UNUSUAL EVENT DETECTION IN CROWDED SCENES BY TRAJECTORY ANALYSIS

Shifu Zhou, Wei Shen, Dan Zeng, Zhijiang Zhang

Key Laboratory of Specialty Fiber Optics and Optical Access Networks, Shanghai University

ABSTRACT

Anomaly detection in crowded scenes is a challenge task due to variation of the definitions for both abnormality and normality, the low resolution on the target, ambiguity of appearance, and severe occlusions of inter-object. In this paper, we propose a novel statistical framework to detect abnormal behaviors of the crowded scene by modeling trajectories of pedestrians. First, the trajectories are acquired by Kanade-Lucas-Tomasi Feature Tracker (KLT). Then trajectories are grouped to form representative trajectories, which characterize the underlying motion patterns of the crowd. Finally, trajectories are modeled by Multi-Observation Hidden Markov Model (MOHMM) to determine whether frames are normal or abnormal. The experiments are conducted on a well-known crowded scene dataset. Experimental results show that the proposed method can capture abnormal crowd behaviors successfully and achieves state-of-the-art performances.

Index Terms— Anomaly detection, crowded scenes, trajectory cluster, Multi-Observation Hidden Markov Model, pattern recognition

1. INTRODUCTION

Detecting abnormal activities in crowded scenes is one of the most challenging tasks in computer vision. Videos of crowded scene present significant difficulties for detecting abnormalities due to the large number of pedestrians in close proximity, the volatility of individual appearance, and the frequent partial occlusions that they produce. In addition, there are potential dangerous activities in crowded areas, including crowd panic, stampedes, and accidents involving a large number of individuals, which make automatic video analysis in the most need.

Most methods [1, 2, 3, 4] for detecting unusual events in video sequences are limited to the scenes with a few objects. Therefore, one common drawback among these methods is they are unable to handle crowded scenes. Once the density of objects in the scene increases, a degradation in their performance in terms of unusual events detection, is observed. A possible reason of the degradation is that they focus on representing the motion and appearance of an object or an individual in scenes. In fact, the motion and appearance of individuals change frequently in crowded scenes due to the complexity of crowd activities.

There are two main types of approaches for understanding crowd behaviors. The approaches of the first type are refereed as object-centric techniques, in which a crowd is treated as a set of individuals [5, 6]. Hence, to understand the crowd activities, it needs to perform segmentation or object detection. However, these approaches face considerable complexity in detecting objects, tracking trajectories, and recognizing behaviors in crowded scenes. On the other hand, the approaches [7, 8] of the second type consider the crowd as a set of particles in analysis of high density scenes. These approaches avoid the hard task of segmenting or tracking individuals in crowded scene.

To detect anomaly in crowded scenes, several methods have been proposed. Ramin Mehran *et al.* [9] used the social force model, which estimates the interaction forces of moving particles, to detect unusual events. However, it can be observed that the use of bag-of-words method may encounter difficulty in locating anomalies. In [10], a cuboid based windowing strategy is used to capture motion patterns in local areas and the unusual activities are detected as statistical deviations. However, the coherently meaningful features are likely to separate in different cuboids which may result in information loss. Mixtures of dynamic textures (MDT) was studied in [11] to model normal crowd behavior and outliers detected by the model are labeled as anomalies.

In this paper, we propose a method for anomaly detection in crowded scene by modeling trajectories, which characterize underlying motion information of the crowd. Different with cuboid based windowing strategy [10], trajectories of pedestrians preserve important global motion information of the crowd. Moreover, due to the stability of distribution of trajectories, we can decrease the false alarm rate of detecting unusual events. The experimental results demonstrate the proposed outperform the approach based on social force model [9]. Fig.1 shows the framework of our method.

