Modular Approach for Multiple Event Detection in Surveillance Videos

Imran Ahmed, Ghulam Masood, Qazi Nida-Ur-Rehman, Muhammad Nawaz, Zahoor Elahi

Abstract— In the prevailing law and order situation video surveillance have widespread applications from public places to monitoring and security. In this paper a modular approach has been proposed to detect multiple events in videos. We have divided these events into three broad categories i.e. intrusion, loitering and slip and fall. The proposed approach is divided into primary and secondary analysis. Videos for surveillance have to be passed through the primary analysis, which can be used as an input for the secondary analysis. The proposed system achieved higher accuracy (90 %) than state of the art (85 %) for the aforementioned events. The results have been verified using publically available benchmark datasets.

Index Terms— video surveillance, abnormal events, Linear Kalman filter.

I. Introduction

ENORMOUS steps have been taken to tackle with the law and order situation by the Law Enforcement Agencies especially after 9/11 terrorist attack, wherein, video surveillance can play an important role. Utilizing the technology based on computer vision can provide an additional feature to our human resources to handle such situations. By analyzing these factors, an automatic system has been proposed to identify regions or behaviors based on motion leading to some fatal events, could draw the attention of security personnel for more effective and proactive video surveillance. Video surveillance for public safety is gaining attention worldwide and needs efficient mechanisms to monitor and track the events occurring in public spaces.

Abnormal events are referred to number of ways in literature as irregular, anomaly, uncommon, suspicious, and unusual outliers. Video analytics are computer vision programs designed to monitor real time or recorded videos to analyze the events and find the abnormal events.

Manuscript received January 03, 2016; revised February 06, 2016.

Imran Ahmed, Assistant Professor, Center of Excellence in Information Technology, Institute of Management Sciences, Peshawar, KPK, Pakistan. (E-mail: imran.ahmed@imsciences.edu.pk).

Ghulam Masood is corresponding author and working as a Research Associate in the same organization (E-mail: ghulam.masood@ gmail.com).

Qazi Nida-Ur-Rehman is corresponding author and working as a Research Associate in the same organization. (E-mail: qazinida48@ gmail.com).

Muhammad Nawaz, Assistant Professor, Center of Excellence in Information Technology, Institute of Management Sciences, Peshawar, KPK, Pakistan. (Email: m.nawaz@imsciences.edu.pk).

Zahoor Elahi is co-author and working as research associate in the same organization. (Email: zahoores@gmail.com).

Nowadays sensing devices such as cameras are integrated video analytics programs to detect abnormal events. The goal of analytics solutions and smart surveillance is to prolong the capability of orthodox surveillance is to spot, and categorize objects in the view, but also to deduce the activity or event taking place.

In our work modular approach has been used to detect multiple events in videos. We have divided these events into three broad categories i.e. intrusion, loitering and slip and fall. The proposed approach is divided into primary and secondary analysis. Video for surveillance has to be passed through the primary analysis, which can be used as an input for secondary analysis. The object motion detected tracked by using Linear Kalman Filter is then passed to the secondary analysis module for a particular event.

