VIDEO ANOMALY DETECTION IN SPATIOTEMPORAL CONTEXT

Fan Jiang[1], *Junsong Yuan*[2], *Sotirios A. Tsaftaris*[1], *and Aggelos K. Katsaggelos*[1]

[1]Dept of EECS, Northwestern University
2145 Sheridan Rd, Evanston, IL 60208
{fji295, stsaft, aggk}@eecs.northwestern.edu

[2]School of EEE, Nanyang Technological University
50 Nanyang Ave, Singapore, 639798
jsyuan@ntu.edu.sg

ABSTRACT

Compared to other approaches that analyze object trajectories, we propose to detect anomalous video events at three levels considering spatiotemporal context of video objects, i.e., point anomaly, sequential anomaly, and co-occurrence anomaly. A hierarchical data mining approach is proposed to achieve this task. At each level, the frequency based analysis is performed to automatically discover regular rules of normal events. The events deviating from these rules are detected as anomalies. Experiments on real traffic video prove that the detected video anomalies are hazardous or illegal according to the traffic rule.

1. INTRODUCTION

Detecting anomalous events from surveillance video is a challenging problem due to the inherent difficulty in defining anomaly explicitly. The more practical approach is to detect normal events first (as they follow some regular rules) and treat the rest as anomalies. In many cases, however, no prior knowledge about the regular rules exists and no training data for normal video events are available.

To address this problem, the clustering-based approach has been investigated in the literature [1–7], which is based on the fact that normal events appear frequently and dominate the data, while anomalies are different from the commonality and appear rarely. For instance, a car moving against the direction of most other moving vehicles could indicate an anomalous event. Therefore, unsupervised clustering can be performed on all video events. Those events clustered into large groups can be identified as normal, while those outliers distant from all cluster centers are defined as anomaly.

Despite the success of clustering-based approaches for anomaly detection, there are several limitations. Most clustering approaches consider a video event as the motion trajectory of one single object [2–4, 7]. However, this definition neglects some spatial and temporal context information. On one hand, video anomaly may not correspond to the whole trajectory, just to a part of it. On the other hand, anomaly can arise due to the inappropriate interactions among multiple objects (i.e., multiple trajectories), even though their individual behaviors are normal. Thus, anomaly detection based on trajectory clustering can result in miss detection.

Instead of analyzing solely trajectories, in our work we define video events at different levels considering both spatial and temporal context. At each level, frequency based analysis is performed. Events appearing with high frequency are automatically discovered and declared to be an explicitly description of the regular rules. The events deviating from these rules are detected as anomalies. We test the proposed approach on real traffic videos, where vehicles have been detected and tracked. The task is to discover anomalous events from a collection of movement trajectories of vehicles. The results show that our approach can automatically infer regular rules of traffic motion of the specific scene (corresponding to the real traffic rules) and detect anomalous events at three levels: motion of one vehicle at any time instance, motion of one vehicle within a time range, and co-occurrence of multiple vehicles. Most of the detected video anomalies are proved to be hazardous or illegal, according to vehicular traffic rules.

2. POINT ANOMALY DETECTION

In a video scenario with moving objects (vehicles, humans, etc.), the most easily observed activity is the instant behavior of any single object i at any time instance t, which we categorize as an *atomic event* $e_a(i,t)$. Typically, an atomic event describes the location, moving direction, and velocity of the object at each video frame. It is the basic unit for describing more complicated events and interactions. As most object follow some regular motion rules, anomalous atomic events (referred to as point anomaly) can be detected based on their low frequency of appearance. After this step we can readily detect some obvious anomalies from the video, and exclude them from subsequent analysis.

3. SEQUENTIAL ANOMALY DETECTION

A video anomaly may not only consist of instantaneous behavior, but may also be characterized by the ordering or transition of instantaneous behaviors. For example, in a traffic scenario, two atomic events, such as entering an intersection from straight-only lane and making a left turn within the intersection, can be normal. But their combination is anomalous (illegal). In order to exploit this temporal context, we define a *sequential event* $e_s(i)$ as a sequence of atomic events associated with the trajectory of an object i. Note that the same atomic event appearing continuously is regarded as only one item in the sequence. For example, $e_s(i)$ is represented by the sequence $\big(e_a(i,1), e_a(i,2), e_a(i,4), \cdots\big)$, if $e_a(i,3) = e_a(i,2)$.

