Performance Analysis of Spatio-temporal Human Detected Keyframe Extraction

Victoria Priscilla C[1]
Associate Professor & Head, PG Department of Computer Science, Shrimathi Devkunvar Nanalal Bhatt Vaishnav College for Women, University of Madras, Chennai (TamilNadu), India, aprofvictoria@gmail.com

Rajeshwari D[2]
Research Scholar, Research Department of Computer Science, Shrimathi Devkunvar Nanalal Bhatt Vaishnav College for Women, University of Madras, Chennai (TamilNadu), India, rajisowji2011@gmail.com

Abstract

Closed circuit television (CCTV) surveillance for detecting the humans involves an expanded research analysis especially for crime scene detection due to various restraints such as crowded annotation, night footages, and rainy (noisy) clips. The main visualization of the crime scene is to recognize the person in particular obtained in all frames is a challenging task. For this occurrence, Content-Based Video Retrieval (CBVR) method refines the collection of these video frames resulting keyframes to reduce the burden of huge storage. Here, Spatio-Temporal classifiers method as an added advantage with frame differencing and edge detection method reports the human detected keyframes without the termination of background regions in order to negotiate the crime scene more efficiently. The main objective of this paper is to analyze the obtained keyframes with Human detection pointing a distinctive between Spatio-Temporal HOG-SVM and HAAR-like classifier to survey the optimum. Finally, the resulting keyframes mutated with the canny edge detection method by HOG-SVM sequel with greater accuracy level of 98.21% compared to HAAR-like classifier.

Keywords: *CCTV surveillance, HOG (Histogram of Oriented Gradients) – SVM (Support Vector Machine), HAAR-like Cascade Classifier, Keyframe Extraction, Spatio-Temporal feature extraction, Human Detection, CBVR (Content Based Video Retrieval).*

I. Introduction

In the current circumstances, CCTV surveillance is used in many public areas such as official sectors, airport authorities, railway stations etc. The surveillance video footage documents only the images or the videos of the instance occurred with huge stored video data of day-to-day instance. The security room further receives the captured videos of crime scene only after the event expired with no other supplementary information. The main requisites of investigator have to sit for a long stretch to suspect the suspicious activity or some abnormal activities from these footages to accumulate the evidence. The suspected frames with the human criminal activity in the video surveillance still a challengeable task. Therefore it is necessary to determine an algorithm to detect the human and object in the surveillance footages to quote the crime scene frames.

Human detection in video surveillance plays a crucial role in diverse applications including the abnormal event perception, crime scenes,

etc., Many researchers followed various methods to detect the human has a growth in many image processing applications and deep learning algorithms. The area of application of human detection in Video surveillance is more enhanced in Machine Learning, Artificial Intelligence and Deep Learning.

The proposed work grasps the current scenario of CCTV footages as dataset for the crime scene investigation focused on background detection, motion detection and moving object classification. The final resultant keyframes by frame-differencing method and edge detection method makes a comparative on human detected key frames with HOG-SVM and HAAR-like cascade classifier to reach out the optimum.

The paper work is systemized as follows: Section 2 briefly explains the related work of the human detection in surveillance. The proposed approach is elaborated detailed in Section 3. The implementation and the experimental results are described in Section 4 followed by the conclusion work in Section 5.

II. LITERATURE REVIEW

The literature review proposes the various approaches by the authors for detecting the human from the video surveillance. As the scope is to determine the human from other non-human objects in a video surveillance, the Content Based Video Retrieval deserves a best source for the proposal. The initial stage of detecting the humans from the frames can be performed using background subtraction, optical flow or spatio-temporal filtering.

In the first phase the background subtraction in general is the distinction of motion. This detects the region in motion by deducting the required person image as pixel-by-pixel with elimination of the background [1]. The moving objects are detected by methods with the combination of background subtraction using pixel intensity [2], adaptive background mixture [3], and single change detection using wavelet transform [4] through background subtraction but these methods works on with low complexity and achieving better results. Also, the background subtraction has the fewer tendencies on evidence collection for crime scene investigation owing to the loss of background scene.

The next phase is the study on optical flow, which tracks the person for a required period of time to make exact information, is a major drawback for investigation. This method is processed for the human motion flow of current frame and the consecutive frames of stable video scenes [5] using optical flow model [6] and optical flow gradient based [7] etc. The detected objects from the optical flow reports well in the simple background but not on complex background scenes.