2. METHODOLOGY

2.1. Trajectory acquirement

Trajectory, the path followed by an object moving through space, is a high-level information in visual surveillance applications. The conventional methods that capture trajectories

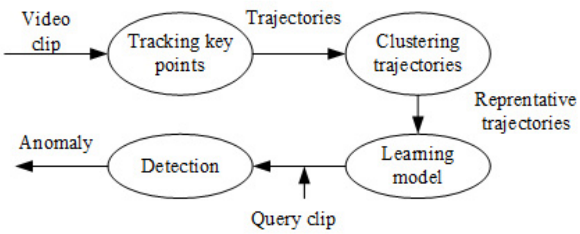

Fig. 1. The framework for anomaly detection.

of moving objects assume a static background or easily discernible moving objects, and, as a result, are limited to scenes with relatively few objects. However, in a crowded scene, there are hundreds of pedestrians in each frame, and possibly thousands throughout a video. Consequently, it is a hard task to segment and track pedestrians in crowded scene. To acquire trajectories of particles in crowded scene, Wu [12] estimated the positions of moving particles by optical flow. However, the trajectories may become unreliable when illumination changes. To obtain reliable trajectories, we adopt KLT tracker [13] to track keypoints of pedestrians in the crowded scene. For a given video sequence of a crowded scene, we first divide it into a set of clips with respect to time domain. Each clip can be represented by a matrix whose size is $T \times W \times H$, where T is the number of frames, W and H are the width and height of the frames, respectively. The trajectory of a keypoint is represented by a vector:

$$a = \{(X^t, Y^t)|t \in T\}, \tag{1}$$

where (X^t, Y^t) is the location of the keypoint at time t. We call the trajectory of a keypoint a keypoint trajectory. All the keypoint trajectories of a clip are denoted by

$$A = \{a_i|i = 1, 2, \ldots N\}. \tag{2}$$

Fig. 2. Keypoint trajectories parked on three crowded scenes (keypoints are not shown).

2.2. Trajectory Clustering

In crowded scenes, the motion of people may be influenced by their neighborhoods or the layout of scene, but they have their sinks or destinations. Hence, the crowd can be divided into several small groups whose members move coherently. A small group may include one or more pedestrians. Fig.2 shows the keypoint trajectories of three crowded scenes. In a crowded scene, many trajectories may belong to a single object, and their motion patterns are consistent with the objects. Keypoint trajectories distributed in one or more objects may be merged to offer a representative trajectory that describes the potential motion pattern of the small group. To capture the underlying motion patterns of the crowd, we apply clustering to trajectories.

Before clustering trajectories, it is necessary to discard motionless trajectories which carry little information due to noise or tracking error. Given a trajectory set $A = a_i{}_{i=1}^{N}$, we remove those with variance smaller than a predefined threshold,

$$cov(a_i) = \frac{1}{T-1} \sum_{t=1}^{T} |(X^t - \overline{X})(Y^t - \overline{Y})| < \epsilon \tag{3}$$

where $\overline{X}$ and $\overline{Y}$ are the mean of coordinate of a_i. Fig.3 shows the result after removing the motionless trajectories.

To obtain representative trajectories that describe the underlying motion patterns of the crowd, we cluster the remaining trajectories. To determinate an optimal number of clusters, we adopt a hierarchical k-means cluster strategy. The k-means algorithm is used by setting *k* to two in order to divide the trajectories into two subsets. Then, the two subsets are divided again into two subsets by setting *k* to two. The recursion terminates when the inter-cluster distance of two subsets of any level is less than a predefine threshold. The reprehensive trajectory, which is defined as the one with minimum sum of distance to all other trajectories in the same cluster, is used to measure inter-cluster distance. By this cluster strategy, we obtain the number of clusters and the reprehensive trajectories. Fig.4 shows an intermediate result of the trajectory clustering, in which the clustering process generates 42 clusters and each group is shown in a random color. The right column of Fig.4 shows the zoom-in view of a part of the scene, where the thin curves are keypoint trajectories and the thick black curve denotes the representative trajectory of the corresponding cluster.