A sequential anomaly can be identified by finding sequences that appear rarely. However, the challenge is that sequential anomaly may appear as part of the whole sequence, thus techniques are needed to deal with variations of time length. Another difficulty is the effect of noise when counting the frequency of similar sequences. We should allow for possible variation of the sequence.

To accommodate this design constraint, we adopt the technique of frequent subsequence mining (CloSpan by Yan et al. [8]). From all the sequential events collected from the video, this data mining technique discovers the frequent subsequences (instead of the whole sequences) with their frequency above a given threshold (short repetitive subsequences are filtered out). The resulting subsequences are regarded as patterns of normal sequential events. Then, we classify every sequential event to one of the normal patterns using the minimum edit distance [9], i.e., the minimum number of operations (insertion, deletion, or substitution of an atomic event) needed to transform one sequence into the other. Finally, those atomic events within a sequence, which need to be deleted to match the normal patterns, are detected as anomalies. Video anomaly detected at this level is referred to as sequential anomaly.

4. CO-OCCURRENCE ANOMALY DETECTION

The highest level of anomaly arises from the co-occurrence of multiple objects. For example, in a traffic scenario, turning left and going straight within the intersection are both normal events when considered independently; however, making a left turn in front of incoming traffic is illegal and thus anomalous. This co-occurrence anomaly usually happens in the area with multiple objects and intensive interactions, e.g., within a road intersection. Therefore, we define a co-occurrence event as an instant event for a specific area A of every video frame. As every object appearing in this area has a label of sequential event pattern (sequential anomaly excluded), a *co-occurrence event* $e_c(t)$ can be represented as a set, i.e., $\{e_s(i)|$ all i appearing in area A at $t\}$.

Treating each co-occurrence event as a transaction, we apply the frequent itemset mining algorithm [10, 11] on all co-occurrence events collected from the video. The discovered frequent subsets of co-occurrences (frequency above a given threshold) are treated as normal patterns of co-occurrences. In order to classify each co-occurrence event to one of the normal patterns, we may simply perform nearest neighbor classification considering the overlapping extent of two sets. Nevertheless, it neglects the temporal consistency of co-occurrence events through video stream. For example, in traffic video of road intersection, as a result of the traffic light signaling there exist a few traffic states and specific ways of transitioning between the states. The classification of co-occurrences at every time is subject to this state transition.

Actually, the co-occurrences at all times $\{e_c(t)\}$ can be considered as an observation sequence Y generated from a hidden Markov model (HMM). The states of the HMM correspond to traffic states, and should be labeled as one of the normal co-occurrence patterns. There exist a certain probability of state transition $\{a_{ij}\}$ (state i to state j) and emission probability $\{b_j(t)\}$ (state j generating co-occurrence $e_c(t)$). Therefore, co-occurrence classification becomes an HMM decoding problem. At first step we need to determine the parameters of the HMM. However, the conventional forward-backward algorithm is not applicable as $\{b_j(t)\}$ cannot be specified. $b_j(t)$ is a discrete probability distribution but with infinite number of observed values, because $e_c(t)$ may consist of an arbitrary number of sequential events at time t. Thus in a forward algorithm, we cannot calculate the exact $\alpha_j(t)$, the probability that the HMM in state j at step t having generated the first t observances.

Fortunately, the Viterbi algorithm does not rely on the exact value of $\alpha_j(t)$ but on the comparison among all α's, because it only needs to choose one path at each step that maximizes α. Thus we propose a model of emission probability and use the Viterbi algorithm iteratively for state labeling.

