

Monitoring of People Entering and Exiting Private Areas using Computer Vision

Vinay Kumar V.

Department of studies in Computer Science
University of Mysore, Mysuru, India

P. Nagabhushan

Indian Institute of Information Technology - Allahabad
Prayagraj, UP, India

ABSTRACT

Entry-Exit surveillance is a novel research problem that addresses security concerns when people attain absolute privacy in camera forbidden areas such as toilets and changing rooms that are basic amenities to the humans in public places such as Shopping malls, Airports, Bus and Rail stations. The objective is, if not inside these camera forbidden areas, from outside, the individuals are to be monitored to analyze the time spent by them inside and also the suspecting transformations in their appearances if any. In this paper, firstly, a pseudo-annotated dataset of a laboratory observation of people entering and exiting the camera forbidden area captured using two cameras as an extension of the state-of-the-art single-camera based EnEx dataset is presented. Conventionally the proposed dataset is named *EnExX*. Next, a spatial transition based event detection to determine the entry or exit of individuals is presented with standard results by evaluating the proposed model using the proposed dataset and the publicly available standard video surveillance datasets that are hypothesized to Entry-Exit surveillance scenarios. The proposed dataset is expected to enkindle active research in Entry-Exit Surveillance domain.

General Terms

Intelligent Video Surveillance, Datasets

Keywords

Entry-Exit Surveillance, EnExX dataset, Private areas, Tracking

1. INTRODUCTION

Intelligent Video surveillance systems [1] contribute to realization of the concept of smart cities by addressing the public safety issues. They provide scope to track every person and analyze every event in their ambit. With the growing availability of economic surveillance cameras and processors, wide area surveillance using multiple cameras bring entire area in public places such as malls, airports, bus and rail stations under surveillance to enhance public safety. However, there are certain places like wash rooms, baby feeding rooms and changing rooms, referred to as private areas where camera surveillance is completely forbidden to ensure absolute privacy to the public. This is considered as a major setback to security systems. Suspicious events such as crimes and medical emergencies are left unnoticed for a longer time. Hence, there is a need for surveillance system that can facilitate monitoring of

persons entering into / exiting from private areas without intruding the privacy of the public by analyzing their time of entry and exit, transformation in their appearances between before and after attaining absolute privacy. The notion of Entry-Exit Surveillance (EES) provides scope for monitoring people entering and exiting private areas by placing the cameras outside. It is essentially a tracking problem where every person who enter private area is tracked for a convinced exit. This involves various individual sub problems as discussed below.

- People detection - Detection of every person who enter the field of view.
- Motion based people tracking throughout their appearance in the field of view to confirm if they enter the private area or come out of the private area.
- Learning and representation of features of every person who enter the private area to re-identify when they exit.

1.1 Progress in Entry-Exit Surveillance and Challenges

The problem has received relatively little attention. Traces of research available in the literature include [2, 3, 4].

- The work in [2] introduces a benchmark dataset- EnEx dataset that demonstrates various dimensions of the problem. The dataset comprise of video sequences captured using single camera in five different locations by placing the camera so as to view the entrances of the private areas. The camera could capture back view of the individuals during entry and front view during exit. This flipped view difference of an individual between entry and exit events made re-identification challenging. State of the art Person re-identification algorithms suffered even without variations in appearances of the individuals due to the asymmetry between the flipped view of the individuals.
- The method in [3] analyze the direction of propagation of distortion in the color and intensity channels of the video signal to detect entry and exit events. However, the method is limited to ideal situations like one individual entering or exiting at an instance which is not always the case in real-time applicability scenarios.
- The model proposed in [4] focus on endorsing exit of every person who had entered private area by re-identifying them using appearance cues.

1.2 Motivations and Contributions

From the above discussion, it can be concluded that the EnEx dataset proposed in the literature is considered as a kick start for addressing Entry-Exit surveillance problem but is limited due to single view that in turn limits the performance of the state of the art algorithms due to the real-time challenges such as occlusions and wide variety of outfits of individuals. Hence, there is a need for a better performance evaluation benchmark that provide 360° view of the individuals using multiple cameras to enhance re-identification accuracies. This also helps in improvising *entry-exit event detection* accuracies as occlusions caused by overlapping individuals is handled using multiple camera views. In this backdrop, following are the major contributions presented in this paper.