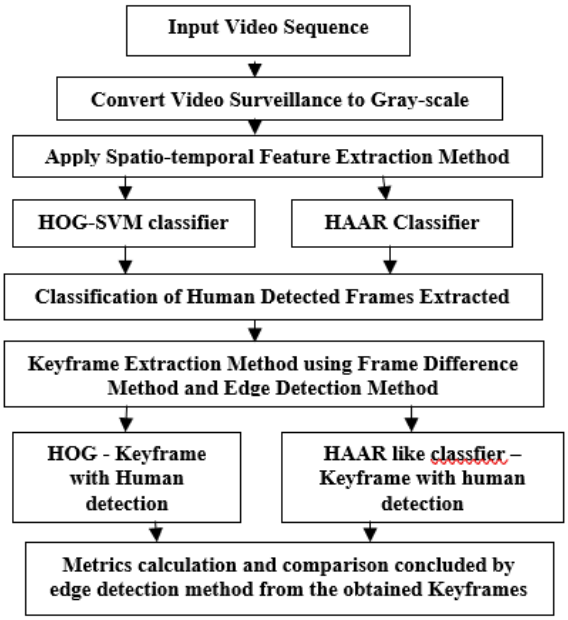
The next level of object detection is using spatio-temporal filtering, where it filters the objects at variable background further a major precedence for crime scene investigation. The spatio-temporal features detects the object using spatio-temporal graph method[8], spatio-temporal correlation with space-time object outline [9], spatio-temporal improving the per-pixel predicting map[10], and also with three dimensional Gabor filter method [11] at varying background determination, further they are good in computational time.

The proposed method uses the Spatio-temporal based technique from which the humans are detected in each frames using the feature extraction method HOG-SVM and HAAR- like cascade classifier. Here, HAAR Wavelet Based Cascade detector selects the most human like region with the spatio difference [12] and on the other side Histogram of Oriented Gradient [HOG] defines a bounding box to describe the person differed to other objects. The extracted keyframes are reported has a summarized video frames supporting the crime investigations. Here, the performance analysis of these two methods when compared substantially proved that HOG features along with linear SVM

(Support Vector Machine) achieve human detection keyframe at greater accuracy [13] level.

III. PROPOSED APPROACH METHOD

Fig. 1. Overall framework of proposed work



The proposed work is carried out by three different steps, to make a comparative discussion on human detection: (1) The Video sequence collected from surveillance. (2) The feature extraction method is applied to these Video collections and (3) the Keyframes are evaluated and compared. The flow chart of the proposed work is given in figure1.

A. VIDEO PREPROCESSING

The CCTV surveillance of the recorded video footages is collected as dataset to capture the human detection. The footages at the initial step are to be sliced as frames. The collected footages are converted into gray-scale and also for faster detection these video frames are resized. The Human highlighted frames are extracted through the feature extraction method from which the humans are classified well.

B. SPATIO TEMPORAL FEATURE EXTRACTION- HUMAN DETECTION

Human detection is classified by these two methods. A detailed analysis has been studied here on the basis for the development of the proposed technology.

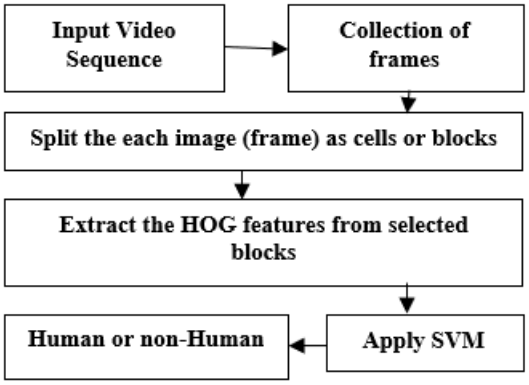
B.1.HOG AND SVM COMBINATION

The human detection in CCTV surveillance was also aimed by pattern recognition as like by the standing, sitting and walking [15]. The human can also be detected by their actions and their movements through individual and through a group of interactions [16]. The humans are much survived a lot in surveillance for crime scenes which can be prevailed by their 2D and 3D head surface model for the internal recognition of human [17]. Not only by pattern their body shapes such as face, skin, and other regions to authorize them as human [18]. In such circumstances, the feature extraction method supports a long way for the Human detection technique using HOG, which was primarily introduced Dalal and Triggs[13].

