IJIRR

International Journal of Information Research and Review
Vol. 04, Issue, 07, pp.4364-4368, July, 2017

RESEARCH ARTICLE

ONLINE EVENT DISCOVERY FROM CAMERA IN INDOOR ENVIRONMENT

***Ayesha Choudhary**

Jawaharlal Nehru University, India

ARTICLE INFO

Article History:

Received 03rd April, 2017
Received in revised form
27th May, 2017
Accepted 19th June, 2017
Published online 26th July, 2017

Keywords:

Event discovery, unsupervised learning,
Kalman tracking,
Background subtraction.

ABSTRACT

In this paper, we propose a novel, real-time unsupervised learning based event discovery framework for indoor environment. We assume that a static camera is continuously observing an indoor area. Our framework is able to find the important regions such as the entrances and exits which we call as hotspots and to find the most popular paths. Our framework also finds the correlations between these hotspots. We develop an incremental clustering mechanism to discover events on-the-fly and define a probability associated with each cluster to ensure the correctness of the clustering stage.

Copyright©2017, Ayesha Choudhary. *This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.*

INTRODUCTION

With the decrease in cost of cameras and computing devices, and an increase in availability of smart cameras, online analysis of camera feed for event discovery is gaining popularity. Automated event discovery in videos finds many applications such as in retail environment, assisted living, security and surveillance, automated people counting, egress planning to name a few (Choudhary, 2016). The aim of event discovery is to learn the events that occur in an environment, analyse the pattern of these events and flag out unusual events in case any occur. In this paper, we focus on discovering events that occur in an indoor environment. The goal of our work is to find the various *hotspots* such as the entrances and exits in the environment, and the patterns of movement between these hotspots. This is essential in finding the normal patterns of movements in the area under observation. Such information is essential for many applications such as, unusual event discovery in case of security and surveillance, finding patterns of movement in retail stores for optimal placement of products and in traffic management and egress planning in smart buildings. In this paper, we assume that the camera is static, and placed in such a manner that all the important areas within the scene are observed by the camera. We also assume that the area is not densely populated at any time and that the indoor environment is well lighted at all times. We propose a probabilistic framework based on incremental clustering for finding the hotspots in the area under observation and the usual patterns of movement in that environment. We use background subtraction (Stauffer, 1999) to detect the moving objects in the scene and Kalman tracking (Kalman, 1960), to track the objects. The rest of the paper is organized as follows. We give a brief discussion on the current state of the art in Section 2 and discuss the techniques used in Section 3. In Section 4 we give the proposed framework and in Section 5 we discuss the experimental results. We conclude in Section 6 with a discussion on future directions.

****Corresponding author: Ayesha Choudhary,***
Jawaharlal Nehru University, India

Literature Survey

In recent times, a great deal of attention has been given to industrial and academic research on automated analysis of surveillance videos. This is primarily because of the need to automate surveillance video analysis which currently requires large amounts of observation by human operators. However, it is challenging to automatically analyze the surveillance video in a real-world scenario because of illumination changes, clutter, background noise, occlusions, weather conditions in case of outdoor environment and various types of activities that occur in the area under observation. It is specially challenging because the same event can be carried out in multiple ways. Moreover, event discovery requires object detection, tracking and recognition, each of which is itself a challenging task.

In the past, systems were limited to recognizing pre-specified activities (Rota, 2000; Hongeng, 2001; Bashir, 2007; Ayers, 2012). The main drawback of this analysis is that it is not always possible to pre-define all possible activities in the area under observation. Authors in (Choudhary, 2008; Choudhary, 2008), define unsupervised learning based methods for event discovery from videos originating from a single camera. In (Choudhary, 2008), a tubular representation of the video data is presented and a novel technique of component based clustering is described. They show that multi-perspective analysis of video data can be carried out using their scheme. In (Choudhary, 2008), the authors create epitomes of videos and use probabilistic latent semantic analysis (pLSA) (Hofmann, 1999), to discover usual and unusual activities in the area under observation. In (Porikli, 2004), hidden Markov model (HMM) (Rabiner, 1986) are used for trajectory analysis, to define if a trajectory is usual or not.