II. Literature Review

The state-of-the-art review and survey literature can be seen in [2][6][9][14][15]. Tang et al [1] implemented the blob tracking approach for object tracking. The author divides the objects into four categories i.e. new, existing, splitting and merging. The categories are defined on the basis of certain distinct rules for each class respectively. Yang et al [3] proposed an improved multi-target tracking algorithm for multi target object tracking. To get good approximation of subsequent distribution, the authors used fuzzy clustering in combination with fitness function. Benezeth et al [4] proposed a location-based method to reveal anomalies in the events and create a activity model for them. Low-level features have been used to classify events and create an activity model. Ellis et al [5] proposed regression based tracking approach, for this purpose the Linear Predictor (LP) tracker has been used, in their work they improve the offline learning to online-LP learning. Ahmed et al [7] developed an algorithm to detect unusual movement in videos. Optical flow computes the interest points in the video. The algorithm then tends to find abrupt change in motion of these interest points. Elarbi-Boudihir et al. [8] presents the system architecture of Intelligent Video Surveillance System (IVSS), based on IP cameras and installed the system in an educational institute for security purposes. The event detection system is based on SVM which is used to create activity model. Tian Wang et al. [10] propose an algorithm for abnormal event detection using image descriptor and online nonlinear classification method; they introduce the covariance matrix of the optical flow and image intensity as a descriptor encoding moving information. Lim et al [12] proposed a framework for multiple abnormal event detection. They defined particular set of attributes for every event which made detection of those events easier. In our paper we will be using their approach to divide the events that we have considered into set of attributes and then detect them accordingly. Patino et al [13] developed an algorithm to detect loitering in videos. Their algorithm tracks moving people by depth maps. Activity zones are automatically learned considering trajectory of the moving objects. Statistical properties of zones and transition between them make it possible to detect the abnormal event. Maeng [16] detected intrusion in public spaces by transforming it into geometrical problem. The movement of a person is monitored by multiple cameras. Bounding box is drawn automatically on the moving person and upon its intersection with the specified portion intrusion alarm is raised.

III. Methodology

In this research we introduce a computational model for events like intrusion, loitering and slip and fall for public safety. In order to track multiple events for surveillance applications we have divided our method into two steps i.e. primary analysis and secondary analysis.

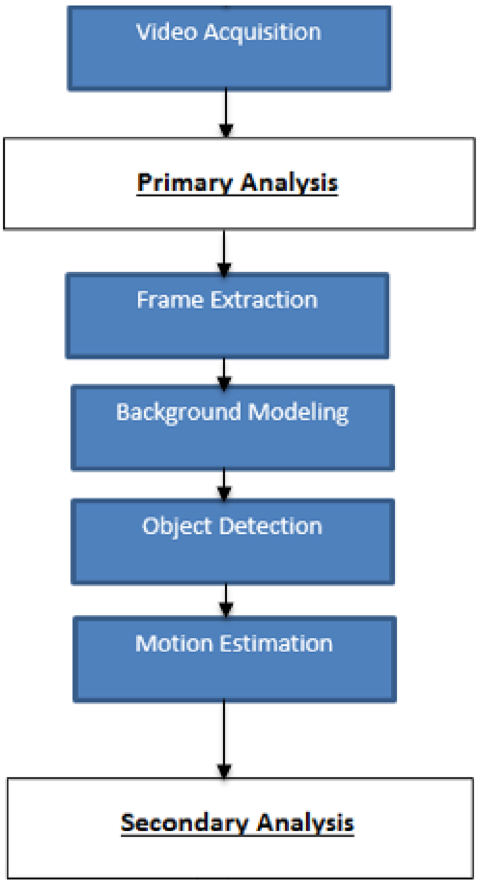

Fig 1. The flowchart of proposed algorithm. In primary analysis, background modeling is done to extract the foreground region. Object detection is done using frame differencing. Kalman filter is used to track the motion of the object. The tracked object is passed to secondary analysis module as an input.

IV. Primary Analysis

As depicted in Fig 1, the primary analysis is comprised of Frame Extraction, Background Modeling, Object Detection and Motion Estimation.

A. Background Modeling

After video data is acquired, background model is computed using frame averaging in order to detect foreground. Frame averaging is a statistical method to create a background frame by the combination of multiple frames. For background modeling greater the number of frames for averaging, the better the results will be as the noise is decreased. In our case we have consider 16 initial frames to model background.

B. Object Detection

In order to detect the object, the background model is subtracted from the current frame, also called frame subtraction. To get distinct object we used an Otsu's based automatic threshold for frame differencing. In the below mathematical equation the time is denoted by k(t) to compare with the background image denoted by P.

$$I[F(t)] = I[k(t)] - I[P] \quad (1)$$

Further we put a threshold T with this subtraction to improve the subtraction, the equation becomes

$$[I[F(t)] - I[f(t+1)]] > T \quad (2)$$

For good results of background subtraction we need to have suitable threshold T. We can apply T either manually or by computing it automatically. In our algorithm we have computed the threshold automatically using Otsu's segmentation. Automatic threshold has an advantage over manual threshold as it finds the suitable value for segmenting the objects based on the intensity of the image. Mathematically,

$$\sigma_w^2(t) = w_0(t)\sigma_0^2(t) + w_1(t)\sigma_1^2(t) \quad (3)$$

Weights $\omega_{0,1}$ show the probability of two classes distinguished by threshold T and $\sigma_{0,1}^2$ represent variance of these classes respectively.