HOG method is used to describe the shape and regional appearance of the object with the intensity allocation of gradients or through the direction of contours.

Here the gradients are derived by evaluating the derivative in form of x and y. The performance of these descriptors on each frame can be acquired by splitting the image (single frame) into small connected sector called as cells or blocks. So prior to the frame extraction of HOG features, an iterative study is accomplished to find the most important blocks. These set of cells are kept in proposal while extracting the HOG features. This will eliminate the unwanted blocks. Then for each blocks the histogram of gradient are computed or the edge orientation include the pixel of the cell are evaluated. The integration of these histogram results the required descriptor as depicted in Fig 2.

Fig.2. Overall View of HOG Human Detection process



The resultant descriptors are classified using Support Vector Machine (SVM) which was initiated by Vapnik which splits the hyperplane between the spaces of two classes [19]. As the human detected through the HOG, the SVM maximize the edge, which is the distance between the separation boundary and the nearest regions. The SVM algorithm transforms the space of the input frames into a space of higher dimension, which exist linear separator [20]. The HOG and SVM results the video surveillance dataset with a high efficiency on each frame holding the picture with the Human or non-human as depicted in Fig. 3.

Fig.3. HOG penetrating Frames

B.2. HAAR-LIKE CASCADE CLASSIFIER

The comparative feature extraction method goes on with HAAR Classifier. The HAAR classifier is one of the effective methods for object detection. This method was introduced by Paul Viola and Michael Jones during year 2001[21]. This method is used not only for human full body detection but also for face detection [22], pedestrian detection [23], and white blood cell detection [24]. The cascade function is tested and trained from an excess of images of both the positive and negative. The trained results are stored in a separate .xml files as cascade function. Now the successive step is to convert the obtained frames to gray scale. The combination of the cascade function as body classifier xml file extracts the bounding boxes of human using the body_classifier.detectMultiScale along with the scale factors. Here the scale factor denotes the parameters of how far the image is reduced to its image scale. Finally, the HAAR-Like Cascade classifier results the rectangular bounding box Human on each frame as pictured in Fig. 3.

Fig.4. Overall View of HOG Human Detection process

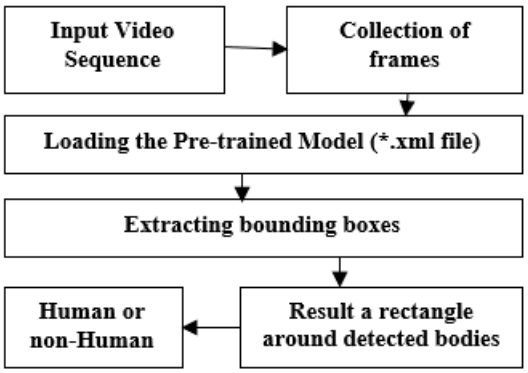
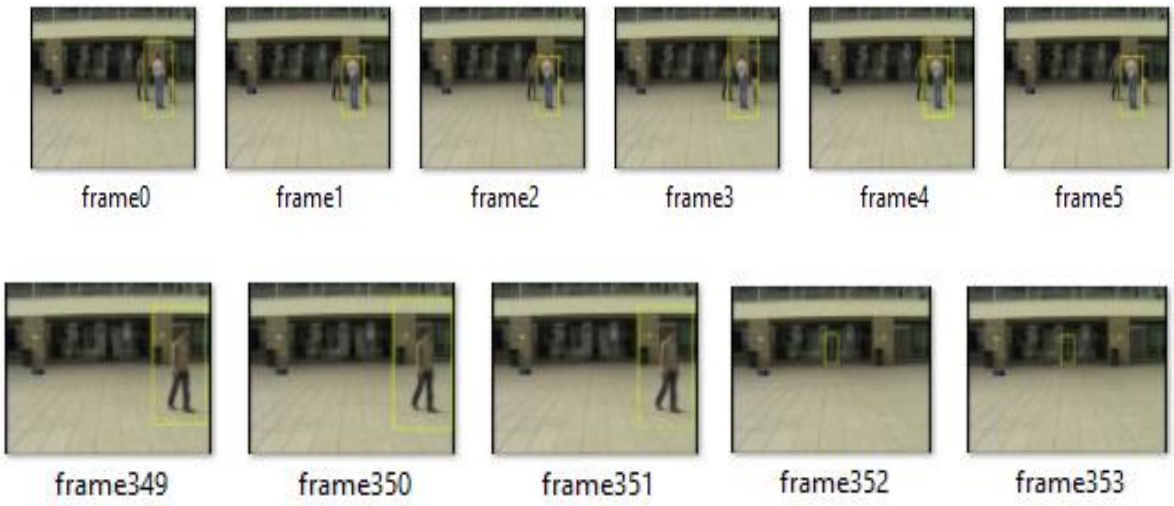