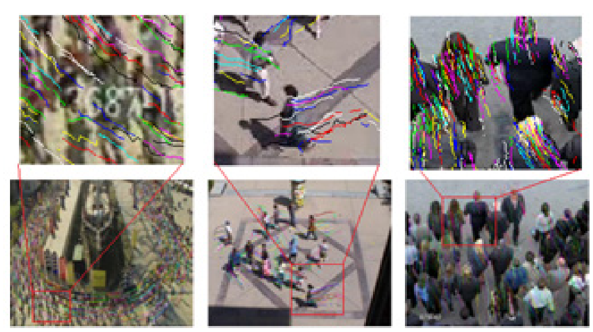

Fig. 3. Trajectories with low variance are removed. Top row shows zoom-in view of parts of each scene.

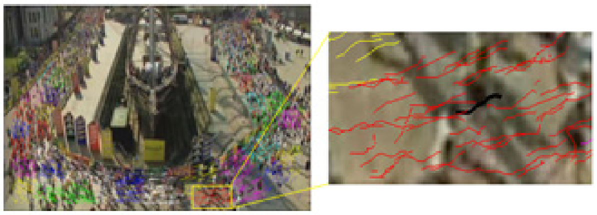

Fig. 4. Trajectory clusters based on spatial distance (left) and representative trajectory (marked in black) of one group (right).

Since representative trajectories characterize the underlying motion patterns of a scene, it's principle to use them to model a crowded scene given a training video sequence of usual activity. Anomalies are detected by identifying the trajectories of clips with low likelihoods. In this paper, we assume that trajectories are evolved from representative trajectories and we model the evolution from a probabilistic perspective. For instance, people moves in the opposite direction against the crowd and the likelihood of representative trajectories transfer to its corresponding trajectories is considerable low. Given a trajectory a_i, we can evaluate the probability of that it belongs to a specific representative trajectory s by using distance measure,

$$d(a_i, s) = \frac{1}{T}||a_i - s||_2. \tag{4}$$

The probability of a trajectory a_i given representative trajectory s is

$$p(a_i|s) = \exp(-\frac{d^2(a_i, s)}{\sigma_s^2}), \tag{5}$$

where σ_s is the variation of representative trajectory s.

2.3. Crowd Modeling

The distribution of trajectories in crowded scenes is complex and chaotic due to the complicated mechanics of a human crowd. In addition, the goal-directed dynamics and psychological characteristics influence how a pedestrian will behave in a crowd [14]. The motion of unstructured crowded scenes are more complicated than those of structured crowded scenes [15]. To detect the anomalies in both structured and unstructured crowded scenes, we employ a Multi-Observation Hidden Markov Model (MOHMM) [16], whose graphic illustration is shown in Fig.5. Ordinary MOHMM are defined by five parameters $\lambda = \{\mathbf{N}, \mathbf{O}, \mathbf{T}, \mathbf{E}, \boldsymbol{\pi}\}$, where $\mathbf{N}$ is the number of hidden states of the model, $\mathbf{O}$ is the possible symbols (values) of observation, $\mathbf{T}$ is a state transition probability matrix, $\mathbf{E}$ is a set of $\mathbf{N}$ emission probability density functions, and $\boldsymbol{\pi}$ is an initial state distribution. We model a crowded scene with unusual activities by a single MOHMM. The set of possible observations $\mathbf{O}$ is a set of trajectories. Generally, continuous or complex observations are often quantized as discrete symbols or values for MOHMM, however, this would significantly discard some important motion information in each trajectory. To avoid this problem, we associate the hidden states $\mathbf{N}$ with the representative trajectories. Furthermore, we use equation (5) to evaluate the emission probabilities. Under this construction allows trajectory to remain a multi-dimensional vector.

Since the number of trajectories of each clips is not the same, we interpolate trajectories to each clip to ensure the number of the trajectories in each of them are equal.

To train the MOHMM, we estimated $\mathbf{T}$ and $\boldsymbol{\pi}$ by Baum-Welch Algorithm. The emission densities are not re-trained due to the equation (5) already offers a good approximation.

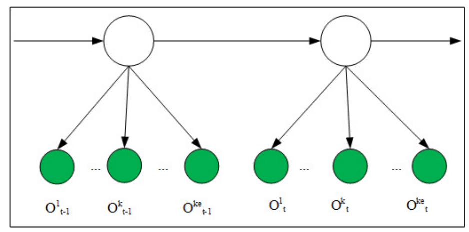

Fig. 5. Illustration of a MOHMM. White nodes are shown as hidden states and green nodes as observations.

