

# Video Events Extraction based on Mixed Feature Modeling

**B. S. Daga**  
Fr. Conceicao Rodrigues  
College of Engineering  
Bandra, Mumbai

**A. A. Ghatol**  
Former VC, Dr Babasaheb  
Ambedkar Technological  
University, Lonere

**Anuprita Daga**  
CISO, Reliance, Santacruz,  
Mumbai

## ABSTRACT

Now a day's increase in access of video-based application has opened need for extracting the content in videos. Unprocessed data and low-level features alone are not sufficient to complete the user's need so deeper understanding of the content at the semantic level is required. Currently, manual techniques which are inefficient, subjective and expensive in time and limit the querying capabilities are used to fulfill the gap between low-level representative features and high-level semantic content. The system that allows the user to query and retrieve associated objects, events, and concepts automatically is proposed. The events can also be representative objects, actions, their impressions, and so on. Here an ontology-based video semantic content model which uses spatial/temporal relations in event and concept definitions is leveraged. Simple & efficient process consideration on main object detection & its common associated mixed direct measurable feature like shape, texture & derived features like co-occurrence & topology is evaluated. An ontology definition provides a wide-domain applicable rule construction standard. In addition to domain ontology's, additional rule definitions to lower spatial relation computation cost are used. This leads to describe some complicated events close to human thinking. The proposed system has been implemented and tested on domains like road accident & sports for precision and recall measures.

## General Terms

Video Object Detection, Video Event Extraction. Semantic Content

## Keywords

Video Content Modeling, Multicue, Spatio-temporal, Event Detection, Event Extraction

## 1. INTRODUCTION

Video is one of the content in Multimedia. Multimedia is playing a crucial role in the current majority of the applications. In all these applications maximum user oriented information can be served. Semantic content-based video retrieval requires many changes in a multimedia database management system which is core component of multimedia applications. Adopting prevailing notion of building multimedia applications [7] is effective in building large scale applications. These changes take effect mainly in modelling and querying techniques as stated in [1]. The internet user's time span is reducing with growing content accessibility and information abundance. Thus, narrowing right down to the matter of video explorations, in future one understands that merely providing the connected search topic for video won't be satisfactory. The developers and video content suppliers must perceive and respect user's instinct interest to remain

alive with the trend of data science as stated in [31]. For instance, let's take into account a case of a user seeking for a cricket match instance of the player striking the boundary. Obviously the user is not curious about going through the whole match that is lasting for nearly a day which cost time and irrelevant information to person. He/She would not have an interest in manually browsing the video timeline to see actual boundary event. Here, a cricket boundary event seeker may be a target user. So for any user information like, e.g Ramp walk accident event[9], Olympic inaugural torch lighting event, actioning sequence of the object, a road accident event, cricket batsman's clean bold wicket. System must give result with associated event extraction in more natural way like human being. The extended analysis dimension directions author is building-up here supporting the precise event extraction from corresponding video data.

Additional, precise video event extraction may be achieved if it is facilitated on content identification intelligence. For simplicity video is conceptualised to be made up of 3 levels of data. These are raw video information, low level options and user's contents. The raw video information would represent the attributes like frame rate, video length, format, etc. The secondary low level options would be of audio, text, visual like colours & their distributions, motion etc. The last and also the highest level of data content would be the semantic contents of events recorded in the video. These levels would be highly appreciated by the humans. Many researcher suggested that event based classification of the video frame sequence is required to get the extraction of the actual event. Content search within the video isn't that simple task. Several researchers had resulted multiple solutions. They have addressed issue depending the video under consideration. It may be strictly audio based or audio assisted frame sequence extraction or visual object motion based extraction. For instance, associate audio based video extraction would merely grip out "Clean Bold" sound within video and audio to track out the corresponding frames in neighbourhood of matching timestamps from video frame sequence.

Unfortunately, many of such techniques won't yield successful results. As it's not necessary that the audio information would be always present in the video. Audio sometimes crackly and distorted in video recording. Besides as literature shows, many of noteworthy researchers had also worked on ontology based semantics information extraction. Where they primarily look for key frames extraction based on object identification and then workout different topological & Spatio-temporal relations alignment to categorize event. Moreover, the problem becomes trickier as a video is a frame sequence without having the direct relation to its semantic contents.

