PARTICLE-BASED TRACKING MODEL FOR AUTOMATIC ANOMALY DETECTION

Erwan Jouneau and Cyril Carincotte

Multitel asbl - 2 Rue Pierre et Marie Curie - 7000 Mons - Belgium

ABSTRACT

In this paper, we present a new method to automatically discover recurrent activities occurring in a video scene, and to identify the temporal relations between these activities, e.g. to discover the different flows of cars at a road intersection, and to identify the traffic light sequence that governs these flows. The proposed method is based on particle-based trajectories, analyzed through a cascade of HMM and HDP-HMM models. We demonstrate the effectiveness of our model for scene activity recognition task on a road intersection dataset. We last show that our model is also able to perform on the fly abnormal events detection (by identifying activities or relations that do not fit in the usual/discovered ones), with encouraging performances.

Index Terms— Video surveillance, activity recognition, anomaly detection, HMM, HDP-HMM, topic models.

1. INTRODUCTION

Nowadays, surveillance systems are deployed everywhere: in shopping malls, parking lots, etc. With this growing number of CCTV systems, there is an increasing demand for automatic analysis of the data generated by such systems.

In the past few years, there had been several attempts to develop methods that can automatically learn the rules which governs a video scene, and automatically classify a video content as normal or abnormal [1, 2, 3, 4]. In the surveillance context, the main challenge of this task lies in the *a priori* unknown type of objects, events or anomalies to be considered for the detection. Anomaly can have various causes (e.g. a single object behavior, the spatial or temporal relation between two or more objects, a combination of them., etc.), which cannot be exhaustively listed to ensure their detection. To cope with this issue, most of the recent approaches build a model that identify the video content structure (i.e. identify mobile objects patterns that correspond to recurrent activities), and then compute a normal/abnormal level by inferring the current ongoing activities on the built model.

Recent research on automatic scene activity modeling can be divided into two main categories: clustering/classification approaches based on object tracking and "topic-like" approaches based on motion between successive frames. Tracking based approaches [4, 5] can identify anomaly involving one object, but need complete/reliable tracks to do so. Indeed, in real CCTV data, object segmentation and occlusion issues are quite frequent and may cause problems. To deal with track failures, non parametric track features have already been tried [6, 7]. The main limitation of such techniques lies in the detection of events involving multiple objects, which can not be handled straightforwardly in the clustering stage. To avoid object tracking, different "topic-like" frameworks using motion between successive frames have been proposed: e.g. variants of HMM [8] but failing to detect events where multiple objects are involved, topic models [9] where no temporal information can be taken into accounts, or even HMM-based model handling temporal correlations of multiple objects [10], but requiring to *a priori* define the number of regions which represent the scene content. More recently, the learning of co-occurring activities with a temporal aspect has been addressed either using the temporal order of words e.g. PLSM [1] or by modeling the temporal order of the activities e.g. MCTM [3] or DDP-HMM [2] (variant of HDP-HMM), all using position and motion direction as input.

In this paper, we propose to combine both approaches through a cascade of HMM and HDP-HMM models; we propose to use a HMM model to classify particle-based trajectories, and to use a HDP-HMM model to identify co-occurring trajectories and their temporal relations. The remaining of the paper is organized as follows; first, next section introduces the cascade model we propose to use. We then present in Sec. 3 quantitative results obtained on road intersection CCTV data; we compare performances with different state-of-the-art models for both activity recognition and abnormality detection. Conclusions and perspectives are last drawn in Sec. 4.

2. MODEL OVERVIEW

This section introduces the cascade of particle-based tracking, HMM and HDP-HMM models we propose to use. For concision matters, this paper does not provide detailed descriptions of models; interested readers may consult [11] and [12] for more details.

