IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-6, Issue-2 E-ISSN: 2347-2693

Entry-Exit event detection from video frames

Vinay Kumar V[1*], P Nagabhushan[2]

[1*]Department of Studies in Computer Science, University of Mysore, Mysuru, India
[2]Department of Studies in Computer Science, University of Mysore, Mysuru, India

[*]*Corresponding Author: vkumar.vinay@ymail.com*

Available online at: www.ijcseonline.org

Received: 31/Jan//2018, Revised: 10/Feb2018, Accepted: 22/Feb/2018, Published: 28/Feb/2018

***Abstract*—** Video surveillance has been one of the ubiquitous aspects of life since few decades. However, there are certain places that demand privacy of an individual like washrooms, changing rooms, baby feeding rooms at airports, etc where cameras cannot be installed / are restricted. Thus, it has raised concerns about safety and security of the public. The objective of our research is to design and analyze the processes and various conceptual models to automate the Entry-Exit surveillance of the people entering into or exiting from the Camera restricted areas. As part of the objective, in this paper, work is carried out to detect or determine the Entry-Exit events using the video frames captured at the entrances of the camera restricted areas by analyzing the variations in histograms of colors-RGB in the video frames using Histogram distance measures. Few grids in the Camera View Scene are selected by continuous learning and are extracted to determine the events happening in the scene thus contributing to improvement in computing time. Confirmation of event happening and classifying it as Entry or Exit or Miscellaneous is presented by temporal analysis of these grids. Experiments are conducted on few standard data sets like SBM datasets transforming them to our scenario, as well as our manual data sets captured in real time with few assumptions to test the techniques proposed.

Keywords—Computer vision, video surveillance, camera prohibited areas ,color histograms, regression lines.

I. INTRODUCTION

A. Context of the Study

Automated Video surveillance has been active research area in the field of Computer vision. Mining of useful information efficiently from a huge collection of videos captured by the surveillance cameras by intelligently detecting, tracking and recognizing the object of interest, and understanding and analyzing their activities is the primary goal of Intelligent Video Surveillance Systems [1].

Video surveillance is applied in wide variety of scenarios both in public and private environments like crime anticipation [3], supervising vehicle traffic [4, 5] there by predicting and detecting accidents, and monitoring of elderly persons [6], children and patients at home. These applications demand surveillances in indoors and outdoors scenes of public places like malls, banks, airports, railway stations, bus stations, parking lots, toll-plazas, offices and apartments. In all these scenarios, the camera(s) would be watching the target.

However, in certain scenarios the target could become out of catchment scope of the camera(s) where cameras are restricted. The major concern of security is in these Camera restricted areas where cameras are restricted due to need of privacy of individuals in places like washrooms, changing rooms, baby feeding rooms at airports, etc. where surveillance is a challenging task as the person move away from the catchment of surveillance cameras.

Many theses [11, 12 and 13] and research papers [1-4, 6-10, and 14-21] have been submitted on automated video surveillance systems and their variety of applications. Video surveillance using single / multiple cameras has been surveyed. The concept of multi-camera calibration, object tracking, and computing the topology of camera networks are studied with the help of a classical review paper by Wang [10]. But, video surveillance in camera restricted areas is still an open problem. However, still there is a scope for video surveillance if we can place the surveillance camera so as to have the view of the entrance of such camera prohibited areas from outside which can facilitate monitoring of persons entering into / exiting from such areas. Based on the possible amount of knowledge that can be gathered such as biometric and behavioural features of the person, his/her time of entry and exit, unusual behaviour if any, the process of monitoring can be automated. This motivates us to carry out research on Entry-Exit surveillance by placing the surveillance cameras facing the entrances of the camera restricted areas.

The fundamental step in any video surveillance systems is to detect event that are happening in the scene. Precise localization of foreground objects by removing the background is a basic but significant task for such applications. The works in this paper focus on event detection at the entrances of Camera restricted areas.

B. State of the art and motivation

Event detection can be achieved using different techniques like temporal difference, background subtraction, optical flow, etc. Stauffer and Grimson [22], in the year 1999, developed the adaptive background mixture models for real-time tracking. Probabilistic method for background subtraction is proposed. Many improvements have been proposed to this method in the later years to accomplish various challenged faced in designing automated video surveillance systems. The classical review paper by Kalpana, et al [23] provides insights on various background subtraction methods using Gaussian Mixture Model (GMM), compares them and evaluates them using various quantitative techniques.

