

Fusion of Optical Flow and Histogram of Oriented Gradient Features in Abnormal Activity Recognition

Mahasweta Joshi¹ and Jitendra Chaudhari²

¹Birla Vishvkarma Mahavidhyalaya, V. V. Nagar, Gujarat, India
Email: mjjoshi@bvmengineering.ac.in

²CHARUSAT University, Changa, Gujarat, India
Email: jitendrachaudhari.ec@charusat.ac.in

Abstract—Nowadays abnormal action and event recognition is very encouraging research topic in surveillance applications. There are three main component of abnormal activity recognition. The components are: (i) Feature representation approaches (ii) pattern recognition models and (iii) performance evaluation strategies. Compare to second and third component, first component: Feature representation is very important. It should have robust appearance and motion information. This information of the features plays very important role in video analysis. In this paper, fusion of features has been performed to find more accurate recognition of abnormal activity. Optical Flow and HOG (Histogram of Oriented Gradients) features has been fused to get more accurate feature vector. Upon this vector rule based recognition method has been applied for classifying abnormal event. This proposed algorithm has been implemented on UCF, BEHAVE and UMN dataset. The result shows that with fusion of features more accuracy has been achieved in finding starting frame of abnormal activity in videos. Compare to existing algorithm, proposed algorithm achieved 16%, 1.8% and 21% more accuracy in UMN, BEHAVE and UCF dataset respectively.

Index Terms— Optical flow, Histogram of Oriented Gradients, UCF dataset.

I. INTRODUCTION

Video surveillance is promising research area in current world for asset and personnel safety. But we need automation in it. It is so because this can reduce the manual workload. Object detection, tracking, person detection, event recognition and activity recognition are different areas where current researchers are focusing. Public area can be monitored with the help of this applications and with less human power. However, detecting an abnormal activity from a video is a challenging problem that is extremely contextual [1]. Abnormal activities are an anomalies of the object behaviour from the usual behaviour. This includes object in uncommon place, uncommon pattern of motion such as movement in incorrect direction, illegal turns in traffic, object access or entry in private zone, fighting among the persons, left bag, unexpected movements, let fallen object or any type of uncommon occurrence. For one event it may be a normal in one scenario and abnormal in another scenario [2].

The activity recognition system is primarily having following steps:

- Optional step: Pre processing
- Important step: Extraction of features

- Identification step: Tracking of objects
- Understanding object step: Understanding behaviour of the object.

Foreground extraction and Noise removal has been done in pre-processing step. In feature extraction, number of features of the object or person has been extracted from the frames of the video. Based on extracted features, object trajectory has been established. Behaviour understanding is a vital and significant step. Here behaviour of the object has been detected and based on that behaviour classification of event has been performed. In last step, it gives normal and abnormal event detection in videos [2]. Applications [3, 4] of abnormal activity detection system are:

- Traffic monitoring
- Crowd analysis
- Automatic forensic video retrieval
- Criminal activity recognition
- Medical applications
- Surveillance activity

It's mainly used for security alarm in surveillance activity. Though it has potential application, it is much popular with making surveillance system more intelligent [3]. In proposed frame work, feature extraction step has been emphasized. Here, two features optical flow and HOG both are extracted and fused for further action. Section 2 describe the related work, section 3 describe the optical flow and HOG features, section 4 describes proposed abnormal activity detection algorithm, section 5 describe the implementation results and discussion and section 6 describes the conclusion.

