Modeling Patterns of Activity and Detecting Abnormal Events with Low-level Co-occurrences

Yannick Benezeth, Pierre-Marc Jodoin, Venkatesh Saligrama

Chapter 1

MODELING PATTERNS OF ACTIVITY AND DETECTING ABNORMAL EVENTS WITH LOW-LEVEL CO-OCCURRENCES

Yannick Benezeth
ENSI de Bourges - Institut PRISME
88 bd. Lahitolle - 18000 Bourges, France
yannick.benezeth@ensi-bourges.fr

Pierre-Marc Jodoin
Université de Sherbrooke
2500 bd. de l'Université Sherbrooke, J1K 2R1, Canada
pierre-marc.jodoin@usherbrooke.ca

Venkatesh Saligrama
Boston University - Department of Electrical and Computer Engineering
8 Saint Mary's Street, Boston, MA 02215, USA
srv@bu.edu

Abstract We explore in this chapter a location-based approach for behavior modeling and abnormality detection. In contrast to conventional object-based approaches for which objects are identified, classified, and tracked to locate objects with suspicious behavior, we proceed directly with event characterization and behavior modeling using low-level features. Our approach consists of two-phases. In the first phase, co-occurence of activity between temporal sequences of motion labels are used to build a statistical model for normal behavior. This model of co-occurrence statistics is embedded within a co-occurence matrix which accounts for spatio-temporal co-occurence of activity. In the second phase, the co-occurence matrix is used as a potential function in a Markov Random Field framework to describe, as the video streams in, the probability of observing new volumes of activity. The co-occurence matrix is thus used for detecting moving objects whose behavior differs from the ones observed during the training phase. Interestingly, the Markov Random Field distribution implicitly accounts for speed, direction, as well as the average size of the objects without any higher-level intervention. Furthermore, when the spatio-temporal volume is large enough, the co-occurrence distribution contains the average normal path followed by moving objects. Our method has been tested on various outdoor videos representing various challenges.

Keywords: video surveillance, abnormality detection, motion detection, MRF

Introduction

In this paper, we present a low-level location-based approach for activity analysis and abnormal detection. In several traditional approaches (e.g. [2]), moving objects are first detected, analyzed and then tracked. Subsequently, behavior models are built based on object tracks and non-conformant ones are deemed abnormal. The main problem with this approach is that in case of complex environments, object extraction and tracking are performed directly on *cluttered* raw video or motion labels. We propose performing activity analysis and abnormal behavior detection first, followed possibly by object extraction and tracking. If the abnormal activity is reliably identified, then object extraction and tracking focus on *region of interest* (ROI) and thus is relatively straightforward, both in terms of difficulty and computational complexity, on account of sparsity and absence of clutter. A question arises: *How to reliably identify abnormalities from a raw video?*

Some approaches have been proposed to perform such low-level abnormality detection (see for instance [3] and [4]). Nevertheless, we point out that these methods process each pixel independently and thus ignore spatial correlation across space and time. These correlations may not only be important in improving false alarms and misses but also in detecting abnormality of event sequences, such as a person in the act of dropping a baggage, tracking the person who dropped the baggage, a car making an illegal U-turn, etc. In our method, we account for these scenarios through spatio-temporal models. Although this model is simple, it results in extremely interesting results (see figures 1.3, 1.6 and 1.7). Note that our scheme does not rely on object tagging, tracking or classification. Furthermore, the co-occurrence can be readily generalized to higher-dimensions and other interesting features can be augmented.

1. Context, Overview and Notations

Context

Motion labels obtained from background subtraction are atomic information often used by surveillance applications. A comparative study of background subtraction methods can be found in [21]. In this paper, we implemented a basic background subtraction method based on a Euclidean distance metric. Although many video analytics methods use it only in early stages of processing (mainly to locate moving objects) we argue that motion labels carries fundamental information on the content of the scene and thus, can be used to perform high-level tasks. Motivated by this perspective, some authors have already shown that low-level motion label can be used to summarize videos [19], recognize human movements [20] and detect abnormalities [4].