One can easily understand that the object extraction is one of most crucial components in the framework, since the objects are used as the input for the event extraction process. Object detection [8,6] from a frame or a video sequence has attracted attention of many engineers working in the field. In current time object detection technology is well recognized. Most of latest cell phone cameras can detect the face, smiles, mountains & many other specialized objects. However, when it comes to not only to detect the object but to relate intrinsically to some naturally occurring relations, is the area where improvements are going on. The precise object identification comes through, improving the object detector [8] with multicue optimizer. Such a multicue optimizer would look for additional information from the video frames associated with the detected object [9]. The objects shape and size are the very basic cues for incorporation in better understanding the situation. Moreover, such additional information may include the detected object position in the frame. It can also be its association with the human appearance in the frame. Not all the time, in all frames, the object can be totally visible. The object detector has to work through occlusions at times. In such situations, it's not enough to look for additional cues from the same video frame. Thereby, an improvised system shall also look at the path traced by the object in the multiple frames in sequence and its association with the reference subject. The reference subject could be the human, vehicle or anything like that relating to the object. So Spatio-temporal characterization of the data is required.

Present paper is targeted to use mixed features for detected object and its Spatio-temporal characterization to develop video model which is supporting event extraction from an unconstrained video sequence. As the authors are still working out in building the final automatic video event extraction system, test initial results and system performance are encouraging to boost further system development. Design and implementation of a system to overcome such challenges is the need of today.

The organization of the paper is made simple as intension is to emphasize the gap in the system design phase. First of all existing system framework is highlighted followed by proposed framework. It is implemented in modular form. In module description part features explanation is given. This is followed by technique & algorithm for video analysis. At the end intermediate results for event retrieval is covered.

## 2. EXISTING SYSTEM

Semantic content extraction process will extract the correct content. It also enables the user to search the desired data from the videos [12]. Normally video is perceived to contain three types of data such as raw video data, low level features and high level concepts. An elementary physical video unit characteristic like format, length and frame rate comes under unprocessed raw video data. While visual content like texture, distributed colors, shapes, and motion are characterized as low level features. Logical content contains high level concepts like objects and events. The lowest level is not useful for making sense out of content. It can't provide core meaningful content from the videos. Hence it probably not sufficient to address the user needs.

Semantic content extraction system with subdued capability in term of retrieval is being developed in number of variants. 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 which limit querying capabilities. The video is given as input to this System. Typically the key frames are generated. Further from these key frames findings in low level future like size, length, format and in other derived feature is done to define content.

The generic architecture form of semantic video content extraction system is shown in Fig1.