Scenario recognition is an important part of surveillance video analysis. In (Vu, 2003), the authors present a scenario recognition module by describing the objects and constraints that connect the sub-scenarios, which are then used to define a scenario. The input to the system are the trajectories, and the pre-defined scene model. Very recently, authors in (Al-Wattar, 2016), have proposed a method for monitoring elderly by recognizing activities using both spatial and temporal data. They use background subtraction to detect the current location of the person and use temporal information to decide which activity is taking place. In our work, we use incremental clustering to find the various hotspots and common paths to enable further analysis of the activities that occur in the area under observation. In the next section, we describe the techniques that we have used in this work, that is, background subtraction and Kalman tracking.

Techniques used

In this section, we briefly describe the background subtraction and Kalman tracking algorithms that we have used in this work.

Background Subtraction

Background subtraction is the technique used for detecting the moving objects in a scene where the scene is relatively static. Therefore, in our work, the camera is static and the scene has mostly stationary items, therefore, we use background subtraction algorithm to detect the moving people in the scene. Although many methods are available for background subtraction (Kim, 2004 and Stauffer, 1999), we used the statistical background model (Stauffer, 1999), in our work. In (Stauffer, 1999) the background is modelled as a Gaussian mixture model, where each pixel is regarded as an independent process. This implies that each pixel is modelled by a separate Gaussian model. The Gaussian models are updated continuously from the video and the static objects in the scene are then treated as the background objects. The moving objects in the scene are then detected as deviations from the learnt models and are therefore, detected as the foreground objects.

Kalman Tracking: Tracking is a fundamental activity required for finding the location of the object of interest for event discovery in videos. Kalman tracking [3] has been extensively used for tracking objects of interest in videos. It provides a simple yet powerful mechanism for recursively estimating the state of the system, even when the state of the system is uncertain and not precisely known. The two main operations in Kalman tracking are the prediction and correction steps that occur recursively. In the prediction step, an a priori estimation of the state that the system will be in the next step is carried out, while in the correction step based on the obtained measurements, an a posteriori estimate of the state of the system is obtained. This is shown in Figure 1.

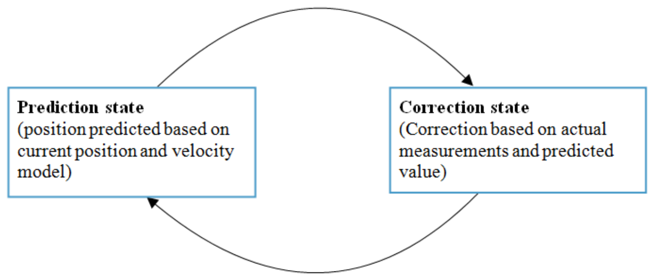

Figure 1. The basic concept of the Kalman tracker

For object tracking, the Kalman filter estimates the next position of the object based on the current position. In the correction step, the prediction result is updated based on the error between the current location found by object detection and the prediction in the previous step. We assume that the object moves with a constant velocity. Therefore, the motion model taken into consideration is given in Equation 1. Let (x_k, y_k) be the position in the k^{th} frame (v_x,v_y) be the constant velocity with which the object moves. Then the position in the $(k+1)^{th}$ frame is computed using the Equation 1.

$$x_{k+1} = x_k + v_x$$
$$y_{k+1} = y_k + v_y \qquad \text{.....................(1)}$$

In our work, after background subtraction is done on a video frame, the detected moving object is enclosed in a bounding box by finding the connected component. We then track the mid-point of the lower horizontal line of the bounding box. We assume that the person is upright and that the height of the bounding box is much more than the width of the bounding box representing the moving object. Since Kalman tracker may fail at times, the tracks that we get may not be long continuous tracks. However, in our work, during the track clustering that we perform, we overcome this shortcoming

