Cumhuriyet Üniversitesi Fen Fakültesi
Fen Bilimleri Dergisi (CFD), Cilt:36, No: 3 Özel Sayı (2015)
ISSN: 1300-1949

Cumhuriyet University Faculty of Science
Science Journal (CSJ), Vol. 36, No: 3 Special Issue (2015)
ISSN: 1300-1949

A Novel Method for Scene Modeling to Detect Unusual Activity

Hamidreza RABIEE[1], Javad HADDADNIA[2*], Omid RAHMANI SERYASAT[1]

[1]*Department of Electrical and Computer Engineering, Hakim Sabzevari University, Sabzevar, Iran*

[2,] *Associate Professor, Electrical and Computer Engineering Department, Hakim Sabzevari University, Sabzevar, Iran*

Received: 01.02.2015; Accepted: 06.06.2015

Abstract. Automated video surveillance is crucial for the security of various sites including airports, train stations, military bases, and many other public facilities. A modern surveillance system is expected to not only perform basic object detection and tracking, but also to interpret object behaviors. This higher level interpretation can have several applications including abnormal behavior detection, analysis of traffic trends, and improving object detection and tracking. In this paper we focus on the problem of interpreting the output of the object detection and tracking module in order to gather knowledge about the scene. This knowledge is used to build a scene model which can be used to detect abnormal motion patterns and to enhance the surveillance performance by improving object detection. We present two novel and complementing models here: first model that is suitable for modeling single object motion, and real-time applications and second model that is useful for learning relationship between concurrently moving object pairs in the scene.

Keywords: Scene Modeling

1.INTRODUCTION

The understanding of human activities in videos has attracted the attention of many in the computer vision research community. This technology can be useful in a variety of applications including, but not limited to, security & surveillance, human computer interaction, robotics, and multimedia. All of these application domains will have a significant impact on various aspects of our everyday lives. Security & surveillance systems can be important for the public safety at airports, train stations, and large parking lots.

In the case of human computer interaction and robotics, a key objective is to automatically recognize different gestures to which the machine then responds to appropriately. For instance, in the recent years there has been an increased interest in developing camera equipped gaming consoles where the goal is to create a more realistic interactive experience. The term "scene modeling" is not used here in context of scene content matching in domain of video matching and retrieval [2]. Buxton [3] provided a detailed review of the models that have been used for learning scene activity. Johnson *et al.* [7] presented a vector quantization based approach for learning typical trajectories of pedestrians in the scene, but they require entry/exit points to be marked manually. Grimson *et al.* [5] used location, velocity and size to classify activities. In [11], Remagnino *et al.* use velocity and aspect ratio to classify different tracks into vehicle or person. They utilize a Bayesian classifier for this task and an HMM model to capture common events in the scene. Saleemi *et al.* [12] proposed a single Kernel Density Estimate (KDE) model for the whole scene, which requires to save all training data. Their approach does not address anomalies due to object size and only focuses on the object velocity. In the past year or two there has been an increased interest in detection of unusual activities in crowded situations[8, 10, 1, 3,11]. Kim *et al.*[8] proposed a space-time Markov Random Field (MRF) model for

*Corresponding author. *Email address: jhaddadnia@yahoo.com*

Special Issue: The Second National Conference on Applied Research in Science and Technology

detecting abnormal activities in the scene. They learn the distribution of local optical flow using a mixture of probabilistic principal component analyzers.

In this paper we presents our approach for learning object motion patterns in a stationary camera. We present results of anomaly detection and scene model feedback to improve object detection. We present two complementing models for learning object motion patterns of single objects, as well as object pairs.

2.MODELLING SINGLE OBJECT ACTIVITIES

2.1. Learning the Scene Model

In this section, we present the details of the structure and learning of the proposed scene model. The visual tracking information serves as the input for our framework. We have used the object detection and tracking system presented in [6]. For a given surveillance video, the tracker produces a set of m tracks $\{T_1,...,T_i,...,T_m\}$, where every track is a set of observations of the same object. For instance, any i_{th} track is a set of n observations $T_i = \{O_1,...,O_j,...,O_n\}$, where $O_j = (t, x, y, w, h)$ contains the time stamp t of observation, location (x, y), width w, and height h of the object. We also use the size (w, h) feature, as it provides useful information for finding anomalous behavior and improving object detection. For instance, this model assists in detecting a pedestrian on the road or a bicyclist on the sidewalk, even when the motion is not very discriminative. Using the set of observations, we want to generate a set of transition vectors that will be used to train the statistical model and provide the details about the motion and size of the objects. Proposed scene analysis approach detects abnormal events and provides scene model

Feedback is shown in figure 1. Traditional object detection is improved by using the pixel-level parameter feedback.