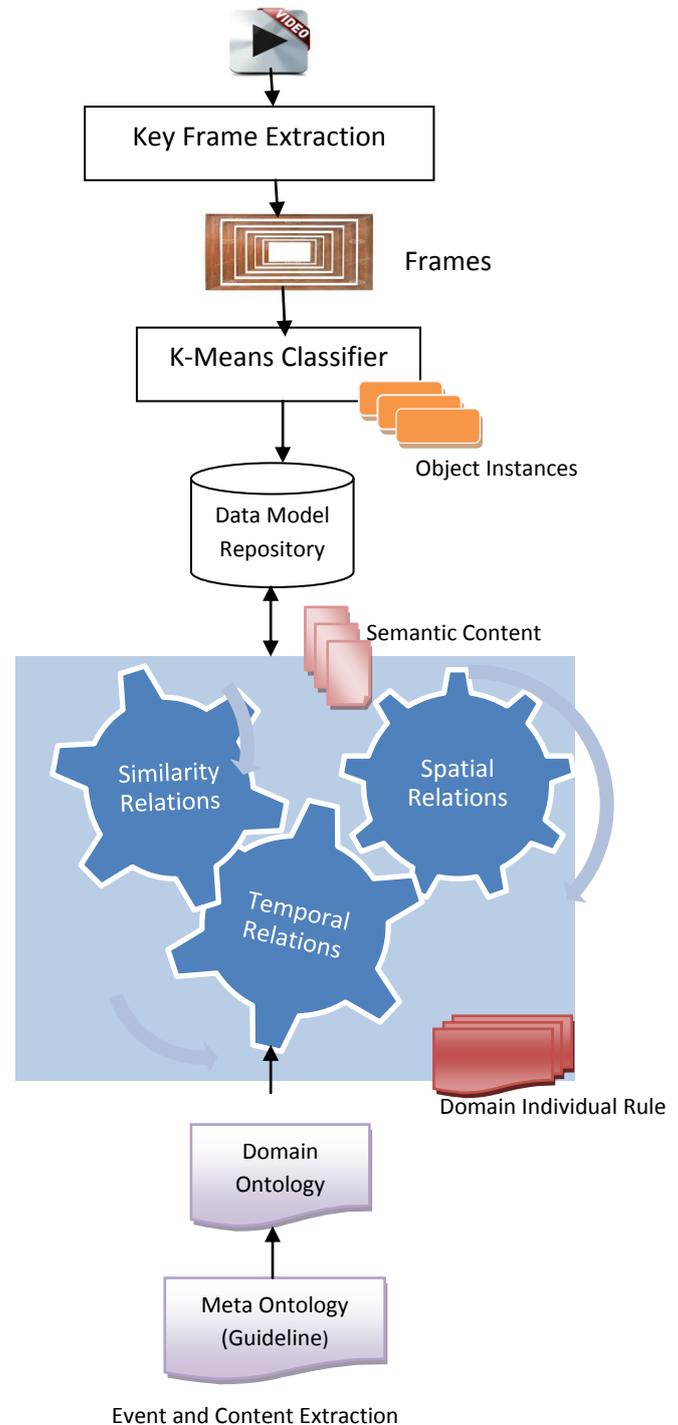


Fig 1: Generic Video Semantic Content Extraction System

## 2.1 Disadvantages

1. It does not produce the output in graphical manner and so performance level cannot be predicted.
2. Semantic content are extracted only in partial manner as rule based methods are used.
3. The raw video information and low-level options alone are not sufficient to meet the user's need. A deeper understanding of the data is needed in several video-based applications which is not done in modeling aspect of the video.
4. It is difficult to extract content directly from raw video as output information. This is often as a result of video is a temporal sequence of frames without relevance to its content.

## 3. PROPOSED SYSTEM

We proposed variant semantic video content extraction system which contributes to video event modeling and thus video content analysis area. This is achievable upto certain degree by specifically focusing on segmentation[19,20,21] & classification aspect of video retrieval process [5] This is often accomplished through the event finding based on ontology and content extraction algorithms [13,14,15].It is variants of linguistics content extraction system that permits the user to question and retrieve objects, events, and concepts that are measured & extracted Automatically[5]

The block diagram of video event extraction system based on prominent object identifier and Spatio-temporal inter objects relation based event semantics is illustrated in Fig. 2 .Initially, we have used a meta-ontology, a rule construction norms that is domain specific to construct domain Ontology. Besides, general meta models it offer solutions for multimedia's structure representations [6,22]. During this study, we have proposed a wide-domain applicable video content so as to model the content in videos. Video content model could be a well-defined meta-ontology for constructing domain ontology's [12,16]. It's alternative to the rule-based and domain-dependent extraction ways. Constructing rules for extraction could be a tedious task and isn't easily extendible. Additionally, to the complexness of handling such distinction, every rule structure will have weaknesses. Besides, video content model provides rule construction ability with the help of its meta-ontology. It eases the rule construction method and makes its use on larger video knowledge. Based on domain different kinds of event in terms of rule can be defined by using event grammar. For the case of Soccer Game simple event representation may be of the form :

has a ball -> ({player, ball}, {}, {}, {distance(player, ball)< threshold}, {during (this, threshold)})

ball is free -> ({player, ball}, {}, {}, {distance(evry(player), ball) > threshold}, {during (this, 10sec)}, {})

Where has a ball, ball is free are some of sample rules for events. Here spatial & temporal relation association is taken into account to define meaningful event. These rules if combined will give complex event definition.[23]