2.1. Particle-based tracking

To cope with the tracking issues mentioned in introduction, we propose to use a tracking algorithm which proved to be

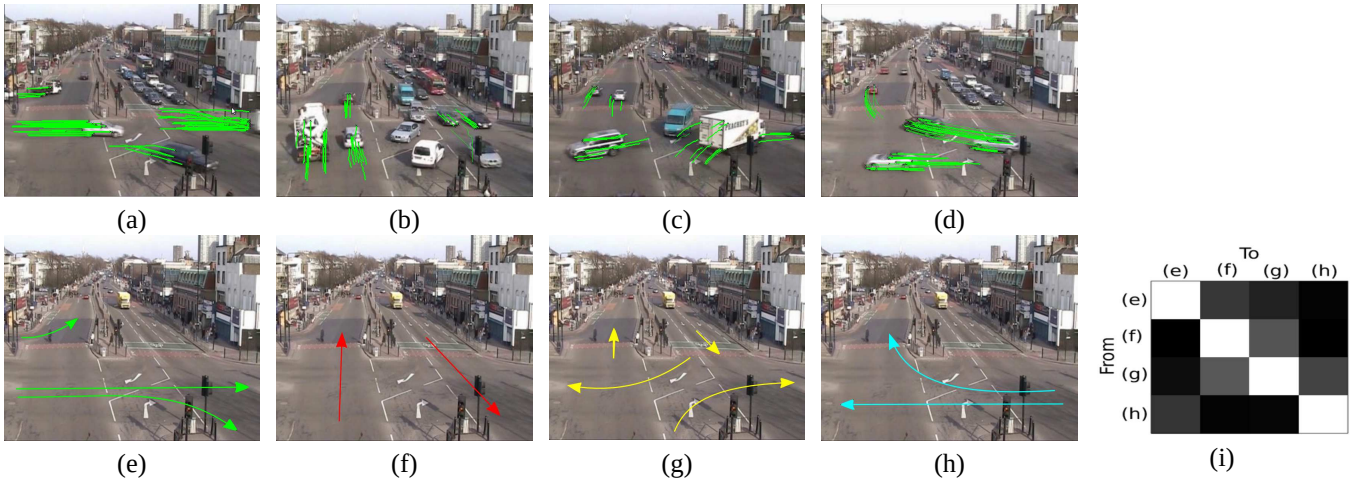

Fig. 1. Sample images for each activities with extracted trajectories (first row), and corresponding scene activity patterns automatically discovered (second row). Transition matrix of learnt patterns (right column).

efficient in various contexts (indoor/outdoor, metro/train...). This approach does not use any background modeling estimation, thus preventing from tracks failures and related issues. The principle of the proposed tracking is to quickly locate moving objects in the scene using randomly distributed particles in the image [11]. Roughly, inactive particles are randomly distributed and activated using a frame-differencing threshold. After activation, a particle continuously tracks the motion of the underlying object using a block-matching algorithm. Particle's trajectories are finally analyzed and filtered by computing a set of various features (e.g. linearity of track, track length,track duration, track direction, start/stop particle location, etc.) to recycle useless ones (e.g. tracking failures) and keep active relevant ones. Occlusions just interrupt trajectories. Fig. 1(a-d) show examples of resulting trajectories.

2.2. HMM/HDP-HMM cascade

The model we propose to use is composed of two stages; a first HMM stage to classify the particle-based trajectories, and a second HDP-HMM stage to identify the co-occurring trajectories and the temporal relation between them.

For the first stage, we adopt a HMM with Gaussian mixture, and use the position of the current active particle (2 features) and corresponding trajectory direction (1 feature) as input features. It worths pointing out that the trajectory direction we use is computed over the whole trajectory, and not only between the last two frames. In addition, in most of state-of-the-art "topic-like" models, an optical flow is computed on a grid, then filtered and sampled to construct a vocabulary with the position and the direction of the movement for each block of this grid called "classic words" in the rest of the paper.

Instead of using such features, we propose to directly use the trajectory classes issued from the HMM. The advantages of our words is thus that it is independent of the resolution, no threshold has to be fixed manually and new information/feature can be added on this first stage without increasing too much the size of the dictionary; remaining drawback is that the number of classes of this first stage HMM still has to be fixed manually.

For the second stage, we adopt a HDP-HMM [12], that automatically determines the different activities in the scene and the temporal relation between them. Compared to standard topic models like LDA [13], the number of activities in the HDP-HMM has not to be fixed but is also discovered. In addition, it allows to identify anomaly with respect to learn activities using a log-likelihood threshold.

3. EXPERIMENTS

3.1. Datasets and settings

Experiments were conducted on one CCTV footage dataset (road intersection), showing flows of cars and people governed by a traffic light (see Fig. 1(e-h)). This dataset contains 50 minutes of video at a resolution of 360x288 (30 fps). 5 minutes is used for training and the remaining data for testing purpose. The cascade configuration is the following; for the first stage, we fix arbitrarily the number of HMM states to 20; clip length were set to 2 seconds (60 frames).

