

Semantic Content Extraction in Video Surveillance System

Pooja Prakash Sutar

P.G. Student, Department of Computer Engineering SIT, Lonavla Savitribai Phule University, Pune Maharashtra, India.

ABSTRACT:Recent increase in the video monitoring and surveillance has urged users to use video based application for extracting content and video semantics from it. The Results of such video monitoring and surveillance can be useful in various ways. Semantic analysis of videos has proven its worth in fields such as sports, national border security, suspicious activity monitoring along the LOC, etc. For fulfilling user's requirement such as an expected outcome in less time period and proper result with minimum storage space, the raw data and low level feature of videos are not sufficient. Though manual technique can prove handy in this regards, they are inefficient costly and time consuming. Increased demand for the video-based applications has revealed the need for extracting the content in video. We require automated system which moderate can surveillance and gives us information about important events happening in the premises. There are 3 layers of video data 1) Low level data (raw data) 2) Middle level data (content information) 3) High level Data (event information). The raw data contains the low level features of video like number of frames, length, pixel number, etc. these low level features are of very little or no importance to the user. The content, information consist information about object and occurrences of that object in video. The proposed system accepts videos as inputs which are divided into frames. Each frame is assigned to a particular number that gives event of frame in specific image from input video. These frames can be used for finding the position and speed of objects. Frame number plays the vital role in any video monitoring system.

KEYWORDS: Object Detection, Object recognition, Event Extraction, Concept Extraction

I. INTRODUCTION

Video surveillance devices have been used to gather information and to monitor objects and their events and activities. The closed-circuit televisions (CCTV), thermal cameras and night vision devices, are the three most widely used devices in Visual surveillance, but these traditional video surveillance systems have disadvantages such as a need of personal monitoring of CCTV or searching digital video recorders (DVR) when necessary. The Second disadvantage of such legacy monitoring systems is storage of video. These monitoring systems also come with a risk of human negligence in monitoring. All these drawbacks of the traditional video monitoring systems have led to increase in demand for intelligent video surveillance systems which can Store the video in compacted form and also give result more accurate and fast [2]. There has been a constant rise in the number of monitoring cameras deployed in urban areas. The main objective of making such installations is to increase the security level in public spaces. The effective and continuous observation of live video transmissions in the monitoring center by a small number of operators can be very difficult. The methods for aiding human work are recommended, because one man usually observes more than one camera image. Intelligent monitoring systems allow real time analysis of videos.

Most of people use manual semantic content extraction methods. Manual extraction approaches are tedious, subjective and time consuming [3] which limit querying capabilities. Besides, the studies that perform automatic or semi-automatic 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, most of these studies propose techniques for specific event type extraction or work for specific cases and assumptions. In simple periodic events are recognized where the success of event extraction is highly dependent on the robustness of tracking. The event recognition methods described in [5] are based on a heuristic method that could not handle multiple-actor events. Event definitions are made through predefined object motions and their temporal behavior.

The organization of this paper is as follows. Section 2, provides a detailed explanation of literature review. Proposed system and detailed system architecture with functional diagram is explained in section 3. Section 4, gives an overview

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2015

of the video semantic content extraction with necessary algorithms. Finally, in section 5, conclusions and future research directions are discussed.

II. LITERATURE REVIEW

Reference No	A technique used	Pros	Cons	Performance
[1]	Temporal normalcy and Spatial normalcy	Spatial anomalies are usually detected the scale of the smallest scene entities, typically people. However, a normal event on this scale may be anomalous at a larger scale, and vice versa	The detection of anomalous events in crowded scenes can be evaluated in a few data sets. These have various limitations in terms of size, the saliency of the anomalies, evaluation criteria	This data set provides both frames-level and pixel-level ground truth, and a protocol for the evaluation of anomaly detection algorithms. The proposed anomaly detector was shown effective in both this and a number of previous data sets.
[2]	Ontology based modeling ,rule based, domain ontology based	Domain ontology is enriched by including additional rule definition which gives handling fuzziness in class and relation definition.	Improve model and extraction capabilities of framework for spatial relation for viewing angle of camera and motion in depth dimension	An automatic semantic content extraction system for videos which can be utilized in a various area such as surveillance, sports events and news video application.
[5]	Optical flow combined with soft computing methods (neural network) is employed to evaluate the type of crowd behavior, and fuzzy logic aids detection of the holdups.	The System is assumed to detect such a situation, helping to indicate the exit that is currently less occupied.	The large amount of computational power needs to be engaged to conduct the contextual analysis of images acquired by monitoring cameras.	The experiments were carried out for fragments of recordings. Each fragment contained two types of behavior. In the beginning the normal behavior appeared, i.e. people were walking, and then abnormal type of behavior occurred.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2015

