International Journal of Engineering & Technology, 7 (3) (2018) 1000-1004

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET
doi: 10.14419/ijet.v7i3.11773
Research paper

To detect abnormal event at ATM system by using image processing based on IOT technologies

Kande Archana [1] *, P Bhaskara Reddy [2]

[1] *Asst. Professor MLR Institute of Technology, Hyd*
[2] *Director HITS Hyd*
**Corresponding author E-mail: kande.archana@gmail.com*

Abstract

Now a day' s ATMs are equipped with money there is possibility of robberies. This paper proposes a framework which will provide high security in ATMs. The Prototype includes ARM controller, Vibration Sensor, GSM and GPS Technique, DC Motor, Stepper motor, Buzz-er, LCD Display, and Keil Tool. Whenever robbery occurs, Vibration sensor is used here which senses vibration produced from ATM machine. This system uses ARM controller based embedded system to process real time data collected using the vibration sensor. Once the vibration is sensed the beep sound will occur from the buzzer. DC Motor is used for closing the door of ATM.
Stepper motor is used to leak the gas inside the ATM to bring the thief into unconscious stage. Camera is always in processing and sending video continuous to the PC and it will be saved in computer. RTC used to capture the robber occur time and sends the SMS and MMS to the nearby police station and corresponding bank through the GSM and GPS. Here LCD displays board using showing the output of the message continuously. This will prevent the robbery and the person involving in robbery can be easily caught. Here, Keil tools are used to implement the idea and results are obtained. keil tools is used for run the DC motor and stepper motor for automatic door lock and also leak the gas inside the ATM. By this system robberies will be stopped and the complaints cases also reduced maximally. Thus the proposed framework results are revealed that the framework can be providing high security to the ATM System's

Keywords: *Use GSM and GPS; LCD; DC and Stepper Motor; Vibration sensor; ARM controller.*

1. Introduction

ATM is a computerized telecommunication device that serves the customer of a financial firmwith a swift access to financial transactions in a public space by exempting the need for a clerk or bank teller. The numbers of ATM installations are increasing dramatically to support the transactions in billions. Increase in nefarious activities like robbery, murder, and other crimes have raised an urgency to install an effective system that can protect people as well as ATM installations [1], [2]. Generally ATM installations are equipped with CCTV cameras that keep a watch on the activities. Unfortunately, CCTV is not sufficient to provide security due to their inability to recognize unusual behaviors themselves [3] and hence monitoring authority needs to monitor these feeds 24 × 7 which is a challenging task. Today, we need an advanced system that can effectively monitor and automatically recognize unusual crime activities in an ATM room and can also report to the nearest monitoring firm before an offender could elope. Another approach to handle this situation could be an alarm system or electrical buzzer. Each ATM premise can be equipped with an electric buzzer. ATMuser can press this buzzer to send signal to response group if any abnormal event takes place. Alarm systems may become ineffective as individual alarm must be responded by a main alarm response group,which shouldfirst examine the type or nature of the event being alarmed before any help signal can be requested. In addition,most alarms require a noticeable effort to operate, presenting an uncertainty that the perpetrator can simply physically stop the victim from triggering the alarm or may take a belligerent action against the victim if the victim is seen to initiate an alarm signal. Absence of automated securitymechanism leads to postincident forensic analysis by the law agencies. Many a time law enforcement authorities become aware of the crime after several hours after the incident.This is a major problem in the urban areas as well as in the rural areas. Recently, a grue some attack on a woman at an ATM located in Bangalore city, India [4], has brought to focus the issue of security at such kiosks (Figure 1(a)). This incident has sent shock waves across the country and highlighted the need to tackle such brutal acts. In some cases, ATM guard is also killed when he tries to save the victim.

Fig. 1: Attack on Person at Various ATM Installations.

Because attackers are generally equipped with weapons like machete, guns, pistols, iron and rod and usually are multiple in numbers. Figure 1(b) [5] depicts the typical scene of a guard tied down with a rope by two attackers in Bangalorecity, India, and Figure 1(c) [6] depicts the attack on a man at Karachi city, Pakistan. Therefore, it is necessary to have an automated system that can proactively identify and generate alarm on unusual behavior.