Garima Mathur, et al [24], in their critical review on Intelligent Video Surveillance techniques for suspicious activity detection, provides insights with their literature survey ranging from 1977 till the year 2015. 57 IEEE papers are thoroughly reviewed and comparative analysis of methods to detect events like unusual human behaviour, tracing abandoned object or deserted baggage, etc are presented.

Zheng, et al [25] in 2006, proposed histogram analysis over time that can quickly extract the background image from traffic video streams. The algorithm could mine closely perfect backgrounds on free highways when traffic was in free-flow condition. It was learnt that the mean and median of the color values were far away from accuracy in determining an object because the background colors were interfered with the most by the colors of the object. The algorithm used mode value of the color values in video frames. This motivated us to work on histograms of video frames captured so as to determine the events happening in the scene.

C. Proposed approach

In this paper, a new technique is introduced to detect entry-exit events happening in the scene. Each video frame extracted from the surveillance footage is represented using cumulative histogram of pixel colors (R, G and B) transformed to first order polynomials i.e., regression lines are drawn for every cumulative histogram. The regression lines are compared in order to differentiate between the background frames and the events frames by thresholding and hence the occurrences of events are detected. Determining if the event occurred is an entry or an exit is done by capturing the variations in levels in one or few particular areas of the scene by selecting few grids instead of considering the whole image. Our aim is to analyze and determine the levels of possibility of Entry-Exit detection with minimum resources i.e., a single static camera and by using low level feature – color in our approach.

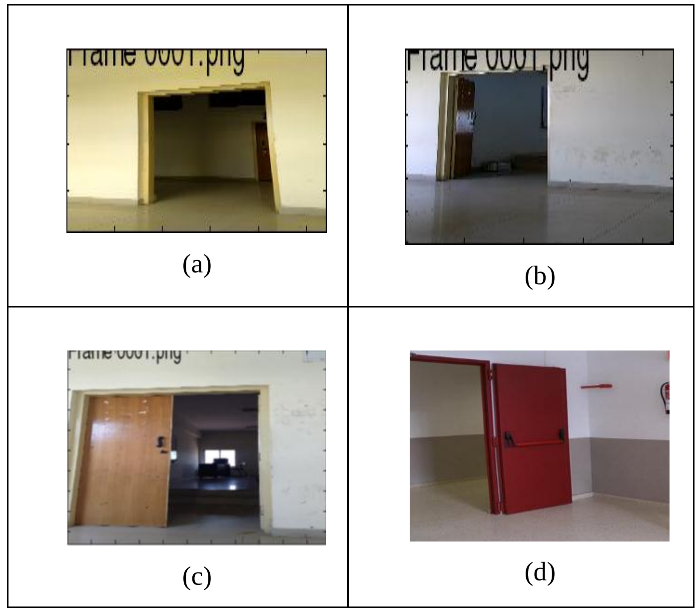


Figure 1. Frames captured by the cameras focusing various entrances (which are regions of interest here). (d) is a part of an image taken from SBM dataset [31] for experimentation purposes to test the efficacy of the system of the proposed model.

D. Apriori Assumptions:

The analysis is done based on the limited experiments conducted with few assumptions that are as follows–

- The camera used is a static camera i.e., without panning, tilting and zooming (PTZ) properties.
- The red, green and blue pixel values in consecutive frames are independent and have same variances [10]. This assumption though isn't certain always, gives advantage to avoid expensive matrix inversion operations at the cost of minute accuracy gain. Hence, either of red, green and blue pixel values is selected heuristically for computations (Comparison of each color pixel values are well tabulated in the experiments section).
- The background view is auto updated by incremental learning using adaptive Gaussian mixture models [26].