[6]	2-D approaches with or without explicit shape models and 3-D approaches.	Inference about “what is going on” in a scene, leading to behavior recognition or classification, is motivated by behavior summarization from pose, movement, gesture, etc.	Salient behaviors are so because they are different from the regular patterns in that context. Thus, anomalies are temporal or spatial outlier events not conforming to learned patterns.	The proceeds with grouping under: 1) Training and learning framework 2) The density of moving targets in the scene 3) contextual types of anomalies and available datasets.
-----	--	---	---	--

III. RELATED WORK

Methodology:

We are implementing the algorithm depicted in following block diagram. Taking video as an input, this algorithm breaks the input video into frames. These video frames are further subjected to Improved Frame Difference Optimization [2] technique to draw preferential conclusions from the video. Objects are detected in key frame by using background subtraction method. After this we check the similarity of these objects with the objects which are already stored in database and depending on that we recognize the objects by using Euclidean distance method [3]. The flow chart represents proposed dissertation work, which consist of three main part i.e. face detection, face extraction and face recognition.

DETAILED DISCRPTION OF FUNCTIONAL DIAGRAM

Frame Extraction –

Each video comprises of set of frames, in our dataset we have frame rate 30frames/s for videos. We are going to process each video frame in our methodology. For more throughputs we can take only 3 or 4 frame per second or apply key frame extraction methodology. Key frame are the frame in video which contain important information [5].

Background-Subtraction-

We use background subtraction method to remove unwanted stable objects from frame. This task is done by an inbuilt function of Open CV. Another option to obtain background subtraction is foreground extraction method which give result as compliment of background subtraction.

Object Detection -

Object is basic entity of any video data. We are using different features of object like shape, size, color etc. The object properties also include the frame no. and time at which it has come into picture. For object extraction we use foreground detection method. After extracting the object we compare it to the database of object and get name of object. There are many type of object which is detected using foreground extraction like humans, vehicles, animals, cycle, birds, gate etc. All such objects and their Images are stored in database and motion of these objects is processed.

Object Recognition –

As we discussed earlier we have detected multiple type of objects from video but as concerning to our college front gate camera we have to concentrate on only two types of objects.

Vehicles (four Wheelers)

Human (face detection)

This type of object is detected by using haar cascade in open cv which works in two different stages: training and detection.

Detection-

After a classifier is trained, it can be applied to a region of interest (of the same size as used during the training) in an input image. The classifier outputs a “1” if the region is likely to show the object (i.e., face/car), and “0” otherwise. To search for the object in the whole image one can move the search window across the image and check every location using the classifier.

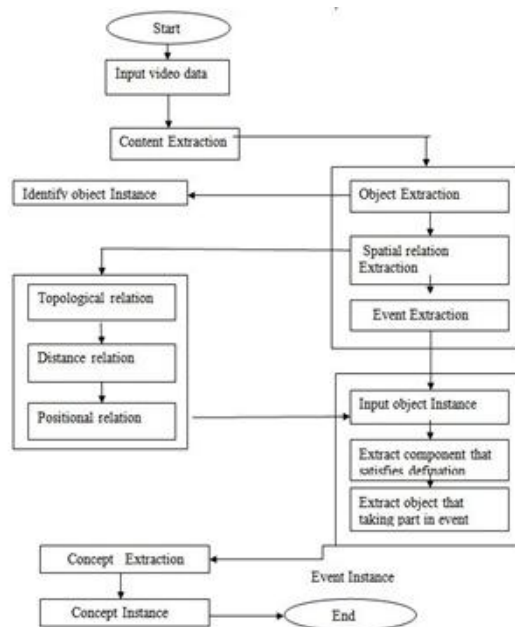


Fig: Block diagram

IV. VIDEO SEMANTIC CONTENT EXTRACTION

The linguistic part of VISCOM contains classes and relations between these classes. Some of the classes represent semantic content types such as Object and Event while others are used in the automatic semantic content extraction process. Relations defined in VISCOM give ability to model events and concepts related with other objects and events.

Semantic Content Extraction-

In the Semantic content extraction model the data is divided into object specific and scene specific. In object specific contain x, y, coordinate of image, area in pixel, speed direction and motion. The object specific gives the learning result with learning motion pattern and learning width distribution and learning path and source.

Event Extraction-

Event:

If there is change in any of above relations, we consider it as an event. 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 addition, temporal event relations can also be used in event.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2015

Definitions- An event has a name, a definition in terms of temporal event relations or spatial or temporal object relations, and role definitions of the objects taking part in the event. Jump ball, rebound, and free throw are examples of events for the basketball domain.