Video based human activity analysis has gained lots of attention amongst the researchers.The goal of human activity recognition is to analyze different activities automatically from an unknown video [7]. Analysis of various activities involves recognition ofmotion pattern and generation of high level description of actions. There are various approaches like manifold approaches, spatiotemporal interest of feature points, motion history images, accumulated motion image, and bag of words model which are recently used by many researchers for effective human action recognition and representation [8– 12]. In this paper, we present a system that can amend the current trends of the surveillance system. The system can automatically recognize different actions or number of persons through a CCTV camera like single normal, multiple normal, and multiple abnormal and generate signal accordingly. Using our system, the offender is more likely to be caught by the police red-handed because they are informed about the crime instantly. In addition, the proposed system can be used to generate automated alarm that can alert security guard deputed at the ATM location as well as other people around the premise to obtain immediate security. The paper is organized as follows. Section 2 presents related works and background of this work. In Section 3, we present our proposed method. Section 4 depicts the results and analysis of the proposed approach. Finally, conclusions are drawn in Section 5.

2. Literature review

The intricacy at ATM booth described by COPS [1, 2] is the main motivation behind this research which has inspired us to develop an effective security system. In this section, we present the related work and research undergone in developing videobased security systemthat helpedus tomake anefficient surveillance system.Various approaches have been proposed by researchers for human action recognition (HAR). Davis and Bobick [13] in their paper have presented the usage of temporal templates for recognizing human actions. References [7, 14] have presented a detailed survey on human motion and behavior analysis using MHI and its variants. Other approaches like Optical Flow and Random Sample Consensus (RANSAC) by [8] decipher the representation and recognition technique of human actions. For feature extraction, Hu has proposed a novel theory popularly known

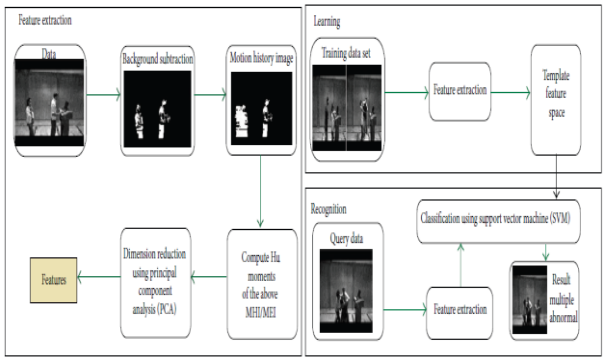
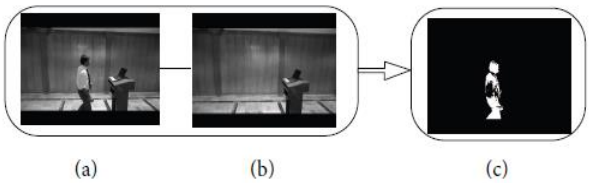


Fig. 2: To recognize abnormal image by using MHI and Hu methods

As Hu moments which are invariant to translation, scale, and rotation [15]. Bobick andDavis [16] in their paper have shown the usage of Hu moments for feature extraction from temporal templates. Humoments are widely used shape descriptors due to its simplicity and less computational approach [17– 19]. Various other descriptors like Fourier descriptors (FD) and Zernike moments have also been proposed. Fourier descriptors prove to be a disadvantage when the image size varies because the number of points also varies and the method becomes computationally high to work at real time. Zernike moments are advanced version of Hu moments whose magnitude is invariant to rotation but their computation time is
also extensive towork at real time[20]. Besides the availability of various methods for feature extraction, we have used the conventional Hu moments method for shape description of theMHI/MEI. It is becauseHu moments are computationally effective as compared to other descriptors. To make machine
learn these features, a classifier has to be used. There are varieties of classifiers available like support vector machine (SVM), neural networks (NN), and Bayesian classifier. Debard et al. [21] have presented the identification of abnormal event, that is, fall using SVM. References [22– 25] have shown a great adaptive learning of support vector machine in video surveillance. SVM, apart from its learning from two classes [16], has shown multiclass classification through SVM which helped us to analyze multiple classes through it. Sometimes it does happen that redundancy in data comes inherently from the video. For instance MHI/MEI formed by the presence of two persons in a video is also formed by the presence of an obese person. This kind of data may reduce the learning accuracy of SVM. So to address this kind of problem, principal component analysis (PCA) has been used. Reference [27] has shown the use of PCAwith SVM in the work in action recognition in video. Another great work from [28, 29] has illustrated the use of PCA in dimension reduction. The main motive of this paper is to build a strong security framework system which can work at real time environment at ATM booth or other similar premises.