- a. A new dataset of a laboratory observation of people entering and exiting the private areas using two cameras is proposed. The dataset is conventionally named **EnExX** to indicate dataset for Entry-Exit Surveillance using two cameras.
- b. Baseline results are presented for Entry-Exit event detection method that is proposed to analyze the three dimensional layout of the camera view as well as the transition of individuals through frames and to determine the type of event as *entry* or *exit* or *just appeared*.

2. THE PROPOSED SYSTEM

This section discusses the newly created dataset in the first part and the proposed Entry-Exit event detection method in the next part.

2.1 EnExX Dataset

The dataset consist of video sequences captured in 2018 using two cameras, simulating entry and exit of people into/from the private area as well as people crossing the field of view of the camera in a continuous corridor without entering the private area. This experiment used Master degree student volunteers from a university in India. Two IP cameras of resolution 1920X1080P are placed opposite to each other at a height of two meters from the ground among which one camera has the entrance of the private area in its field of view. The entrance on top of it has one camera placed and the opposite camera viewing the entrance is placed at a distance of 9.5 meters with oblique view. The camera cannot be placed parallel to the entrance to have the view of the entrance and placing it orthogonal to the entrance makes the public privacy compromised. Hence, oblique view of the entrance is chosen. The different camera views that make up the dataset are shown in Figure . The work presented in this paper is based on the view in the figure 3b.

The working hypothesis is that every individual appearing in the field of view is to be tracked for their origin and shrink positions to study the behavior of individuals in the entry-exit surveillance ambit and its relation with the private area to detect the entry and exit events. Every individual who is detected to have entered the private area is represented with a set of features in the gallery and every individual who is detected to have exited the private area is considered as probe to match with the gallery. The matched individual is to be removed from the gallery. The video clips are of duration one minute each captured at the rate of 20 frames per second (fps). The videos are captured for four hours each on four consecutive days. One major concern in simulating the EES environment is that the individuals might be aware that they are observed by the cameras. Hence videos captured in the initial period, that is first day are not considered and after individuals getting used to the experimen-



(a) Entrance view.



(b) Opposite view.

Fig. 1: Sample background images of video sequences in EnExX dataset

tal set-up, they behaved normally which is evident in the videos in the dataset. Tracks of every individual in the video sequences with their positions in each frame are provided as ground truth. With this the tasks of Entry-Exit event detection and person re-identification after attaining absolute privacy can be evaluated. Tracks are built using intelligent people detectors [6] after subtracting frames from the background [5]. The background are incrementally learnt and are updated to handle illumination variations. Every newly detected individual is tracked in the continuous frames using kalman filters [8].

2.2 Computer Vision analysis

Background

People Detection- The preliminary task in Entry-Exit surveillance is to detect all the individuals in every frame. This is currently performed using static cameras and hence we review works on static camera based surveillance here. In many tracking applications, adaptive gaussian mixture model [10] is used to model the background which proves to be computationally costly in real-time applicability scenarios as well as motion of the individuals is concentrated rather than segmentation of the individual completely. Gradient based features such as histogram of oriented gradient (HOG) in [6] has marked substantial gains over intensity based features. With efficient background subtraction before applying the detector gains faster computation time. With the introduction of Aggregated Channel Features in [7], individual detection has attained higher accuracy even with low resolution video frames.

Tracking- Spatial tracking of every individual in the surveillance ambit is one of the important phases in automated video surveillance. In multi-camera video surveillance systems, tracking is of two types - intra-camera tracking where individuals are tracked within a camera view and inter-camera tracking where individuals are tracked across multiple camera views where the camera views can be overlapping or are adjacent or are disjoint. The main focus in this paper is to track individuals within a camera view. Detailed discussion on individual tracking methods can be found in [9, 13].

