Automatic Semantic Rule Based Coordination for Content Extraction in Videos Using Fuzzy Ontology

Shubham Prasad, Dr. Dinesh Kumar Sahu
1, 2 SRK University, Bhopal M.P., India[1]

Abstract : *Recent advances in digital video analysis and retrieval 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. Raw data 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 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.*
Keywords: *Semantic content extraction, video semantic content model, fuzziness, ontology.*

I. Introduction

There is increase in use of the developing intelligent methods to model and extract the video content. Typical applications in which modeling and extracting video content are crucial include surveillance, video-on-demand systems, intrusion detection, and border monitoring, sport Events, criminal investigation systems, and many others. The ultimate goal is to enable users to retrieve some desired Contents from massive amounts of video data in an efficient and semantically meaningful manner. There are 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 very 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 some low level content of a video, but they hardly provide semantics which is much more useful for users. Users are interested in querying and retrieving the video in terms of what the video contains.

II. Literature Survey

In this reference I understand that, 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 [1]. Therefore due to this many 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. There are many research works in this area. Most of them use manual semantic content extraction methods .Manual extraction approaches are tedious, subjective, and time consuming [2]; here the researcher studies that perform automatic or semiautomatic extraction do not provide a satisfying solution. Although there are several studies employing different methodologies such as object detection and tracking ,multimodality and spatiotemporal derivatives, the most of these studies propose techniques for specific event type extraction or work for specific cases and assumptions In. [3], For the simple periodic events are recognized researcher find the success of event extraction is highly dependent on robustness of tracking. The event recognition methods described in [4] and for to understand this researcher has find the heuristic method that could not handle Multiple-actor events. Event definitions are made through predefined object motions and their temporal behavior. The shortcoming of this study is its dependence on motion detection. In [5], And for this scenario events are modeled from shape and trajectory features using a hierarchical activity representation extended from Hakeem and Shah [6]

In this paper the researcher proposed new method to detect events in terms of a temporally related

chain of directly measurable and highly correlated low level actions by using only temporal relations. Another key issue in semantic content extraction is the representation of the semantic content. Many researchers have studied this from different aspects. A simple representation could relate the events with their low-level features (shape, color , etc.) 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 ontology's 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. Studies such as Bil Video [7],[8],
For the multitier of the object this paper is used [9]

III. System Model

The model represents automatic semantic content extraction process by using following mode
1. Frame Extraction
2. Object Extraction
3. GA Based Classifier
4. Event and Concept Extraction

1) Frame Extraction

It is important module of the dissertation in which we will provide video as an input and then generate the number of frames of that video according to the user need and then we will apply this module and extract the required frames, and then these frame will be used in next module i.e. in GA based classifier to separate each object instances from these key frames.

2) 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 are 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. Here, we proposed that training data generation will be simulated in the proposed system itself.

Algorithm 1. Ontology Construction with VISCOM [1]
Require: VISCOM
Ensure: Domain Ontology
1: define O, E and C individuals.
2: define all possible SR's occurring within an E.
3: define all possible OM's occurring within an E.
4: use SR's and M's to define SC's.
5: describe temporal relations between SC's as TSCC's.
6: make EDs with SC's, SR's and TSCC's.
7: for all E's do
8: if an event can be defined with an event def then
9: define E in terms of ED's.
10: end if
11: if an event can be defined with temporal relations Between other events then
12: define E's in terms of ETR's.
13: end if
14: end for
15: for all C's do
16: construct a relation with the C that can be placed in its meaning.
17: end for
18: define S's. [1]

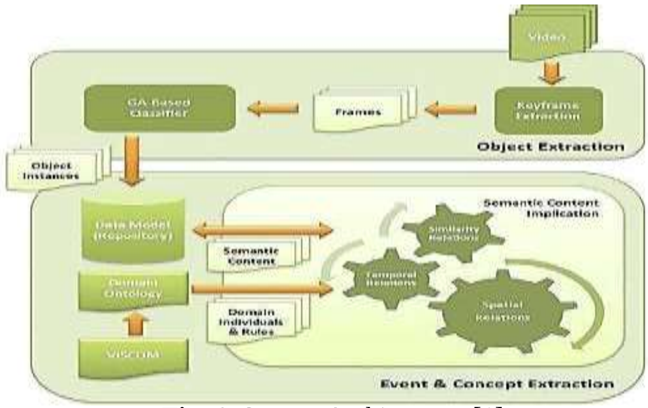

Fig: 1. System Architecture. [1]

3) GA Based Classifier.[1]

As stated in module 2 details, whatever key frames extracted that key frames will be given as input to GA based classifier. GA based classifier then applies the Generic algorithm to get instances of objects which are presented in the input key frame for this module. These object instances are vital to predict semantic contents of events and concepts present in the provided key frame.

4) Event and Concept Extraction

It will take object instance as input which is generated by the GA based classifier and forms semantics. To get the semantics of each such object instance, it will apply the VISCOM algorithm as proposed in this dissertation and will find spatial and temporal relation of that object and generate the semantic content which we expect for that particular object.