3. The proposed methodology

The proposed methodology/system (Figure 2) uses computer vision techniques for recognition of normal and abnormal behavior of a person.The systemconsists of a structure where objects are moving with respect to a fixed background and each frame of video is processed as follows. First, foreground extraction technique is used to obtain clear silhouette of people. Then a fixed size window is used to record the MHI.The MHI is used to generate pattern of a person under different situations. To describe this pattern, Hu moments are used. These dimensions are further reduced by applying principal components analysis (PCA) to remove redundancies and make the system computationally effective. Further we make use of support vectormachine to predict themost likely class and the result is displayed



(a) (b) (c)

Fig. 3: a) Current Image; b) Fixed Background Image; c) Subtracted Image (Binary).

4. Feature extraction

5. Background subtraction and motion history image (MHI)

Background subtraction is used to xtract the foreground objects from video sequence. Generally, ATMs are installed in a closed enclosure where background does not change over the time. We make use of the frame without any moving object as a background frame. Subsequent frames are subtracted from this frame to obtain moving objects. Figures 3(a) and 3(b) represent the two images to be subtracted and Figure 3(c) shows the output binary image received after its conversion from grayscale to binary. After obtaining the foreground objects, we compute the MHI.Motion templates are an efficient way to record general movement or motion and are suitable for human activity recognition [16] and gesture recognition [30]. The MHI is a binary image where pixel intensity is a function of the recency of motion in a video sequence.The pixel intensity is linearly ramping value as a function of time, where brighter (more whiter) values represent the more recent motion locations. As an object moves it leaves behind a motion history of its movements. With the passage of time the old motion histories of object are eliminated to capture the new motion patterns so that old patterns do not get mixed with the new one.MHI at any given time t is given as

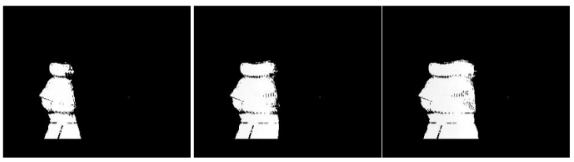
$$M(x,y,t) = \begin{cases} \mathrm{TAU}_{\mathrm{MAX}}, & \text{if } D(x,y,t) \neq 0 \\ M(x,y,t-1) - 1, & \text{else if} \\ & M(x,y,t-1) - 1 > \mathrm{TAU}_{\mathrm{MIN}}, \\ 0, & \text{otherwise,} \end{cases}$$

where $D(x,y,t)$ is intensity of pixel (x,y) at time t of diffImage. $\mathrm{TAU}_{\mathrm{MAX}}$ is a constant representing brighter pixel value. $\mathrm{TAU}_{\mathrm{MIN}}$ is a constant representing less bright pixel value. Consider N (window size) = $\mathrm{TAU}_{\mathrm{MAX}} - \mathrm{TAU}_{\mathrm{MIN}}$. (1)

Original mhi Image starts out as blank image or pixel with all zeroes. Algorithm of feature extraction from MHI using Hu moments is shown in Algorithm 2.

6. Hu moments

Once MHI is obtained, features need to be extracted from it. We have used Hu moments for this purpose. The Hu moments [15], obtained from the templates are known to yield reasonable shape discrimination in a translation and scale invariant manner. Humoments provide seven values as an extracted feature from a given image.These moments are invariant to translation, scale, and rotation of an image. Out of seven invariants, six are absolute orthogonal invariants and the seventh one is skew orthogonal invariant. Hu moments are computed as follows:



(a) (b) (c)

Fig. 4: a) MHI of 5 Fames of Single Walking Person; b) MHI of 10 Fames of Single Walking Person; c) MHI of 15 Fames of Single Walking Person.