2.2.1 Problem Statement - The aim is to spatially track every individual who enter the field of view of the camera till their departure. Two events are marked for each appearance of every individual S_i in the camera view-

- 1 Entry into the scene- Here the task is to analyze the origin point of the individual and classify the event either as Entry into the scene E_A or Exit from the private area E_X .
- 2 Exit from the scene- Here the task is to analyze the sink point of the individual and classify the event either as Entry into the private area E_N or Exit from the scene E_A .

The appearance of every individual who enter the camera view scene is to be marked either as entry or exit or just appeared events based on the following transitions.

- (E_A, E_A) - *Just appeared*.
- (E_A, E_N) - *Entry*.
- (E_X, E_A) - *Exit*.
- (E_X, E_N) - *Re-entry (Just appeared)*.

2.2.2 Event detection . This is illustrated with an example of an entry sequence as shown in figure 2.

Let the area covered by entire area be A with coordinates $A[0, 0, x_A, y_A]$ where x_A and y_A are width and height of the frame respectively. Let the entrance of the private area be represented by E with coordinates $E[x_{Emin}, y_{Emin}, x_E, y_E]$ where x_{Emin} and y_{Emin} are top-left coordinates of the entrance in the frame and x_E and y_E are width and height of the entrance respectively. The entrance is marked in *red* for EnExX dataset as shown in the figure 2.

Let there be n individuals who appears in the camera view scene in the interval $[t_{Start}, t_{End}]$ be represented by $S = \{S_1, S_2, \dots, S_n\}$. For each individual, before creating a track, it is important to determine whether the individual entered the scene from outside or exited from private area. If the individual detected as a new track is bounded in a bounding box B (marked in *yellow* in figure 2) and if

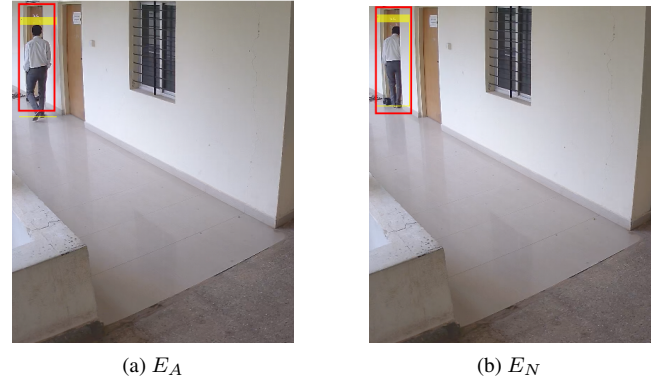


Fig. 2: Illustration of transition of individuals from state E_A to E_N

$B \subset E$, then it can be concluded that the individual has just entered the scene by exiting from private area. However, if $B \not\subset E$, then it can be concluded that the individual has just entered the scene from outside.

Similarly, before marking each individual to have exited the scene, it is important to determine whether the individual exited the scene to outside or entered private area. If the track is unassigned for a certain amount of frames, then the last frame where the individual was assigned has to be examined for its spacial location in the frame. If the individual is bounded in a bounding box B and if $B \subset E$, then it can be concluded that the individual has just entered the private area. However, if $B \not\subset E$, then it can be concluded that the individual has just exited the scene to outside.

Challenges and Solutions: . The important factors that affects Entry/Exit event detection are- overlapping of individuals while entering and exiting, and individuals occluding the entrance that makes it challenging to determine the entry/exit of other individuals into/from the private area. The solutions for these challenges are as follows.

When there are more than one individual in the scene and if one individual is blocking the entrance, then the entrance region has to be updated for that frame for other individuals. If S_i is the individual who blocks the entrance and if B_i is its bounding box, then for every other individual in the scene, the regions $E-B$ and B_i-B' can act as entrance where B_i' is the extended area of B_i .

The multiple view provided by two cameras in the proposed dataset help in locating the individuals in a track that are occluded by individuals overlapping in the entrance facing view using the opposite view. The spatial co-ordinates of the two cameras are mapped using mapping function as part of camera calibration.

The time t_i and t_j has to be marked for each individual as time of entry into the scene and time of exit from the scene respectively and the activity detected has to be updated as *Entry* or *Exit* or *Just appeared* accordingly.

The state transitions of the individuals who appears in the camera view scene in Entry/Exit Surveillance is as shown in figures 3 and 4.