E. Outline of the paper

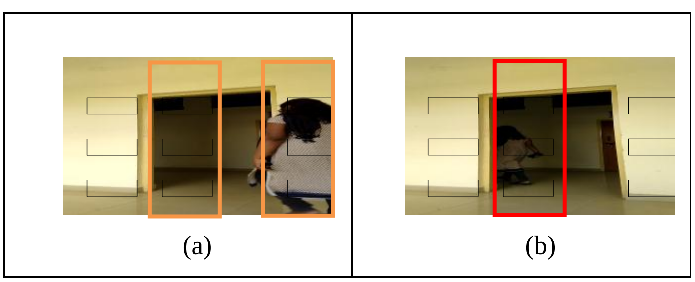
We give an overview of the Camera view scene modeling that includes placement of the camera and strategies to confirm the occurrence of an event in section 2. Methods to distinguish between Entry and Exit or any miscellaneous events are discussed in section 3. In section 4 we present experiments and analysis.

II. Camera View-Scene Modeling

In any Video surveillance system, computing the topology [10] of the camera views in order to extract maximum information is very important. Due to scene structural constraints, quality of the camera used, the topology of the camera view could be complex. An algorithm for optimal placement of camera in surveillance areas has been studied from [27 and 28] and same has been applied to model the camera view scene thus maximizing the amount of knowledge gain. However, a robust system should not depend on careful placement of the camera but should be robust to whatever is in its purview [23]. These scene structures can be automatically learnt from surveillance data or can be manually modeled as discussed in the experimental section.

The Camera view scene from the perspective of Computer Vision can be modeled using the following corollaries-

- Few regions in the scene of the camera view can be selected as grids and can be named as origin (area where objects enter camera view) and end grids (area where objects exit camera view). These grids are called as Grids of Interest.
- For an Entry event, few grids in the camera view scene can be labeled as origin grids O_{EN} and few grids can be labeled as end grids E_{EN}.
- Similarly, for an Exit event, few grids in the camera view scene can be labeled as origin grids O_{EX} and few grids can be labeled as end grids E_{EX}.
- It is to be noted that origin grids O_{EN} for the Entry event would be the end grids E_{EX} for the Exit event and end grids E_{EN} of the Entry event would be the origin grids O_{EX} for the Exit event as shown in figure 2.
- There can be more than one grid considered for either of origin or end grids. In such cases, all of them have to be considered which can still effectively determine the Entry-Exit events with the computational advantages.



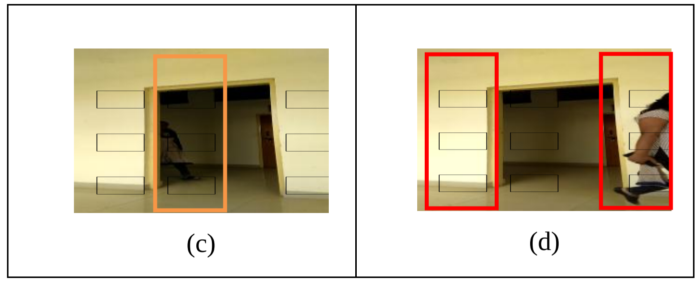


Figure 2. Camera View Scene sample frames during Entry and Exit events (a) Origin points (in green) for entry event (b) End points (in red) for entry event (c) Origin points (in green) for exit event and (d) End points (in red) for exit event

A. Statistical Model

Every frame extracted from a Video sample is a three-dimensional data that comprise of red (R), green (G) and blue (B) components. Color levels for each color component can be represented using Cumulative histograms as shown in figure3 so that no information is lost. A first order polynomial is applied onto these histograms and regression lines can be obtained. The slope of these regression lines (lt) are used to represent the color level of every color component for each frame t.

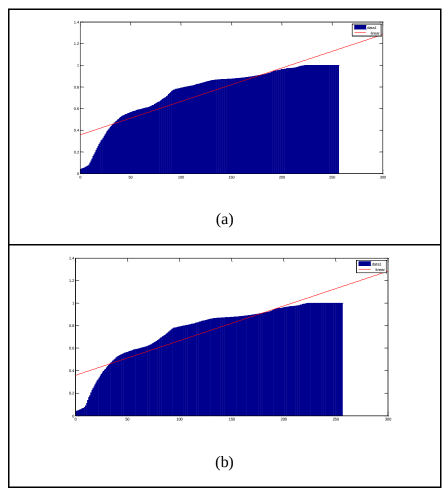


Figure 3. Cumulative histograms of the color levels in two consecutive frames and the first order regression line drawn over the histograms plot

Lalitha and Nagabhushan [29] proposed a technique for pattern classification by reducing dimension of data based on regression lines. The maximum difference between the regression lines is used to find the distance between two samples. Similarly, the distance between any two frames is determined by the distance calculated between their regression lines. The types of relation between two regression lines are – overlapping, non-overlapping and diverted as shown in the figure 4.