$\omega_{0,1}(t)$ Is given by:

$$\omega_0(t) = \sum_{i=0}^{T-1} p(i) \quad (4)$$

$$\omega_1(t) = \sum_{i=T}^{L-1} p(i) \quad (5)$$

In terms of class probability *w* and means μ the equation becomes:

$$\mu_0(t) = \sum_{i=0}^{t-1} ip(i)/\omega_0 \quad (6)$$

ISBN: 978-988-19253-0-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

$$\mu_1(t) = \sum_{i=0}^{L-1} ip(i)/\omega_1 \tag{7}$$

$$\mu_T = \sum_{i=0}^{L-1} ip(i) \tag{8}$$

The following relations can easily be derived:

$$\omega_0\mu_0 + \omega_1\mu_1 = \mu_T \tag{9}$$

$$\omega_0 + \omega_1 = 1 \tag{10}$$

C. Motion Detection

For motion detection in videos we have used linear Kalman filter to track the position of the detected object. Kalman Filter is recursive hence draws fresh readings which can be processed as they arrive. The reason for taking Kalman filter for object tracking is real time processing and implementation is easy.

The Kalman filter makes an assumption that state at time k comes from the state at (k − 1).
Mathematically,

$$x_k = F_k X_{k-1} + B_k\mu_k + w_k \tag{11}$$

Where F_k represents state transition model applied to the previous state $x_{(k-1)}$; B_k shows control-input model applied to vector μ_k; w_k is noise assumed to be drawn from a zero mean with covariance Q_k.

$$w_k \sim N(0, Q_k) \tag{12}$$

Z_k For a particular state X_k can be calculated using

$$Z_k = H_k x_k + v_k \tag{13}$$

Where H_k represents observation model and v_k is the observation noise with covariance R_k.

$$v_k \sim N\,(0, R_k) \tag{14}$$

In Fig 2, the motion of an object tracked by linear Kalman filter can be seen, the red box on the object is showing the tracked objects.

Fig 2. Shows the object tracking using Kalman filter, the object with the red rectangle shows the tracked object.

V. Secondary Analysis

After the primary analysis is complete the secondary analysis is invoked. The secondary analysis perform actions on three different scenarios i.e. intrusion, loitering and slip and fall. Each event has its own particular attributes. We have divided these classes based on these attributes, which increases the performance of the proposed algorithm. Attributes for a particular event may vary according to the nature of the scenario. The model that we have developed is applied for public surveillance scenarios. The complete module for the secondary analysis is depicted in Fig 3.

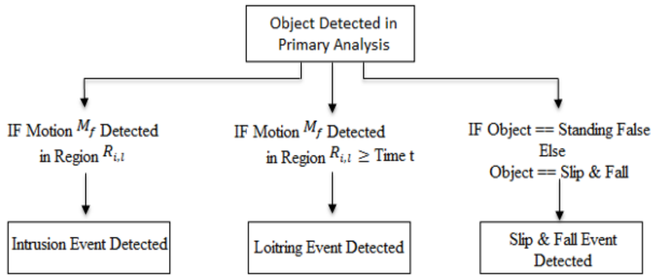

Fig 3. Shows the frame work of secondary analysis

VI. Experimental Evaluation

In this work, we analyze multiple abnormal events occurring in public places. The proposed method is applied on the events which are (a) Intrusion, (b) Loitering, (c) Slip and Fall.

A. Intrusion Detection

Intrusion is defined as movement of an object in restricted area. Intrusion detection is suitable for applications such as security, perimeter analysis, and border control. For intrusion detection we have considered the following attributes based on which a moving object is considered to be intruder in a video.

