape, motion, etc. are relative features. And high level concepts such as objects, events are logic content. First two levels are inefficient to extract semantic content as the extraction is based on only low level features. They hardly provide semantics to the user in which he is interested. So, it is necessary to develop intelligent methods for semantic content extraction in videos.

VISCOM (Video Semantic Content Model) is proposed for domain independent ontology construction. It is meta-ontology. For object extraction k-means clustering algorithm is used. It gives more appropriate objects in video. As, clustering is used only query related and identical objects are get extracted. Events and concepts are extracted using spatial and temporal relations between objects. Also, VISCOM which is ontology is used in the extraction process. Rest of the paper is organized as II gives literature survey, III describes proposed system, IV explains mathematical model of system, and V conclude the paper and gives future scope.

RELATED WORK

We categorize video content extraction evolution process in manual techniques, spatiao-temporal relations, ontology based video content extraction.

In content-based video retrieval modeling and querying capabilities are provided [2]. Video modeling makes distinction between the structure and content of video. On the basis of either feature-based or semantic-based modeling can be done. System is unable to integrate both the feature-based and semantic modeling. A framework for mapping from features to high-level concepts is used in layered video data model [3]. Flexibility is provided for using different video processing techniques. For mapping low-level features to high-level concepts a new video data model which supports the integrated use of two different approaches. Model is unable automatically extract high level semantics from video data.

The VSAM proposed detection and tracking module to extract moving objects trajectory from video streams [4]. The event detection and analysis is based on the detection of moving objects and by estimating their speed and trajectory and inference of their behavior. The limitation of this model is stability as it is induced manually by observer.

To recognize the behavior of moving objects system is implemented [5]. An activity is to be considered composed of action threads and in which each thread is executed by a single actor. Multi-agent event is represented by an event graph composed of several action threads related by logical and temporal constraints [6]. The complexity depends on number of moving objects and complexity of scene. Video event graph used to learn the event structure from training videos [6]. Graph composed of temporally correlated sub-events, which is used to automatically encode the event correlation graph. Problem with graph formulation because sub-events are depend on their predecessor.

BilVideo is the video database management system provides full support for queries on spatio-temporal, low level features and semantic on the video data [7][8]. It is domain independent and to handle the spatio-temporal queries a knowledge-base is used that consists of a fact-base and comprehensive set of rules which reduce the number of facts. To reduce the work for manual selection and labelling of objects significantly by detecting and tracking the salient objects. Extended Advanced video information system (AVIS) is data model that allows efficient and effective presentation of spatial-temporal properties of objects which focused on the semantic content of video streams [9]. It is object based video data model and application independent. It supports fuzzy spatial queries including querying spatial relationships between objects.

Multilvel video database model provides a reasonable approach to bridging the gap between low-level representation features and high level semantic concepts from a human point of view [14][15]. A hierarchical semantics, sensitive video classifier is proposed to shorten the semantic gap between the low-level visual features and high level semantic concepts. Semantic content analysis framework based on domain ontology which is used to define high level semantic concepts and their relations in the context of the examined domain [10]. To enrich analysis low-level features and video content analysis algorithms are integrated into the ontology. Description logic is used to describe how video processing methods and low-level features should be applied according to different semantic content. Linguistic and dynamic visual ontology presented in [11]. The structure with proposed ontology can be used to perform higher level annotation of the video clips to generate complex queries. Logical structure of domain ontology defined in terms of linguistic concepts. Event description framework (EDF) is to capture event semantics that enables storage, inference and retrieval of events from lower level event observations [12]. EDF identify a set of classes for semantic annotation of multimedia data and describe their properties and relationships between events and entities. It is ontology for

spatial-temporal relationships. But, event extraction is manually done in this framework.
Video Event Recognition Language describes event ontology [13]. It is formal language for representing events for describing ontology for application domain and for annotating data with those ontology categories. To address the problem of designing ontology for visual activity recognition [16] proposes a system. On general ontology design principles and adapt them to the specific domain of human activity ontology. Qualitative evaluation principles and provide several examples from existing ontology and how they can be improved upon are discussed. Genetic Algorithm based object extraction and classification mechanism is proposed for extracting the content of the videos [17]. The object extraction is defined as a classification problem and a Genetic Algorithm based classifier is proposed for classification.
But, manual techniques are both cost and time they are subjective also. The extraction process is manual. In some approaches automatic or semiautomatic methods with spatio-temporal relations are used. But it fails to give appropriate solution. In ontology based systems ontology is constructed for semantic content representation. Different languages are designed for representation purpose.

