GAUSSIAN PROCESS FOR ACTIVITY MODELING AND ANOMALY DETECTION

Wentong Liao[a], Bodo Rosenhahn[a], Michael Ying Yang[b]

[a] Institute for Information Processing, Leibniz University Hannover, Germany
[b] Computer Vision Lab, TU Dresden, Germany
liao@tnt.uni-hannovder.de, ying.yang1@tu-dresden.de

Commission WG III/3

KEY WORDS: Gaussian Process regression, activity modeling, anomaly detection

ABSTRACT:

Complex activity modeling and identification of anomaly is one of the most interesting and desired capabilities for automated video behavior analysis. A number of different approaches have been proposed in the past to tackle this problem. There are two main challenges for activity modeling and anomaly detection: 1) most existing approaches require sufficient data and supervision for learning; 2) the most interesting abnormal activities arise rarely and are ambiguous among typical activities, i.e. hard to be precisely defined. In this paper, we propose a novel approach to model complex activities and detect anomalies by using non-parametric Gaussian Process (GP) models in a crowded and complicated traffic scene. In comparison with parametric models such as HMM, GP models are non-parametric and have their advantages. Our GP models exploit implicit spatial-temporal dependence among local activity patterns. The learned GP regression models give a probabilistic prediction of regional activities at next time interval based on observations at present. An anomaly will be detected by comparing the actual observations with the prediction at real time. We verify the effectiveness and robustness of the proposed model on the QMUL Junction Dataset. Furthermore, we provide a publicly available manually labeled ground truth of this data set.

1 INTRODUCTION

Activity modeling and automatic anomaly detection in videos have become active fields in computer vision and machine learning because of the wide deployments of the surveillance cameras. These tasks remain difficult challenges in crowded and complicated scenes because of frequent occlusions among objects over time and space, different types of activities occurring simultaneously, noise caused by low video quality and the ambiguous definition of anomaly.

Modeling activities and connecting them to each other is one of the most important problems because moving agents normally have neither explicit spatial nor temporal dependencies. Traditionally, many researchers have concentrated on analyzing motion trajectories to model activities and interactions (Wang et al., 2011; Sun and Nevatia, 2013; Talha and Junejo, 2014; Wang et al., 2014). By means of tracking, the co-occurring activities are separated from each other. However, tracking-based approaches are very sensitive to tracking errors. If detection, tracking or recognition fails only in some frames, the future results could be completely wrong. They are only appropriate in a simple scene with a only few objects and clear behaviors. Hence, tracking does not work well in complex scenes of crowded motion, as indicated above.

