"Features Preserving Video Event Detection using Relative Motion Histogram of Bag of Visual Words"

Arifa Mulani[1] Varsha Mahajan[2] Swati Wadghule[3] Radha Shirbhate[4]
[1,2,3]Student [4]Assistant Professor
[1,2,3,4]Department of Computer Engineering
[1,2,3,4]JSPM'S BSIOTR, Pune, India

Abstract— Incident discovery in video is a investigate vicinity which attempts to build up a computer system with the capability to robotically read the video and locate the happening from images. Now a day there is a huge demand to find the Event depending on motion relativity and feature selection. In this paper, we propose our predictive based on movement relativity and feature selection for video incident discovery. Furthermost, we recommend a new movement feature, namely The concept of Expanded Relative Motion Histogram of Bag-of-Visual-Words (ERMH-BoW) is used for motion relativity and event detection. In ERMH-BoW, by representing what aspect of an event detection with Bag-of-Visual-Words(BoW), we construct relative motion histograms between different visual words to show the objects activities or how aspect of the event. ERMH-BoW thus integrates both what and how aspects for a absolute event description. Meanwhile, we show that by employing motion relativity, it is invariant to the varying camera movement and able to honestly describe the object activities in an event. We further propose a approach based on information gain and in formativeness weighting to select a cleaner. Our experiments are carried out on several challenging datasets provided by TRECVID for the MED (Multimedia Event Detection) task. It demonstrate that our proposed approach outperforms the state-of-the-art approaches for video event detection.

Key words: Feature selection, motion relativity and video event detection

I. INTRODUCTION

Now-a-days, there is necessity of security in motion detection and event analysis of different video file" there is growing demand for the software systems to recognize Event in computer system when video information as it is not in proper sequence " To detect and recognize events of user interest from video streams, audio and texts a lot of efforts have been put to event-based video analysis including unusual event detection. In this paper, we focus on extracting effective features from video sequences for event detection. In a video clip, an event is usually described from two aspects: i) what are the objects scenes that participate in the event, e.g. people, objects, buildings, etc.; ii) how the event evolves in temporal domain, i.e. the course of the event.

The features are to describe what aspect have been intensively studied, including global features such as color moment, wavelet texture, edge histogram, local features such as SIFT, Color SIFT and semantic features concept score. In summary, the effectiveness of ERMH-Bow in describing an event occurrence lies in three aspects: i) It directly integrate both the static (Bow with SIFT) and the active information (motion) of an occasion in one feature; ii)It is invariant to the unstable camera group and thus able to find out the ordinary motion pattern in the videos contain the equivalent happening; iii) It depicts not only the action of the matter, but also the interactions sandwiched between dissimilar matter scene which is important in describing event occurrences.

II. LITERATURE SURVEY

A. *Detecting Irregularities In Images And In Video*

Authors: O.Boiman and M.Irani.

He try to expose a new video segments using chunks of data extracted from previous databases or examples. Regions in the observed data which can be composed using large continuous chunks of data from the database are considered very likely whereas regions in the observed data which cannot be composed from database are regarded as unlikely suspicious.

B. *Mosift: Recognizing Human Actions In Sueveillance Videos*

Author: M.Chen

He proposed the new goal of this approach is to build robust human action recognition for real world surveillance videos. Current approaches tend to extend spatial descriptions by adding a temporal component for the appearance descriptor, which only implicitly captures motion information. We propose an algorithm called MoSIFT, which detects interest points and encodes not only their local appearance but also explicitly models local motion. The idea is to detect distinctive local features through local appearance and motion. We construct MoSIFT feature descriptors in the spirit of the well-known SIFT descriptors to be robust to small deformations through grid aggregation.

C. *Statastical Analysis of Dynamic Actions*

Author: Lihi Zelnik-Manor

He define Real-world action recognition applications require the development of systems which are fast, can handle a large variety of actions without a priori knowledge of the type of actions, need a minimal number of parameters, and necessitate as short as possible learning stage. In this paper, we suggest such an approach. We regard dynamic activities as long-term temporal objects, which are characterized by spatio-temporal features at multiple temporal scales. Based on this, we design a simple statistical distance measure between video sequences which captures the similarities in their behavioral content. This measure is nonparametric and can thus handle a wide range of complex dynamic actions. Having a behavior-based distance measure between sequences, we use it for a variety of tasks, including: video indexing, temporal segmentation, and action- based video clustering. These tasks are performed without prior knowledge of the types of actions, their models, or their temporal extents.