Fig.5.HAAR-like Classifier penetrating Frames



As according to the requirements the humans are detected from video surveillance dataset with the trained xml files are resulted in Fig.5.

The proposed work gives an intention from the obtained frames of both these methods. The collection of frames from both these methods derives that HOG with SVM outcomes with an increasing level of frames contrast to HAAR-like classifier. As for crime scene detection even a single progress has to be tapered from which the redundancy can be eradicated. So, the obtained frames through HOG along SVM determine the best for the further progress of extracting the keyframe to classify the human detected keyframes.

C. KEY FRAME EXTRACTION USING FRAME DIFFERENCE WITH EDGE DETECTION METHOD

The keyframes can be extracted by many terms such as background subtraction, optical flow, inter-frame difference etc. The frame difference is evaluated as the absolute difference between the current frame and previous frame by their pixel calibration.

The intention of choosing the frame difference method as revealed from other methods is depicted in table 1.

Table 1 Merits and de-merits of motion detection methods

Methods	Advantages	Disadvantages
Frame Difference[25]	Low Computing Complexity and the processing speed is high	The human detection is more sensible but rarely affected by background luminance.
Background Subtraction[26]	Detects the human without background with self-adaptive updates.	The scene cannot be justified without background
Optical flow[27]	Motion features are highly concentrated to obtain a statistical measure	Low processing speed with high complexity

The humans are classified by the feature extraction method, while their motion and the redundancy in frames can be reduced by the frame difference method. To the extent we reach to the objective of the proposed work of declaring the best keyframes by the frame difference method.

The Canny edge detection is a method used to detect the edges with noise suppressed. Hence the canny edge detection along with frame difference method supports to provide an efficient keyframes. Thus a comparative metrics results for human detected frames by HOG and HAAR produce a distinguished result when compared to keyframes without detecting Humans.

IV. RESULTS AND DISCUSSION

The proposed work is implemented in Open CV image processing using python. Here, the resulted keyframes from the CCTV footage datasets are depicted in table 2. The performance measures of the obtained keyframes are processed for metrics evaluation through compression ratio, precision and accuracy.

$$\textbf{Compression ratio}=\left\{1-\frac{N_{spk}}{N_f}\right\}*100 \quad (1)$$

reports the compactness of the obtained keyframe.

$$\textbf{Precision }=\frac{N_k}{N_{spk}}*100 \quad (2)$$

reports the number of keyframes retrieved

$$\textbf{Recall}=\frac{N_{af}}{N_{af}+(N_{spk}-N_k)}*100 \quad (3)$$

reports the relevant keyframes from the obtained frames.

Here, Nkis the total keyframe without detecting human, Nf is the total human detected frames through HOG/HAAR, Nspk is the total keyframe obtained by HOG/Haar.

Table 2.and Fig 6. illustrates the results obtained from HOG-SVM, where gradients highlighted the humans as rectangular bounding box for the complete footages. Still there exists a lack of detecting some of the humans in crowded areas. This can be rectified by non max suppression techniques or by some other techniques but the results proved better for the summarization of video keyframes.