2.4. Anomaly detection

The trajectory set of the video clips of normal crowded scene are used for training the MOHMM. The normal and abnormal events are identified based on the comparison of the current observations probability given by the learned MOHMM. The observation of a test video clip A is considered as anomaly if:

$$p(A|\lambda) < L_{th}, \tag{6}$$

where L_{th} is the threshold to identify the clip is normal or abnormal.

3. EXPERIMENTAL RESULTS

We evaluate the proposed method on the publicly available dataset of normal and abnormal crowd videos from University of Minnesota [17]. The dataset comprises eleven video sequences of three different scenarios and each video consists of an initial part of normal behavior and ends with sequences of abnormal panic.

To evaluate the performance of our method, each sequence of the three scenarios is divided into 10-frame clips as in [9]. The trajectories of each clip are acquired by employing KLT tracker. According to our statistics, most of trajectories whose variance less than 0.44 carry little motion information. hence, we choose $\epsilon = 0.44$ for the threshold to filter out the motionless trajectories.

For each scenario, we learn a normal model by randomly selecting $2/3$ of the normal clips as training samples. The trajectories in these clips are clustered by hierarchical k-means cluster strategy to obtain representative trajectories. since we adopt hierarchical clustering strategy, we don't need to determine the number of clusters and we only need to choose a threshold of inter-cluster distance for clustering. It can be observed that the proposed method achieves the best performance when we adopt 15 as threshold of inter-cluster distance. Then, the representative trajectories are associated with the hidden states of MOHMM. The parameters $\mathbf{T}$ and $\boldsymbol{\pi}$ are initialized uniformly and $\mathbf{E}$ is estimated by equation (5). Since training MOHMM requires the same number of trajectories of each clip, we interpolate trajectories to let the number of them in each clip up to 300. Fig.6 shows the trajectories (after interpolation) of clips 30 and 61 in scenario 3.

All the rest clips (including normal and abnormal frames) of a sequence are used for anomaly detection. We determine whether the clips are normal or abnormal by the probabilities outputted by the learned MOHMM. As an instance, Fig.7 shows the change of probability of clips of a sequence in scenario 3. From Fig.8, it can be observed (between clips 59 and 65) that there is an obvious decline corresponding to an anomaly, which agrees with the ground truth.

For performance evaluation, we compare our method with three baseline approaches: optical flows [9], the social force model [9] and lagrangian particle trajectory [12]. we have reconstructed the methods of the three baselines and the comparison was performed on the dataset [17]. The experimental results are shown in Fig.8, which depicts that our method achieves a higher detection accuracy than social force model [9] and pure optical flow [9]. Moreover, the performance of our method is slightly better than that of lagrangian particle trajectory [12].

Fig. 6. Trajectories (after interpolation) of two clips in a sequence, the left shows normal behavior and the right one is anomaly

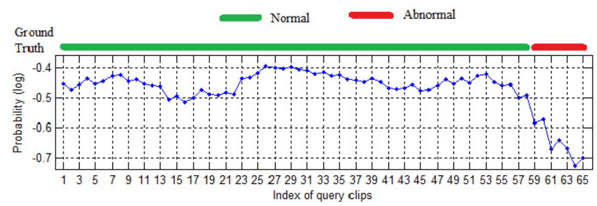

Fig. 7. Probabilities of query clips and corresponding ground truth.