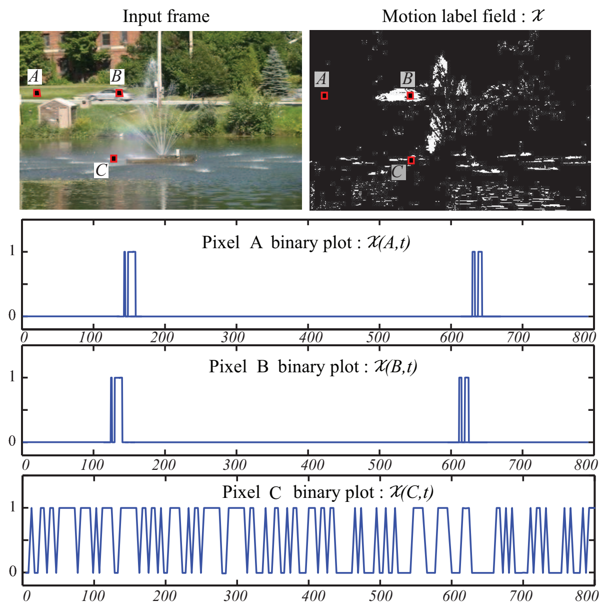


Figure 1.1. Binary signature for three pixels, two being highly correlated (A and B). The abscissa axis is *w.r.t.* time.

In general, motion label sequences provide valuable information for characterizing "usual behavior" observed at each pixel. For instance, consider patterns associated with random activity (shaking tree), regular activity (highway traffic), bursty activity (due to traffic light), or simply inactivity. All of these scenarios are characterized by patterns of motion label sequences at the pixel-level (or in general location). Consequently, abnormal behavior can be detected using low-level features whenever the observed pattern is unlikely under the normal activity model. In these cases, object identification and tracking can be circumvented for detecting abnormal behavior.

However, the pure pixel-by-pixel approach is insufficient in applications where abnormality is manifested spatially as, for instance, cars running against traffic flow, cars making illegal U-turns, etc. Consequently, we need a strategy for incorporating spatial patterns in addition to the temporal patterns of motion label sequences. The shortcomings of characterizing purely temporal behavior is further depicted in Fig. 1.1, which shows two pixels with identical signatures (except for a time-shift arising from cars going from right to left). Normal/Abnormal behavior arising from the pattern of activity between the two pixels cannot obviously be captured through a purely pixel-by-pixel analysis. For instance, a burst of activity occurring at pixel A before pixel B would mean that a car now runs from left to right.

Overview and Notation

The reader can follow the upcoming exposition through Fig. 1.2. Let $I_{\vec{x},k}$ be the luminance (or color) of a video sequence sampled on a 2-D lattice of size $W_0 \times H_0$ at discrete times k, i.e., $\vec{x} \in W_0 \times H_0 \subset R^2$, $k \in Z^+$. To simplify notation, we use s to denote the pixel location $\vec{x}$ at time t. X_s is the motion label estimated through simple background subtraction where $X_s \in \{0, 1\}$, 0 and 1 denoting the "inactive" and "active" states. We also define the motion labels sequence centered at $s = (\vec{x}, t)$ as being $\vec{X}_{\vec{x},t} = [X_{\vec{x},t-\eta}, \ldots, X_{\vec{x},t+\eta}]$ where $2\eta + 1$ is the length of the vector $\vec{X}_s$. In short, $\vec{X}_{\vec{x},t}$ is a one-dimensional binary sequence at pixel $\vec{x}$ and time t as shown in Fig.1.1. A contiguous sequence of ones denotes a busy period and is associated with a passing object while a sequence of zeros corresponds to an idle period of activity. The entire spatio-temporal sequence can be alternatively defined over a 3D lattice $\mathcal{S}$ of size $W_0 \times H_0 \times T_0$ with $s \in \mathcal{S}$ being a point in the spatio-temporal space, I_s being the corresponding luminance (or color) and X_s the corresponding motion label.

Now lets consider for each pixel $\vec{x}$ at time t, a spatio-temporal neighborhood centered at $s = (\vec{x}, t)$. This neighborhood is a 3D lattice $\mathcal{M}_s \subset \mathcal{S}$ with size $W \times H \times T$, $W < W_0$, $H < H_0$ and $T << T_0$, centered on $s \in \mathcal{S}$. Let us also consider a location $r = (\vec{y}, \tau) \in \mathcal{M}_s$ in the spatio-temporal neighborhood of $s = (\vec{x}, t)$. The spatial neighborhood of a pixel $\vec{x}$ is the set of all pixels $\vec{y}$ such that $s = (\vec{x}, t)$ and $r = (\vec{y}, \tau)$ are both in $\mathcal{M}_s$ for all t.