IV. Proposed Methodology

General Algorithmic steps used for Frame Extraction Module

1. Input video file for which semantic contents to be identified.
2. For Input Video File, extract attributes consisting of General Attributes (File Name, File Format, Duration, Bitrate) with Audio (Format, Bitrate, Channels, Sampling) and Video attributes (Format, Bitrate, Frame rate and Frame size) .
3. Based on duration value (consider 1 Frame of 1 second), extract desired number of frames.
4. Save the Frames for further computation.
5. Stop

V. Snapshots

Figure 2 shows the initial window developed to input the video file as data contents for the proposed system. As stated earlier, with the help of provided provision as shown in Figure 1, user can upload desired file contents by browsing the desired storage location on the terminal.

Once, the video file is loaded, it can be used for further computation and hence, initially, associated attributes are extracted of each loaded video file, where, general, audio-related and video-related attributes are extracted as shown in Figure. Figure 3 shows information about files successfully sent to neighbor i.e. from system to system.

Fig: 2 Uploading Video

Fig: 3 No. of Frame

VI. Results And Analysis

The first objective of the proposed system is to retrieve particular object from extracted frames of video file. So as main component of whole system, in frame extraction module, where we provided video as an input and then generated the number of frames of that video according to the video length specified by duration attribute divided into seconds and for each second, frames are extracted from input video file and then these frame will be used in next module i.e. in GA based classifier to separate each object instances from these key frames.

Following Fig. 4 shows the 30 frames obtained on input video files like wildlife.wmv of duration 30 seconds and 04 frames obtained with capture_2012121.jpg file of 04 seconds video file. Similarly, the frame extraction for 10 video files is given in Table 1.

Fig 4. Images of Frames

Table :1. Extracted Frames Details

Sr. No.	Video File	Duration (In. Seconds.)	Number of Extracted Frmaes
1	Wlidlife.wmv	30	30
2	Capture_201221	04	04
3	Capture_201222	05	05
4	Capture_201223	03	03
5	Capture_201224	10	10
6	Capture_201225	15	15
7	Capture_201226	18	18
8	Capture_201227	23	23
9	Capture_201228	22	22
10	Capture_201229	20	20

VII. Conclusion

The primary aim of this dissertation 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 ontology generated with a domain-independent ontology-based semantic content met ontology model and a set of special rule definitions. Automatic Semantic Content Extraction Framework con- attributes in several ways to semantic video modeling and semantic content extraction research areas. In this I have finding the number of frame from the given video and store the imagesof that video.

VIII. Future Scope

In our proposed design we not only retrieve the semantic content but also removed the gap betweenlow level features and high level logic content. In future, one can check whether these features can be extracted from the crashed video and also can predict the viewing angle and depth dimension of the camera.

References

[1]. M. Petkovic and W. Jonker, “An Overview of Data Models and Query Languages for Content-Based Video Retrieval”, In Intl Conf. Advances in Infrastructure for E-Business, Science, and Internet, Aug. 2021.

[2]. M. Petkovic and W. Jonker, “Content-Based Video Retrieval by Integrating Spatio-Temporal and Stochastic Recognition of Events”, In Proc. IEEE Intl Workshop Detection and Recognition of Events in Video, pp. 75-82, 2020.

[3]. L.S. Davis, S. Fejes, D. Harwood, Y. Yacoob, I. Haratoglu, and M.J. Black, “Visual Activity”, In Proc. Third Asian Conf. Computer Vision (ACCV), vol. 2, pp. 267-274, 2020.

[4]. G.G. Medioni, I. Cohen, F. Bremond, 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. 2020.

[5]. S. Hongeng, R. Nevatia, and F. Bremond, Video-Based Event Recognition: Activity Representation and Probabilistic Recognition Methods, Computer Vision and Image Understanding, vol. 96,no. 2, pp. 129-162, 2015.

[6]. A. Hakeem and M. Shah, Multiple Agent Event Detection and Representation in Videos, Proc. 20th Natl Conf. Artificial Intelligence (AAAI), pp. 89-94, 2018.

[7]. M.E. Donderler, E. Saykol, U. Arslan, O . Ulusoy, and U.Gudu kbay, Bilvideo: Design and Implementation of a Video Database Management System, Multimedia Tools Applications, vol. 27, no. 1, pp. 79-104, 2019.

[8]. T. Sevilmis, M. Bastan, U. Gudu 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, 2018.

[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, 2021.

[10]. I. Horrocks, P.F. Patel-Schneider, H. Boley, S. Tabet, B. Grosof, and M. Dean, “Swrl: A Semantic Web Rule Language,” technical report, W3C, http://www.w3.org/Submission/SWRL/, 2019.

[11] C. Xu, J. Wang, K. Wan, Y. Li, and L. Duan, “Live Sports Event Detection Based on Broadcast Video and Web-Casting Text,” MULTIMEDIA ’06: Proc. 14th Ann. ACM Int’l Conf. Multimedia,pp. 221-230, 2020.

[12] P.-W. Huang and C.-H. Lee, “Image Database Design Based on 9D-SPA Representation for Spatial Relations,” IEEE Trans. Knowledge and Data Eng., vol. 16, no. 12, pp. 1486-1496, Dec. 2019.

[13] I. Horrocks, P.F. Patel-Schneider, H. Boley, S. Tabet, B. Grosof, and M. Dean, “Swrl: A Semantic Web Rule Language,” technical report, W3C, http://www.w3.org/Submission/SWRL/, 2018