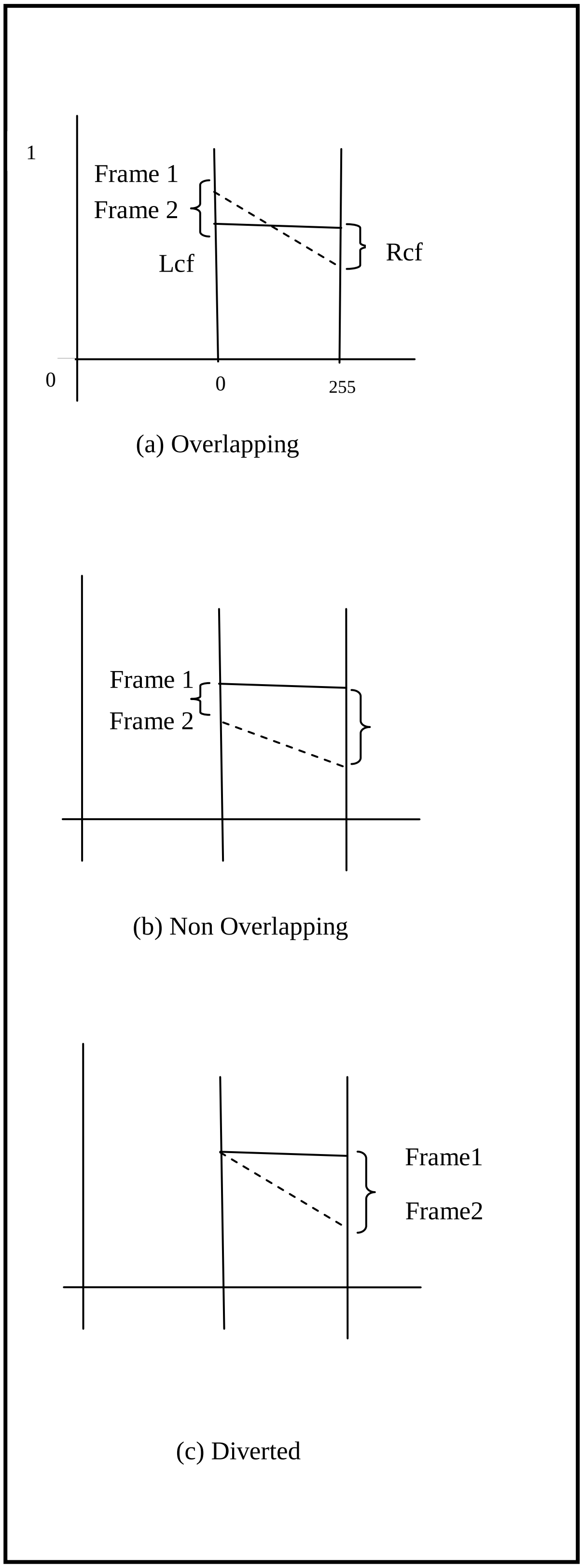


Figure 4. Types of relation between two regression lines drawn over the histograms plots.

Let Lcf and Rcf be the respective minimum and maximum points in regression lines constructed on the cumulative histograms for every frame f for color c. The distance between two frames m and n can be denoted by dis (m, n) = Max { LrmLrn, RrmRrn, LgmLgn, RgmRgn, LbmLbn, RbmRbn} [29]. The threshold value for the statistical model is initially determined by taking the average of the distances between two consecutive frames for a background video footage with f frames. The threshold is then updated every 30 frames (1 second) by incremental learning assuming it to be background. However, once the frames start exhibiting variations more than the threshold, the incremental learning is frozen until the system returns back to the background state.