II. PROPOSED SYSTEM

The integration of both ontology-based and spatio-temporal approaches for semantic content extraction in videos is needed. So, the system using ontology based on fuzzy logic and spatial temporal relations is proposed. As shown in fig.1 video is given as input for extracting semantic content from it. Then key-frames are extracted from video which gives the representation of video. Key-frames are processed for segmentation and feature extraction. By applying k-means clustering algorithm on it objects are extracted. Using spatial relations and domain ontology events are extracted. Concepts are extracted by temporal relations between events and domain ontology. Domain ontology is constructed by using VISCOM and object instances.

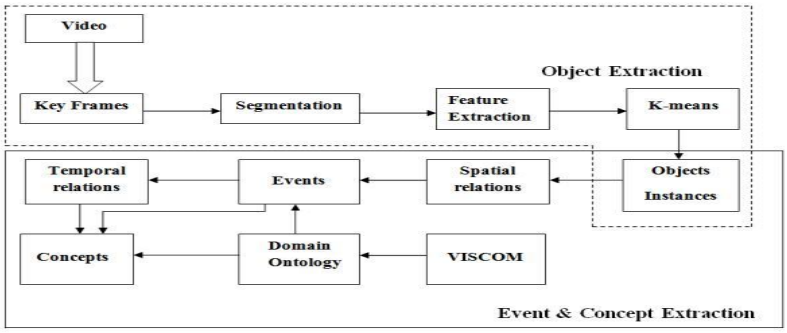

Figure1. System Architecture

3.1. Video Semantic Content Model (VISCOM) and Domain Ontology

It is domain independent meta-ontology used for domain ontology construction [1]. It is solution to rule-based and domain dependent extraction methods. VISCOM provides standardized rule construction ability with the help of its meta-ontology. Rule construction process gets easier and makes its use possible on larger video data. VISCOM ontology can cover most of the event definitions for a wide variety of domains. VISCOM is composed of VISCOM classes (VC) and domain-independent VISCOM class individuals (DII). $VC_x = (VC_{xname}, VC_{xprop})$ where, VC_{xname} is the name of the class and VC_{xprop} is the set of relations and properties of class. Domain-independent VISCOM class individuals are grouped under movement, temporal, structural, and spatial relation types. DII = MRI U TRI U OCTI U SRI. 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.
Using VISCOM domain ontology is constructed [1]. Basically, domain specific semantic contents are defined as individuals of VISCOM classes and properties. Each event definition uses different spatial and temporal relations between objects in order to define the event. The ontology developer always has a chance to add a new definition that will cover cases where existing definitions are not sufficient enough. Also has an opportunity to add new individual definitions, modify, or delete them at any time.

3.2. Object Instances Extraction

The first task in video semantic content extraction is to efficiently extract the objects in video. Object extraction process is shown in fig.1. Video is given as input key-frames extraction in video gives images. Then segmentation is done. Features are then extracted. In the system k-means clustering algorithm is used for object extraction which gives more specific objects in video. The basic approach is first to train a k-means clustering representation, using the input training data. K-Means Algorithm:

I. Compute the Euclidean distance.
II. Order the labeled examples by increasing distance.
III. Find a heuristically optimal number k of nearest neighbors, based on RMSE. This is done using cross validation.
IV. Calculate an inverse distance weighted average with the k-nearest multivariate neighbors.

3.3. Event and Concept Extraction

Events are extracted using spatial and temporal relations between object instances. An extracted event instance is represented with a type, a frame set representing the events interval, a membership value and the roles of the objects taking part in the event. Frame Set is used to represent the frame interval of instances. Each extraction process outputs instances of a semantic content type defined as an individual in the domain ontology. 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.