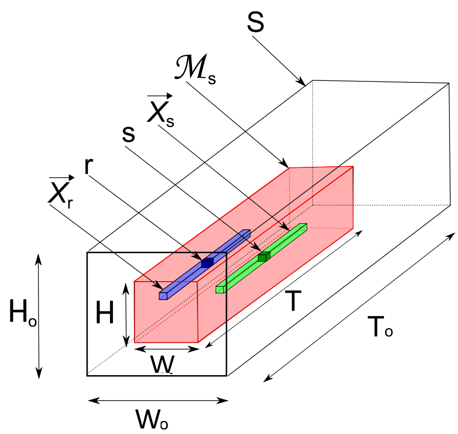


Figure 1.2. 3D lattice S with spatio-temporal neighborhood $\mathcal{M}_s$.

As we mentioned previously, whenever a moving object passes in front of $\vec{x}$ at time t, it leaves a spatio-temporal trace as some sites $r = (\vec{y}, \tau) \in \mathcal{M}_s$ co-occur with $s = (\vec{x}, t)$. Interestingly, several moving objects exhibiting regular behavior (think of cars on a highway going in the same direction) leave, after a while, similar traces in the spatial neighborhood of $\mathcal{M}_s$. Interestingly, the co-occurrence of two spatio-temporal neighbors s and r is not only due to the position and orientation of the camera in the scene, but also due to the shape, velocity and direction of the moving objects passing in front of a given spatial location $\vec{x}$. In this context, the goal of the co-occurence matrix is to estimate how frequently a site r co-occurs with s given a training video sequence exhibiting normal activity. We next define the notion of co-occurrence. A site $r \in \mathcal{M}_s$ co-occurs with s whenever their corresponding motion vector $\vec{X}_s$ and $\vec{X}_r$ exhibit a similar signature. The similarity between motion vectors at s and r is expressed using the mutual information defined as:

$$sim(\vec{X}_s, \vec{X}_r) = \sum_{x \in \{0,1\}} \sum_{y \in \{0,1\}} P_{\vec{X}_s \vec{X}_r}(x, y). \log \left(\frac{P_{\vec{X}_s \vec{X}_r}(x, y)}{P_{\vec{X}_s}(x) P_{\vec{X}_r}(y)} \right) \quad (1.1)$$

where $\vec{X}_s(i) = x$ and $\vec{X}_r(i) = y$, x and $y \in \{0, 1\}$, $i = [t - \eta, \ldots, t + \eta]$, $P_{\vec{X}_s \vec{X}_r}(x, y)$ is the joint probability of discret variables $\vec{X}_s$ and $\vec{X}_r$ and $P_{\vec{X}_s}(x)$ and $P_{\vec{X}_r}(y)$ are the marginal probabilities.

The mutual information is a useful tool for determining whether two motion labels sequences contain the same activity. For example, a temporal sequence of motion labels containing random values due to noise of false detections (caused, say, by an unstable background) will have a low mutual information with almost any other sequence. On the other hand, two sequences containing the trace left by the same moving object will have a large mutual information. In this way, the mutual information criteria minimizes the influence of spurious false detections and noisy environments.

2. Our Method

In this section, we present how, for a given site s, a co-occurrence matrix and the associated statistical model can be estimated from a training video sequence. Our statistical model is a Markov-Random Field (MRF) model that accounts for the likelihood of the co-occurrences. We later present how abnormal events can be detected and how low-level connected graphs can be used to follow relevant moving objects.

Training Phase

Nominal Model. Let O_s denote a motion label volume in the spatial-neighborhood of location s, *i.e.* $O_s = (X_r : r \in \mathcal{M}_s)$. We are interested in modeling the likelihood of the normal observations, *i.e.*, $P_N(O_s)$. We do this using an MRF model parameterized through co-occurrences:

$$P_N(O_s) = \frac{1}{Z} \exp\left(\sum_{u,v \in \mathcal{M}_s} \alpha_{uv} sim(\vec{X}_u, \vec{X}_v) \right) \tag{1.2}$$

where $sim(\vec{X}_u, \vec{X}_v)$ is the mutual information between motion label vectors $\vec{X}_u$ and $\vec{X}_v$. α_{uv} is the co-occurrence potential between site u and v determined in a learning phase as it will shortly be described (for the remainder of the paper, α_{uv} will be refereed to as the co-occurrence matrix). Z is the usual partition function, which is a normalization constant to ensure that the right hand side sums to one.