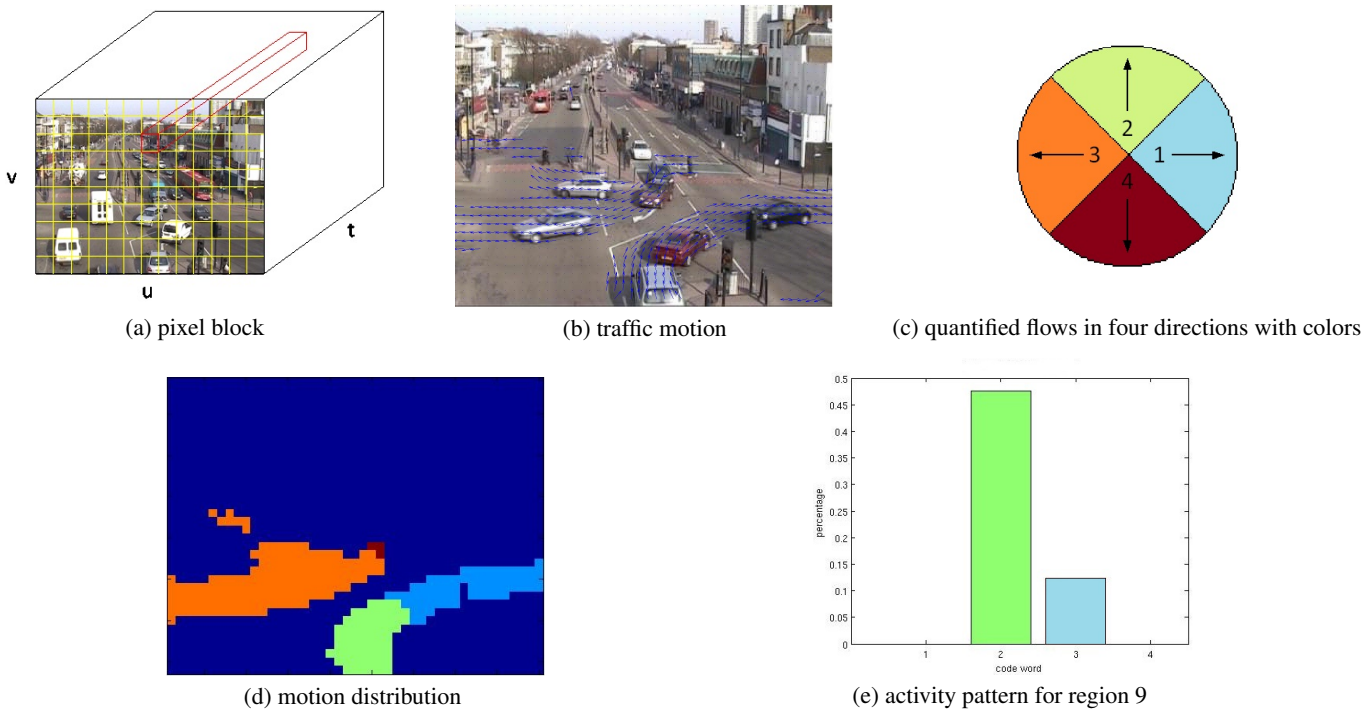
To tackle the problems of missed detection and broken tracks, many recent studies have focused on directly using low-level visual features (Kuettel et al., 2010; Saleemi et al., 2010; Hospedales et al., 2011). These studies develop more sophisticated statistical models to cluster typical activities: a generative statistic model firstly learns statistical information of typical activities. An abnormal activity is detected if it has low likelihood under a criterion. Typical approaches include Dynamic Bayesian Networks (DBNs) (Swears et al., 2014; Vo and Bobick, 2014) such as Hidden Markov Models (HMM) (Banerjee and Nevatia, 2014). The probabilistic topic models (PTMs) (Kinoshita et al., 2014) such as Latent Dirichlet Allocation (LDA) (Hospedales et al., 2011) or Hierarchical Dirichlet Process (HDP) (Kuettel et al., 2010) are powerful methods to learn activities in surveillance videos. However, inference in topic models is computationally consuming. Moreover, they are unsupervised that means the limitation of accuracy and precluding classification of activities.

Feature-based (descriptor) approaches are useful to model and recognize activities. They in general use an appropriate classifier to classify the learned activities and detect anomalies. The typical classifier includes GP classifier and SVM (Chathuramali and Rodrigo, 2012; Althloothi et al., 2014; Hasan and Roy-Chowdhury, 2014) are widely adopted because of their advantage in terms of high classification accuracy and relative simpler learning process. However, they are supervised models and a training data set with manually assigned labels is necessary in advance. Moreover, they have high requirement in the applicability and the preciseness of features to ensure their performance. The most widely used features include HOG features, optic flow based features, etc. However, in a complicated and crowded scene the low-level features are more reliable.

In this paper, we propose a novel approach to model activities and detect anomalies using non-parametric Gaussian Process (GP) regression models (Rasmussen and Williams, 2006). We model activities based on semantic region. The spacial dependencies among activities are connected using GP regression models and the temporal dependencies are modeled by using a one-step ahead prediction strategy. The understanding of these relationships is crucial for detecting subtle anomalies. After training, our model makes a predictive distribution of regional activities on each semantic region and compares with the actual observation. If the intensity of an observation is larger enough than the predictive distribution, which means that the relationship learned from the training data between different activity patterns is broken, some abnormal activities are occurring in the observed region.

Contributions. We propose a novel method based on non-parametric GP models to exploit spacial-temporal dependencies among ac-

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
Editors: A. Yilmaz, C. Mallet, M. Ying Yang, and Y. Vizilter
doi:10.5194/isprsannals-II-3-W5-467-2015



(a) pixel block (b) traffic motion (c) quantified flows in four directions with colors

(d) motion distribution (e) activity pattern for region 9

Figure 1: Representation of motion.

tivity patterns using Gaussian Automatic Relevance Determination (ARD) kernel function and model activities. Activities are predicted by learned GP regression models and anomalies are detected by comparing actual observations with prediction at real time. We furthermore provide a publicly available ground truth of the popular Traffic Junction Dataset (Hospedales et al., 2009). This ground truth has been manually labeled.