Proposed Work

In this paper, we propose an unsupervised learning based system for detection of hotspots and patterns of movement from a camera on-the-fly. We assume that the camera is placed in an indoor environment in such a manner that the important areas of the scene are always visible in the camera. Moreover, we assume that the camera is a static camera. This allows us to apply background subtraction technique for detecting the moving objects in the scene. The objects of interest are the moving objects in the scene. Since it is an indoor environment, the objects of interest are mainly human beings. We use Kalman tracking for tracking the detected objects. Both background subtraction and Kalman tracking are applied one after the other in an online manner on each frame and therefore, we are able to process the camera feed on-the-fly. As the information is gathered, we perform incremental clustering on various parameters separately. Incremental clustering on the location of first appearance is done to find all the entrance locations. Similarly, when a person gets out of the scene, the location when the person is last seen is clustered to find all the exit locations. As a person moves in the scene, Kalman tracker is used to track the person. We incrementally cluster the tracks to find the events that occur in the scene. We find that by clustering the tracks we can find the correlations between the hotspots of the scene and find the locations in the scene that are most often visited. Therefore, events in an indoor environment are the movement of people between these hotspots. Incremental clustering for hotspots is carried out in the following manner as explained with reference to the entry locations. The same algorithm is carried out for the exit locations. The first time a person enters the area under observation, its location is treated as the first element of the first cluster. It also forms the cluster center of the entry cluster, C_1. From the next time onwards, if the entry location of a person is within a certain threshold from any of the entry cluster centers, the location becomes an element of that cluster and the cluster center is re-calculated as the mean position of all its elements, otherwise a new cluster is created. In this manner, we get all the clusters that may be locations of an entrance to the indoor environment. The same procedure is carried out for the exit locations, every time a person exits from the scene. However, this may lead to a few large clusters and many small clusters. We apply probabilistic measure on a cluster to decide whether a cluster depicts a true entrance/ exit. Moreover, it eliminates all the small clusters.

Let P_{CNi}be the probability that the i^{th} entrance cluster CN_i is a true entrance. Then, P_{CNi} is given by Equation 2.

P_{CNi} = no. of elements in CNi/ Total no. of elements in all entrance clusters (2)

If $P_{CNi} > 0.8$, then we consider CN_i to be a true entrance cluster.

Similarly, if P_{CEi} is the probability that the i^{th} exit cluster CEi is a true exit. Then, P_{CEi} is given by Equation 3.

P_{CEi} = no.of elements in CEi/ Total no.of elements in all exit clusters (3)

If $P_{CEi} > 0.8$, then we consider CE_i to be a true exit cluster.

Therefore, in this manner, we are able to find the hotspots in the scene. This can be extended beyond entrance and exit locations to find other locations of interest also. We also incrementally cluster the tracks in the scene, to find the most commonly taken paths. The track clusters are incrementally formed in a similar manner as the entrance and exit clusters. These track clusters also give the correlations between the hotspots. The correlations between the hotspots give us the information about the way people move around in the scene. Many a time, the tracker fails and when a person walks from one hotspot to the other. Since we may not have proper tracks between two or more hotspots, we use our incremental clustering algorithm in two steps to form the track clusters and overcome the failure of the tracker to give long, unbroken tracks. In the first step, we check whether the distance of the starting point of a track is within a threshold of the end point of another track. If so, then we temporarily consider these two tracklets to be part of a longer track and join them. We then find whether the endpoints of this longer track lie within a hotspot cluster. If so, then we consider the newly formed track as one track and perform the next step of track clustering. Since this happens in real-time, the number of tracklets to be considered are very few and therefore, we consider all the tracklets whose endpoints do not belong to a hotspot cluster. For track clustering, we consider the tracks whose endpoints belong to two different hotspots. We cluster the tracks incrementally as the video frame is processed based on the hotspots that the endpoints belong to. This gives the paths in the scene. Equation 4 is used to compute the probability of a track cluster to find out the true track clusters. Let P_{CTi}be the probability that the i^{th} track cluster CT_i is a true path. Then, P_{CTi} is given by Equation 4.

P_{CTi} = no. of elements in CTi/ Total no. of elements in all track clusters (4)

If $P_{CTi} > 0.8$, then we consider CT_i to be a true track cluster.

EXPERIMENTAL RESULTS AND DISCUSSION

We conducted experiments inside a home with the future plan of using our framework for an assisted living application for the elderly. A static camera was placed such that it continuously (24x 7) viewed a part for the indoor environment such that the ground plane was clearly visible. A computer with Intel i7 processor with 8GB RAM was attached with the camera and as each frame arrived, it was processed in C++. First, background subtraction is applied to the frame, the foreground object is detected, and the connected components is enclosed in a bounding box as shown in Figure 1. The midpoint of the lower horizontal line of the bounding rectangle has been taken as the foot position of the object. The Kalman tracker is initialized with this foot position in the first instance when the object first appears. Then in each subsequent frame, background subtraction and Kalman tracking is performed.