III. DESIGN PROCESS

In projected system with the consideration of the existing system we are proposing some techniques to represent the system by different way which includes:

- It strongly integrates both the stagnant (Bow with SIFT) and the vibrant information (motion) of an incident in one attribute;
- It is invariant to the altering camera movement and thus able to find out the frequent motion patterns in the videos containing the same incident;
- It explains not only the activity of the substance, but also the communications amid dissimilar substance/scenes which is vital in relating incident occurrences. The proposed system explains that ERMH-Bow can considerably progress the performance of visual incident uncovering compared with the state-of-the-art way by characteristic assortment, the computation time for incident discovery.
- With the ERMH-Bow feature is reduced while the accuracy is also slightly improved.
- Effective and efficient event detection.

A. System Architecture

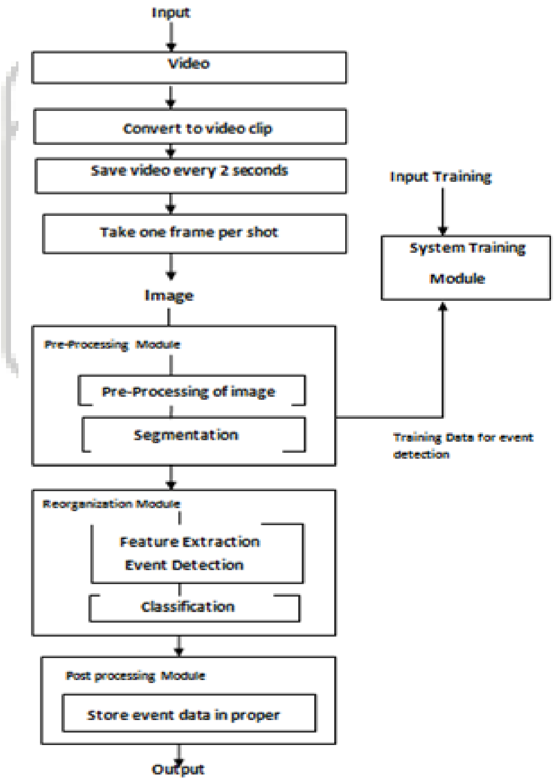

Fig. 1: System Architecture

IV. MODULE DESCRIPTION

The way based on activity relativity and characteristic assortment for visual incident discovery System Implementation consist of a variety of parts describe as underneath. Efforts are put in implementation of project by the edge of Java Technology which is extensible, portable and platform independent. The database technology that we are using is MySQL database. Different components of our system include;

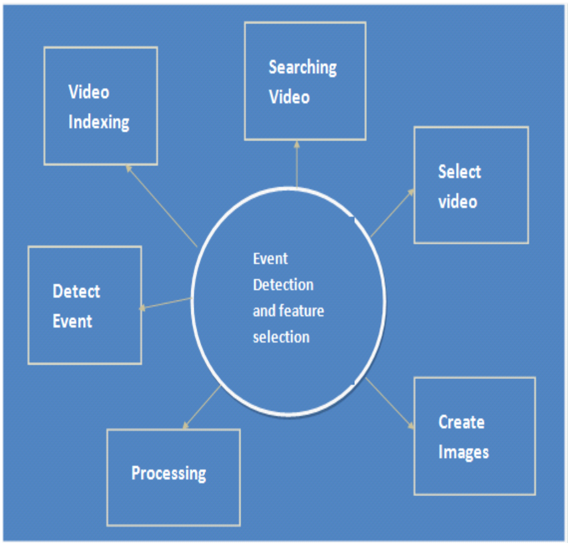

Fig. 2: System Components

A. Video To Image Conversions

This section includes the Video conversion. The video transformed into the clip, for doing this we are saving video for every two seconds and take one frame per shot. This will gives us the outputs which create images of video.

B. Pre-Processing Stage

The pre-processing stage includes several steps which includes Pre-processing of image and Segmentation

C. Pre-Processing Of Image

The pre-processing of image includes following mechanism to convert image in to proper format. These also called structured documents such as Grayscale and binary format.

D. Segmentation

After the pre-processing of image step the next and vital step is the segmentation is the most vital process in motion detection. Segmentation is done to make the separation between the individual events of an image.

E. Reorganisations

The next to Pre-processing step is the reorganization of the data, which includes the various specific steps which includes Feature Extraction, Event Detection and Classification.

F. Feature Extraction

Feature extraction is the process to retrieve the most important data from the raw data. The most important is data can be represented accurately. In this we use matrix feature extraction method. In this method first we convert the image to binary matrix i.e. black and white image convert to matrix form.

G. Event Detection

The Event Detection is depending on the motion relativity and feature and sequence of motion find the Event of video.

H. Classification

In the classification stage we identify the motion of images and depending on feature and motion created new cluster.

I. Post Processing

The Post processing stage includes the functionality that the output is in the images format so store it in the proper format.