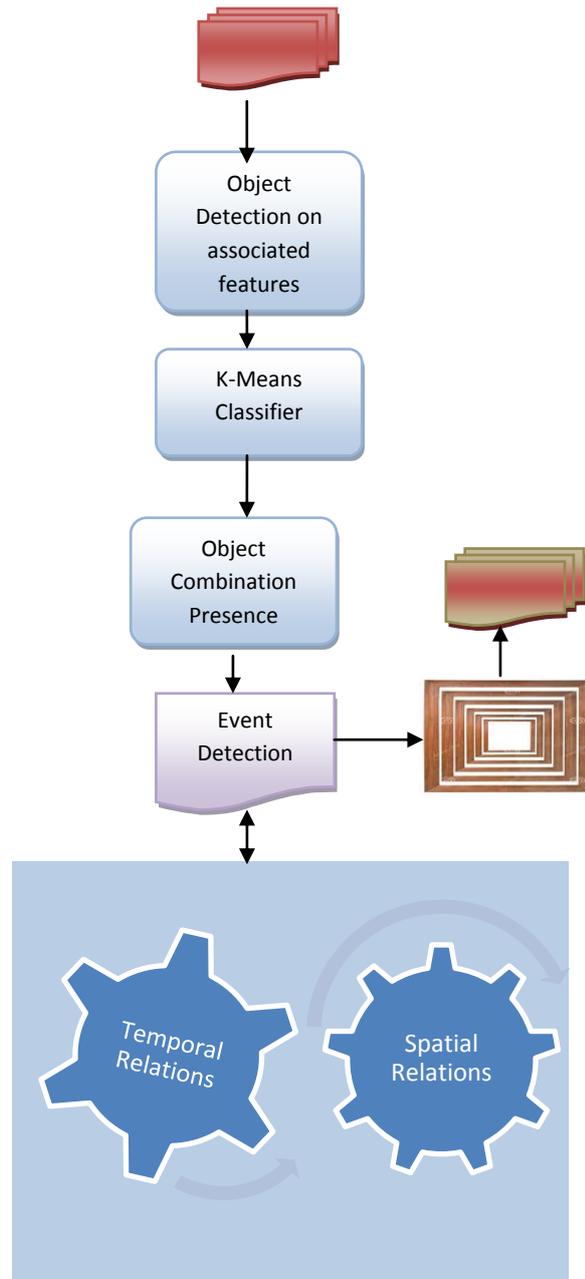


Fig 2: Proposed Video Semantic Content Extraction System

### 3.1 Advantages

1. Ontology provides several adoptions and capabilities for content modeling.[17] But the majority of videos content modeling studies propose domain specific ontology models limiting its use to a specific domain.
2. The ontology based video models and content extraction methods are capable to address uncertainty [26]by use of fuzzy logic.
3. It separates the domain knowledge from the operational knowledge which eases the task of customizing a product from its components according to a required specification.

## 3.2 Modules

- 3.2.1 Input Video
- 3.2.2 Feature Extraction
- 3.2.3 Domain Ontology
- 3.2.4 Relation Extraction
- 3.2.5 Spatial Relation Finder
- 3.2.6 Semantic Video Content Model

## 3.3 Module Description

### Input Video

The user browses the video from the database and it will be given as input to the System. High level user orientation of video in terms of its understanding is shown in Fig 3. By using the ontology and semantic algorithm the content will be extracted. One can understand that user's video content model requires supports of semantic objects and events. Hence, Video model is developed coarsely on meta-ontology based framework which provides guidelines to construct domain ontology. These results are effective to domain specific semantic contents generation which are described as individuals of Video model's classes and properties.

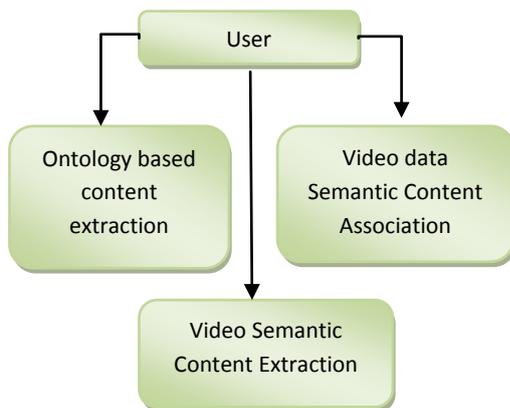


Fig 3. User's Video Content Model

### 3.2.2 Feature Extraction

Generic feature identification & their use in modeling the event is important task done in this module. It is based on the detected event moments in the video, a temporal range containing the event moment and event boundary range is detected [2]. This is due to the following considerations:

- 1) Dealing with live video only which contain close event moment
- 2) Event structure follows certain temporal patterns due to the production rules.