B. Occurrence of an event

With the building of statistical model of a scene considering all possible constraints that causes distortion in the video scene, a threshold is computed at regular intervals that significantly determine the occurrence of an event. Figure 5 shows the gradual variation in the color levels in the camera view scene that takes cross over from the threshold computed and hence determine the occurrence of an event. It should also be noted that spontaneous variation in one frame do not determine an event as there are chances of external agencies like illuminations, shadows and reflections affecting the pixel values in a frame. Occurrence of an event is not defined by the sudden transition in the values of one frame but in series of consecutive frames with the movement of objects. Video footages that are captured at the rate of 30fps are considered to be sufficient enough to capture any significant event that can happen how much ever fast it maybe [30]. Hence, the variations in the pixel values in consecutive frames should be continuous in order to determine it as an event. This can be extracted using the distance between the consecutive. Though the variations in the color levels in the camera view scene determine the occurrence of an event, it is also necessary to determine if it is an entry or an exit or any miscellaneous event which is discussed in the next section.

III. UNDERSTANDING OF EVENTS

Considering the change in the color levels, only the occurrence of an event can be determined whereas focusing on few particular areas of the scene taken as grids, the possibility of it being an entry or an exit can be determined.

A. Entry event

Consider an event between the time interval ti and tj. For an Entry event, the origin grids O_{EN} considered shall have more variations in color levels at the beginning of the time interval as compared to the variations in the color levels at the end grids E_{EN} and gradually comes to the saturation as defined in the statistical model for background whereas pixel values in the end grids E_{EN} shall maintain the uniformity at the beginning of the event and gradually shows variations in the color levels. This can clearly indicate an Entry event. However, at the origin grids O_{EN}, once the uniformity is attained, it shall not have any more variations beyond the threshold level till the end of that event.

B. Exit events

For an Exit event, the end grids E_{EN} considered for an Entry event shall act as origin grids O_{EX} and shall have more variations in the beginning of the time interval as compared to the variations in the end grids E_{EX} and gradually comes to the uniform level as defined in the statistical model for background whereas end grids E_{EX} considered shall maintain the uniformity at the beginning of the event and gradually shows variations in the color levels. This can clearly indicate an Exit event. However, at the origin grids O_{EX} once the uniformity is attained, it shall not have any more variations beyond the threshold level till the end of that event.

C. Miscellaneous events

With reference to the statistical model built for every scene, the variations in the color levels in the frames that crosses over the determined threshold value is considered as an event. The clear distinction between Entry and Exit events has been well explained above and the event must satisfy the constraints as defined for Entry/Exit events. However, there are certain instances where the appearance of an object happens in the scene leading to variations in the color levels. Let us consider an event between the time interval ti and tj. If there are variations in the color levels in the grids O_{EN} at the beginning of an event, it must reach the saturation i.e., on par with the statistical model for background, before the end of the event ensuring E_{EN} exhibit variations till the end of the same event in order to classify it as an Entry event. Similar is the case for Exit event. Hence, an event which doesn't satisfy the constraints defined in either of Entry and Exit events can be considered as a miscellaneous event, cases for which are discussed in the experiments section ahead.

IV. Experiments and Results

The Camera View scene modeling and understanding of various events discussed in previous sections is illustrated as follows. Figure 5 shows frames extracted from a video clip of 16.3 seconds at the rate of 30 frames per second (fps) captured when no event occurred. Figure 6 shows the distances between the regression lines of histograms of color levels (for blue component selected heuristically) constructed for consecutive frames which clearly say that when no event occurs, the difference in pixel values between any two consecutive frames is not more than the threshold fixed by continuous learning, 2.0 in this case as plotted for the background footage.