The remainder of this paper is organized as follows. In Section 2, we reflect related work in this application. In Section 3, we firstly review GP models, and then present how to model activities and detect anomalies with GP regression models in details. In Section 4, experimental results are presented and discussed. Finally, we conclude this paper in Section 5.

2 RELATED WORK

GP models have been widely used to tackle vision problems such as motion flow analysis (Kim et al., 2011), human motion analysis(Wang et al., 2008), tracking (Urtasun et al., 2006) and action recognition (Gong and Xiang, 2003). However, the visual data used in these cases are collected in controlled or good environments. They are relatively cleaner and simpler than the practical surveillance videos, which are much noisier. Most existing approaches use GP models to model activities and detect anomalies based on trajectory analysis (Kim et al., 2011) or feature-based methods. Although they have been proved to be effective methods for this task, their drawbacks caused by tracking are also obviously as discussed in Section 1.

To our best knowledge, only the work (Loy et al., 2009a) has attempted to use GP models for complex activity modeling and anomaly detection in a crowded traffic scene. It has proved that GP models outperform HMM on both sensitivity to anomaly and robustness to noise. However, their proposed model has drawbacks on activity representation and activity modeling with GP. (1) Two different motion features are assumed independent to each other. In practice there are implicit spacial-temporal dependencies between them. (2) Activity patterns of a region are represented only by the sum of the motion vectors in this region. For instance, a slight abnormal motion hardly affects the result of composition of motion vectors if there are relative strong motions in this region. This anomaly is covered by the motion stream and therefore ignored.

We adopt quantified directions to represent local motions instead of directly using optic flow. A regional activity is modeled using the distribution of quantified motions. The relationships between different activities in different regions are learned by GP regression model using the Automatic Relevance Determination (ARD) kernel (Chu and Ghahramani, 2005).

3 ACTIVITY MODELING AND ANOMALY DETECTION

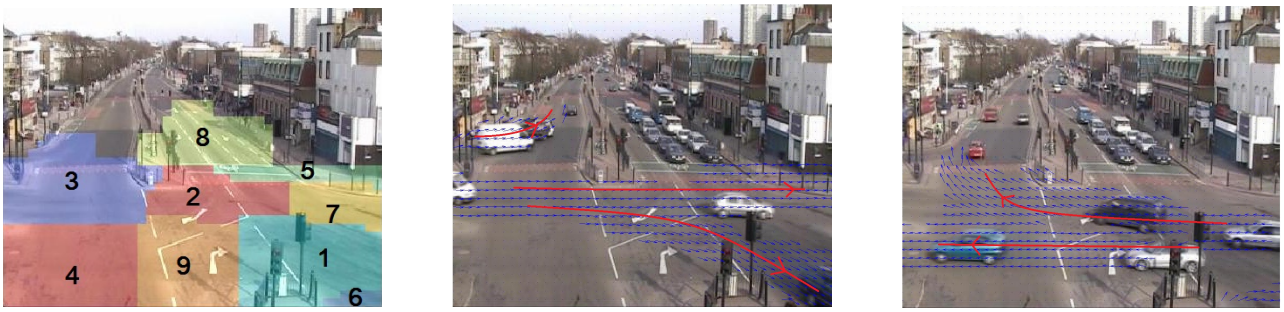
3.1 Gaussian Process Regression

A regression model $y = f(\mathbf{x}) + \epsilon$ is considered, where $\epsilon \sim N(0, \sigma^2)$ is an independent Gaussian white noise. Gaussian process $f(x)$ is specified by its mean function $m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})]$, in this study we assume the mean value as zero, and covariance function $K(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))] = \mathbb{E}[f(\mathbf{x})f(\mathbf{x}')]$. Then the observation vector $\mathbf{y} = \{y_1, \ldots, y_n\}$ distributes as a zero-mean multivariate Gaussian distribution and its covariance matrix is $\mathbf{K}^* = \mathbf{K} + \sigma^2\mathbf{I}$, where $\mathbf{K}$ denotes a $n - by - n$ covariance matrix between all pairs of training points and $\mathbf{K}_{ij} = K(\mathbf{x}_i, \mathbf{x}_j)$. For a test point $\mathbf{x}^*$, the posterior density $p(y^*|\mathbf{x}^*, \mathbf{x}, \mathbf{y})$ is a uni-variate normal distribution with the mean and variance as follow (Rasmussen and Williams, 2006):