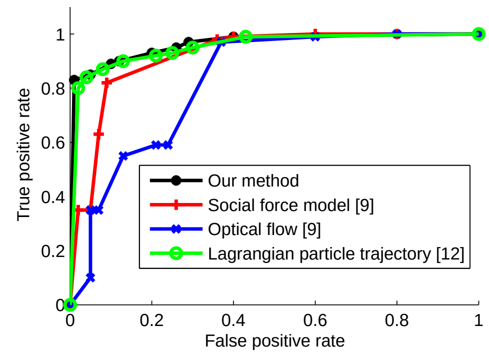

Fig. 8. The ROC curves of our method, social force model [9], optical flow [9], and lagrangian particle trajectory [12]

4. CONCLUSION

In this paper, we have proposed a novel framework to detect anomaly in crowded scenes. By modeling trajectories of the crowd of normal activities, we identify anomalies by probability perspective. Experimental results on the well-known dataset have demonstrated that the proposed method can improve the accuracy of detecting anomalies in crowded scenes. The drawback of the proposed method is that it can not locate the abnormal regions in the crowded scene. we will handle the problem in later research.

5. ACKNOWLEDGEMENTS

This work was supported in part by the National Natural Science Foundation of China under Grant 61303095, in part by Research Fund for the Doctoral Program of Higher Education of China under Grant 20133108120017, in part by Innovation Program of Shanghai Municipal Education Commission under Grant 14YZ018, in part by Innovation Program of Shanghai University under Grant SDCX2013012 and in part by Cultivation Fund for the Young Faculty of Higher Education of Shanghai under Grant ZZSD13005.

6. REFERENCES

[1] Bin Zhao, Li Fei-Fei, and Eric P Xing, "Online detection of unusual events in videos via dynamic sparse coding," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 3313–3320.

[2] Tao Xiang and Shaogang Gong, "Video behavior profiling for anomaly detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 30, no. 5, pp. 893–908, 2008.

[3] Hanning Zhou and Don Kimber, "Unusual event detection via multi-camera video mining," in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*. IEEE, 2006, vol. 3, pp. 1161–1166.

[4] Mark Jager, Christian Knoll, and Fred A Hamprecht, "Weakly supervised learning of a classifier for unusual event detection," *Image Processing, IEEE Transactions on*, vol. 17, no. 9, pp. 1700–1708, 2008.

[5] Peter Tu, Thomas Sebastian, Gianfranco Doretto, Nils Krahnstoever, Jens Rittscher, and Ting Yu, "Unified crowd segmentation," in *Computer Vision–ECCV 2008*. 2008, pp. 691–704, Springer.

[6] Mikel Rodriguez, Ivan Laptev, Josef Sivic, and J-Y Audibert, "Density-aware person detection and tracking in crowds," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 2423–2430.

[7] Bolei Zhou, Xiaoou Tang, and Xiaogang Wang, "Measuring crowd collectiveness," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 3049–3056.

[8] Bolei Zhou, Xiaoou Tang, and Xiaogang Wang, "Measuring crowd collectiveness," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pp. 3049–3056. 2013.

[9] Ramin Mehran, Alexis Oyama, and Mubarak Shah, "Abnormal crowd behavior detection using social force model," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 935–942.

[10] Louis Kratz and Ko Nishino, "Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 1446–1453.

[11] Vijay Mahadevan, Weixin Li, Viral Bhalodia, and Nuno Vasconcelos, "Anomaly detection in crowded scenes," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 1975–1981.

[12] Shandong Wu, Brian E Moore, and Mubarak Shah, "Chaotic invariants of lagrangian particle trajectories for anomaly detection in crowded scenes," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 2054–2060.

[13] Bruce D Lucas, Takeo Kanade, et al., "An iterative image registration technique with an application to stereo vision.," in *IJCAI*, 1981, vol. 81, pp. 674–679.

[14] Saad Ali and Mubarak Shah, "Floor fields for tracking in high density crowd scenes," in *Computer Vision–ECCV 2008*. 2008, pp. 1–14, Springer.

[15] Mikel Rodriguez, Saad Ali, and Takeo Kanade, "Tracking in unstructured crowded scenes," in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 1389–1396.

[16] Xiaolin Li, Marc Parizeau, and Réjean Plamondon, "Training hidden markov models with multiple observations-a combinatorial method," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 4, pp. 371–377, 2000.

[17] "Unusual crowd activity dataset made available by the university of minnesota at: http://mha.cs.umn.edu/movies/crowdactivity-all.avi.5," .