II. RELATED WORK

Abnormal activity detection can be performed in two kind of category. First one is individual activity video and second one is crowd activity video. In former category of videos, segmentation and object tracking has to be performed on individuals but it suffers from occlusion problem. In crowd activity videos, features has been extracted and algorithm works on these features [6]. Habib Ullah et. al [7], represents the unsupervised method for detecting individual movement from a crowd. Researchers had used feature tracker to generate trajectories. To make cluster unsupervised method – Bayesian clustering is used. Gabriel J. Brostow and Roberto Cipolla [8] describe the unsupervised method which detect individual entities using Bayesian clustering algorithm. The limitation of this algorithm is if an individual is present in corners or absents then clustering will ignore that individual. As a part of recognition, it has two categories: one is machine learning methods [9, 10] and second is thresholding [11, 12]. When the event activities are well defined, machine learning based methods are helpful to recognize the activities but it is very difficult to define and model so many activities for unconstrained environment. In threshold based method, one has to extract few indicators only which will become helpful to identify activities. When threshold value cross the indicator value, it will generate the alert. The advantages of this method are: (i) No training data (ii) No learning process (iii) very few computation.

III. FEATURE EXTRACTION

Feature extraction is key step in procedure of abnormal activity recognition. There are various methods to extract features. Here optical flow and HOG have been used in proposed algorithm.

A. Optical Flow

Optical flow [4, 5] is a pattern of apparent 2D motion of objects in sequences of time-ordered images. Video frames can be consider as time-ordered images. From these images feature vector has been generated. These vectors show motion of objects from one image to next consecutive image. Pixels travels from one video frame to another video frame which indicates optical flow. The apparent motion of the brightness pattern calls optical flow [5]. Relative motion of objects and the viewer creates optical flow. Optical flow is a distribution of seeming velocities concerning movement of brightness patterns in a video frame. The important, useful and necessary information of the motion, which is relative has been stored by optical flow. Suitable actions can be generated when the movement has been extracted. It is done for controlling object. The brightness patterns are moving with respect to movement of the object or camera. When the motion is very small and the illumination of the video frame keeps uniform in space as well as it steadies over time, it can be expected that the particular pixel remains constant throughout the motion [13, 14, 15]. The equation of classical brightness constancy constraint should be known as

$$E(a + \Delta a, b + \Delta b, p + \Delta p) = E(a, b, p) \quad (1)$$

Here the brightness of a point (a, b) represented at time p by E (a, b, p) on the image plane. Intensifying the left part and ignoring high-order terms, the equation is going to reduce in following form:

$$E_p + E_a u + E_b v = 0 \quad (2)$$

Here u represent component of the optical flow along the x axis and v represents component of the optical flow along the y axes, respectively. It has been assumed that for every pixel of the video frame has a unique optical flow. A cost function of the sum of squares of the brightness variation can be framed. The optimum solution of optical flow (u, v) has been easily determined using a least-squared method which creates minimum cost function. The point based calculation has been resulted in unconventionalities of the optical flow. It occurs due to noise available in real video frames. So the above equation has been extended to patch based formulation. Using averaging, the estimation error has been reduced significantly. Let's assume that the image patch Ω and its pixels have own a similar optical flow. Patch-based brightness constancy is represented by following equation:

$$\sum_{\Omega} (E_p + E_a u + E_b v) = 0 \quad (3)$$

Here for Ω , u and v signify the values of the optical flow. This method enhances the accuracy of optical flow and also improves computational efficiency [16].

B. Histogram of Oriented Gradient

The histogram of oriented gradient (HOG) descriptors had been proposed by Dalal and Triggs for human activity detection. HOG was derived based on the assessment of standardized local histograms of image gradient trends in a dense network by considering fine-scale gradation, fine-directional correlation, relatively coarse spatial clustering, and high-quality local contrast standardization in nested descriptor block [17]. The HOG has significant advantages:

- HOG can captures the edge or gradients structure
- Local representation of information
- Invariance to translation and rotation

For human detection, moderately coarse spatial sampling, fine-orientation sampling and strong local photo normalization turned out to be the best strategy, possibly because it allows limbs and body parts to change appearance and move from side to side frequently provided that you maintain the orientation nearly upright [18].

IV. PROPOSED ALGORITHM

As mentioned in [19], an abnormal activity has been detected using Harris corners, optical flow and classification by dynamic thresholding. This algorithm has been implemented here with fusion of optical flow and HOG features. The block diagram of proposed algorithm has been depicted as shown in Fig.1. These are the steps of proposed algorithm.