$$\mathbf{m}(y^*) = \mathbf{k}(\mathbf{x}^*)^T\mathbf{K}^{*-1}\mathbf{y} \quad (1)$$

$$\mathbf{var}(y^*) = K(\mathbf{x}^*, \mathbf{x}^*) - \mathbf{k}(\mathbf{x}^*)^T\mathbf{K}^{*-1}\mathbf{k}(\mathbf{x}^*) \quad (2)$$

where $\mathbf{k}(\mathbf{x}^*) = [K(\mathbf{x}^*, \mathbf{x}_1), \ldots, K(\mathbf{x}^*, \mathbf{x}_n)]^T$ is a covariance vector between the test points $\mathbf{x}^*$ and the training points, and the



(a) scene decomposition (b) rightward flow (c) leftward flow

(d) vertical flow (e) left and right turn

Figure 2: (b)-(e) four main different traffic flows, they are sorted according to the traffic light cycle, (a) result of the semantic scene decomposition.

covariance of the test point $\mathbf{x}^*$ is $K(\mathbf{x}^*, \mathbf{x}^*)$.

A covariance matrix is given by the covariance function, which is also known as the kernel function. It is a crucial ingredient in GP models because it encodes our assumption on continuity and smoothness of the GP function $f(\mathbf{x})$. There are many popular functions (Rasmussen and Williams, 2006), such as linear, rational quadratic and neural network. In this paper, we choose ARD kernel function since it automatically determines the strength of relevance among different activities. We adopt the Conjugate Gradient method (Nocedal and Wright., 2006) to optimize the hyperparameters by maximizing the marginal log-likelihood.

3.2 Activity Representation

We divide the whole video into a sequence of 50-frame (2 seconds) clips and quantize the position by dividing the image space into $8 * 8$ blocks (see Figure 1(a)). We extract optical flow in each pair of consecutive frames using (Liu, 2009) over time with a threshold to reduce noise as shown in Figure 1(b). We accumulate the optical flow in each clip over its frames. The optical flow vectors of pixel blocks are the mean of the accumulated flow of its memberships. A local pixel block activity pattern is represented as a dual time series signal: $\mathbf{u} = \{u_t : t \in \tau\}$, $\mathbf{v} = \{v_t : t \in \tau\}$, u_t and v_t are horizontal and vertical flow components at clip t respectively, and τ is the total number of clips used for training.

(Loy et al., 2009b) present an approach to automatically decompose a complex scene into multiple semantic regions according to the spatial-temporal distribution of activity patterns. By means of this approach local activity patterns are modeled in each region without tracking or analyzing motion trajectories. With the extracted motion features, we use a method similar to (Loy et al., 2009b) to group blocks into semantic regions with the help of a spectral clustering algorithm (Zelnik-Manor and Perona, 2004) and we remove any region which has more than 90% of zero-activity blocks (background) (see Figure 2(a)).

In (Loy et al., 2009a), an activity in a region is represented by a moving vector (U, V), which are $U = \Sigma u_b$ and $V = \Sigma v_b$ respectively, where (u_b, v_b) is the flow vector of a pixel block in this region. However, this method does not separate different activities in a region. Furthermore, the sum of all motion vectors in a region may cause a zero result. In this case, there are actual objects moving in this region, but the motion feature U or V equals 0 or is close to 0. It will be considered as nothing happens here. On the other hand, an abnormal activity with small moving velocity is ignored when the normal motions in this region are strong. For instance, a jay-walker crosses a strong vehicle flow, but (U, V) cannot reflect this anomaly.

Our model represents activity patterns explicitly. The block optic flow vectors are quantified using a codebook with words, $\omega = (x, y, u, v)$, where (x, y) represents the position and (u, v) is the displacement, quantified into 4 directions (see Figure 1(c)). In such crowded scene, we rather care about objects moving direction than their velocity. We obtain a spacial distribution of these 4 motion features (see Figure 1(d)). The regional activity patterns are represented as the percentage of these 4 motion features $[d_i^1, d_i^2, d_i^3, d_i^4]$, where d_i^1 is the percentage of direction 1 in $i-th$ region and $d_i^1 + d_i^2 + d_i^3 + d_i^4 = 1$ or 0 (nothing happens). An example of the corresponding histogram for a regional activity pattern is shown in Figure 1(d). In this way, any existing motion is separated and detected.