For Motion M in frame F, in region $R_{i,l}$

$$R_{i,l} = F_i + F_l \tag{15}$$

F_i Represent the initial frame, F_l represents the end frame and $R_{i,l}$ is the restricted region for Intrusion detection.
If motion M_f is:

ISBN: 978-988-19253-0-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

$$M_f < F_i \tag{16}$$
(No intrusion)

OR

$$M_f > F_l \tag{17}$$
(No intrusion)

If motion M_f is detected in the restricted region $R_{i,l}$, then the intrusion will be detected:

$$M_f == R_{i,l} \tag{18}$$
(Intrusion detected)

B. Loitering Detection

Loitering is similar activity as compared to intrusion but former is the one in which an alert is being triggered after some interval of time. Both intrusion and loitering are correlated, they differ in the context but in loitering detection we considered the time t.

If M_f in the restricted region R_{il} is more than time t it will triggered an alert that loitering in video has been occurred:

$$M_f == R_{i,l} < t \tag{19}$$
(No loitering)

Else

$$M_f == R_{i,l} \geq t \tag{20}$$
(Loitering detected)

Raise alert and draw bounding box against that object.

C. Slip and fall

Slip and fall is often occurring event in public places and much attention needs to be paid in order to handle the situation then and there. This particular event will be detected by changing of the position of an object during motion M_o such as change from vertical position V to horizontal position H. And if the object in the changed position is static for some time interval T, then an alert has been generated for the slip and fall.

Mathematically:

IF

$$Mo == V \tag{21}$$

OR

$$Mo == H < T \tag{22}$$
(No slip & fall)

Else

$$Mo == H > T \tag{23}$$
(Slip & fall detected)

VII. RESULTS AND DISCUSSION

The main aim of this study is to discuss the efficiency of the proposed method in detection of aforesaid events occurring in public places. Events mentioned above have been tested on benchmark datasets such as the PETS [18], Weizmann [19] and YouTube [17]. The results are based on 50 videos taken from the mentioned datasets. The table 1 shows the overall video samples obtained from the benchmark dataset.

TABLE I
VIDEOS DETAILS

Activity	No of Videos	Total Frames	Average Duration /Video	Average Size /Video	fps
Intrusion	24	2543	3.2s	6MB	30
Loitering	20	1403	2.51s	5.1 MB	30
Slip & Fall	6	510	2.24s	4.42 MB	30
Total	**50**	**4456**	**2.65s**	**5.17 MB**	**30**

The detailed results for the each event are discussed individually below:

A. Intrusion Detection

The module of Intrusion detection straight away detects an event upon the entry of a moving object in the restricted zone. Fig 4, illustrates sample outputs of the intrusion detection event. The Fig 4 (a) and (c) shows that object is moving smoothly in the non-restricted zone but Fig 4 (b) shows that upon entrance to the restricted zone a bounding box is drawn on the moving object and an alert is generated.

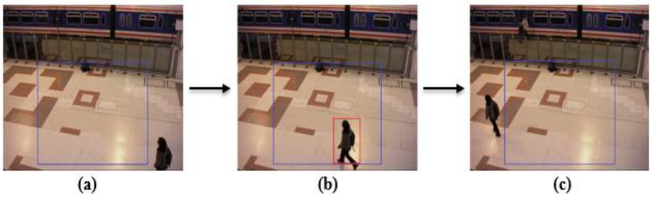

Fig 4. No intruder object in (a) and (c), the object detected as intruder in (b), the blue rectangle area represent the restricted region with the intruder object enter into this region which is representing with the red rectangle in (b).

B. Loitering Detection

The module of Loitering detection detects a moving object upon persistent in the restricted region after some specific interval of time. The timespan to trigger the loitering event may vary accordingly to the condition and requirements. Fig 5, illustrates the output from loitering detection event. Fig 5 (a), shows the moving object detected but not marked as loitering as it is in the specified timespan but as soon as the timespan of the moving object is elapsed we can see from Fig 5 (b) and (c) that the bounding box across the object has appeared and an alert for loitering is generated.