- Step 1: Here pre-processing has been applied on images. As a part of it, using averaging method background subtraction has been applied.
- Step 2: Now feature extraction has been done. Using Harries corners, key points has been identified. It will be taken feature points. This feature points represents moving object. The object has been track by method called optical flow. Using it motion intensity has been estimated. On Harris corner points HOG has been applied and created HOG feature vector. Create a final feature vector by making fusion of Optical flow feature vector and HOG feature vector.
- Step 3: Classification has been performed by dynamic thresholding method [19]. Using it, input has been classified in two classes. One is normal activity and another is abnormal activity.

V. RESULTS OF IMPLEMENTATION

Three dataset has been used for implementing this algorithm. First dataset is UMN dataset, where in each video signifies crowd escape panic. Second dataset has been used is BEHAVE dataset. It has been taken from [20]. The dataset includes various situation of persons behaving different interaction. Third dataset has been used is UCF crime dataset, which is very large dataset and it contains 1900 videos. It includes videos of normal and abnormal activities. There are 13 abnormal activities in this dataset [21]. Table 1 shows the comparison of [19]

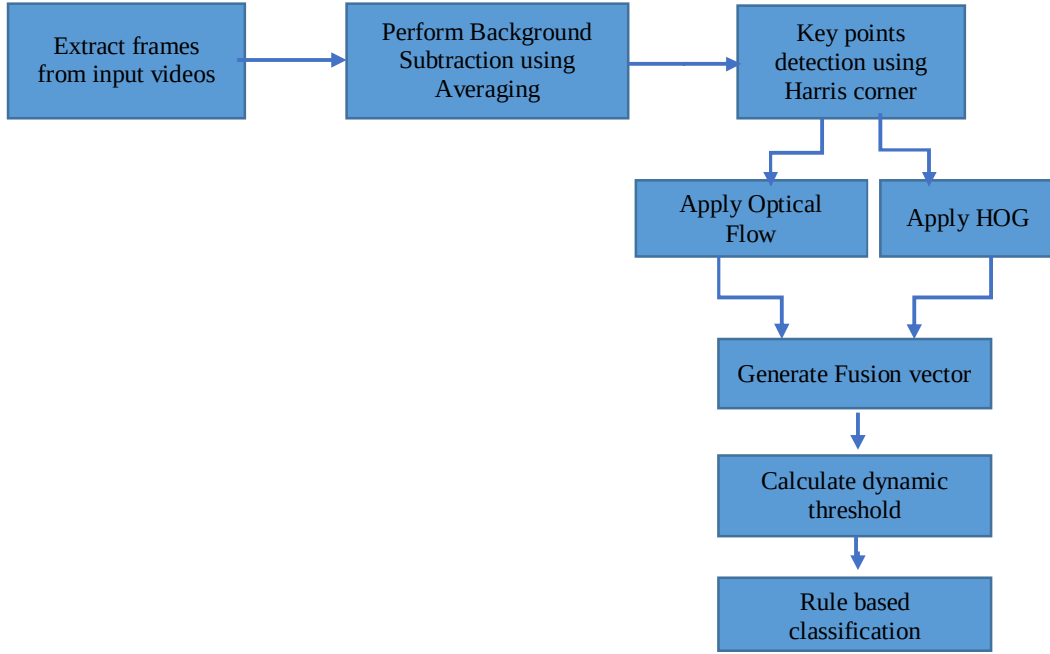


Figure 1: Proposed algorithm

and proposed algorithm where starting frame of abnormal activity has been detected and the results shows that in proposed algorithm gives more accurate results.