Learning the Co-Occurence Matrix. As mentioned previously, the co-occurence matrix α_{uv} accounts for how many times sites u and v co-occur during the training phase. Two sites are said to co-occur whenever their motion signature $\vec{X}_u$ and $\vec{X}_v$ exhibit a similar profile. In this paper, we measure the similarity between two sites based on their mutual information.

The co-occurrence matrix α_{uv} of two spatio-temporal locations, $u, v \in \mathcal{M}_s$ can be empirically computed as follows:

$$\alpha_{uv} = \frac{\beta_{uv}}{T_0 - T} \sum_{t=T/2}^{T_0 - T/2} sim(\vec{X}_u, \vec{X}_v) \tag{1.3}$$

where T_0 is the total number of frames in the training video sequence and β_{uv} is a constant that can depend on distance between the locations u and v (in this paper we assume $\beta_{uv} = 1$). Note that by definition, α_{uv} does not depend on the time index t. Therefore,

$$\alpha_{uv} = \alpha_{(\vec{y}_1, t+\tau_1),(\vec{y}_2, t+\tau_2)} = \alpha_{(\vec{y}_1, \tau_1),(\vec{y}_2, \tau_2)}. \tag{1.4}$$

A Specific Case for Co-Occurrence. Benezeth et al. [18] show that the co-occurence between two sites s an r can be determined by considering motion labels values X_s and X_r instead of the motion labels sequences $\vec{X}_s$ and $\vec{X}_r$. In this way, two sites co-occur whenever $\vec{X}_s = \vec{X}_r = 1$. In this case α_{uv} can be easily computed. However, this formulation is sensitive to noise and spurious false positives caused by unstable background. As can be seen in Fig. 1.7, accounting for plane co-occurrence between motion labels (third row) generates a large number of false positives and poor detection of true moving objects. This clearly shows how mutual information allows for *essential co-occurences*, *i.e* co-occurences caused by real moving objects only.

Complexity Issues & Conditional Independence. The main issue is the cost of computation of all of the edge potentials, since they are combinatorially many. In our practical implementations, we typically only consider a sparse number of well-separated locations for testing abnormalities. In many of our applications, abnormalities are typically associated with patterns of abnormal activity as opposed to inactivity. Motivated by this perspective, we make the following simplifying assumption: for any spatio-temporal neighborhood, $\mathcal{M}_s$ centered around $s = (\vec{x}, t)$, the co-occurrences are conditionally independent given X_s is active (namely $X_s = 1$). It will become clear why this assumption is not meaningful when $X_s = 0$. In other words, given X_s the values realized at the spatio-temporal locations X_v and X_u are statistically independent. Alternatively, one may think of this assumption as an instantiation of a naive Bayes perspective, namely, we assume that the pairwise co-occurrences in the spatial neighborhood of a location s are all independent. Practically, this assumption implies that we must have,

$$\alpha_{uv} = 0, \; u \neq s, \; v \neq s \tag{1.5}$$

In practice we have found this assumption does not severely degrade performance in our applications. Note that from a pure implementation perspective, the co-occurrence matrix $[\alpha_{uv}]$ is a 3D array with each component accounting for the number of times each site u co-occur with v while translating $\mathcal{M}_s$. In other words, Eq. 1.5 reduces the complexity of the method from $(W \times H \times T)^2$ pairwise co-occurences to consider down to $W \times H \times T$.

Observation Phase

Abnormal Model. It is generally difficult to describe an abnormality model except to say that abnormality is anything that does not look normal. However, from a classification perspective it becomes necessary to make some implicit assumptions about abnormality. Several researchers implicitly assume that abnormal observations are uniformly distributed in the feature space [1]. In this paper, we consider an action to be suspicious when its occurrence is rare (or simply inexistent) in the training sequence. We also assume that abnormal observations are independent and identically distributed across the different pixels. This assumption amounts to a multinomial distribution. For simplicity, let $N_0 = |\mathcal{M}_s|$ be the total number of spatio-temporal locations and N_1 the total number of co-occurring pixels, i.e.,

$$N_1 = \sum_{u \in \mathcal{M}_s} f(\vec{X}_u, \vec{X}_s) \tag{1.6}$$

with

$$f(\vec{X}_u, \vec{X}_s) = \begin{cases} 1 & \text{if } sim(\vec{X}_u, \vec{X}_s) > \tau \\ 0 & \text{otherwise} \end{cases} \tag{1.7}$$