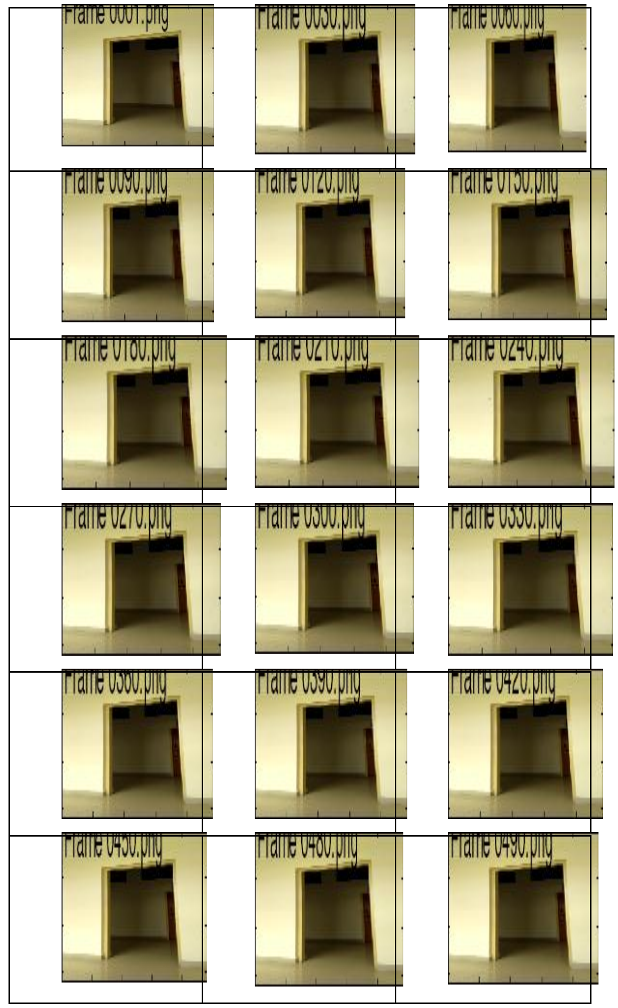


Figure 5. Frames extracted every second from a video clip of 16.3 seconds at the rate of 30 frames per second (fps) captured when no event occurred.

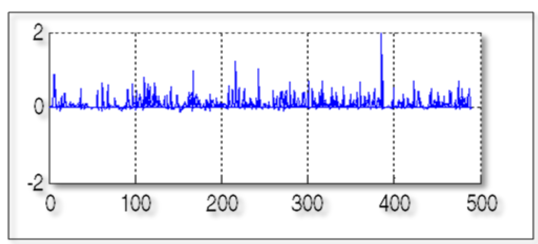


Figure 6. Distance plot between every two consecutive frames extracted from a video captured when no event occurred

Various scenarios have been considered for the analysis of events. The distance plot between consecutive frames in the video captured when an event occurred is as shown in figure 7. It clearly shows that the event to have occurred between frames 70 and 120.

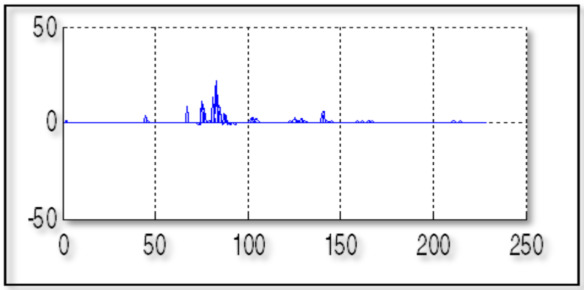


Figure 7. Distance plot between every two consecutive frames extracted from a video captured when an event occurred for the duration of 7.66 seconds that accounts to 230 frames.

Along with the confirmation of occurrence of an event, it is also important to analyze and understand the type of event. This is performed by computing the distance between grids defined and is shown in figure 8. Considering the grids 3, 6 and 9 to be the Origin grids O_{EN} for an entry event, the variations is only at the beginning of the event. During the course of the event, the other grids take over and the variations are seen till the end of the event. As there is no more variation in the origin grids after attaining the saturation well before the other grids, it can be confirmed as an 'Entry event'.

The experiments are conducted on standard [31] as well as the home-grown data sets where videos captured at three different entrances. The data sets consist of all possible cases that include Casual Entry, Rapid Entry, Slow Entry, Casual Exit, Rapid Exit, Slow Exit and Non-Entry-Exit cases.

The attires of the people captured in the videos have many varieties like colors closer to the Camera view scene background which may result in camouflaging and colors in contrast with the Camera view scene background.

One of the important observations that were made is in the probable grids to capture the head of a person Entering or Exiting. Grid 3 in figure 8 shows very high variation as compared to other grids as the color feature in the head part (especially women with long hair) is contrast to the Camera View scene background provided the person has not covered his / head. Also, the flooring parts of entrances are always to be considered as Origin grids for Entry event and End grids for Exit event as the variations captured on floors are on par with the O_{EN} and E_{EX}.