$$\begin{aligned} \mathrm{hu5} &= (n_{30} - 3n_{12})(n_{30} + n_{12}) \\ &\cdot \left[(n_{30} + n_{12})^2 - 3(n_{21} + n_{03})^2\right] \\ &+ (3n_{21} - n_{03})(n_{21} + n_{03}) \\ &\cdot \left[3(n_{30} + n_{12})^2 - (n_{21} + n_{03})^2\right], \end{aligned} \quad (6)$$

$$\begin{aligned} \mathrm{hu6} &= (n_{20} - n_{02})\left[(n_{30} + n_{12})^2 - (n_{21} + n_{03})^2\right] \\ &+ 4n_{11}(n_{30} + n_{12})(n_{21} + n_{03}), \\ \mathrm{hu7} &= (3n_{21} - n_{03})(n_{30} + n_{12}) \end{aligned} \quad (7)$$

$$\begin{aligned} &\cdot \left[(n_{30} + n_{12})^2 - 3(n_{21} + n_{03})^2\right] \\ &- (n_{30} - 3n_{12})(n_{21} + n_{03}) \\ &\cdot \left[3(n_{30} + n_{12})^2 - (n_{21} + n_{03})^2\right], \end{aligned} \quad (8)$$

$$m_{rs} = \iint_{-\infty}^{\infty} (x^r y^s) f(x,y)\, dx\, dy, \quad r,s = 0,1,2,\ldots,$$

$$u_{rs} = \iint_{-\infty}^{\infty} (x-\bar{a})^r (y-\bar{o})^s f(x,y)\, dx\, dy, \quad (9)$$

$r, s = 0, 1, 2, \ldots$ where

$$\bar{a} = \frac{m_{10}}{m_{00}}, \quad \bar{o} = \frac{m_{01}}{m_{00}}, \quad (10)$$

$$n_{rs} = \frac{u_{rs}}{u_{00}^{\gamma}} \quad \gamma = \frac{(r+s+2)}{2}, \; r+s = 2,3,\ldots, \quad (11)$$

Where$m\ rs$ are the two-dimensional $(r+s)$th ordermoments of the image function $f(x, y)$. (a, o) are the centroidof the image $f(x, y)$. urs are the central moments of the image $f(x, y)$. $n\ rs$ are the normalized central moments.Since HI can effectively record a motion or an activity that occurred in a small time interval t as shown in Figure 4 and Hu moments can uniquely describe an image by generating a set of seven values, we have used MHI for recording activity and Hu moments for the purpose of describing that activity; Figure 9 and Table 3 support our above statement.The MHI algorithm presented in Section 3.1.2 by [13] has been applied to generate MHI on our training and testing data set followed by computation of Hu moments using Algorithm 2 based on (2)– (8). The eight attributes, seven Hu moments along with an area, are used to describe an image pattern. These eight attributes are fed to principal component analysis for further reduction in the attribute set. Principal component analysis is an efficient technique for reducing high-dimensional data, by computing dependencies between the attributes to represent it in a more tractable, lower-dimensional form, without losing much information. WEKA [23] is used for applying PCA.

Table 1: Training Data Set.

Number of MHI frames	5	10	15
Single	961	478	318
Multiple	961	479	320
Multiple abnormal	960	478	315
Total	**2882**	**1435**	**953**

Table 2: Testing Data Set.

Number of MHI frames	5	10	15
Single	270	135	89
Multiple	110	55	36
Multiple abnormal	340	169	113
Total	**720**	**359**	**238**

7. Action classification using support vector machine

Normal task of machine learning is to learn from a, usually very large, space of data to classify the one that will best fit the data based on prior knowledge. SVM is a machine learning tool and a widely used classifier in computer vision, bioinformatics, and so forth, due to its ability and high accuracy to deal with the high dimensions of data. Support vector machine is a popular machine learning method for classification, regression, and other learning tasks. LibSVM [14] is a library for support vector machines used by us. A typical use of LibSVM involves two steps: first, training a data set to obtain a model file and second, using themodel file to predict information of a testing data set. RBF Kernel is used for both training and testing purpose.

Table 3: Distinct Values of Hu Moments under Different Activities in Figure 9.

Class	Value 1	Value 2	Value 3	Value 4	Value 5	Value 6	Value 7
Single	5.54E – 1	3.74E – 2	2.64E – 1	1.12E – 1	1.90E – 2	2.16E – 2	2.36E – 3
Multiple	2.44E – 1	3.06E – 5	3.36E – 3	1.28E – 3	2.52E – 6	1.76E – 6	–8.57E – 7
Multiple abnormal	2.73E – 1	2.91E – 3	1.63E – 3	1.81E – 4	7.55E – 8	3.55E – 6	–6.29E – 8

Table 4: TP, TN, FP, and FN of All Classes of Our Testing Data Set.