Thus, we first extract some generic features from the video and use these features to model the event structure. The feature extraction is done in real time. The ultimate goal is to extract the semantic content and generate the extracted features automatically. This is used in given key frame extraction algorithm. The relation establishment between the objects and events is done by use of ontology methods [11,27]. Generic ontology defines the semantics, i.e., meaning of the key entities and represents the knowledge about events in a machine friendly way. It is also used to facilitate the solutions for multimedia representations. The extracts like format, length, frame rate in the process are stored in the

database. At the end spatial and temporal relations defined in database are used to extract semantic information.

### 3.2.3 Domain Ontology

The main parts of system are classes and relations. Few classes indicate logic content types such as instances and Event [12][15]. The remaining classes can be used in semantic content extraction process. Relations defined in system gives ability to model the events and concepts related with other objects and events. Concept rule definitions are used to achieve the concept instances. Object instances and event instances are grouped under concept instances based on similarity evaluation. As mentioned[12] for the demonstration of our system an application in sports[30] & road accident 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 modeling and detecting the events in cricket video. The features associated with each sequence and object comprises their definitions in terms of low-level features. The category of sequences and objects and the selection of features are based on domain knowledge. Cricket sport domain ontology is constructed and the definitions are used as described in this section.

### 3.2.4 Relation Extraction

To add temporality to sequence of spatial changes the definition of event individual's, temporal relations are used. The well-known formalisms planned for temporal reasoning is Allen's temporal interval pure mathematics that captures true temporality. Thus temporality is added to sequence & spatial changes or events individuals in the definition of event individuals. This is referred from Allen's interval algebra [24]. in which he proposed thirteen basic relations between time intervals that are distinct, exhaustive, and qualitative.

- distinct because no pair of definite intervals can be related by more than one of the relationships
- exhaustive because any pair of definite intervals are described by one of the relations
- qualitative (rather than quantitative) because no numeric time spans are considered

These relations and the operations on them form Allen's interval algebra. It is used to express parallelism and mutual exclusion between model components of Video model.

### 3.2.5 Spatial Relation

Conceptual Semantics is a formal approach to natural language meaning development. The relation instances having these relations are extracted by using the rule definitions. The rule definitions are used to extract the spatial relations [9]. First, the computation time for spatial relation is calculated when no rule definition is created. The rules are defined one by one and then computation times are computed after adding each rule definition to the system. Object instances are represented with the MBR.. 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. 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.

### 3.2.6 Semantic Video Content Model

Video model is an ontology-based structure. Semantic content types and relations between these types are collected under Video model Classes. The classes and constants are associated by Data model Properties. Model's Object Properties define relations between classes. The proposed ontology-based method is good for both event and concept extraction. Event modeling is the process of using user logic to aggregate over event-level data to produce 'modeled' data that is simpler for querying. Events are central elements in the representation of data from a variety of domains. Examples of these domains are history, cultural heritage, geography and multimedia. Event-centered modeling captures the time and place aspects of a domain. In addition, events provide a way to describe complicated relations between people, places, actions and objects.

At first object instances should be obtained correctly. If object instance is misclassified which is reflected in the object instance set, then success of event and concept extraction process decreases. Its use is taken for defining Meta ontology which is used for constructing the domain ontology.

As the video-event complexity grows with growing system demand and the data sets because of more and more generalizations of requirement, a class of researchers also opting for employing fuzzy systems for defining the complex semantic based ontological system, which is to be used for accurate and more general event extraction systems.

## 4. KEY FRAMES AND FEATURE EXTRACTION

The process is well studied as object extraction from images since videos are a set of images (key frames)[10] [11]. Before starting with the image segmentation and object extraction from images, key frames are obtained by using a key frame extraction algorithm.

**Algorithm:** Key Frame Extraction

**Input:** Video

**Output:** Key Frames

1. Select video file
2. Analyze video file (Extracting information about number of frames & their linking.)
3. For all available frames from specified video file
4. Form shortlisted frame list

After key frames extraction extracted frames are proceeds for segmentation and feature extraction. For that purpose, segmentation algorithm is proposed. Firstly, each segment obtained is assumed to be a candidate object. Also all combinations of nearby segments are taken as candidate objects. All candidate objects are tried to be classified by the k-means neighboring based algorithm [4][25]. Using ranking filters real objects are obtained. Segmentation, the features of the segments are extracted as training data. These features are MPEG descriptors, such as color, texture and shape descriptors. Segmentation and feature extraction algorithm given below.