then, the probability distribution of observations under the abnormal distribution is given by,

$$P_A(O_s) = p^{N_1}(1-p)^{N_0 - N_1} = \left(\frac{p}{1-p}\right)^{N_1} (1-p)^{N_0} \tag{1.8}$$

where, p is the probability that $f(\vec{X}_u, \vec{X}_s) = 1$

Abnormality Detection. Consider now a test video sequence $\mathcal{S}$ defined on a 3D lattice of size $W_0 \times H_0 \times T_{test}$, a spatio-temporal neighborhood $\mathcal{M}_s$ with $s = (\vec{x}, t)$ in the test video, and its corresponding motion-label observations O_s. The goal now is to detect every time instant $t \in [0, T_{test}]$ for which the observations O_s has a low probability under nominal distribution in comparison to likelihood of abnormality. It is well-known that the likelihood ratio test (LRT) is the optimal test for deciding between the two hypothesis: nominal vs. abnormal. The likelihood ratio $\ell(O_s)$ is the ratio of the probability of observations under nominal and abnormal hypothesis, from Eq. (1.2), (1.5) and (1.8), it follows:

$$\begin{aligned} \ell(O_s) &= \frac{P_N(O_s)}{P_A(O_s)} \qquad (1.9) \\ &= \frac{(1-p)^{N_0}}{Z} \exp\left(\sum_{r \in \mathcal{M}_s} \alpha_{sr} sim(\vec{X}_s, \vec{X}_r) - \log \frac{p}{1-p} \left(\sum_{r \in \mathcal{M}_s} f(\vec{X}_r, \vec{X}_s) \right) \right) \end{aligned}$$

where, as before, N_0 is the number of spatio-temporal locations and Z is a normalization constant.

The likelihood ratio test is to decide between nominal and abnormal hypothesis based on a global threshold η:

$$\ell(O_s) = \exp\left(\sum_{r \in \mathcal{M}_s} \alpha_{sr} sim(\vec{X}_s, \vec{X}_r) - \tau \sum_{r \in \mathcal{M}_s} f(\vec{X}_r, \vec{X}_s) \right) \underset{abnormal}{\overset{nominal}{\gtrless}} \eta \qquad (1.10)$$

where $\tau = \log(p/1-p)$. Here we have absorbed Z, p^{N_0} into η. A related test obtained by choosing $\eta = 1$ above reduces to a test for positivity or negativity of the argument of the exponential function. This reduces to the following simple test:

$$\frac{\sum_{r \in \mathcal{M}_s} \alpha_{sr} sim(\vec{X}_r, \vec{X}_s)}{\sum_{r \in \mathcal{M}_s} f(\vec{X}_r, \vec{X}_s)} \underset{abnormal}{\overset{nominal}{\gtrless}} \tau. \qquad (1.11)$$

3. Experimental results

We present in this section some results obtained on various outdoor sequences representing different challenges. For each sequence, a co-occurence matrix of size ranging between $130 \times 70 \times 300$ and $210 \times 210 \times 150$ have been used. The number of frames T used to estimate P_N (Eq. 1.2) varies between 2000 and 7000 (*i.e.* from 1 and 4 minutes of video) depending on the sequence. Note that results are presented in thumbnails of Fig. 1.3, 1.6 and 1.7. The green moving objects are ones classified as being normal and the red moving objects are those classified as being abnormal, i.e., whose trace is significantly different from the co-occurence matrix Eq. (1.11).

The first example (see Fig. 1.3) shows normal traffic and a car making an illegal U-turns. In Fig. 1.4(a), a co-occurrence matrix associated with a normal traffic flow is presented. As shown in Fig. 1.4(c), the trace left by the U-turn significantly differs from the usual traffic flow illustrated in the Fig. 1.4(b). Cars following the regular path are tagged in green and cars making an illegal U-turn are tagged in red.