Number of MHI frames	5				10				15			
	TP	TN	FP	FN	TP	TN	FP	FN	TP	TN	FP	FN
Single	259	440	10	11	126	219	5	9	71	149	0	18
Multiple	54	593	17	56	47	293	11	8	17	190	12	19
Multiple abnormal	324	324	56	16	163	183	7	6	111	98	27	2

Table 5: Comparison of the Recognition Accuracy, Recall, and Precision for Various MHI Frames on our Testing Data Set

Number of MHI frames	5			10			15		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall	Accuracy	Precision	Recall
Single	97.08%	96.30%	95.90%	96.10%	96.20%	93.30%	86.97%	100.00%	79.80%
Multiple	89.86%	76.10%	49.10%	94.71%	81.00%	85.50%	92.44%	58.60%	47.20%
Multiple abnormal	90.00%	85.30%	95.30%	96.38%	95.90%	96.40%	87.82%	80.40%	98.20%

8. Experimental results and analysis

The system has been trained and tested using java [14] and opencv [10], [23] on a computer having Intelcore i3, 2.13GHz processor with 2 GB RAM on a video of 320 × 240 resolution for different number of MHI frames. The system was tested against three classes, single normal, multiple normal, and multiple abnormal, over six videos (2 single of 10 seconds each, 2 multiple normal of 27 seconds each, and 2 multiple abnormal of 29 seconds each). We have made our own data set for training and testing purpose by taking seven actors (five boys and two girls) for frame size 320 × 240 and 25 fps frame rate. The system is trained using these videos for different number ofMHI frames (Table 1). Testing is done on a different video from the one used for training purpose.The system was tested for different number of MHI frames (5, 10, and 15) (Table 2). Consider

$$\text{accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})},$$
$$\text{precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})},$$
$$\text{recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})}. \quad (12)$$

TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative values, respectively. Table 4 shows the values of TP, TN, FP, and FN for three different classes on different window size of MHI used on our testing data set. Table 5 gives the accuracy, precision, and recall of three classes calculated using (12), by LibSVM over our testing data set. The comparison is made among
three different numbers of MHI frames taken as 5, 10, and 15. Advantage of MHI representation is that a range of time (in terms of frames) is encoded in a single frame. Selection of number of frames to formMHI is very important because variation in number of frames may provide different information regarding event. Hence for effective recognition system it becomes necessary to identify suitable window size (number of frames) for MHI. The system is tested over a video of 1 minute and 46 seconds typically consisting of all three classes. Total number (all classes) of sampleMHIvalues for testing in three different MHI frames (5, 10, and 15) is 720, 359, and 238, respectively. We have observed that window size ten is most appropriate for recognizing abnormal events. Apart from this testing, a 10-fold cross-validation of [6] data set has been done to support the correctness of the proposed methodology.The results are represented in Table 7. Color code for prediction results in colored images: one person (normal working), green color; multiple persons (normal working), blue color; multiple persons (abnormal working), red color. MHI and corresponding prediction results are shown in Figures 5, 6, and 7 for different window size (5, 10, and 15). Table 6 shows the value of AUC for different classes andMHI frames. Figure 8 shows the corresponding ROC curve on the testing data set.

9. Conclusion

In this paper, we have presented a system for security framework at ATM that can also be used in similar premises. In particular, this paper presents the recognition of normal and abnormal events at the ATM. The need of developing such security systemis the increasing number of crime rates at the ATM booth and also the lack of prevailing video surveillance system in themarket.Thesystem accuracy differs for different MHI frames. In our case, the overall prevision accuracy.

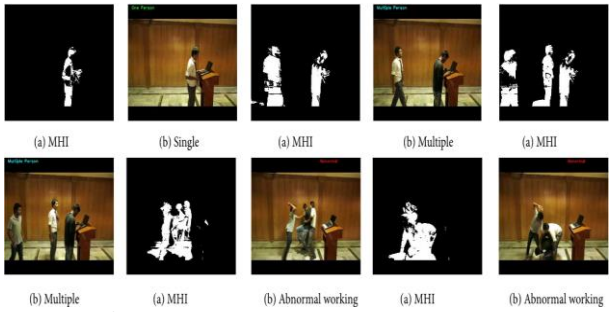


Fig. 5: MHI of 5 Frames (Black and White (A)) and Prediction by Libsvm (Colored (B)) for Each Class (Left to Right).