TABLE I. COMPARISON OF [19] AND PROPOSED ALGORITHM FOR DETECTING STARTING FRAME OF ABNORMAL ACTIVITY FROM VIDEO (SAMPLE VIDEO)

Dataset	Actual starting frame of abnormal activity	Starting frame of abnormal activity in existing [19] algorithm	Starting frame of abnormal activity in proposed algorithm
UMN_1	410	370	146
UMN_2	270 and 940	-	-
UMN_3	159	104	150
BEHAVE_1	458	428	448
BEHAVE_2	645	683	-
BEHAVE_3	421	-	-
BEHAVE_4	222	-	233
UCF_Fight_1	100	36	98
UCF_Fight_2	350	-	-
UCF_Fight_3	150	92	137
UCF_Fight_4	620	313	486
UCF_Fight_5	305	313	317
UCF_Fight_6	335	194	218
UCF_Shooting_1	245	285	262
UCF_Shooting_2	730	-	-
UCF_Shooting_3	480	-	462
UCF_Shooting_4	180	158	172
UCF_Shooting_5	160	145	158

TABLE II. COMPARISON OF ACCURACY

Dataset	Existing Algorithm[19]	Proposed algorithm
UMN	51.88424	67.7067
BEHAVE	49.83531	50.69289
UCF	55.16881	76.14356

Table II shows the comparison of accuracy achieved by proposed algorithm in UMN, BEHAVE and UCF dataset. Figure 2 (a),(b),(c), and (d) shows the starting frame of sample video taken from UMN, BEHAVE and UCF – Crime dataset. Figure 3 represents accuracy measurement in dataset: UMN, BEHAVE and UCF. This chart describes the accuracy achieved in simulation of Existing algorithm and proposed algorithm.



Figure 2: Implementation result of proposed algorithm on sample of (a) UMN (b) BEHAVE (c) and (d) UCF Crime dataset

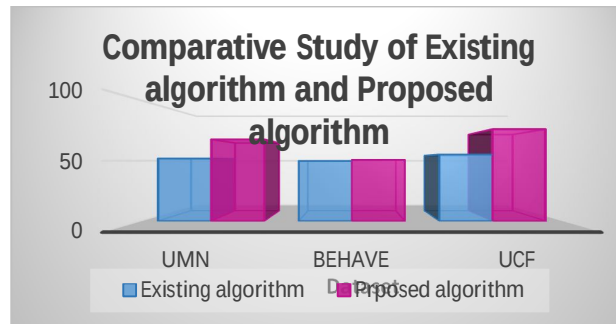


Figure 3: Comparative chart of Existing algorithm [19] and proposed algorithm

Table I, Table II, Fig.2 and Fig.3 shows the quantitative and qualitative analysis. It shows that when HOG has been added in part of feature extraction, the algorithm gives better result than the existing algorithm [19]. In case of UMN and UCF dataset, proposed algorithm achieved 16% and 21% more accuracy respectively. But in Behave dataset due to combination of multiple similar activity like walking and walking with fighting on same place proposed algorithm achieved only 1.8% more accuracy. So our proposed algorithm will not much good for BEHAVE dataset.

VI. CONCLUSION

The velocity, speed and motion of human action can be represented by optical flow features. Optical flow mostly used when the action has similar shape with different speed. When only shape or gradient has to be recognized then HOG works better. The HOG and optical flow have been fused to combine the advantages of both. Using fusion of these features, the experiment results shows that improvement has been noticed in result of

classification. It represent better activity classification than other existing methods. The proposed algorithm has been assessed with different datasets like UMN, BEHAVE and UCF crime dataset. A more accuracy has been reported in finding starting frame of abnormal activity videos. The main reason of increase in performance is due to considering motion and shape of the actions efficiently.