3.3 Activity Modelling with Gaussian Process

Four GP regression models are constructed for each region to model features of 4 directions separately. Therefore we have $4N$ GP models for N decomposed regions. However, we firstly need to construct a reasonable input and output pair $(y_i, \mathbf{x}_i)$, where $\mathbf{x}_i$ is the input feature vector, y_i is the corresponding output value of a GP regression model and i is the index of data point.

In (Loy et al., 2009a) two GP regression models are constructed for each region for two features U and V separately. They

(a) emergency of ambulance

(b) emergency of fire engine

(c) emergency of fire engine 2

(c) emergency of police vehicle

Figure 3: Examples of detected anomaly caused by emergency vehicles using our GP models (abnormal regions are denoted in red whilst the corresponding abnormal object is highlighted with a box).

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
Editors: A. Yilmaz, C. Mallet, M. Ying Yang, and Y. Vizilter
doi:10.5194/isprsannals-II-3-W5-467-2015

(a) illegal u-turn

(b) jaywalking 1

(c) jaywalking 2

Figure 4: Examples of detected anomaly using our GP models (abnormal regions are denoted in red whilst the corresponding abnormal object is highlighted with a box).

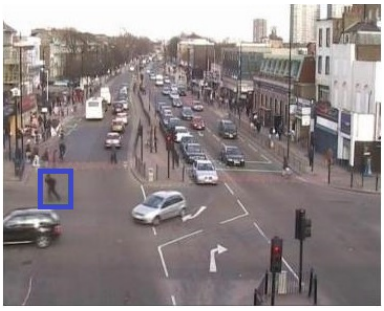
assume that the two motion features are mutually independent. Each of them is only affected by the same feature from other regions. For instance, $y_{i,t} = U_{i,t}$ is a output value of the model, where $U_{i,t}$ is observed at time interval t from the ith region r_i. $\mathbf{x}_{i,t} = \{U_{j,t-1} | j \neq i \& i, j \in N\}$ is the corresponding input feature vector, where $\{U_{j,t-1} | j \neq i, j \in N\}$ are observed at time interval $t-1$ from all other regions. The models for feature V are in the same way: $y_{i,t} = V_{i,t}$ and $\mathbf{x}_{i,t} = \{V_{j,t-1} | j \neq i, j \in N\}$. But actually U and V spacially and temporally relate to each other.

Our proposed method exploits the implicit spacial-temporal dependencies among different motion features of all regions in a novel way. For each motion feature $\{d_i^m | m \in [1,2,3,4]\}$, we select one of them, which can most affect d_i^m, from all other regions to construct the input feature vector: $\mathbf{x}_i = \{d_j^m | j \neq i \& i, j \in N\}$. They are determined by ARD kernel function as follow: $d_{i,t}^m$ is an output value of a GP model, which is observed at time interval t from ith region. The input feature vector for $d_{i,t}^m$ consists of four motion features observed at previous time interval $t-1$ from jth region $[d_{j,t-1}^1, d_{j,t-1}^2, d_{j,t-1}^3, d_{j,t-1}^4]$ ($j \neq i$). Therefore $d_{i,t}^m$ is one-step ahead predicted by this GP regression model from observed activity patterns in jth region at previous interval. The feature with the lowest length-scale in ARD kernel is selected because of its highest influence on $d_{i,t}^m$. All these selected features observed at interval $t-1$ from all other regions construct the global input feature vector $\mathbf{x}_{i,t}^m = \{d_{j,t-1}^m | j \neq i \quad and \quad i, j \in N, m \in [1,2,3,4]\}$ and $d_{i,t}^m$ is its corresponding output value of the GP regression model. Each GP regression model thus predicts one motion feature of the regional activity patterns at next interval using most relevant motion feature of other activity patterns observed at present in other regions.

3.4 Anomaly Detection

The output value of GP regression model has the highest possibility, when it equals the mean of the Gaussian distribution given by Eq. (1). And the output value has 95% of certainty to fall into area $(\mathbf{m} - 1.96\sigma^2, \mathbf{m} + 1.96\sigma^2)$. If the observed value (non-negative) of activity patterns at interval t is larger than $\mathbf{m} + 1.96\sigma^2$, this activity is detected as anomaly because an abnormal activity is caused by a suddenly increased motion strength. In contrast, a motion strength lighter than expectation is viewed as normal in this scene. In each detected interval, if any of the four motion features is detected as anomaly, its region is identified as abnormal.