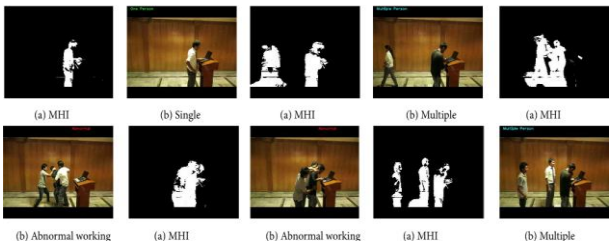


Fig. 6: MHI of 10 Frames (Black and White (A)) and Prediction by Libsvm (Colored (B)) for Each Class (Left to Right).

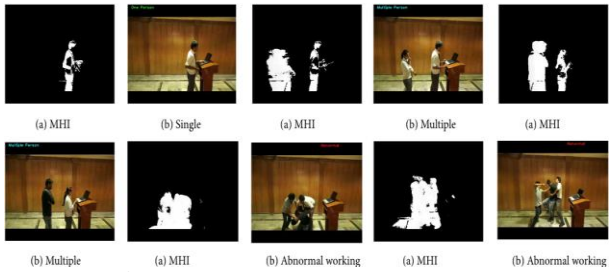


Fig. 7: MHI of 15 Frames (Black and White (A)) and Prediction by Libsvm (Colored (B)) for Each Class (Left to Right).

Was 92.31% for 5MHI frames, 95.73% for 10MHI frames, and 89.07% for 15MHI frames on our testing data set using LibSVM. The main reason of low accuracy of system in 5MHI frames is due to the fact that a very few number of frames (5) contribute to the formation of MHI where only a small part of an activity pattern is recorded in an image thus affecting the recognition rate of that activity, whereas in case of 15MHI frames as a large number of frames (15) contribute to the formation of MHI, it is more likely that the previous activity pattern is hindered by subsequent activity pattern when the motion is frequent thus causing a formation of distorted pattern resulting in low recognition rate. The above stated problems were less in case of 10MHI frames; hence the accuracy rate was better. Our system' s overall accuracy would have been higher if we had removed the transition frames between normal and abnormal activities. Since at real time this could not be eliminated hence we have included this scenario. The future scope of this paper is wide open in research aspect. Various other feature extraction methods can be applied to test the accuracy of the system. Also, other classifiers like SVM can be used for the same purpose. Since our system is restricted to work for video only, our future aspect will be to focus on audio based recognition also.

References

[1] M. S. Scott, *Robbery at Automated Teller Machines*, US Department of Justice, Office of Community Oriented Policing Services, 2001.

[2] N. Sharma, " Analysis of different vulnerabilities in auto teller machine transactions," *Journal of Global Research in Computer Science*, pp. 38– 40, 2012.

[3] R. S. Shirbhate, N. D. Mishra, and R. P. Pande, " Video surveillance system using motion detection: a survey," *International Journal Advanced Networking and Applications*, vol. 3, no. 5, pp. 19– 22, 2012.

[4] IBN LIVE, 2014, ttp://ibnlive.in.com/news/bangalore-atmattack-womans-skull-fractured-still-in-icu/435190-62-129.html.

[5] NDTV, 2014, http://www.ndtv.com/article/cities/bangalore-thisbrave- atm-guard-grabs-machete-from-robbers-hits-one-ofthem-465052?curl=1422275786.

[6] MAN ATTACKED, January 2014, http://www.youtube.com/ watch?v=RGaupYx2fpQ.

[7] R. Poppe, " Asurvey on vision-based human action recognition," *Image and Vision Computing*, vol. 28, no. 6, pp. 976– 990, 2010. https://doi.org/10.1016/j.imavis.2009.11.014.

[8] U. Mahbub,H. Imtiaz, andM. A. R. Ahad, " Action recognition based on statistical analysis fromclustered flow vectors," *Signal, Image and Video Processing*, vol. 8, no. 2, pp. 243– 253, 2014. https://doi.org/10.1007/s11760-013-0533-3.

[9] W. Gong, J. Gonz`alez, and F. X. Roca, " Human action recognition based on estimated weak poses," *EURASIP Journal on Advances in Signal Processing*, vol. 2012, article 162, 2012.

[10] M. Paul, S. M. Haque, and S. Chakraborty, " Human detection in surveillance videos and its applications—a review," *EURASIP Journal on Advances in Signal Processing*, vol. 2013, article 176, 2013.