Table 2. Comparison of With and Without Human Detected Keyframes

Video	Frames obtained without human detection	Keyframes without human detection	Human detected Frames through HOG-SVM	Keyframes through HOG-SVM	Human detected Frames through HAAR classifier	Keyframes through Haar-like classifier
CCTV1	520	7	1006	18	353	7
CCTV2	579	5	1369	25	502	6
CCTV3	499	7	672	14	498	7
CCTV4	629	8	2497	48	1077	18
CCTV5	677	13	895	33	720	25

Fig. 6. Human detected HOG-SVM Keyframes

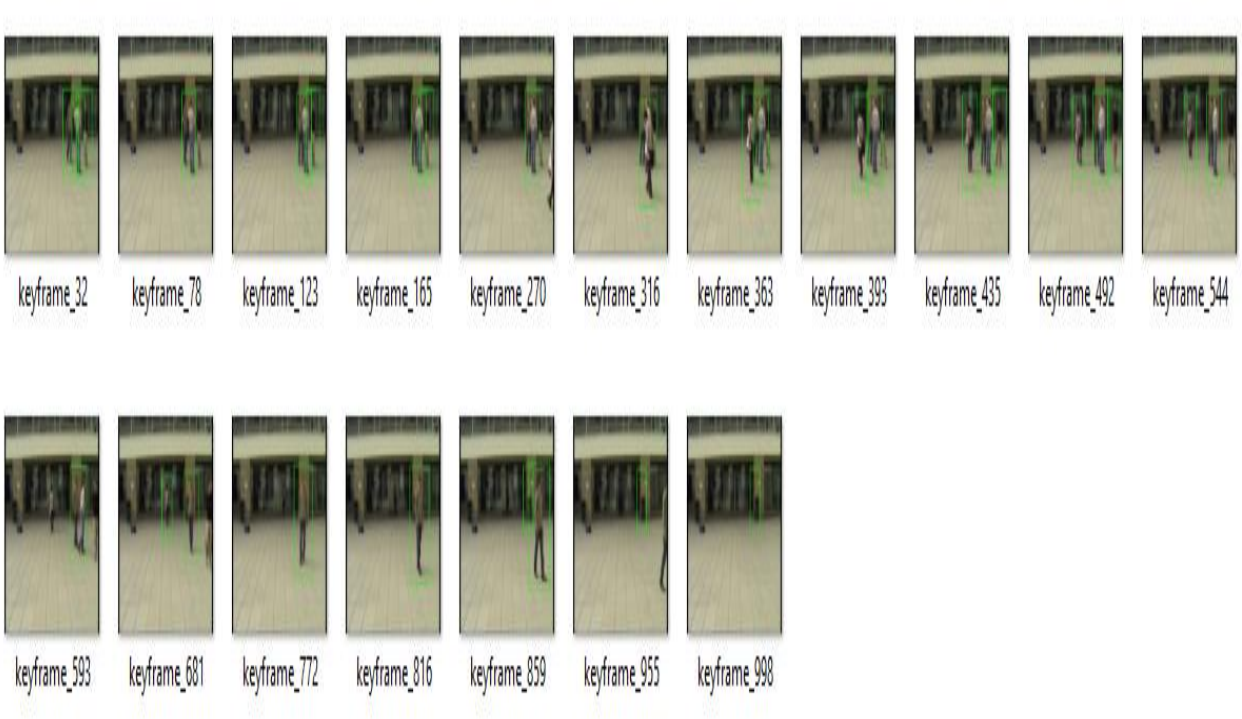


Table 3. Spatio Result – Human Detection – HOG Key Frame Extraction

VIDEO	FRAMES OBTAINED	KEYFRAMES	PRECISSION	RECALL	CR
CCTV1	1006	18	38.89	97.93	98.21
CCTV2	1369	25	20.00	96.67	98.18

CCTV3	672	14	50.00	100.31	98.34
CCTV4	2497	48	16.67	94.03	98.08
CCTV5	895	33	22.02	96.54	98.24

Table 2 and Fig 7.illustrates the results obtained using Haar-like classified detector. It proves the accuracy level of detecting the humans perfect at different scales. However, it lacks because of the intensity range of retrieving the frames when compared to HOG. Even a small change is oriented in HOG which is not enough in Haar. The overall summarization produced by the Haar-like classifier fails to report in keyframes obtained.