First, we assume that for any co-occurrence $e_c(t)$, the probabilities that state i or j having generated it are proportional to the number of items in $e_c(t)$ that belong to i or j. In other words, the emission probabilities of different states i, j for the same $e_c(t)$ satisfy $\frac{b_i(t)}{b_j(t)} = \frac{m_i(t)}{m_j(t)}$, where $m_j(t)$ is the number of items in $e_c(t)$ that belong to pattern j. For example, $e_c(t) = \{1,1,2,2,2,3,4,5,5\}$ ($1,2,3,\cdots$ are sequential event labels), the states $i = \{1,2,3\}, j = \{3,4,5\}$. We have $\frac{b_i(t)}{b_j(t)} = \frac{6}{4}$. Thus, for any state j, $b_j(t) = c(t) \cdot m_j(t)$, where $c(t)$ is a constant for different states. Based on this modeling of $\{b_j(t)\}$, applying the Viterbi algorithm, we have

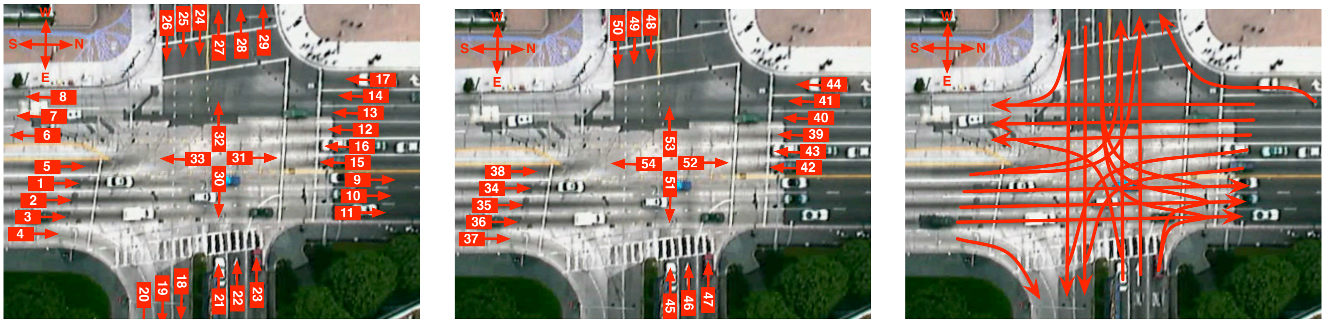
$$\alpha_j(1) = b_j(1) = c(1) \cdot m_j(1), \tag{1}$$

$$\begin{aligned}\alpha_j(2) &= b_j(2) \sum_i \big(\alpha_i(1) a_{ij}\big) \\ &= c(1)c(2) \cdot m_j(2) \sum_i \big(m_i(1) a_{ij}\big),\end{aligned} \tag{2}$$

$$\cdots$$

$$\alpha_j(t) = \Big(\prod_{i=1}^{t} c(i)\Big) \cdot M\Big(\{m_j(i)\}_{i=1}^{t}, \{a_{ij}\}\Big) \tag{3}$$

Note that the $c(t)$ term is constant and M is only related to $\{m_j(i)\}$ and $\{a_{ij}\}$. Therefore, $\alpha_j(t)$ can be compared for all



(a) Frequent atomic events (moving)

(b) Frequent atomic events (stopping)

(c) Frequent sequential events

Fig. 1. Frequent atomic events (a)(b) and frequent sequential events (c)

j at any time t without knowing $\{b_j(t)\}$ exactly.

Then, we use the following iterative method to determine $\{a_{ij}\}$ and to label states, i.e.,

1. Set initial $\{a_{ij}\}$ to uniform distribution;
2. Estimate states by Viterbi algorithm based on (3) ;
3. Estimate $\{a_{ij}\}$ by taking the ratio between the number of transitions from state i to state j and the total number of any transitions from state i. Go to 2) until convergence is reached.

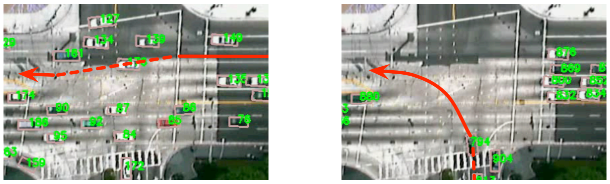
Once each co-occurrence event is labeled, anomalies can be easily detected by finding those items that are different from its corresponding normal co-occurrence pattern.