(a) undetected jaywalking 1



(b) undetected jaywalking 2

(c) undetected bicycle in improper region

(d) undetected u-turn 1

(e) undetected u-turn 2

Figure 5: Examples of undetected anomaly using our GP models (the key abnormal object is highlighted with a box).

4 EXPERIMENTS

We conduct the experiments on the popular QMUL Junction Dataset (Hospedales et al., 2009), which contains a video of a crowded, traffic junction in London ($360 * 288$, 25fps, 89999frames). We used the first 10000 frames (200 clips) of the video for scene decomposition and training the GP regression models and the rest (1599 clips) for testing. The main challenges in this scene are: 1) low quality surveillance video and different kinds of noise, 2) irregular and violent varying of traffic volume over time, in particular at the entrances and exits, 3) large numbers of motion-agents and complex interactions among them, 4) some kind of ambiguous activities.

This traffic scene is controlled by traffic lights and dominated by four traffic flows (see Figure 2(b)-(e), they are sorted according to the traffic light cycle). Figure 2(a) shows the result of scene decomposition into 9 semantic regions. The decomposition agrees with the distribution of the main four traffic flows and human understanding. For example, regions 2 and 9 in the middle of the scene are the areas for vehicles from top and bottom, respectively, to wait for an interruption of vertical flow and make a left or right turn (as shown in Figure 2(e)). Region 5 covers the pedestrian crossing and so is the region 6 in the right bottom of the scene. Region 8 is only for vehicles which are driving downward or waiting for the traffic light for vertical flow. Regions 1 and 7 are entrance for leftward flow and exit for rightward flow, while regions 3 and 4 are exit for leftward flow and entrance for rightward flow (as shown in Figure 2(b) and (c), respectively). Region 4 is the entrance for vertical flow (as shown in Figure 2(e)). Interactions in these four regions are complex because during different phases of the traffic cycle they contain different interactions. But the activities within the same region are similar to each other and distinguish themselves from the others in other regions.

We adopt ARD kernel function to automatically learn the strength of correlation among activities in different regions and select the most relevant one, which have the lowest length-scale. For instance, the hyper-parameter of the GP model for feature d_3^2 (upward moving in region 3) shows that the length-scale of the four activity features in region 7 are $[3.7863, 0.8610, 0.03950, 2703.4]$, respectively. Feature d_7^3 has the lowest length-scale, interpreting that the leftward moving in region 7 has the highest influence on the upward moving in region 3. This understanding agrees with our human understanding of the traffic activity patterns in this scene because, if a vehicle drives towards left in region 7, it will cross region 2 and then turn to the upper region in region 3, as illustrated in Figure 2(d). The feature d_2^3 (leftward moving in region 2) has the length-scale of activity in region 7 as $[8.8726, 1.7201, 0.0259, 24.6840]$, respectively. Feature d_7^3 affects d_2^3 mostly, because the right-to-left vehicle flow in region 2 is always from region 7. It further proves that our method to construct the spacial-temporal relationship is appropriate and effective.

Jaywalking, illegal U-turn and emergency vehicles driving in an improper lane are the main anomalies in this scene. We have manually labeled all anomalies that appear in the whole video as ground truth and they will be made publicly available. Figure 3 shows some examples of typical anomalies caused by emergency vehicles. The ambulance in Figure 3(a) is driving in a definitely wrong direction in the lane for a normal case. It is the easiest kind of anomaly to detect. In Figure 3(b) and (c), the fire engine interrupts the current traffic state (vertical flow, and left and right turn, respectively). It increases the percentage of corresponding motion features in the regions which it passes by. Hence, these regions are detected that, there is an anomaly arising. The police vehicle is driving conversely as shown in Figure 3(c) and easy to be detected because the leftward motions are impossible in current traffic state. In the experiments, all of the anomalies caused by ambulances, fire engines and police vehicles were successfully detected by our GP models.