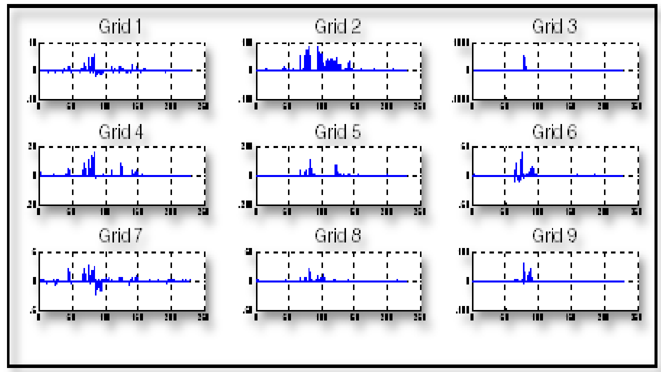


Figure 8. Distance plot between grids marked on every two consecutive frames extracted from a video captured when an event occurred for the duration of 7.66 seconds that accounts to 230 frames.

It is to be noted that the proposed methodologies are ideal for narrow entrances. The cases of occlusions are to be handled in a different way as two events- an entry and an exit occurring simultaneously may be concluded as a miscellaneous event. However, the distinction between the entry and exit confirmation for the assumed scenarios works extremely well. The misclassification happened mainly for rapid Entry and Exit cases that accounted to 22%. The proposed methods performed best in classifying miscellaneous events. None of the miscellaneous events considered were misclassified to be either of Entry and Exit events.

V. Conclusion and Future Scope

The problem of Entry-Exit surveillance in the areas where Cameras are prohibited using color features is addressed in a novel way. This can be considered as a kick-start to solve the problem that has to be analyzed in many dimensions. In the proposed method, a low-level feature - color levels in each frame are represented as a regression line. The color values are no more just ordinal values as storage and operations for every frame are complex tasks. Experiments are conducted on real data that have demonstrated the ability of this method in distinguishing between events with the aforementioned assumptions. However, surveillance using multiple cameras may provide scope to achieve more accuracy. The satisfactory result using the proposed method leads to the further exploration to analyze the possibility to distinguish between various individuals entering and exiting Camera prohibited areas.

Acknowledgment

The first author acknowledges MHRD, Govt of India for providing financial grants to carry out this research work.

REFERENCES

[1] Collins, R.T., Lipton, A.J., Fujiyoshi, H., Kanade, T., 2001. Algorithms for cooperative multisensor surveillance. Proc. IEEE 89, 1456–1477.

[2] Maria-Florina Balcan, Avrim Blum, Patrick Pakyan Choi, John Lafferty, Brian Pantano, Mugizi Robert Rwebangira and Xiaojin Zhu, Person Identification in Webcam Images:An Application of Semi-Supervised Learning, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213 USA

[3] Agustina, J.R., Clavell, G.G. The impact of CCTV on fundamental rights and crime prevention strategies: The case of the Catalan Control Commission of Video surveillance Devices. computer law & security review. 2011, 27, 168-74.

[4] Jeong, J., Gu, Y., He, T., Du, D.H.C. Virtual Scanning Algorithm for Road Network Surveillance. IEEE Transactions On Parallel And Distributed Systems. 2010, 21, 1734-49.

[5] Leotta, M.J., Mundy, J.L. Vehicle Surveillance with a Generic, Adaptive, 3D Vehicle Model. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2011, 33, 1457-69.

[6] Rougier, C., Meunier, J., St-Arnaud, A., Rousseau, J. Robust Video Surveillance for Fall Detection Based on Human Shape Deformation. IEEE Transactions on Circuits and Systems for Video Technology. 2011, 21, 611-22.

[7] Yigithan Dedeoglu, Moving Object Detection,Tracking And Classification For Smart Video Surveillance, a thesis submitted to the Department of Computer Engineering and the Institute of Engineering and science of Bilkent University.