**Algorithm:** Segmentation and Feature Extraction

**Input:** Key Frames

**Output:** Segments and its features

1. Analyze key frame
2. Extract low level information and segmentation information
3. Loop all available segmentation from the specified key frame
4. Extract first segmentation
5. Analyze segmentation
6. Extract all feature information about specific segmentation
7. Go to step 3
8. Analyze feature information specified by key frame segmentation (apply k-means)
9. Store to database
10. Go to next key frame

## 5. OBJECT EXTRACTION WITH K-MEANS ALGORITHM

K-means Algorithms methodology is a powerful search technique for solving problems in many different research areas[25] [28]. K-means are mostly used for improving the performance. It is alternative to genetic algorithm.[18] Instead of testing all combinations, k-means ensure the most fitting-ones to be obtained in fewer tries. K-means is tried & used in different levels/phases of object/image retrieval[4]. K-means are used for many different purposes like object extraction, selecting representative key frames, generating coordinated objects, feature formation, pixel selection, etc.

**Algorithm:** Object Extraction

**Input:** Segmented and feature extracted key frames

**Output:** Object

1. Extracting current frames
2. Analyze feature information available on all segmentation
3. Compare segmentation into key frames (pixel wise, color wise, shape/area)
4. Check for key frame annotation
5. Create segmentation for annotation
6. Specify annotation type
7. Compare new segmentation to current segmentation
8. Store to database

### K-Means Clustering

Let  $X = \{x_1, x_2, x_3, \dots, x_n\}$  be the set of data points and  $V = \{v_1, v_2, \dots, v_c\}$  be the set of Centre's.

1. Randomly select 'c' cluster Centre's.
2. Calculate the distance between each data point and cluster Centre.

3. Assign the data point to the cluster Centre whose distance from the cluster Centre is minimum of all the cluster centers.
4. Recalculate the new cluster Centre using
5. Recalculate the distance between each data point and new obtained cluster centers.
6. If no data point was reallocated then stop, otherwise repeat from step 3.

Where,  $\|xi - vj\|$  ' is the Euclidean distance between  $xi$  and  $vj$ .  
' $ci$ ' is the number of data points in  $ith$  cluster.  
' $c$ ' is the number of cluster centers.

## 6. TEST RESULTS

The video event extraction system built has been tested on mixed scenarios. It is implemented in Matlab & Portege[29] as ontology development environment.[19,20] The first case is where target is to extract the exact event footage from the cricket match video. There are many cricket match recorded videos available on internet which is being used for study purpose. Whereas, in second case the target is to extract the event related to accident. Fig. 4, 5, 6 &7 demonstrate the events extraction in the two mentioned domains with the help from actual frames as captured during the program run.



Fig. 4 Bold Event Extraction Demonstration



Fig. 5 Boundary Event Extraction Demonstration

In cricket domain case, the event of 'Bold' has the semantics definition based on the spatio-temporal conditions where, the video frames are scanned in sequence to detect the simultaneous appearance of the objects from pre-defined object class, 'ball', 'Bat' and the 'Stump'. Wherein, the temporal paths of bat and the ball are being tracked by object identifier automatically, the data model engine keeps a check on their relative distancing. So that, as the three objects, 'stump', 'ball' and the 'Bat' appear at the same time where the Spatio-temporal paths of object 'Ball' and the object 'Bat' meet or cross for a threshold frames number, indicating that their relative distance going to zero and the two object boxes overlapping on each other. Fig. 4 demonstrates this with the help from actual extracted frames with object identifiers. This is when the event 'bold' is identified. The system thereby then extracts out the fixed number of frames upstream of the identified event frame and downstream of it to compile the extraction footage from the video. Similarly fig. 5 shows event boundary related frames extraction



Fig. 6 Auto Rickshaw Accident Event Extraction Demonstration



Fig. 7 Car Accident Event Extraction Demonstration

In other case, the similar semantic model for the event describing the auto-rickshaw accident is defined by the spatio-temporal relation between the objects detected, 'auto-rickshaw' and 'car'. Thereby the object identifier hunts for object classes of model auto-rickshaw and car. When the accident happens, the auto-rickshaw came and collides to car in action slowly, frame by frame, so that the auto-rickshaw comes in close circle to car. Fig. 6 demonstrates this with the help from actual extracted frames with object identifiers. Thus the accidental auto-rickshaw accident event gets identified and extracted. Similarly fig 7 shows car accident event demonstration in result.