3.2. Activity recognition performance

Fig. 1 shows the co-occurring trajectories and activities discovered; each activity corresponds to one particular behavior in the scene. Fig. 1(a) to (d) are sample images explained by all the discovered activities ((e) to (h)). Fig. 1(e) represents the horizontal flow (of cars and people) from left to right, Fig. 1(f) the vertical flows of cars, Fig. 1(g) the vertical flows

of cars that turn on left or right, and Fig. 1(h) the horizontal flows (of cars and people) from right to left. Fig. 1(i) correspond to the transition matrix of the learnt Markov chain, which allow to clearly distinguish the different phases and their usual temporal ordering; i.e. the cycle (e)-(f)-(g)-(h) managed by the traffic light, and the cycle (f)-(g) for turning vehicles when it is possible. With the current configuration, pedestrians share the same activities as cars.

So as to perform a quantitative evaluation of the recognition rate for the different activities, we selected ten minutes of the video in the testing dataset and manually associated one or more activities to each frame. We then inferred these ten minutes of annotated video on the trained model, varied the decision threshold, and compared results to the ground truth (GT) to obtain precision-recall curves (see Fig. 2). The end of turning activity (g) can be confused with the end of activity (e) and (h), that's why the result of (g) are slightly worse.

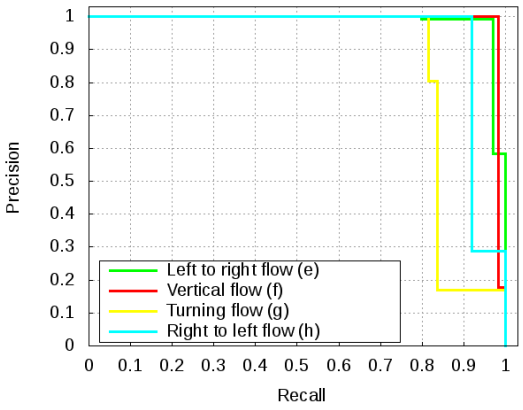

Fig. 2. Precision-Recall curve for activity recognition.

3.3. Anomaly detection performance

The result of the anomaly detection tasks is presented in this section. Obvious categories of anomalies have been defined, namely jaywalking, illegal u-turn or drive wrong way. Jaywalking category is composed of pedestrians crossing the intersection without respecting the zebra crossing, or pedestrians crossing on the zebra but without respecting the red light. Illegal u-turn concerns the vertical flow of vehicles. The drive wrong way category regroups all behaviors that we should never see like vehicles in the wrong lane. Fig. 3 illustrates such anomaly detection results. From top to bottom, the log-likelihood defining the level of normality/abnormality is represented by the graphic, the most probable activity for each clip by the colored bar, and the detected anomaly by the linked sample image.

Similarly to previous section, we perform a quantitative evaluation of the anomaly detection rate for the different anomalies, using the whole testing video (45 min) manually

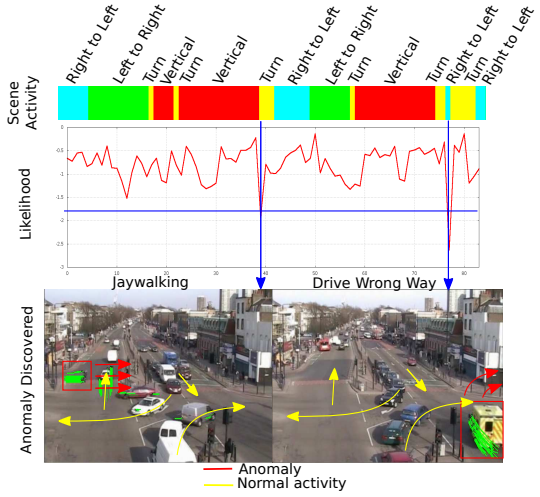

Fig. 3. Activity recognition and anomaly detection results.

annotated. Fig. 4 and Table 1 present respectively the interpolated precision-recall curves and the quantitative results corresponding to the best operating point. In this evaluation, we also compare our particle-based features with "classic words" as described in section 2.2, for both the HDP-HMM approach and the classical Latent Dirichlet Allocation (LDA) model with k topics (for fair comparison purpose k is fix to the number of activities discovered by the HDP-HMM).