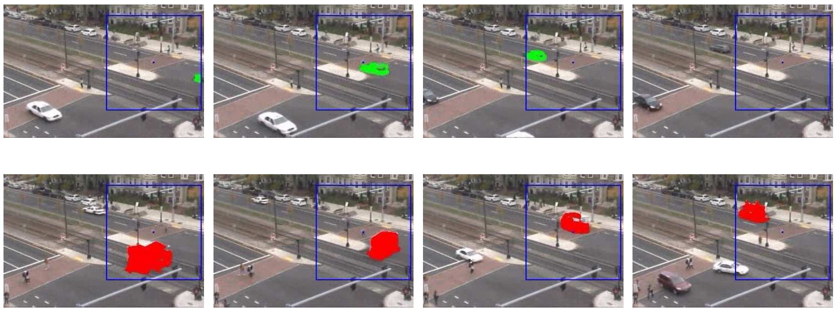
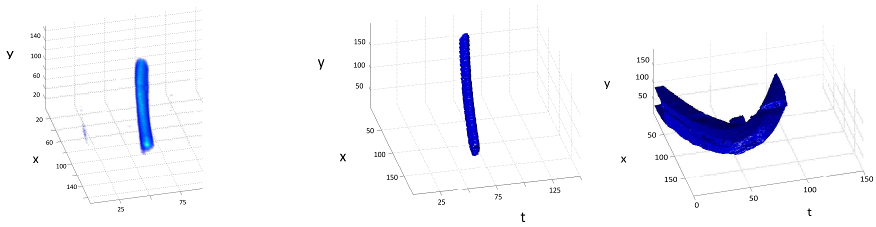


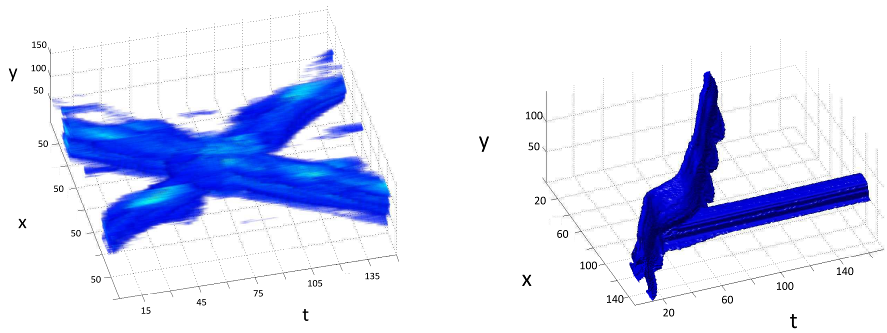
Figure 1.3. Example video in which cars following the regular traffic flow are tagged in green while the car making an illegal U-turn have been picked up by our algorithm and tagged in red.



(a) (b) (c)

Figure 1.4. (a) Co-occurrence matrix of a regular traffic flow (b) one car moving using the regular path (c) the trace left by a car making an illegal u-turn.

The second example shows a person dropping a baggage and abandoning it. In this video, pedestrians usually walk from left to right and from right to left, hence the X shape of the co-occurrence matrix (see Fig. 1.5(a)). When the person drops the bag, the abandoned package leaves a straight elongated line which differs from the co-occurrence matrix and thus causes this situation to be suspicious (see Fig. 1.5(b) and Fig. 1.6). Note that the likelihood ratio test (in Eq. 1.11) is computed only when the key-pixel is active. When considering Fig. 1.6, that is when the person passed through the key pixel, not before. This being said, with a connected component analysis (in both space and time) we are able to tag the overall action as being ”normal” or abnormal. That is the reason why the person dropping the bag is tagged in red during the overall action.



(a) (b)

Figure 1.5. (a) co-occurence matrix of pedestrians walking from left to right and from right to left and (b) the trace left by a person dropping a bag.

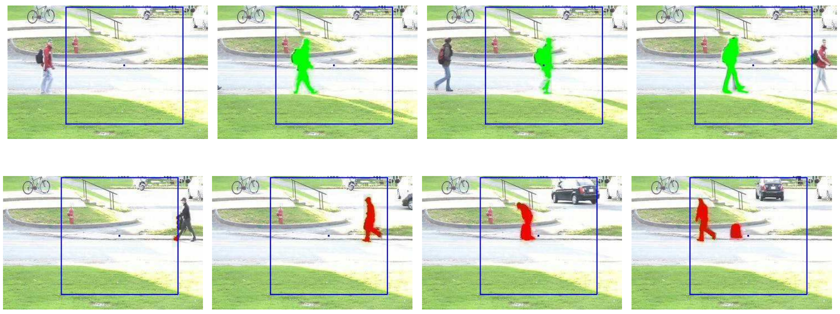


Figure 1.6. Example video in which people walking are tagged in green while the person dropping a bag is tagged in red.