3. EVALUATION

The proposed Entry-Exit event detection model is evaluated using the proposed dataset as well as the standard EnEx dataset which is captured using single camera that has entrance of the private area in its field of view. In addition, the CAVIAR dataset [11] that

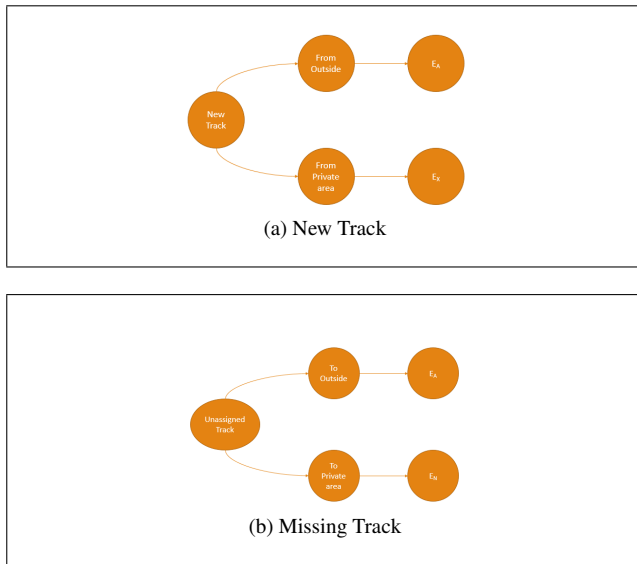


Fig. 3: State of each individual based on their position in the frames

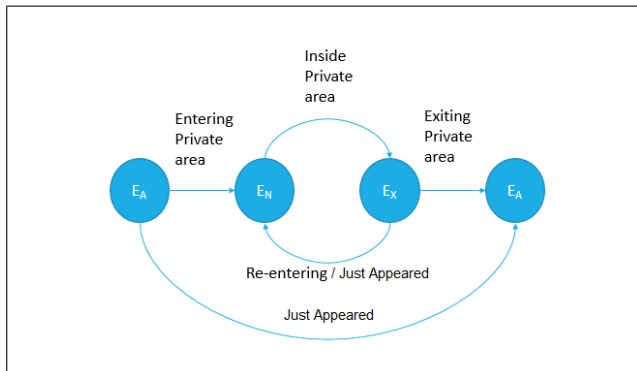


Fig. 4: State transition diagram for tracking of individuals in Entry-Exit surveillance

comprise of video sequences captured in a shopping mall corridor where people move in and out of a shopping outlet and the PAMELA-UANDES dataset [12] where people enter and exit out of metro trains are hypothesized to Entry-Exit scenarios and the proposed method is evaluated. As the problem is novel to compare with the current state of the art, the performance of the model is evaluated across the mentioned datasets.

Given the video stream as input to the system, individuals are detected in every frame using Histogram of Oriented Gradient based deeply learned people detection method [6]. For every new individual in the field of view, a new track is created and is assigned throughout its appearance in the field of view and every unassigned track after a thresholded number of frames is labeled as missing track and the frame in which the individual last appeared is considered last frame of the track.

Each track is given as input to the proposed model. The initial and final frames of the track are analyzed and are labeled as E_N , E_X and E_A based on their position in the frames as discussed in the previous section. The initial and final state of the individuals are

Table 1. : People detection results.

Dataset	Recall	Precision	F1
EnEx [2]	60.4	80.3	68.94
EnExX	80.43	92.23	85.92
CAVIAR [11]	78.2	91.3	84.24
PAMELA-UANDES [12]	80.76	90.05	85.09

Table 2. : Entry/Exit event detection results.

Dataset	Entry	Exit	Just appeared
EnEx [2]	84.4	86.4	88.3
EnExX	86.2	88.1	88.7
CAVIAR [11]	83.2	84.3	83.3
PAMELA-UANDES [12]	70.6	62.5	85.09

represented as ordered pair and the transition of state of the individuals during their presence in the field of view determines the type of event of the track as shown in figure 4.

The performance evaluation analyses of the proposed method that was evaluated on the above mentioned datasets are as follows.