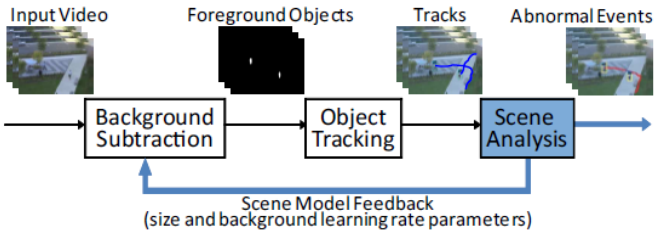


Figure1. Proposed scene analysis approach detects abnormal events and provides scene model feedback. Traditional object detection is improved by using the pixel-level parameter feedback.

For every observation, we compute a set of transition vectors that capture the transition from the given observation to future observations along the same track. Relative velocity is computed for the next observation, as well as a set of subsequent observations. In order to keep the problem computationally tractable, we limit the computation to a temporal window with τ observations. Figure 2 shows a synthetic track with marked observations and transition vectors from a particular observation O_j. This provides a means to detect abnormal tracks through the *global* analysis.

For any observation O_j, relative velocity is computed against all $\{O_{j+1},...,O_{j+\tau}\}$ to generate a set of transition vectors $\{\gamma_j^{j+1},...,\gamma_j^{j+\tau}\}$, where transition vector $\gamma_j^{j+\tau} = (x_{j+\tau}, y_{j+\tau}, \tau, w_j, h_j)$.

A multivariate GMM is used to model the *pdf* of the random variable Γ_l. The probability of an observation γ belonging to the GMM is given by

$$p(\Gamma_l = \gamma | \theta_l) = \sum_{i=1}^{n} \alpha_l^i \, p(\gamma | \theta_l^i) \tag{1}$$

where n is the number of components detected in the mixture, θ_l^i is the set of parameters defining the i_{th} component with weight α_l^i and $\theta_l = \{\theta_l^1, ..., \theta_l^n, \alpha_l^1, ..., \alpha_l^n\}$ defines the complete set
of parameters required to specify the mixture model. Each component is modeled as a Gaussian distribution of the form

$$p(\gamma | \theta_l^i) = \frac{1}{(2\pi)^{d/2} \left| \sum_l^i \right|^{1/2}} e^{\frac{-1}{2(\gamma - \mu_l^i)^T \sum_l^{i-1} (\gamma - \mu_l^i)}} \tag{2}$$

where d is the dimensionality of the model and $\theta_l^i = \{\mu_l^i, \sum_l^i\}$ are the parameters of the model. The computation of the GMM parameters is performed through an improved Expectation Maximization (EM) based algorithm, which was proposed by Figueiredo and Jain [4]. This particular approach provides a solutions to three major limitations of the basic EM algorithm. First, the number of components does not have to be fixed. This algorithm estimates the number of components by removing the components that are not supported by the data. Second, this approach does not require careful initialization and starts with a large number of components which are spread throughout the data. Third, this algorithm also avoids convergence towards a singular estimate near the boundary of the parameter space. The details of the algorithm are available in [4,9].

After learning of the complete scene has been performed, the GMM parameters for every pixel location are stored as the scene model. For a given observation, if we only update the pdf of the pixel at the centroid of the bounding box, then the created models could be spatially sparse.

To achieve better spatial smoothing of the motion models in the neighboring pixels, we update all the pixels in the bounding box. Note that unlike most of the previous approaches, learning of the proposed scene model does not rely on merging track to estimate the main paths in the scene. his reduces possible sources of error due to incorrect path estimation or ambiguity of track membership between two or more paths. Another strength of the proposed structure of the scene model is the ability to perform online learning of motion patterns and adaptation to the changing object behaviors in the scene.

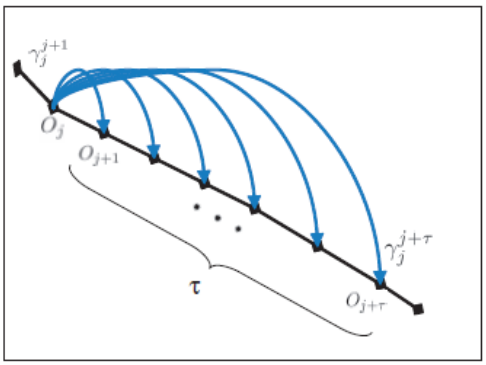


Figure 2. A set of observations with transition (blue) vectors connecting them are shown on a synthetic track. O_j and O_k represent two observations of the same object along the track. γ_j^k *is* the transition vector between O_j *and* O_k

3. ABNORMAL BEHAVIOUR DETECTION

The training phase generates a scene model Θ using the observed motion patterns. This model is a set of GMM parameters $\Theta = \{\theta_l\}$, where *l* is the location of all the pixels with sufficient training observations. We propose an online approach for detecting anomalies in the latest observation O_t from the test track *T*.