5. EXPERIMENTAL RESULTS

The proposed anomaly detection approach was tested on an one hour long surveillance video, monitoring traffic at a road intersection (from http://ngsim.camsys.com/). Example frames are shown in Fig. 1. Traffic motion is guided by traffic lights within the intersection and detailed trajectory information for every vehicle is available. However, with the information of traffic signaling unknown, our goal is to discover traffic rules followed by most vehicles in this area and to detect anomalies at different levels.

For the point anomaly detection, we represent each atomic event by three discrete feature, i.e., the position of the vehicle (the specific lane or intersection it occupies), the driving direction (north, south, west, east), and the velocity (either moving ($v \gg 0$) or stopping ($v \approx 0$)). A 3-D histogram for all the atomic events throughout the video is established. By applying a threshold (10% of the average bin height in the experiments), we detect 54 frequent (normal) behaviors, as shown and numbered in Fig. 1(a)(b). The atomic events that do not fall in any of them are detected as point anomaly.

For sequential anomaly detection, using frequent subsequence mining (the threshold is set to 1% of the total sequence number), we detect 44 frequent (normal) sequential patterns, with some of them shown in Fig. 1(c). It is observed that all the possible traveling routes permitted in this area are included. Fig. 2 illustrates two examples of detected sequential anomaly based on these routes. Fig. 2(a) shows a vehicle changing lane within the intersection. Fig. 2(b) shows a vehicle making a left turn from a no-turn lane. The anomalous events are shown with dashed lines.



(a)

(b)

Fig. 2. Examples of sequential anomaly

Finally, a co-occurrence event is represented by all sequential pattern labels of the vehicles appearing within the intersection at each frame. By applying frequent itemset mining on all co-occurrences (the threshold is set to 1% of the total co-occurrence number), we detect 5 frequent co-occurrence events, actually corresponding to the 5 states generated from traffic light signals. Fig. 3 depicts the driving directions allowed for each state. In the iterative approach of state labeling, we observe the evolution shown in Fig. 5(a). $P(Y)$, the probability of the whole sequence Y generated from the HMM keeps increasing, while the error of transition probability keeps decreasing, until they both converge. Finally, the co-occurrence anomaly is detected, with examples shown in Fig. 4. Fig. 4(a) shows a vehicle turning right while there is left-turning traffic going to the same lane. Fig. 4(b) shows a vehicle going beyond the waiting line trying to turn left while there is incoming traffic going straight. The anomalous parts are shown in dashed lines too.

For the three types of anomaly detection, determining the threshold is an important issue. To further test the robustness of our approach, we vary the threshold and plot ROC curves. The ground truth is acquired by manually labeling all the

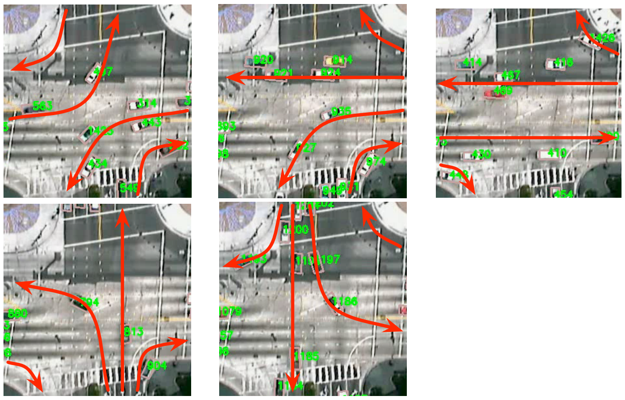
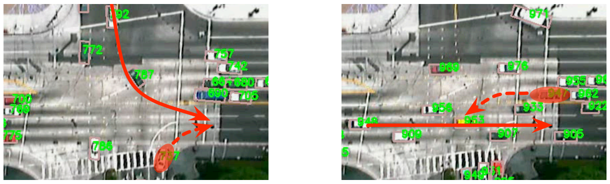