The typical abnormal activities- illegal U-turn and jaywalking in this scene are shown in Figure 4. When a vehicle wants to make a U-turn from region 2, it needs to firstly drive horizontally into region 3. This activity leads to a larger percentage of motion feature d^3 than the prediction in region 3. This is the principle,

how a U-turn is detected by our method. However, some of the illegal U-turn are undetected. For instance, a U-turn, as shown in Figure 5(d), is undetected because the pedestrian (highlighted in the red box) are crossing the road, which belongs to the second case of jaywalking, as mentioned above. The U-turn vehicle may be treated as a motion of pedestrian as well. Good training samples and object detection also can help us to overcome this problem. The U-turn is undetected either, if a large vehicle covers it (see Figure 5(e)). Those methods based on trajectories analysis outperform our methods in terms of detecting illegal U-turn because of their advantage in separation and analysis of an individual activity among the others. However, our model is more flexible because GP models does not require prior knowledge in advance.

Abnormal activities of Jaywalking are successfully detected as shown in Figure 4(b) and (c). The principle for this detection is also because of the percentage of the corresponding motion features higher than expectation of our GP models. However, Jaywalking is undetected in two cases: 1) the vehicle flow suspends and pedestrians cross the road during this time margin (Figure 5(a)). 2) the jay-walker is walking along the vehicle flow (like Figure 5(b)). Its motion does not obviously increase the percentage of any motion features because it has the same motions as the vehicle flow and its motion submerges in the flow; The second case can be avoided by means of object detection because pedestrians will be detected moving in an improper area or lane. The undetected anomaly of bicycle riding in a improper region (like Figure 5(c)) also can be detected with the help of object detection. The understanding of the second case is a challenge. Because suspending vehicle flow may be viewed as the end of current traffic flow. Thus, people crossing the road without vehicle is viewed as a normal activity. Furthermore, the bad training samples for this case are the other main reason. The pedestrians in this video sample do not obey the traffic rule so strictly. They usually cross the road in such case, even the traffic light is red for pedestrians. Therefore, the dynamic models such as HMM could work better for this case.

A human interpreted summary of the categories of abnormal events is given in Tab. 1. Notice that, each entire abnormal event is counted as one event, no matter how many clips it spans. The false detection means that, a clip is detected as an abnormal clip, but there is not any abnormal event of interest. The overall false positive rates is defined as:

$$FPR = \frac{\text{Number of falsely detected clips}}{\text{Number of test clips}}.$$

From the table we know, about 56% of all the abnormal events of interest were correctly detected, while 282 of all the test clips were falsely detected as abnormal. (Loy et al., 2009b) obtained a higher true positive rate 76.43% because our ground truth is different from theirs. For instance, they have only 8 cases of jaywalking while we have 14. Maybe the jaywalking in region 6 (Figure 2(a)) are not taken into account in their ground truth. Therefore, the comparison of experimental results with others methods maybe not objective because of subjective understanding about abnormal events, unless a uniform ground truth is public provided. By implementing the method in (Loy et al., 2009b) we found that, our method performed better in detecting subtle abnormal events such as the jaywalking in region 6, because the week motion vectors of the pedestrian were not cancelled by the strong motion vectors of the vehicle flow in our method.

Types	Detected	Total
Jaywalking	11	14
Emergency	5	5
Illegal U-turn	8	12
Strange driving	0	1
Improper region	0	2
False detect	282	\
Overall TPR	55.88%	
Overall FPR	15.6%	

Table 1: Summary of discovered abnormal events. Overall true positive (TPR) and false positive rates (FPR) are also given.

5 CONCLUSION AND FUTURE WORK

In this paper, we have proposed a novel method using GP regression models to model activities and detect anomalies beyond detection and tracking in a complex scene of crowded motion. In particular, our GP regression models are constructed to learn spacial-temporal relationship among regional activity patterns and predict its next-step distribution. Anomalies are detected by comparing prediction with actual observation. In the experiment we have demonstrated that, our method works reliably to detect anomalies.

However, we also see some limitations of our current model. Firstly, our detection is a semantic region based method. The spacial relationship is rough among activities. It depends on the results of decomposition of semantic regions. The abnormal activities cannot be exactly localized. Secondly, our model can only detect anomalies in the semantic regions. It does not support high-level semantic queries on activities and interactions. For example, what are the typical activities and interactions in this scene. Finally, because of no detection, our model cannot detect abnormal activities according to the categories of agents. For example, if a people riding a bicycle along the path of vehicles, his motions cannot be distinguished from those of vehicles and this activity will not be detected as an anomaly. Furthermore, the motion speed is not taken into account. The over-speed activities will not be detected either. In the future work, other features such as appearance and speed, are worth considering in these scenarios for better performance.