The third example, in Fig. 1.7, shows how our method deals with noisy environments. The third row presents results obtained considering co-occurrences with motion labels [18] while the fourth row presents resuts obtained considering motion label vectors and mutual information. Clearly, the use of mutual information reduce the sensitivity to noise as the boat is clearly detected. The spatio-temporal trace left by the boat is shown in Fig. 1.8.

4. Conclusion

We propose in this chapter a method to perform behavior modeling and abnormality detection based on low-level characteristics. We use the spatial and temporal dependencies between motion label vectors obtained with simple background subtraction. To do this, we built an MRF model parameterized by a co-occurrence matrix. Although simple, this

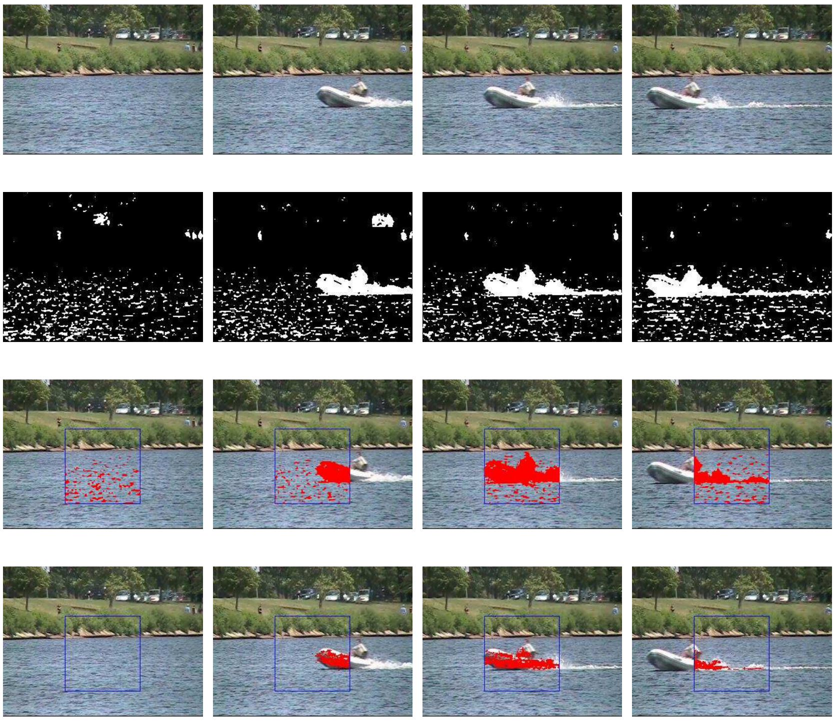


Figure 1.7. Illustration of detection in a challenging environment. The first row present the input images sequence, the second present the result of a background subtraction detection followed by the detection of abnormal activities (the boat displacement) considering co-occurrences with motion labels in the third row ([18])or considering motion label vectors in the fourth row.

matrix contains the average behavior observed in a training sequence. It also implicitly contains information about direction, speed and size of objects usually passing through one (or more) key-pixel(s). Equipped with the co-occurrence matrix, we can detect abnormal events by detecting traces which significantly differ from our nominal model following a likelihood ratio test.

The main advantages of our method are threefold. First, in contrast to conventional object-based approaches for which objects are identified, classified and tracked to locate those with suspicious behavior, we proceed directly with event characterization and behavior modeling using low-level characteristics and thus avoid the risk of errors propagation (e.g. due to the tracking algorithm limits in complex environments). Second, our method does not require any *a priori* knowledge about the abnormal event detection. We learn the usual behavior of moving objects

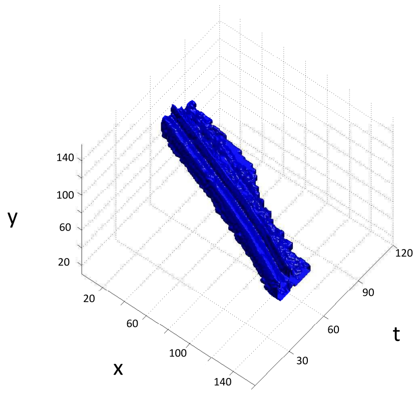


Figure 1.8. Trace left by the displacement of the boat detected using Eq. 1.7.

in a scene and detect activity which significantly differ from usual ones. Third, our method is robust to noise and can detect unusual activities using very noisy background subtraction masks.