V. MATHEMATICAL MODEL

In order to represent proposed system in mathematical form, we used set theory along with the functions and relationship, because it's easy to use and understand the system thoroughly. System problem definition can be framed in the form of a set theory. Let S be a automatic semantic content extraction system.
S = {I, F, O, C} where,
I represent set of inputs;
I = {OI, VISCOM} OI = Object Instances VISCOM = Video Semantic Content Model
F is the set of functions;
F = {F1, F2, F3, F4} F1= Object Extraction, F2 = Domain Ontology Construction, F3 = Event Extraction, F4 = Concept Extraction
C is the set of constraints applied to functions;
C = {C1, C2}
C1 = Domain Individuals and rules, C2 = Spatio-temporal relations

O is the set of outputs;
O = {E, C, O1, O2} E = Event C = Concept

CONCLUSION

Video semantic contents are extracted using fuzzy ontology and rule based model. VISCOM gives meta-ontology which is domain independent. Spatial and temporal relations are applied on extracted objects to get events and concepts related to user query. K-means clustering algorithm gives more relevant and exact object instances. To accomplish goal efficiently and effectively extract semantic content in video. A generic ontology-based semantic meta-ontology model for videos (VISCOM) is proposed. By adding fuzziness in class, relation, and rule definitions the semantic content representation capability and extraction success is improved. An automatic k-means based object extraction method is integrated to the proposed system to extract objects from video. System contributes in several ways to the semantic video modelling and semantic content extraction. Adding fuzziness to domain ontology construction more meaningful and relevant results for semantic search are obtain. Wrong, extra, or missing definitions in the constructed ontology can decrease the extraction success.

REFERANCES

[1] Y. Yildirim, Adnan Yazici, Turgay Yilmaz "Automatic Semantic Content Extraction in Video Using a Fuzzy Ontology and Rule-Based Model", *in IEEE Trans. Knowledge and Data Engineering*, vol. 25, no. 1, pp. 47-61, Jan. 2013.
[2] M. Petkovic, 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 on the Internet*", Aug. 2000.
[3] M. Petkovic "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.
[4] 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.
[5] S. Hongeng, R. Nevatia, "Video-Based Event Recognition: Activity Representation and Probabilistic Recognition Methods,"*Computer Vision and Image Understanding*, vol. 96, no. 2, pp. 129-162, 2004.
[6] 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.
[7] M.E. Do¨nderler, E. Saykol, U. Arslan, OUlusoy, 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.
[8] T.Sevilmis, M.Bastan, U.Gudukbay, 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.
[9] 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.
[10] 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.
\[11] A.D. Bagdanov, M. Bertini, A. Del Bimbo, C. Torniai, and G. Serra, "Semantic Annotation and Retrieval of Video Events Using Multimedia Ontologies," *Proc. IEEE Int'l Conf. Semantic Computing (ICSC)*, Sept. 2007.
[12] R. Nevatia and P. Natarajan, "EDF: A Framework for Semantic Annotation of Video," Proc. *10th IEEE Int'l Conf. Computer Vision Workshops (ICCVW '05)*, p. 1876, 2005.
[13] R. Nevatia, J. Hobbs, and B. Bolles, "An Ontology for Video Event Representation," *Proc. Conf. Computer Vision and Pattern Recognition Workshop*, p.119.
[14] J. Fan, W. Aref, A. Elmagarmid, M. Hacid, M. Marzouk, and X. Zhu, "Multiview: Multilevel Video Content Representation and Retrieval," *J. Electronic* Feb. 2004. *Imaging*, vol. 10, no. 4, pp. 895-908, 2001
[15] J. Fan, A.K. Elmagarmid, X. Zhu, W.G. Aref, and L. Wu, "Classview: Hierarchical Video Shot Classification, Indexing, and Accessing," *IEEE Trans. Multimedia*, vol. 6, no. 1, pp. 70-86.
[16] Y. Yildirim and A. Yazici, "Ontology-Supported Video Modeling and Retrieval," *Proc. Fourth Int'l Conf. Adaptive Multimedia Retrieval: User, Context, and Feedback (AMR)*, pp. 28-41, 2006.
[17] T. Yilmaz, "Object Extraction from Images/Videos Using a Genetic Algorithm Based Approach," master's thesis, Computer Eng. Dept., METU, Turkey, 2008.