V. RESULTS

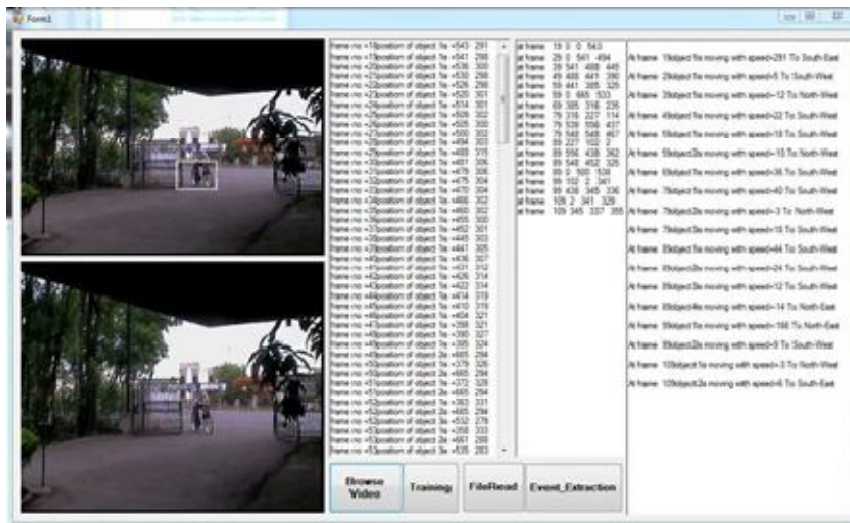


Fig i]: Result of object Detection in college campus video

Frame Processed per Second-

There are 30 frames per second as per stranded ratio which can be in our setup we process each frame. With speed 20 frames per second to achieve more throughput we can use key frame extraction methodology the result for frame 30 will be as below.

Temporal relation:-

Object 2 come into camera 1 at 12.24.23(hh.mm.ss) pm Positional relation Change in position of object 2 in video is from $x=0$ to $x=575$. Distance relation this denotes the position of object from other object.

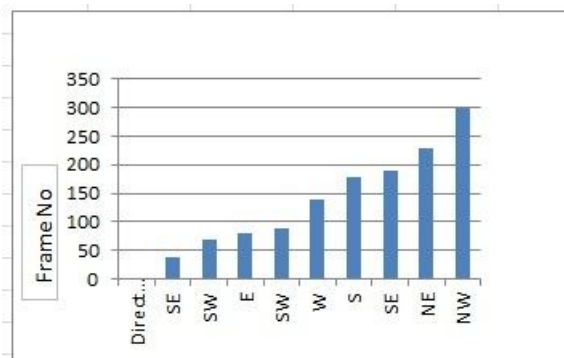


Fig ii]: Result with respect to frame no and direction

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2015

VI. 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.

REFERENCES

- [1] Weixin Li, Student Member, IEEE, Vijay Mahadevan, Member, IEEE, and Nuno Vasconcelos, Senior Member, IEEE "Anomaly Detection and Localization in Crowded Scenes" IEEE transactions on pattern analysis and machine intelligence, vol. 36, no.1, january 2014.
- [2] Yakup Yildirim, Adnan Yazici, Senior Member, IEEE, and Turgay Yilmaz, "Automatic Semantic Content Extraction in Videos Using a Fuzzy Ontology and Rule-Based Model," J. Computers, VOL. 25, NO. 1, JANUARY 2013.
- [3] Tianzhu Zhang, Member, IEEE, Si Liu, Student Member, IEEE, Changsheng Xu and Hanqing Lu, "Mining Semantic Context Information for Intelligent Video Surveillance of Traffic Scenes," VOL. 9, NO. 1, FEBRUARY 2013.
- [4] Gang Yu, Student Member, IEEE, Junsong Yuan, Member, IEEE, and Zicheng Liu, Senior Member, IEEE, "Action Search by Example Using Randomized Visual Vocabularies," VOL. 22, NO. 1, JANUARY 2013.
- [5] Oluwatoyin P. Popoola, Member, IEEE, and Kejun Wang "Video-Based Abnormal Human Behavior Recognition—A Review" IEEE transactions on systems, man, and cybernetics—part c: applications and reviews, vol. 42, no. 6, november 2012.
- [6] Maciej Szczodrak, Jozef Kotus, Krzysztof Kopaczewski, Kuba Lopatka, Andrzej Czyzewski, Henryk Krawczyk "Behavior analysis and dynamic crowd management in video surveillance system" 2011 22nd International Workshop on Database and Expert Systems Applications.
- [7] 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. 2010.
- [8] T. Sevilimis, M. Bastan, U. Gudu, IEEE, and Turgay Yilmaz, "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] S. Hongeng, R. Nevatia, and F. Bremond, "Video-Based Event Recognition: Activity Representation and Probabilistic Recognition Methods," J. Computers, VOL. 27, NO. 1, JANUARY 2004.
- [10] Gang Yu, Student Member, IEEE, Junsong Yuan, Member, IEEE, and Zicheng Liu, Senior Member, IEEE, "Action Search by Example Using Randomized Visual Vocabularies," Computer Vision and Image Understanding in 2004.