Our goal is to determine if the current observation O_t is abnormal or not by analyzing the trail of observations in the track. Therefore, we use the minimum transition probability

$$\beta_t = \min_i \{p(\Gamma_{l(t-i)} = \gamma_{t-i}^t)\} \tag{3}$$

for $i = 1,...,\tau$ and the observation O_t is declared abnormal if following condition is true $\beta_t \le \lambda$, where threshold λ is applied to the least probable transition.

This provides a means of detecting atypical transitions that originated from any one of these higher order transitions. Hence, both local and global anomalies can be detected through this framework. Our approach performs online analysis of the motion patterns to detect anomalies as soon as they occur.

We use this framework to detect various types of anomalous behaviors. Figure 3 presents various types of detected anomalies in a real video. These include pedestrians on the road and grass, skateboarder and bicyclist on the sidewalk, pedestrians sitting down, etc.

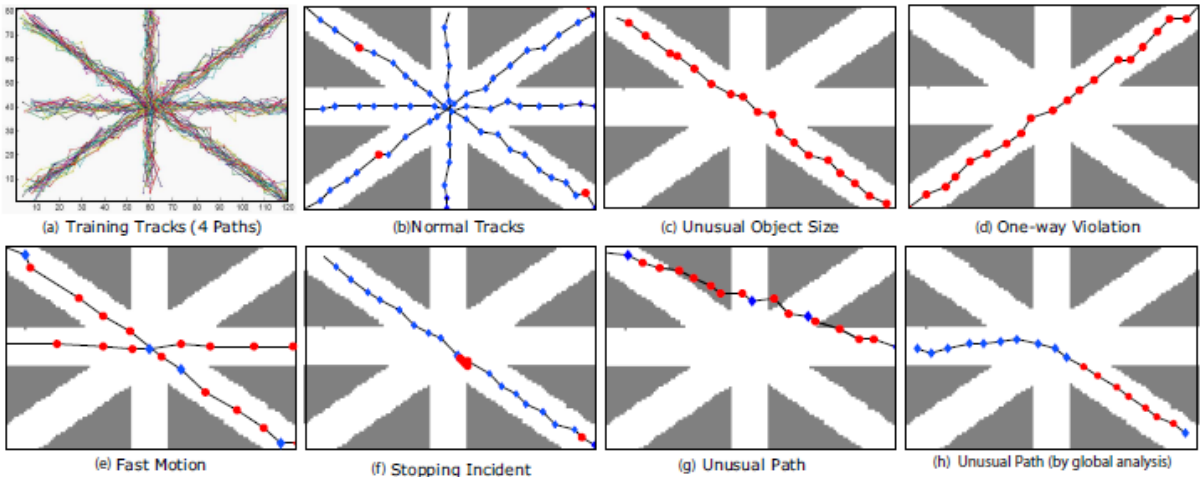
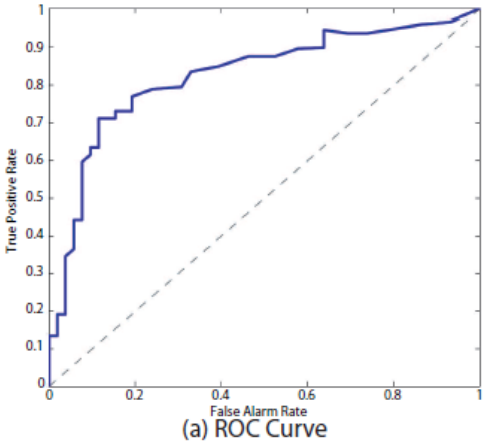


Figure 3. Global anomaly: when the tracks are not allowed to change paths, global analysis detects the violations. Every observation is labeled either normal (blue diamond) or abnormal (red circle). Gray background is the region without motion model. (a) Training set of random unidirectional tracks (along four paths). (b) Local analysis fails to identify anomaly, while (c) global analysis highlights the observation that take an unusual path.

4. EXPERIMENTAL RESULTS

The performance of the proposed framework was tested on real sequences captured from three different surveillance cameras. A typical scene observed from the first camera is shown in Figure 3



(a) ROC Curve

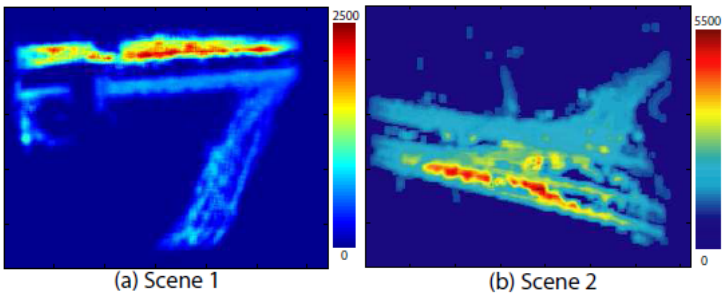
	Training	Testing
Normal Tracks	1342	217
Abnormal Tracks	0	31