The precision and recall are the two measures to quantify the success rate of system. These are the commonly reported parameters for such systems as observed from literature. Precision means the fraction of true positive objects detections to overall object detection. Whereas recall means, fraction of true positive object detection to actual expected true positive object detection.

$$Precision = \frac{True\ positive\ detections}{Overall\ detections} \quad (4)$$

$$Recall = \frac{True\ positive\ detections}{Actual\ expected\ True\ positive\ detections} \quad (5)$$

It might be early to calculate the precision and recall measures for the testing results alone. However, as the testing was tried out over four events namely 'bold', 'boundary', 'car accident', 'auto-rickshaw accident' with 4 sample videos in each. The precision-recall statistic has been manually calculated by pre-identifying the individual event counts in each test videos in the absence of any post-processing automation. The performance outcome has been tabulated in Table 1 and Table 2.

Table 1. System Performance – Event Bold and Boundary

	Event - Bold		Event – Boundary	
	Precision	Recall	Precision	Recall
Video 1	0.9247	0.8625	0.9369	0.8671
Video 2	0.9669	0.9265	0.9217	0.8312
Video 3	0.9044	0.8315	0.9932	0.9249
Video 4	0.8842	0.8113	0.9637	0.8947

	<i>Event - Bold</i>		<i>Event – Boundary</i>	
	<i>Precision</i>	<i>Recall</i>	<i>Precision</i>	<i>Recall</i>
<b>Average</b>	<b>92.00%</b>	<b>85.79%</b>	<b>95.38%</b>	<b>87.94%</b>

**Table 2. System Performance – Car and Auto rickshaw Accident**

	<i>Event – Car Accident</i>		<i>Event – Auto rickshaw Accident</i>	
	<i>Precision</i>	<i>Recall</i>	<i>Precision</i>	<i>Recall</i>
<b>Video 1</b>	0.9045	0.8422	0.9167	0.8369
<b>Video 2</b>	0.9467	0.9063	0.8915	0.8014
<b>Video 3</b>	0.8842	0.8113	0.9637	0.8947
<b>Video 4</b>	0.9045	0.8422	0.9167	0.8369
<b>Average</b>	<b>91.18 %</b>	<b>85.33 %</b>	<b>92.40 %</b>	<b>84.43</b>

## 7. CONCLUSION AND FUTURE WORK

This system is used in various fields such as inspections, sport actions and informative video applications. Our goal is to utilize a Generic-ontology model and rule based method. The main advantage of the system is to extract the content automatically and predict the performance level of the extraction process. The newness is to utilize domain ontology generated with a domain-independent content and a collection of special rule definitions. Content Extraction System contributes in many ways to video modeling and content extraction analysis areas. Initial of all, the content extraction method is finished automatically. Additionally, generic ontology model for videos is used. Moreover, the content exploration capability and extraction success is improved considering fuzzy relation, and rule definitions. K-means based object extraction methodology is integrated to the planned system to capture more relevant content.

The results support the hypothesized relationships based on mixed features which can further improved by pioneering advanced techniques in image processing, pattern recognition, and classification algorithms. Video event understanding is the highest level task in computer vision among pixel-based object-based and logic-based abstraction.

## 8. REFERENCES

[1] 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.

[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.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.

[4] Omar Kettani, Benaissa Tadili, Faycal Ramdani “A Deterministic K-means Algorithm based on Nearest

Neighbor Search” International Journal of Computer Applications (0975 – 8887) Volume 63– No.15, February 2013

[5] B. S. Daga, and A. A. Ghatol, “Detection of Objects and Activities in videos using Spatial Relations and Ontology based approach in Video Database System” International Journal of Advances in Engineering and Technology, Volume 9, Issue 6, pp. 640-650, December 2016, ISSN NUMBER: 22311963

[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] B.V.Patel, B.S.Daga, B.B.Meshram, “Building Multimedia Applications” International Journal on Computer Engineering & Information Technology, Vol. 14, no.19, pp. 10-15, 2010 ISSN NUMBER: 0974-2034

[8] T. Sevilmis, M. Bastan, U. Guido bay, 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.