References

[1] W. Polonik. "Minimum volume sets and generalized quantile processes". Stochastic Processes and their Applications, 69:1-24, 1997.

[2] W. Hu, T. Tab, L. Wang, S. Maybank, "A Survey on Visual Surveillance of Object Motion and Behaviors", transaction on System Man and Cybernetics Part C: Applications and Reviews 34 (3) (2004) 334-352.

[3] A. Adam, E. Rivlin, I. Shimshoni, D. Reinitz, "Robust Real-Time Unusual Event Detection Using Multiple Fixed-Location Monitors", Transaction on Pattern Analysis and Machine Intelligence 30 (3) (2008) 555-560.

[4] P.-M. Jodoin, J.Konrad, V. Saligrama, "Modeling Background Activity for Behavior Subtraction", International Conference on Distributed Smart Cameras.

[5] W. Zhao, R. Chellappa, P. Phillips, A. Rosenfeld, "Face recognition: A literature survey", ACM Computing Surveys 35 (4) (2003) 399-458.

[6] J. Konrad, "Motion detection and estimation", in: A. Bovik (Ed.), Handbook of Image and Video Processing, 2nd Edition, chap. 3.10, Academic Press, 253-274, 2005.

[7] N. Friedman, S. Russell, "Image Segmentation in Video Sequences: A Probabilistic Approach", international conference on Uncertainty in Articial Intelligence (1997) 175-181.

[8] I. Haritaoglu, D. Harwood, L. Davis, "W4: Real-Time Surveillance of People and Their Activities", Transaction on Pattern Analysis and Machine Intelligence 22 (8) (2000) 809-830.

[9] K. Smith, P. Quelhas, D. Gatica-Perez, "Detecting Abandoned Luggage Items in a Public Space", Performance Evaluation of Tracking and Surveillance Workshop (PETS) (2006) 75-82.

[10] S.-N. Lim, H. Fujiyoshi, R. Patil, "A One-Threshold Algorithm for Detecting Abandoned Packages Under Severe Occlusions Using a Single Camera", Tech. Rep. CS-TR-4784, University of Maryland, 2006.

[11] T. Chen, H. Haussecker, A. Bovyrin, R. Belenov, K. Rodyushkin, A. Kuranov, V. Eruhimov, "Computer vision workload analysis: case study of video surveillance systems", Intel Technology Journal 9 (2) (2005) 109-118.

[12] H. Boxton, "Learning and Understanding dynamic scene activity: A review", Image and Vision Computing 2003.

[13] N. Johnson, D. Hogg, "Learning the Distribution of Object Trajectories for Event Recognition", Image and Vision Computing 14 (8) (1996) 609-615

[14] I. Junejo, O. Javed, M. Shah, "Multi Feature Path Modeling for Video Surveillance", International Conference on Pattern Recognition (2004) 716-719.

[15] C. Stauer, E. Grimson, "Learning Patterns of Activity Using Real-Time Tracking", Transaction on Pattern Analysis and Machine Intelligence 22 (8) (2000) 747-757.

[16] W. Hu, X. Xiao, Z. Fu, D. Xie, T. Tan, S. Maybank, "A System for Learning Statistical Motion Patterns", Transaction on Pattern Analysis and Machine Intelligence 28 (9) (2006) 1450-1464.

[17] O. Boiman, M. Irani, "Detecting Irregularities in Images and in Video", International Journal on Computer Vision 74 (1) (2007) 17-31.

[18] Y. Benezeth, P.-M. Jodoin, V. Saligrama, C. Rosenberger, "Abnormal Events Detection Based on Spatio- Temporal Co-occurences", international conference on Computer Vision and Pattern Recognition (2009) 2458-2465.

[19] Y. Pritch, A. Rav-Acha, S. Peleg, "Non-Chronological Video Synopsis and Indexing", Transaction on Pattern Analysis and Machine Intelligence 30 (11) (2008) 1971-1984.

[20] A. F. Bobick, J. W. Davis, "The Recognition of Human Movement Using Temporal Templates", Transaction on Pattern Analysis and Machine Intelligence 23 (3) (2001) 257-267.

[21] Y. Benezeth, P.-M. Jodoin, B. Emile, H. Laurent, C. Rosenberger, "Review and evaluation of commonly-implemented background subtraction algorithms", International Conference on Pattern Recognition (ICPR), 2008.