V. SUPPORT VECTOR MACHINE ALGORITHM

The Support Vector Machine algorithm that is SVM has some finite steps of processing which are defined as follows;

- Training set - The set of images that will be used to train our SVM.
- Test set - At the end of the SVM training we will use these images for classification.
- Labe: It will use Faces and Airplanes these are two objects so we will give them two "labels" for Face and Airplane respectively.
- Classify: Distinguish our test set images. Each normal peer has some processes like data loading and data indexing. In normal peer there are two data flows first is an offline data flow and an online data flow. In fig 2 event detection and feature selection step are shown.
- Video indexing: In indexing provided index to the video .indexes useful to search video.
- Search video: Depending on index video can be search.
- Conversion of video to image: Selected video is converted in images. Video is saving each two seconds and then find shot for each clip.
- Detect Event: Depending on motion relativity and feature selection find the Event of video which is occurring in video.

VI. VIDEO MOTION DETECTION

We explain moving object detection method by using both motion vector :

A. Motion Vector (MV)-Based Motion Detection

1) FF mpeg is a popular open source tool to process video stream. It provides tools to encode, decode, transpose, multiplex, demultiplex, and filter video stream.
2) It contains libraries that can be used to build customized multimedia applications. FFmpeg also provides interfaces to access video compression features.
3) In this case, during decoding, we can parse the compression features easily from video stream for analysis. One of such features is motion vector (MV).
4) To perform feature extraction, we can use either FF mpeg or Open CV. However, since the detection algorithm hasbeen implemented in Open CV, we decide to use Open CV to do the feature extraction process.
5) Open CV is a well-known open source tool for image processing.
6) It has plenty of tools and libraries to do typical image processing jobs, e.g., filtering and segmentation.

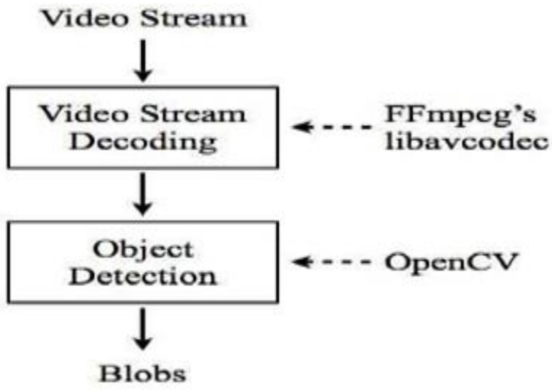

Fig. 3: Motion Detection

Fig. shows the simplified diagram of this detection process. FFmpeg is used only to decode the video stream.

Then, the object detection function uses the decoded data to detect any moving object by using the available programming interfaces provided by Open CV.

We have presented video motion detection methods, motion vector (MV) based and, using open source tools, FFmpeg and OpenCV libraries. With the existing tools, we could build moving object detection application by simply extending the video decoding functions.

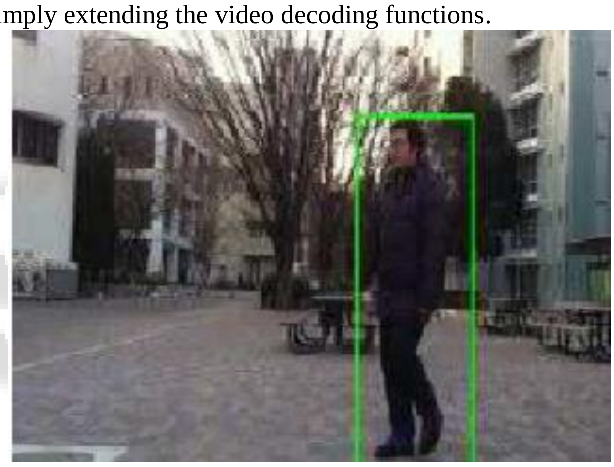

Fig. 4: Output

Above fig show the output of the input with green rectangle.

Graph Frame Size vs Video Size

Where:

Y Axis=frame size

X Axis=video size (resolution)

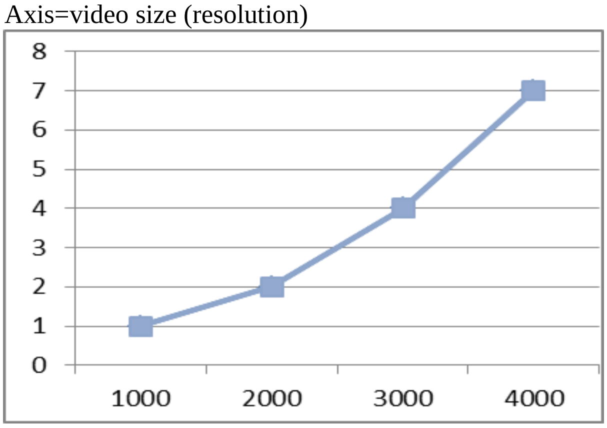

Fig. 5: Graph of Frame Size vs Video Size

Graph Time Vs Number of Frame

Where:

X Axis Number of frame

Y Axis Time for frame Creation

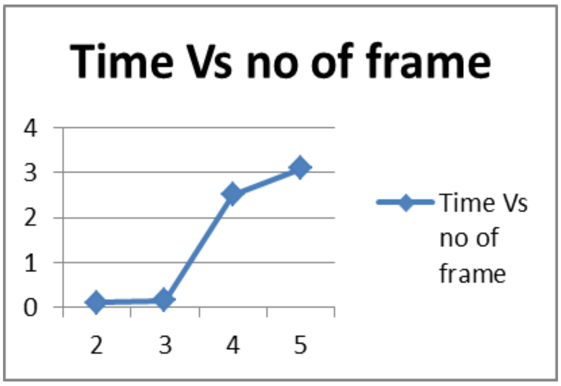

Fig. 6: Graph of Time vs no of Frames

B. *Histogram*

Histogram is created according to the RGB Values.

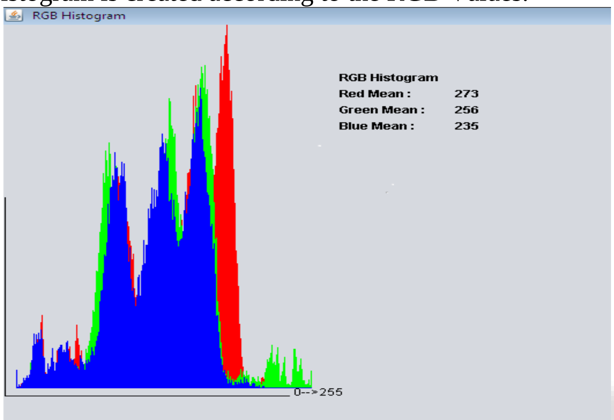

Fig. 7: Histogram

As shown in the above figure, histogram is created as per RGB Values.

C. *Final Output with Color Histrogram*

In Final output Finally Event is detected.

Fig. 6.6: Final Output

VII. Conclusion

ERMH-Bow addresses event detection in open video domains by proposing a new motion feature. Bag-of-visual-words are adopted to represent what aspect of an event, while the relative motion histograms between visual words are used to capture the object activities or how aspect of the event. The derived feature ERMH-Bow can thus provide a complete description of an event by closely integrating what and how aspects. In future we consider the spatial information to detect event.

References

[1] R.Aly, K.McGuinness, and S. Chen et al., "AXES at TRECVid 2012:KIS, INS, and MED," in Proc. NIST TRECVID Workshop, 2012.
[2] E. Ardizzone, M. L. Cascia, A. Avanzato, and A. Bruna, "Video indexing using MPEG compensation vectors," in Proc. IEEE Int. Conf. Multimedia Computing and Systems, 1999, vol. 2.
[3] A. Basharat, A. Gritai, and M. Shah, "Learning object motion patterns for anomaly detection and improved object detection," in Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2008.
[4] H. Bay, T. Tuytelaars, and L. V. Gool, "SURF: Speeded up robust features,"in Proc. Eur. Conf. Computer Vision, 2006.
[5] O. Boiman and M. Irani, "Detecting irregularities in images and in video," in Proc. Int. Conf. Computer Vision, 2005.
[6] L. Cao, S. F. Chang, andN.Codella et al., "IBMresearch and Columbia University TRECVID 2012 multimedia event detection (MED), multimedia event recounting (MER), and semantic indexing (SIN) systems," in Proc. NIST TRECVID Workshop, 2012.
[7] C. Chang and C. Lin, "LIBSVM: A library for support vector machines". 2001[Online]. Available: http://www.csie.ntu.edu.tw/cjlin/libsvm.
[8] M. Chen and A. Hauptmann, "Mosift: Recognizing human actions in surveillance videos," in CMU-CS-09-161, 2009.
[9] H. Cheng, J. Liu, and S. Ali et al., "SRI-sarnoff aURORA aystem at TRECVID 2012: Multimedia event detection and recounting," in Proc. NIST TRECVID Workshop, 2012.
[10] N. Cristianini, J. Kandola, A. Elisseeff, and J. S-Taylor, "On kernel target alignment," Adv. Neural Inf. Process. Syst., vol. 14, pp. 367–373, 2002.
[11] N. Dalal, B. Triggs, and C. Schmid, "Human detection using oriented histograms of flow and appearance," in Proc. Eur. Conf. Computer Vision,2006.
[12] J. W. Davis, "Hierarchical motion history images for recognizing human motion," in Proc. IEEE Workshop Detection and Recognition of Events in Video, 2001.
[13] D. Lowe, "Distinctive image features from scale-invariant keypoints," Int. J. Comput. Vision, vol 60, no. 2, 2004.