ISBN: 978-988-19253-0-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

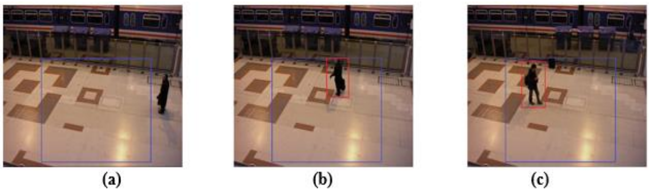

Fig 5. In (a) No loitering of object is detected in the restricted area representing (within the blue rectangle), the loitering of object is detected in (b) and (c), the loitering object is representing with the red rectangle (within the blue rectangle).

C. Slip and fall

The slip and fall event was tested on YouTube videos [17]. In this module we have tracked the motion of the object and if the object remains inactive for a while (say 5 seconds) then slip and fall event is generated. The timespan to trigger slip and fall event can be changed according to the requirements. Fig 6, shows the results of slip and fall event. Fig 69 (a), (c) and (e) shows that no slip and fall event is occur and the object remain still for certain time, when the object slip and fall then the object is detected in fig 6 (b), (d) and (f), the slip and fall message was created using the red rectangle.

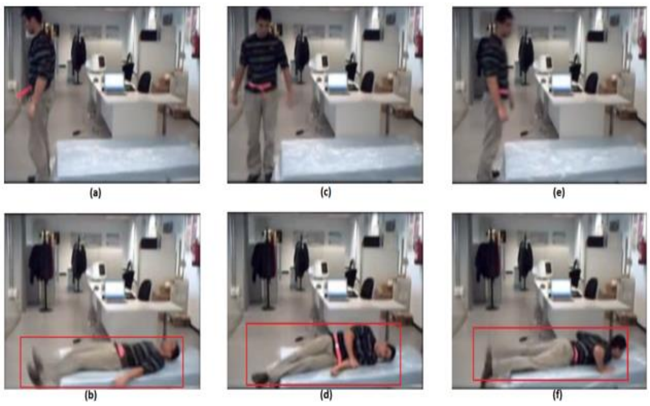

Fig 6. In first row (a), (c) & (e) shows no slip & fall event is occurred, in second row (b), (d) & (f) the slip & fall event is occurred, during the slip and fall event the object is detect with the red rectangle.

The table II shows the overall results for the three events intrusion, loitering and slip & fall with the accuracy rate achieved by the proposed method.

TABLE II
EXPEREMENTAL RESULTS

Event	Accuracy Rate	Computation Time	False Detections
Intrusion	88%	85 MS	12%
Loitering	90%	86 MS	10%
Slip & Fall	92%	89 MS	8%
Overall Result	90%	87 MS	4.7%

VIII. Comparison with State of The Art

We have compared the proposed method with state of the art [12] for the mentioned three events in video surveillance, the proposed method gives an overall accuracy rate of 90%. Although we do not report any state of the art. The proposed method gives better accuracy, as it can be seen in table III.

TABLE III
COMPARISION WITH THE STATE-OF-THE-ART

State-of-the-Art Papers	Accuracy Rate		
	Intrusion	**Loitering**	**Slip & Fall**
Lim, M [12]	83.33 %	83.33 %	88.46 %
Patino, L. [13]	N/A	90 %	N/A
Fernandez-Caballero, A. [11]	N/A	N/A	N/A
Proposed Method	88 %	90 %	92%

IX. Conclusion

Modular approach for detection of three different events in videos is presented in the paper. The model is further divided into two components i.e. primary analysis and secondary analysis. For performance evaluation, we tested our algorithm on a set of 50 videos obtained from publically available benchmark datasets and YouTube videos. The results show the effectiveness of the proposed algorithm. Proposed algorithm achieved a higher accuracy of 90 percent as compared to the state of the art [12]. The proposed work can further be enhanced for detection of multiple events in the same video.