(b) Ground truth

Figure 4. Anomaly detection performance on the scene shown in Figure 3. (a) ROC curve for the 30 mins test video. (b) Table with ground truth number of tracks used in training and testing

Realtime object detection and tracking was performed using the UCF KNIGHT system [6]. Initial training is performed off-line and testing for anomalous behavior detection was performed using the tracking results from a 30 minute test video. Figure 4(b) shows the details of the training and testing sets used for this experiment. Matlab implementation runs at approximately 26 fps for this module on a 3GHz Pentium D PC machine. Figure 3 presents the output of abnormal behavior detection in the test sequence.

The experiments of improving object detection are performed on video from two other surveillance cameras. Results of the improvement in the object detection using the size parameter feedback are presented in Figure 6. Two real scenarios are shown here that support the claim that the proposed size map outperforms the case with fixed s value. In the case of (b), the lowest value of $s = 50$ is chosen and in both scenarios, false positive objects are detected. In the first scene, a small broken part of the pedestrian's shadow is detected as a valid object and in the second case, a noisy observation on the lamp post is declared as a valid object. In the case of (c), a comparatively higher value of $s = 150$ is chosen and it clearly misses the pedestrians that are farther away from the camera. Finally, (d) presents the improved object detection using the proposed size map which provides a different s value at each pixel location. All the actual objects are detected without any noisy detections. The automatically learnt size map proves to be very useful in accurately capturing the perspective distortions in the scene.



(a) Scene 1 (b) Scene 2

Figure 5. The object size maps are computed for scene 1 (Figure 3) and scene 2 (Figure 6). Intensity at every pixel location is the most probable size of the object observed at that location. The highest intensity is observed for the vehicles along the road. Note the gradually reducing sizes due to perspective effect.

Figure 6. Scene 2. Improvement in object detection by the proposed size model. Each row presents an instance in the same video. Column (a) shows the manually extracted patches of the objects currently present in the scene. Column (b) is the output when a uniform global value of $s = 50$ is used. Noisy foreground blobs are also detected as valid objects (red ellipses). (c) presents output when $s = 150$ is used throughout the scene. Individuals are not detected (red ellipses) when the object size is small. (d) presents results of the proposed size model. In both scenarios the valid objects are detected and the noisy observations are avoided.

5.CONCLUSIONS

In this paper we have presented a novel approach for coarse level activity modeling in a scene. In this approach, we adopt an unsupervised learning based approach, which models object motion and size at every pixel location. The proposed framework provides a means of performing higher level analysis to augment the traditional surveillance pipeline. The pdf of motion patterns at every pixel is modeled as a GMM, which is learned through EM based approach. Experiments on real videos have proven the effectiveness of the proposed approach for local and global anomaly detection. This framework does not require explicit extraction of the main paths in the scene. This approach can easily benefit from online learning and can also be used for conventional applications like predicting object path and scene exit points. In summary, the proposed framework is novel, robust, and can be generalized to more features than just motion and size.

REFERENCES

[1] S. Ali and M. Shah. A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. In *CVPR*, 2007.

[2] Basharat, Y. Zhai, and M. Shah. Content based video matching using spatiotemporal volumes. *Comput. Vision Image Understanding*, 110(3):360–377, 2008.

[3] H. Buxton. Generative Models for Learning and Understanding Dynamic Scene Activity. *Workshop on GMBV*, 2002.

[4] M. Figueiredo and A. Jain. Unsupervised learning of finite mixture models. *PAMI, IEEE Transactions on*, 2002.

[5] W. Grimson, C. Stauffer, R. Romano, and L. Lee. Using adaptive tracking to classify and monitor activities in asite. *CVPR*, 1998.

[6] O. Javed and M. Shah. Tracking and object classification for automated surveillance. *ECCV*, 2002.

[7] N. Johnson and D. Hogg. Learning the distribution of object trajectories for event recognition. *BMVC*, 1995.J. Kim and K. Grauman. Observe locally, infer globally: a space-time mrf for detecting abnormal activities with incremental updates. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 2009.

[8] L. Kratz and K. Nishino. Anomaly detection in extremely crowded scenes using spatiotemporal motion pattern models. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 2009.
[9] R. Mehran, A. Oyama, and M. Shah. Abnormal crowd behavior detection using social force model. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 0:935–942, 2009.
[10] P. Remagnino and G. Jones. Classifying Surveillance Events from Attributes and Behaviour. *BMVC*, 2001.
[11] Saleemi, K. Shafique, and M. Shah. Probabilistic modeling of scene dynamics for applications in visual surveillance. *Accepted for Publication in TPAMI*, 2008.