People detection on all the four datasets are found to be satisfactory. The lower people detection rate in EnEx dataset as compared to the other two datasets is because of the camera placements where complete body of the individuals is not available in the frames. The missing rate in caviar dataset is due to frequent occlusion of individuals overlapping each other. However, the method performs best on CAVIAR dataset among the four datasets in classifying the events as Entry/Exit/Miscellaneous. Low Entry-Exit event detection in PAMELA- UANDES dataset is due to the top view of the scene where the visibility of the walking path of the individuals is limited. Overall, the method proposed in this paper gives satisfactory results in detecting the events in Entry-Exit surveillance.

4. CONCLUSION

In this paper, a pseudo-annotated dataset *EnExX* that consist of video sequences simulating Entry-Exit Surveillance scenarios using two oppositely placed cameras is presented. In addition, a novel Entry-Exit event detection method which is a fundamental step in Entry-Exit surveillance is proposed. Baseline results obtained by evaluating the proposed method on the proposed dataset as well as other publicly available dataset is presented. High Entry-Exit event detection accuracy on EnExX dataset demonstrates the importance of a second camera for Entry-Exit Surveillance and can be further improvised using prediction methods that makes use of walking directions of the individuals in the beginning and ending frames of the track.

5. REFERENCES

- [1] Dufour, Jean-Yves, Intelligent Video Surveillance Systems, John Wiley Publisher (2012)
- [2] V, Vinay Kumar, Nagabhushan,P, (2019) Entry-Exit Video Surveillance: A benchmark dataset. In: Chaudhuri, B.B., Nakagawa, M., Khanna, P., Kumar, S. (Eds.) Proceedings of 3rd International Conference on Computer Vision & Image Processing. Advances in Intelligent Systems and Computing, vol 1022. Springer, Singapore
- [3] Vinay Kumar V, P Nagabhushan, Entry-Exit event detection from video frames, International Journal of Computer Sciences and Engineering, Vol.6, Issue.2, pp.112-118, 2018.

- [4] Vinay Kumar V, P Nagabhushan. Appearance invariant Entry-Exit matching using visual soft biometric traits. arXiv preprint arXiv:1909.05145 (2019).
- [5] Cheng J., Yang J., Zhou Y. (2005) A Novel Adaptive Gaussian Mixture Model for Background Subtraction. In: Marques J.S., Prez de la Blanca N., Pina P. (eds) Pattern Recognition and Image Analysis. IbPRIA 2005. Lecture Notes in Computer Science, vol 3522. Springer, Berlin, Heidelberg
- [6] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In CVPR, 2005
- [7] Dollar, P., R. Appel, S. Belongie, and P. Perona. "Fast Feature Pyramids for Object Detection." IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 36, Issue 8, 2014, pp. 1532-1545.
- [8] Welch, Greg, and Gary Bishop, An Introduction to the Kalman Filter, TR 95041. University of North Carolina at Chapel Hill, Department of Computer Science.
- [9] V. Garcia, E. Debreuve and M. Barlaud, "Region-of-Interest Tracking Based on Keypoint Trajectories on a Group of Pictures," 2007 International Workshop on Content-Based Multimedia Indexing, Bordeaux, 2007, pp. 198-203.
- [10] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149), Fort Collins, CO, USA, 1999, pp. 246-252 Vol. 2. doi: 10.1109/CVPR.1999.784637
- [11] R. B. Fisher, PETS04 Surveillance Ground Truth Data Set, Proc. Sixth IEEE Int. Work. on Performance Evaluation of Tracking and Surveillance (PETS04), pp 1-5, May 2004.
- [12] Belloc, M.; Velastin, S.A.; Fernandez, R.; Jara, M.: 'Detection of People Boarding/Alighting a Metropolitan Train using Computer Vision', IET Conference Proceedings, 2018, p. 5 (6 pp.)-5 (6 pp.), DOI: 10.1049/cp.2018.1281
- [13] Rui Yao, Guosheng Lin, Shixiong Xia, Jiaqi Zhao, Yong Zhou.2019. Video Object Segmentation and Tracking: A Survey. arXiv preprint arXiv:1904.09172 (2019).