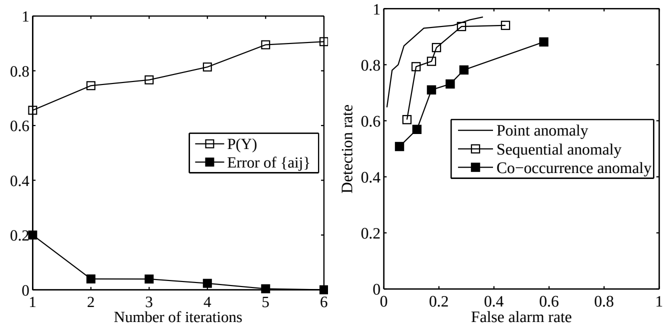
Fig. 3. 5 normal co-occurrence patterns with their example frames



(a) (b)

Fig. 4. Examples of co-occurrence anomaly

events. From Fig. 5(b) we observe that our detection performs well when the threshold is properly set, with a typical detection rate above 90% for point anomaly, above 80% for sequential anomaly, and above 70% for co-occurrence anomaly.



(a) (b)

Fig. 5. Convergence of iteration (a) and ROC curves (b).

6. CONCLUSION

With no prior knowledge about anomalous behaviors in a specific video scenario, we must resort to an unsupervised approach to automatically detect video anomalies from the data. Our approach is data-driven and applicable to many different scenarios. Rules of normal motion, considering the spatiotemporal context of visual objects, are discovered. Video anomalies detected based on these rules proved to be real hazardous or illegal behaviors in our experiments.

7. ACKNOWLEDGEMENT

This work was supported in part by a grant from CCITT/USDoT. Junsong Yuan is supported in part by the Nanyang Assistant Professorship (SUG M58040015).

8. REFERENCES

[1] H. Zhong, J. Shi, and M. Visontai, “Detecting unusual activity in video,” in *Proc. IEEE Conf. on Comput. Vision and Pattern Recognition*, June 2004, vol. 2, pp. 819–826.

[2] F. Porikli and T. Haga, “Event detection by eigenvector decomposition using object and frame features,” in *Proc. IEEE Conf. on Comput. Vision and Pattern Recognition Workshops*, June 2004, vol. 7, pp. 114–124.

[3] D. Makris and T. Ellis, “Learning semantic scene models from observing activity in visual surveillance,” *IEEE Trans. Syst., Man, Cybern. B*, vol. 35, no. 3, pp. 397–408, 2005.

[4] W. Hu, X. Xiao, Z. Fu, D. Xie, T. Tan, and S. Maybank, “A system for learning statistical motion patterns,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 9, pp. 1450–1464, 2006.

[5] T. Xiang and S. Gong, “Video behavior profiling for anomaly detection,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 5, pp. 893–908, 2008.

[6] B. T. Morris and M. M. Trivedi, “Learning, modeling, and classification of vehicle track patterns from live video,” *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 3, pp. 425–437, 2008.

[7] F. Jiang, Y. Wu, and A. K. Katsaggelos, “A dynamic hierarchical clustering method for trajectory-based unusual video event detection,” *IEEE Trans. Image Process.*, vol. 18, no. 4, pp. 907–913, 2009.

[8] X. Yan, J. Han, and R. Afshar, “Clospan: Mining closed sequential patterns in large datasets,” in *Proc. IEEE Int’l Conf. on Data Mining*, 2003.

[9] V. I. Levenshtein, “Binary codes capable of correcting deletions, insertions, and reversals,” *Soviet Physics Doklady*, vol. 10, no. 8, pp. 707–710, 1966.

[10] T. Uno, M. Kiyomi, and H. Arimura, “Lcm ver.2: Efficient mining algorithms for frequent/closed/maximal itemsets,” in *Proc. IEEE Conf. on Data Mining Workshop on FIMI*, 2004.

[11] J. Yuan, Y. Wu, and M. Yang, “From frequent itemsets to semantically meaningful visual patterns,” in *Proc. ACM SIGKDD Conf. on Knowl. Discovery and Data Mining*, 2007, pp. 864–873.