[9] B. S. Daga, and A. A. Ghatol, “Multicue Optimized Object Detection for Automatic Video Event Extraction”, International Journal of Science and Technology, Volume 9( 47), December 2016, ISSN NUMBER: 0974-6846

[10] Huangzhou, Zhejiang, China International Conference on “Image Analysis and Signal Processing - Video shot segmentation and key frame extraction based on SIFT feature” IEEE 2012 International Conference on Image Analysis and Signal Processing (IASP).

[11] “Feature extraction for human action classification using adaptive key frame interval,” IEEE 2014 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA) - Chiang Mai, Thailand.

[12] 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.

[13] 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.

[14] 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.

[15] R. Nevatia, J. Hobbs, and B. Bolles, “An Ontology for Video Event Representation,” Proc. Conf. Computer Vision and Pattern Recognition Workshop, p. 119, [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?renewal=1384914](http://ieeexplore.ieee.org/xpls/abs_all.jsp?renewal=1384914), 2004.

[16] U. Akdemir, P.K. Turaga, and R. Chellappa, “An Ontology Based Approach for Activity Recognition from Video,” Proc. ACM Int’l Conf. Multimedia, A. El-Saddik, S. Vuong, C. Griwodz, A.D. Bimbo, K.S. Candan, and A. Jaimes, eds., pp. 709-712,

- <http://dblp.unitrier.de/db/conf/mm/mm2008.html#AkdemirTC08>, 2008.
- [17] M. Abinaya, R. Kiruthiga, A. Kiruthika “Automatic Mining in Ontology Based Fuzzy Video Semantic Content Model” *International Journal of Communication and Computer Technologies* Volume 02 – No.15 Issue: 02 March 2014 ISSN NUMBER: 2278-9723
- [18] T. Yilmaz, “Object Extraction from Images/Videos Using a Genetic Algorithm Based Approach,” master’s thesis, Computer Eng. Dept., METU, Turkey, 2008.
- [19] 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.
- [20] Y. Yildirim, T. Yilmaz, and A. Yazici, “Ontology-Supported Object and Event Extraction with a Genetic Algorithms Approach for Object Classification,” *Proc. Sixth ACM Int’l Conf. Image and Video Retrieval (CIVR ’07)*, pp. 202-209, 2007.
- [21] V. Mezaris, I. Kompatsiaris, N.V. Boulgouris, and M.G. Strintzis, “Real-Time Compressed-Domain Spatiotemporal Segmentation and Ontologies for Video Indexing and Retrieval,” *IEEE Trans. Circuits Systems Video Technology*, vol. 14, no. 5, pp. 606-621, May 2004.
- [22] Huayong Liu; Tao Li, “Key frame extraction based on improved frame blocks features and second extraction,” *IEEE 2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)2015.8.15-2015.8.17*
- [23] W. Chen and D.S. Warren, “C-logic of Complex Objects,” *PODS ’89: Proc. Eighth ACM SIGACT-SIGMOD-SIGART Symp. Principles of Database Systems*, pp. 369-378, 1989.
- [24] J.F. Allen, “Maintaining Knowledge about Temporal Intervals,” *Comm. ACM*, vol. 26, no. 11, pp. 832-843, 1983.
- [25] Nameirakpam Dhanachandra,, Khumanthem Manglem and Yambem Jina Chanu, “Image Segmentation using K-means Clustering Algorithm and Subtractive Clustering Algorithm,” *Eleventh International Multi-Conference on Information Processing-2015 (IMCIP-2015)*.
- [26] M. Vazirgiannis, “Uncertainty Handling in Spatial Relationships,” *SAC ’00: Proc. ACM Symp. Applied Computing*, pp. 494- 500, 2000.
- [27] 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. 2004
- [28] P. Pedda Sadhu Naik, T. Venu Gopal, “A Novel Approach for Color Image Segmentation Using Iterative Partitioning Mean Shift Clustering Algorithm” *IEEE ICCSP 2015 conference 978-1-4799-8081-9/15/\$31.00 © 2015 IEEE*.
- [29] “Portege’ Ontology Editor,” <http://protege.stanford.edu/>, 2012.
- [30] 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, 2006.
- [31] Xiang Zhai, “The Key Event Extraction Algorithm Based on Shot Events in Soccer Video,” *International Journal of Multimedia and Ubiquitous Engineering*, Vol. 11 No. 1, pp. 33-42, 2016. ISSN No – 1975-0080