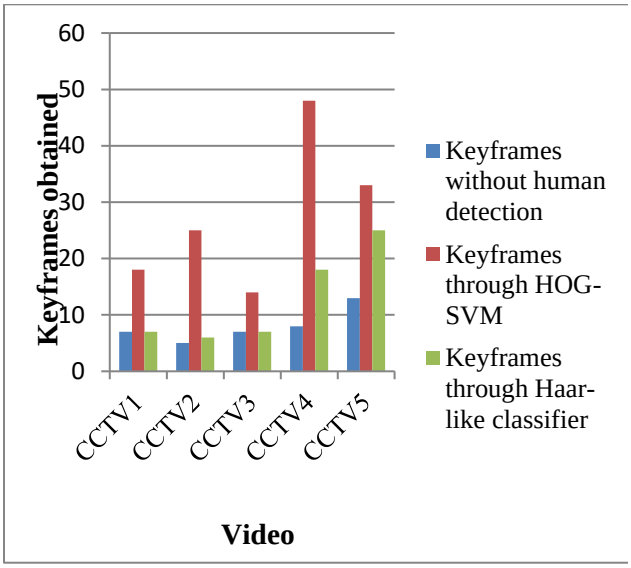
Fig. 7. Human detected Haar-like Keyframes

Table 4. Spatio Result – Human Detection – Haar Key Frame Extraction

VIDEO	FRAMES OBTAINED	KEYFRAMES	PRECISSION	RECALL	CR
CCTV1	353	7	100	100	98.02
CCTV2	502	6	83.33	99.83	98.16
CCTV3	498	7	99.91	99.99	98.06
CCTV4	1077	18	44.44	98.44	98.33
CCTV5	720	25	84.32	95.45	98.03

The comparative study on the proposed work by the frame-differencing method along with the canny edge detection method is represented by graphical representation work as derived in fig. 8 from table 2. The spatio-temporal feature detection method metrics results that HOG-SVM with 98.21% in contrast to HAAR-Like Cascade Classifier metrics results with 98.12%. Hence it is proven that HOG-SVM achieves with greater accuracy as from table 3 and table 4.

Fig 8. Edge Detected keyframes



V. RESULT

Despite the considerable progression on research in video retrieval, CBVR has a little impact on CCTV surveillance for keyframe extraction. The contrast study on Human detection using the spatio-temporal feature extraction methods between HOG and HAAR-like classifier of CCTV surveillance data runs to reach its excellence. Thus the HOG and Haar-like feature extraction supports a lot to determine the humans from a huge stored footages. In human detection process of retrieving the keyframes used by HOG-SVM proves the best, because HOG-SVM approach reports with the increased level of frames of pointing the keyframes with best summarized report when compared to HAAR-like classifier, drops the relevant frames needed for the summarization. Thus the proposed work of detecting the Humans using HOG reports with the necessitate keyframes. The future work of this is to analyze the video footages for human detection method is to be fine-tuned in all circumstances.

References

[1]Bouwmans, Thierry; Porikli, Fatih; Höferlin, Benjamin; Vacavant, Antoine, "Background Modeling and Foreground Detection for Video Surveillance Traditional Approaches in Background Modeling for Static Cameras" , 10.1201/b17223(), 1-1–1-54. doi:10.1201/b17223-3, Jul2014.

[2]Mahalingam T., Subramoniam M., "A robust single and multiple moving object detection, tracking and classification", Appl. Comput. Inf., 2018, to appear, doi: https://doi.org/10.1016/j.aci.2018.01.001

[3]Stauffer C., Grimson W.E., "Adaptive background mixture models for real-time tracking". IEEE Computer Society Conf. on Computer Vision and Pattern Recog., Fort Collins, CO, USA, June 1999, vol. 2, pp. 246–252.

[4]Khare M., Srivastava R.K., Khare A.,"Single change detection based Moving Object Segmentation by using Daubechies complex wavelet Transform", IET Image Proc., 2014, 8, (6), pp. 334–344.

[5]Kurnianggoro L., Shahbaz A., Jo K.H.: 'Dense optical flow in stabilized scenes for moving object detection from a moving camera'. 16th Int. Conf. on Control, Automation and Systems (ICCAS), Gyeongju, Republic of Korea, October 2016, pp. 704–708

[6]Kurnianggoro L, Shahbaz A, Jo K H. "Dense optical flow in stabilized scenes for moving object detection from a moving camera[C]", International Conference on Control, Automation and Systems. IEEE, 2016:704-708.

[7]Li X, Xu C." Moving object detection in dynamic scenes based on optical flow and superpixels", [C] IEEE International Conference on Robotics and Biomimetics. IEEE, 2016:84-89.