REFERENCES

- [1] Kothapalli Vignesh, Gaurav Yadav, Amit Sethi, "Abnormal Event Detection on BMTT-PETS 2017 Surveillance Challenge", IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 4321-4328, 2017.
- [2] Sarita Chaudhary, Mohd Aamir Khan, Charul Bhatnagar, "Multiple Anomalous Activity Detection in Videos", Procedia Computer Science, pp. 336-345, 2018.
- [3] Zhang, Y., Lu, H., Zhang, L., Ruan, X., "Combining motion and appearance cues for anomaly detection", Pattern Recognition, pp. 443-452, 2016.
- [4] Hampapur, A., Brown, L., Connell, J., Pankanti, S., Senior, A., Tian, Y., "Smart surveillance: applications, technologies and implications", Information, Communications and Signal Processing 2, pp. 1133-1138, 2003.
- [5] Mohammad Afrooz Mehr and Maziar Haghpahan, "Optical Flow Features for Event Detection", Halmstad University, 2013.
- [6] Yang Liu, Xiaofeng Li and Limin Jia, "Abnormal Crowd Behavior Detection Based on Optical Flow and Dynamic Threshold", 11th World Congress on Intelligent Control and Automation Shenyang, pp. 2902-2906, 2014.
- [7] H. Ullah, M. Uzair, M. Ullah, A. Khan, A. Ahmad, and W. Khan, "Density independent hydrodynamics model for crowd coherency detection," Neurocomputing, vol. 242, pp. 28-39, 2017.
- [8] G. Brostow and R. Cipolla. Unsupervised bayesian detection of independent motion in crowds. Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, 1:594-601, June 2006.
- [9] L. Kratz and K. Nishino, "Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models", in IEEE Conference on Computer Vision and Pattern Recognition, pp. 1446-1454, 2009.
- [10] Y. Benezeth, P.-M. Jodoin, V. Saligrama and C. Rosenberger, "Abnormal Events Detection Based on Spatio-Temporal Cooccurrences", in IEEE Conference on Computer Vision and Pattern Recognition, pp. 2458-2466, 2009.
- [11] J. S. Xie, L. Guo, Y. B. Chen and L. Zhao, "The Detection of Unusual Events in Video Based on Bayesian Surprise Model", in 2010 2nd International Conference on Information Science and Engineering, pp. 4784-4788, 2010.
- [12] D. Y. Chen and H. Po-Chung, "Dynamic Human Crowd Modeling and Its Application to Anomalous Events Detection", in 2010 IEEE International Conference on Multimedia and Expo, pp. 1582-1587, 2010.
- [13] Anshuman Agarwal, Shivam Gupta, Dushyant Kumar Singh, Review of Optical Flow Technique for Moving Object Detection, 2nd International Conference on Contemporary Computing and Informatics (ic3i), pp. 409-413, 2016.
- [14] Florence Simon and Prof. Ujwal Harode, —Human Action Recognition using Optical Flow Estimation International Research Journal of Advanced Engineering and Science, Volume 4, Issue 2, pp. 164-168, 2019.
- [15]] Milin P. Patel, Shankar K. Parmar, "Moving Object Detection with Moving Background using Optic Flow", IEEE International Conference on Recent Advances and Innovations in Engineering ,2014, pp. 1-6.
- [16] Chi-Cheng Cheng, and Hui-Ting Li, "Feature-Based Optical Flow Computation", International Journal of Information Technology Vol.12 No.7, pp. 82-90, 2006.
- [17] Shian-Ru Ke, Hoang Le Uyen Thuc, Yong-Jin Lee, Jenq-Neng Hwang, Jang-Hee Yoo, Kyoung-Ho Choi, "A Review on Video-Based Human Activity Recognition", Computers, pp. 88-131, 2013.
- [18] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 1, pp. 886-893, 2005.
- [19] Mahasweta J. Joshi, Jitendra P. Chaudhari, "An Abnormal Activity Detection Using Harris corner, Optical Flow and Dynamic Thresholding", Turkish Online Journal of Qualitative Inquiry (TOJQI) Volume 12, Issue 7, pp. 1144 -1150, 2021.
- [20] S. J. Blunsden, R. B. Fisher, "The BEHAVE video dataset: ground truthed video for multi-person behavior classification" , Annals of the BMVA, vol 4, pp. 1-12., 2014.
- [21] Waqas Sultani, Chen Chen, Mubarak Shah, Real-world Anomaly Detection in Surveillance Videos, IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018.