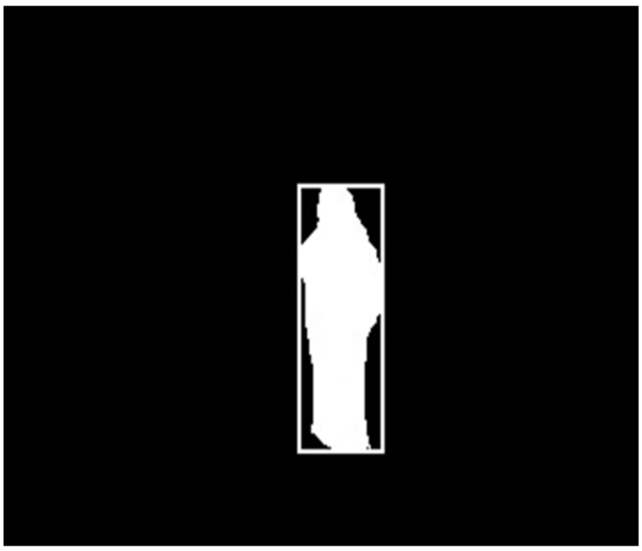

Figure 2. Foreground object detected afterbackground subtraction

To form the clusters of the entrances, the location of the object when it first appears is used for the clustering. Similarly, for all exit clusters the last location of the object is used. The difficulty is that the tracker may fail and therefore, not all the clusters formed by entrance and exit locations may bethe true entrances. However, as seen in Figure 3, the entrance and exit locations are correctly foundusing the probabilities computed as per Equations 2 and 3.

Figure 3. The red circles represent the hotspots in the scene, which are the entrances and exits located by our method. Hotspot 5 is discovered as a hotspot because when people come into the scene or leave the scene, this location gets clustered many times

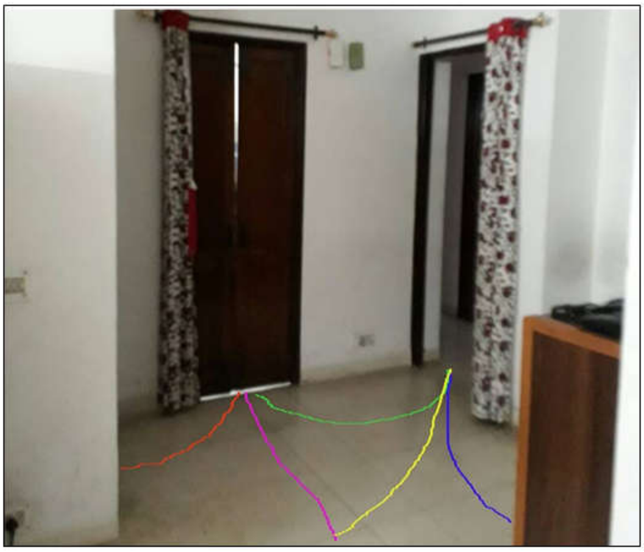

Figure 4. The most common paths taken in the scene are shown in this image. Each color represents a separate path

Our system also finds the most common paths that the people take in the indoor environment as shown in Figure 4. These are the most probable paths found by using the probability measure over the track clusters. Each color represents a separate path in the scene. Figures 3 and 4 also show the correlation between the hotspots. It is most common for people to move between hotspots 1 and 2, 2 and 3, 3 and 4, 3 and 5, 2 and 5. Although people also move between 1 and 5, 1 and 3, however, these paths are not as common. This information also leads to the observation that there may be an alternate entrance to the space beyond hotspot 1. Thus, analysing the video of a region gives rise to various different intelligent information that may not be directly observable.