[8] Mrinali M. Bhajibhakare, Pradeep K. Deshmukh, To Detect and Track Moving Object for Surveillance System, International Journal of Innovative Research in Computer and Communication Engineering Vol. 1, Issue 4, June 2013

[9] Kuihe Yang, Zhiming Cai, Lingling Zhao, Algorithm Research on Moving Object Detection of Surveillance Video Sequence, Optics and Photonics Journal, 2013, 3, 308-312

[10] Xiaogang Wang , "Intelligent multi-camera video surveillance: A review", Pattern Recognition Letters 34(2013) 3-19

[11] Domenico Daniele Bloisi, "Visual Tracking and Data Fusion for Automatic Video Surveillance", SAPIENZA Universita di Roma, September 2009

[12] Arun Hampapur, Lisa Brown, Jonathan Connell, Ahmet Ekin, Norman Haas, Max Lu, Hans Merkl, Sharath Pankanti, Andrew Senior, Chiao-Fe Shu, and Ying Li Tian, "Smart Video Surveillance", IEEE Signal Processing Magazine, March 2005

[13] Zhiyuan Shi, Timothy.M.Hospedales, Tao Xiang, "Transferring a Semantic Representation for Person Re-Identification and Search", CVPR 2015.

[14] Dong Seon Cheng and Marco Cristani Person, "Re-identification by Articulated Appearance Matching", Dept. of Computer Science & Engineering, HUFS, Korea

[15] C.-H. Kuo, S. Khamis, and V. Shet. Person re-identification using semantic color names and rankboost. In WACV, 2013.2

[16] R. Layne, T. Hospedales, and S. Gong. Person Reidentification, chapter Attributes-based Re-identification. Springer, 2014. 1, 2, 3, 6, 8.

[17] S. Khamis, C.-H. Kuo, V. K. Singh, V. Shet, and L. S. Davis. Joint learning for attribute-consistent person reidentification. In ECCV Workshop on Visual Surveillance and Re-Identification, 2014. 2

[18] D. Gray and H. Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In ECCV, 2008. Chapter 2, 6.

[19] H. Wang, S. Gong, and T. Xiang. Unsupervised learning of generative topic saliency for person re- dentification. In BMVC, 2014. 2, 3, 5, 6

[20] Y. Xu, L. Lin, W.-S. Zheng, and X. Liu. Human reidentification by matching compositional template with cluster sampling. In ICCV, 2013. 2

[21] Cheng, D.S., Cristani, M., Stoppa, M., Bazzani, L., Murino, V.: Custom pictorial structures for re-identi_cation. In: Proc. BMVC. (2011)

[22] Stauffer C, Grimson W (1999) Adaptive background mixture models for real-time tracking. IEEE Comput Soc Conf Comput Vis Pattern Recogn 2:246–252

[23] Goyal, K. & Singhai, J. Review of background subtraction methods using Gaussian mixture model for video surveillance systems, Artif Intell Rev (2017). https://doi.org/10.1007/s10462-017-9542-x

[24] Garima Mathur, Mahesh Bundele, Research on Intelligent Video Surveillance techniques for suspicious activity detection critical review, International Conference on Recent Advances and Innovations in Engineering (ICRAIE), 2016.

[25] Zheng J, Wang Y, Nihan N, Hallenbeck E (2006) Extracting roadway background image: a mode based approach. J Transp Res Rep 1944:82–88

[26] Pinto RC, Engel PM. A fast incremental Gaussian mixture model. PLoS ONE 2015;10(10):e0139931 doi: 10.1371/journal.pone.0139931 [PMC free article] [PubMed]

[27] S.Indu and S. chaudhury, "Optimal sensor placement for surviellence of large spaces," Proceedings of ACM/IEEE International conference on Distributed Smart Cameras, Sepember 2009.

[28] R. Al-Hmouz, S. Challa, "Optimal Placement for Opportunistic Cameras Using Genetic Algorithm", ICISSNIP, pp. 337-341, 2005.

[29] Rangarajan, L., Nagabhushan, P., 2000. A new method for pattern classification through dimensionality reduction based on regression analysis. In: Proc. Indian Conf. Computer Vision, Graphics Image Process., December 20– 22, 2000.

[30] S. Gulliver, G. Ghinea, "Stars in their eyes: What eye-tracking reveals about multimedia perceptual quality", IEEE Trans. Syst. Man Cybern. A Syst. Humans, vol. 34, no. 4, pp. 472-482, Jul. 2004

[31] M. Camplani, L. Maddalena, G. Moyà Alcover, A. Petrosino, L. Salgado, A Benchmarking Framework for Background Subtraction in RGBD videos, in S. Battiato, G. Gallo, G.M. Farinella, M. Leo (Eds), New Trends in Image Analysis and Processing-ICIAP 2017 Workshops, Lecture Notes in Computer Science, Springer, 2017