[8]H. Sabirin and M. Kim, ”Moving object detection and tracking using a spatio-temporal graph in H.264/AVC Bitstreams for Video Surveillance”, IEEE Trans. Multimedia, vol. 14, no. 3, pp. 657 - 668, Jun. 2012.

[9]Y. Monma, L. S. Silva, and J. Scharcanski, ”A fast algorithm for tracking moving objects Based on spatio-temporal video segmentation and cluster ensembles”, In Proc. IEEE Int. Instru. andMeasur. Tech. Conf. (I2MTC), May 2015.

[10] Xinchen Yan, Junsong Yuan, and Hui Liang, “Efficient Online Spatio-Temporal Filtering for Video Event Detection”, Springer International Publishing Switzerland , pp. 769–785, 2015

[11] Kumar S. Ray, Vijayan K. Asari, Sr., and Soma Chakraborty, “Object Detection by Spatio-Temporal Analysis and Tracking of the Detected Objects in a Video with Variable Background”, Journal of Visual Communication and Image Representation, April 2017.

[12] P.Viola, M.Jones and D.Snow, “Detecting Pedestrians Using Patterns of Motion and Appearance”, IJCV, Vol.63, No.2, pp.13-161, 2005.

[13] N.Dalal and B.Triggs, “Histograms of Oriented Gradients for Human Detection”, CVPR, pp.886-893, 2005

[14] Y. Luo, H. Zhou, Q. Tan, X. Chen and M. Yun, “Key frame extraction of surveillance video based onmoving object detection and image similarity,” Pattern Recognition and Image Analysis, vol. 28, no. 2,pp. 225–231, 2018.

[15] Stone, E.E, and Skubic, M, (2015)“Fall Detection in Homes of Older Adults using the Microsoft Kinect”, IEEE, Vol. 19, No. 1, pp. 290-301.,

[16] Cheng Z, Qin L, Huang Q, Yan S, and Tian Q,(2014) “Recognizing Human Group Action by Layered Model with Multiple Cues”, Elsevier, Vol. 136, pp. 124-135.

[17] Xia, L., Chen, C.-C., Aggarwal, J.K., “Human detection using depth information by kinect”. In: IEEE Computer Vision and Pattern Recognition (CVPR) Workshops, pp. 15–22 (2011)

[18] Choi, W., Pantofaru, C., Savarese, S, “A general framework for tracking multiple people from a moving camera”. In: IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), vol. 35, no. 7, pp. 1577–1591 (2013)

[19] Vapnik, V. “Statistical Learning Theory”. Wiley- Interscience, New York, (1998).

[20] M. Kachouane, S. Sahki, M. Lakrouf, N. Ouadah “HOG based fast Human Detection”, 2012 24th International Conference on Microelectronics (ICM).

[21] Jones, V, “Rapid object detection using a boosted cascade of simple features”. In Proceedings Computer Vision and Pattern Recognition (CVPR) (2001)

[22] Landesa-Vázquez I, Alba-Castro J, “The role of polarity in Haar-like features for face detection”, in Proceedings of International Conference on Pattern Recognition, 2010, pp. 412–415

[23] Li Y, Lu W, Wang S, Ding X,” Local Haar-like features in edge maps for pedestrian detection”, in Proceedings of International Congress Image and Signal Processing, vol. 3, 2011, pp. 1424–1427

[24] Budiman, R.A.M., Achmad, B., Faridah, Arif, A., Nopriadi, Zharif, L., “Localization of white blood cell images

using Haar Cascade Classifiers", in Proceedings of 2016 1st International Conference onon Biomedical Engineering: Empowering Biomedical Technology for Better Future (IBIOMED 2016), 2016, pp. 1-5

[25] H. Wang, R. Lu, X. Wu, L. Zhang, and J. Shen." Pedestrian detection and tracking algorithm design in transportation video monitoring system". In Information Technology and Computer Science, International Conference on, volume 2, pp 53–56, July 2009.

[26] Z. Jiang, D.Q. Huynh, W. Moran, and S. Challa. "Combining background subtraction and temporal persistency in pedestrian detection from static videos". In Image Processing, 20th IEEE International Conference on, pages 4141–4145, Sept 2013.

[27] Q. Liu, O. Osechas, and J. Rife, ".Optical flow measurement of human walking. In Position Location and Navigation Symposium", IEEE/ION, pages 547–554, April 2012