Hotspot Name	Cluster Probability (average of entrance cluster and exit cluster probabilities)	True Entrance/ Exit (Y/N)	Remarks
1	0.87	Y	There exists a door and people enter and exit through it
2	0.96	Y	A door can be seen
3	0.95	Y	A door can be seen
4	0.92	Y	There exists door and a telephone is also kept near it. People walk in towards the phone.
5	0.9	N	Although this is not an entrance/exit in the indoor environment, but the people enter and exit through this area into the area under the camera's view. Therefore, this gets recognised as a hotspot

This also shows the importance of analysis of surveillance videos. Table 1 gives the cluster probabilities for the hotspots detected and shown in Figure 3. In the indoor area under observation, all entrances are exits also and therefore, we consider the average probability of the entrance and exit clusters to label a location cluster as a hotspot.

Conclusion

In this paper, we propose a novel, simple unsupervised learning framework for discovering hotspots and events in an indoor area that is continuously observed by a camera. The moving objects are detected and tracked in real-time as the camera captures the images. The entry and exit locations of the moving objects are used in an unsupervised learning framework to learn these locations in the scene. We use a probabilistic framework for removing noise and deciding which clusters represent the true entrances and exits. We also find the correlations between the hotspots to find the most common paths.

Future Directions

A vision based surveillance and analytics application requires detection and tracking of objects of interest as well as discovering the common activities and events that occur in the area under observation. Our work in this paper provides a simple and fast method that finds the interesting areas such as the entrances and exits in the scene, which we call hotspots and the events that correlate these hotspots. This can be further used for video analytics applications in various areas such as egress planning, surveillance and security and assisted living applications.

REFERENCES

Al-Wattar M., Khusainov R., Azzi D. and Chiverton J., "Activity Recognition from Video Data Using Spatial and Temporal Features," *2016 12th International Conference* *on Intelligent Environments (IE)*, London, 2016, pp. 250-253

Ayers D. and Shah M.,“Monitoring human behavior from video taken in an offce environment.”, Image and Vision Computing, 19(2):833–846, 2001.

Bashir F., Khokhar A., and Schonfeld D., “Object trajectory-based activity classification and recognition using hidden Markov models.”, IEEE Transactions on Image Processing, 16:1912–1919, 2007

Choudhary A. and Chaudhury A., “Video analytics revisited,” in *IET Computer Vision*, vol. 10, no. 4, pp. 237-247, 6 2016

Choudhary A., Chaudhury S., and Banerjee S., “A Framework for Analysis of Surveillance Videos”, In Proceedings of the 6thIndian Conference on Computer Vision, Graphics and Image Processing, 2008.

Choudhary A., Pal M., Banerjee S., and ChaudhuryS.,“Unusual Activity Analysis using Video Epitomes and pLSA”, In Proceedings of the 6thIndian Conference on Computer Vision, Graphics and Image Processing, 2008.

Cucchiara R., GranaC., Piccardi M. and A. Prati, "Detecting moving objects, ghosts, and shadows in video streams," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, pp. 1337-1342, Oct. 2003.

Hofmann T.,“Probabilistic latent semantic analysis”, In Proceedings of the Conference on Uncertainty in Artificial Intelligence, 1999.

Kalman R. E., "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering,* vol. 82, no. 1, p. 35– 45, 1960.

Kim K., Chalidabhongse T. H., Harwood D. and Davis L., "Background modeling and subtraction by codebook construction," *Image Processing, 2004. ICIP '04. 2004 International Conference on*, 2004, pp. 3061-3064 Vol. 5.

Porikli F.,“Trajectory distance metric using hidden Markov model based representation.”, In Proceedings of the IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS), 2004

RabinerL. R.,“A tutorial on hidden Markov models and selected applications in speech recognition.”, Proceedings of the IEEE, 77(2):257–286, 1989.

Rota N. and Thonnat M.,“Video sequence interpretation for visual surveillance.”, In Proceedings of the IEEE International Workshop on Visual Surveillance, pages 59–68, 2000.

Hongeng S. and Nevatia R.,“Multi-agent event recognition.”, In Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2:84–91, 2001.

Stauffer C. and Grimson W. E. L. “Adaptive background mixture models for real-time tracking.”, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2:2246–252.

VuV. T., Brmond F., and ThonnatM., “Automatic video interpretation: A novel algorithm for temporal scenario recognition.”, In Proceedings of the International Joint Conference on Artificial Intelligence, 2003.

****