[11] W. Kim, J. Lee, M. Kim, D. Oh, and C. Kim, " Human action recognition using ordinal measure of accumulated motion," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, Article ID 219190, 2010. https://doi.org/10.1155/2010/219190.

[12] M. Ahmad and S.-W. Lee, " Human action recognition using shape and CLG-motion flow from multi-view image sequences," *Pattern Recognition*, vol. 41, no. 7, pp. 2237– 2252, 2008. https://doi.org/10.1016/j.patcog.2007.12.008.

[13] J. W. Davis and A. F. Bobick, " The representation and recognition of human movement using temporal templates," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 928– 934, June 1997. https://doi.org/10.1109/CVPR.1997.609439.

[14] M. A. R. Ahad, J. K. Tan, H. Kim, and S. Ishikawa, " Motion history image: its variants and applications," *Machine Vision and Applications*, vol. 23, no. 2, pp. 255– 281, 2012. https://doi.org/10.1007/s00138-010-0298-4.

[15] M.-K. Hu, " Visual pattern recognition by moment invariants," *IRE Transactions on Information Theory*, vol. 8, no. 2, pp. 179– 187, 1962. https://doi.org/10.1109/TIT.1962.1057692.

[16] A. F. Bobick and J. W. Davis, " The recognition of human movement using temporal templates," *IEEE Transactions on Pattern Analysis andMachine Intelligence*, vol. 23,no. 3, pp. 257– 267, 2001.

[17] A. Amato and V. D. Lecce, " Semantic classification of human behaviors in video surveillance systems," *WSEAS Transactions on Computers*, vol. 10, no. 10, pp. 343– 352, 2011.

[18] Q. Chen, R. Wu, Y. Ni, R. Huan, and Z. Wang, " Research on human abnormal behavior detection and recognition in intelligent video surveillance," *Journal of Computational Information Systems*, vol. 9, no. 1, pp. 289– 296, 2013.

[19] P. Srestasathiern and A. Yilmaz, " Planar shape representation and matching under projective transformation," *Computer Vision and Image Understanding*, vol. 115, no. 11, pp. 1525– 1535, 2011. https://doi.org/10.1016/j.cviu.2011.07.004.

[20] S. Bourennane and C. Fossati, " Comparison of shape descriptors for hand posture recognition in video," *Signal, Image and Video Processing*, vol. 6, no. 1, pp. 147– 157, 2012. https://doi.org/10.1007/s11760-010-0176-6.

[21] G. Debard, P. Karsmakers, M. Deschodt et al., " Camera based fall detection using multiple features validated with real life video," in *Intelligent Environments Workshops*, vol. 10, pp. 441– 450.

[22] E. B. Nievas, O. D. Suarez, G. B. Garcia, and R. Sukthankar, " Violence detection in video using computer vision techniques," in *Computer Analysis of Images and Patterns*, vol. 6855, pp. 332– 339, Springer, Berlin, Germany, 2011. https://doi.org/10.1007/978-3-642-23678-5_39.

[23] H. Wang, A. Finn, O. Erdinc, and A. Vincitore, " Spatialtemporal structural and dynamics features for Video Fire

[24] Detection," in *Proceedings of the IEEEWorkshop on Applications of Computer Vision (WACV ' 13)*, pp. 513– 519, Tampa, Fla, USA, January 2013.

[25] S. Oh, A. Hoogs, A. Perera et al., " A large-scale benchmark dataset for event recognition in surveillance video," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR ' 11)*, pp. 3153– 3160, Providence, RI, USA, June 2011. https://doi.org/10.1109/CVPR.2011.5995586.

[26] P. Rota, N. Conci, and N. Sebe, " Real time detection of social interactions in surveillance video," in *Computer Vision—ECCV 2012. Workshops and Demonstrations*, pp. 111– 120, Springer, Berlin, Germany, 2012.

[27] D. Tosato,M. Farenzena, M. Spera, V.Murino, and M. Cristani, " Multi-class classification on Riemannian manifolds for video surveillance," in *Computer Vision—ECCV 2010*, vol. 6312 of *LectureNotes in Computer Science*, pp. 378– 391, Springer, Berlin, Germany, 2010. https://doi.org/10.1007/978-3-642-15552-9_28.