ACKNOWLEDGEMENTS

The work is partially funded by DFG (German Research Foundation) YA 351/2-1. The authors gratefully acknowledge the support.

References

Althloothi, S., Mahoor, M. H., Zhang, X. and Voyles, R. M., 2014. Human activity recognition using multi-features and multiple kernel learning. Pattern Recognition 47(5), pp. 1800–1812.

Banerjee, P. and Nevatia, R., 2014. Pose filter based hidden-crf models for activity detection. In: Computer Vision–ECCV 2014, Springer, pp. 711–726.

Chathuramali, K. M. and Rodrigo, R., 2012. Faster human activity recognition with svm. In: 2012 international Conference on Advances in iCT for Emerging Regions (iCTer), pp. 197–203.

Chu, W. and Ghahramani, Z., 2005. Preference learning with gaussian processes. In: ICML.

Gong, S. and Xiang, T., 2003. Recognition of group activities using dynamic probabilistic networks. In: ICCV, pp. 742–749.

Hasan, M. and Roy-Chowdhury, A. K., 2014. Incremental activity modeling and recognition in streaming videos. In: CVPR, pp. 796–803.

Hospedales, T., Gong, S. and Xiang, T., 2009. A markov clustering topic model for mining behaviour in video. In: ICCV, pp. 1165–1172.

Hospedales, T. M., Li, J., Gong, S. and Xiang, T., 2011. Identifying rare and subtle behaviors: A weakly supervised joint topic model. IEEE Trans. PAMI 33(12), pp. 2451–2464.

Kim, K., Lee, D. and Essa, I., 2011. Gaussian process regression flow for analysis of motion trajectories. In: ICCV, pp. 1164–1171.

Kinoshita, A., Takasu, A. and Adachi, J., 2014. Traffic incident detection using probabilistic topic model. In: EDBT/ICDT Workshops, pp. 323–330.

Kuettel, D., Breitenstein, M. D., Van Gool, L. and Ferrari, V., 2010. What's going on? discovering spatio-temporal dependencies in dynamic scenes. In: CVPR, pp. 1951–1958.

Liu, C., 2009. Beyond pixels: exploring new representations and applications for motion analysis. PhD thesis, Citeseer.

Loy, C., Xiang, T. and Gong, S., 2009a. Modelling multi-object activity by gaussian processes. In: BMVC, pp. 1988–1995.

Loy, C., Xiang, T. and Gong, S., 2009b. Multi-camera activity correlation analysis. In: CVPR, pp. 1988–1995.

Nocedal, J. and Wright., S., 2006. Numerical operation. Springer pp. 101–134.

Rasmussen, C. and Williams, C., 2006. Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning). The MIT Press, Boston.

Saleemi, I., Hartung, L. and Shah, M., 2010. Scene understanding by statistical modeling of motion patterns. In: CVPR, pp. 2069–2076.

Sun, C. and Nevatia, R., 2013. Active: Activity concept transitions in video event classification. In: ICCV, pp. 913–920.

Swears, E., Hoogs, A., Ji, Q. and Boyer, K., 2014. Complex activity recognition using granger constrained dbn (gcdbn) in sports and surveillance video. In: CVPR, pp. 788–795.

Talha, A. M. and Junejo, I. N., 2014. Dynamic scene understanding using temporal association rules. Image and Vision Computing 32(12), pp. 1102–1116.

Urtasun, R., Fleet, D. and Fua, P., 2006. 3d people tracking with gaussian process dynamical models. In: CVPR, pp. 238–245.

Vo, N. N. and Bobick, A. F., 2014. From stochastic grammar to bayes network: Probabilistic parsing of complex activity. In: CVPR, pp. 2641–2648.

Wang, J., Fleet, D. and Hertzmann, A., 2008. Gaussian process dynamical models for human motion. IEEE Trans. PAMI 30, pp. 283–298.

Wang, X., Ma, K., Ng, G. and Grimson, W., 2011. Trajectory analysis and semantic region modeling using nonparametric hierarchical bayesian models. IJCV 95(3), pp. 287–312.

Wang, Y., Wang, D. and Chen, F., 2014. Abnormal behavior detection using trajectory analysis in camera sensor networks. International Journal of Distributed Sensor Networks 2014, pp. 9.

Zelnik-Manor, L. and Perona, P., 2004. Self-tuning spectral clustering. In: NIPS, pp. 1601–1608.