References

[1] Tang, S. L., Kadim, Z., Liang, K. M., & Lim, M. K. (2010). Hybrid blob and particle filter tracking approach for robust object tracking. Procedia Computer Science, 1(1), 2549-2557.

[2] Jo, H., Chug, K., & Sethi, R. J. (2013). A review of physics-based methods for group and crowd analysis in computer vision. Journal of Postdoctoral Research, 1(1), 4-7.

[3] Yang, J., Ji, H., & Liu, J. (2011). Particle swarm optimization algorithm for passive multi-target tracking. Procedia Engineering, 15, 2398-2402.

[4] Benezeth, Y., Jodoin, P. M., & Saligrama, V. (2011). Modeling Patterns of Activity and Detecting Abnormal Events with Low-Level Co-occurrences. InDistributed Video Sensor Networks (pp. 137-148). Springer London.

[5] Ellis, L., Dowson, N., Matas, J., & Bowden, R. (2011). Linear regression and adaptive appearance models for fast simultaneous modelling and tracking.International journal of computer vision, 95(2), 154-179.

[6] Yang, H., Shao, L., Zheng, F., Wang, L., & Song, Z. (2011). Recent advances and trends in visual tracking: A review. Neurocomputing, 74(18), 3823-3831.

[7] Ahmed, M. S., Sharif, M. H., Ihaddadene, N., & Djeraba, C. (2008). Detection of Abnormal Motions in Video. In International Workshop Multimodal Interactions Analysis of Users in a Controlled Environment (MIAUCE) (pp. 1-4).

[8] Elarbi-Boudihir, M., & Al-Shalfan, K. A. (2012). Intelligent Video Surveillance System Architecture for Abnormal Activity Detection. In The International Conference on Informatics and Applications (ICIA2012) (pp. 102-111). The Society of Digital Information and Wireless Communication.

[9] Padilla-López, J. R., Chaaraoui, A. A., & Flórez-Revuelta, F. (2014). A discussion on the validation tests employed to compare human action recognition methods using the MSR Action3D dataset. arXiv preprint arXiv:1407.7390.

[10] Wang, Tian, Jie Chen, and Hichem Snoussi. "Online detection of abnormal events in video streams." Journal of Electrical and Computer Engineering2013 (2013): 20.

[11] Fernández-Caballero, A., Castillo, J. C., & Rodríguez-Sánchez, J. M. (2012). Human activity monitoring by local and global finite state machines. Expert Systems with Applications, 39(8), 6982-6993.

ISBN: 978-988-19253-0-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

[12] Lim, M. K., Tang, S., & Chan, C. S. (2014). iSurveillance: Intelligent framework for multiple events detection in surveillance videos. Expert Systems with Applications, 41(10), 4704-4715.
[13] Patino, L., Ferryman, J., & Beleznai, C. (2015, August). Abnormal behaviour detection on queue analysis from stereo cameras. In Advanced Video and Signal Based Surveillance (AVSS), 2015 12th IEEE International Conference on (pp. 1-6). IEEE.
[14] Lim, C. H., Vats, E., & Chan, C. S. (2015). Fuzzy human motion analysis: A review. Pattern Recognition, 48(5), 1773-1796.
[15] Junior, J. S. J., Musse, S., & Jung, C. (2010). Crowd analysis using computer vision techniques. IEEE Signal Processing Magazine, 5(27), 66-77.
[16] Maeng, S. (2015). A Geometric Approach to Video Surveillance. Indian Journal of Science and Technology, 8(21).
[17] Fall Detection Tag by Gema and Visonic Technologies. https://www.youtube.com/watch?v=V8BLV4Wt3gA
[18] PETS: Performance Evaluation of Tracking and Surveillance http://ftp.pets.rdg.ac.uk/pub/PETS2006/
[19] Moshe Blank, Lena Gorelick, Eli Shechtman, Michal Irani & Ronen Basri. Actions as Space-Time Shapes, The Tenth IEEE International Conference on Computer Vision (ICCV'05), 1395-1402, Beiging,2005. http://www.wisdom.weizmann.ac.il/~vision/SpaceTimeActions.html