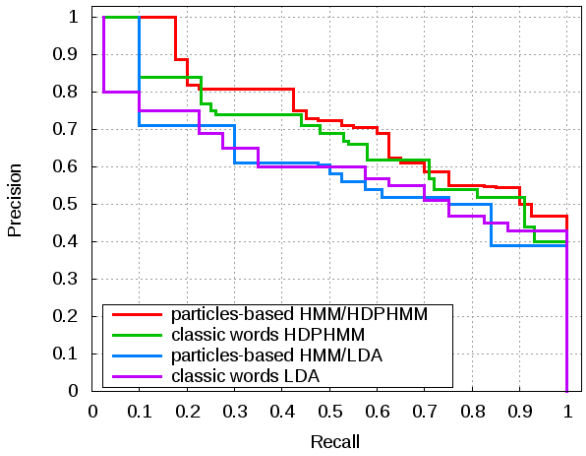

Fig. 4. Precision-Recall curve for anomaly detection.

Several conclusions can be drawn from these experiments; first, the HDP-HMM based approach always reach better performances than LDA one, whatever the input features. Secondly, the use of particle-based tracking features and HMM classifier as input of the HDP-HMM model allows to reach the highest performances in almost all tested configuration. Indeed, contrary to classic words, our particle-based tracking features are less influence by noise in the scene,

	GT	HDPHMM		LDA		MCTM
		classic words	Tracks	classic words	Tracks	
Illegal u-turn	11	4	3	4	3	2
Drive wrong way	16	8	11	11	9	15
Jaywalking	13	7	9	6	8	1
Uninteresting[1]	-	9	8	14	13	29

Table 1. Summary of anomaly discovered.

e.g. local motion or noise acquisition/encoding. Regarding the uninteresting clips discovered, they can be classified in 2 groups: clips with vehicles that start (too) quickly at the traffic light, and clips with objects going in very rare direction not encountered in the 5 min training set.

To conclude, we also compare our HMM-based cascade with the Markov Clustering Topic Model (MCTM) applied on the same dataset [3] (see MCTM column in Table 1). While these results cannot be strictly compared given the difference between the interpretation of the clip discovered, especially for the jaywalking category, the precision and recall rate were respectively 38% and 45% of recall for MCTM, while for the same recall rate, we reach a precision rate of 72%.

4. CONCLUSION

In this paper, we presented a two-stages HMM-based cascade using particle-based tracking features as input, for automatic scene activity modeling and anomaly detection in video footage. The first HMM stage is used to classify particle-based trajectories, while a second HDP-HMM stage is used to identify the co-occurring trajectories and the temporal relation between them, as well as to detect abnormal activities. Compared to features usually used in such context (optical flow or inter-frames motion related features), our particle-based tracking features are resolution independent, automatically filtered and do not require thresholds definition. We demonstrated the robustness of our model for scene activity recognition task on a road intersection scene. We also show that our model is able to perform on the fly abnormal events detection (by identifying activities or relations that do not fit in the usual/discovered ones), with ecnouraging results. Current experiments are focused on the validation of the proposed model for less structured type of CCTV footage (videos of metro). As perspectives, adding a measure at the first stage of the cascade would allow to compute different abnormality level, especially on single trajectories themselves. The use of features not directly related to object motion (e.g. background subtraction outputs) would also be interesting to cope with temporal behaviors including static stage (e.g. cars stopped at traffic light, or people standing by) or features like speed would help to distinguish pedestrians and vehicles activities. Last, strict comparison should be performed with [3], and is planned with [1].

[1]Detected clips with no salient anomaly are labeled "uninteresting".

5. ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Community's Seventh Framework Programme FP7/2007-2013 - Challenge 2- Cognitive Systems, Interaction, Robotics - under grant agreement n° 248907-VANAHEIM.

6. REFERENCES

[1] J. Varadarajan, R. Emonet, and J.M. Odobez, "A sparsity constraint for topic models application to temporal activity mining," *NIPS*, 2010.

[2] D. Kuettel, M.D. Breitenstein, L.V. Gool, and V. Ferrari, "What is going on? Discovering spatio temporal dependencies in dynamic scenes," in *CVPR*, 2010.

[3] T. Hospedales, S. Gong, and T. Xiang, "A markov clustering topic model for mining behaviour in video," in *International Conf. on Computer Vision*, 2009.

[4] I. Ivanov, F. Dufaux, T.M. Ha, and T. Ebrahimi, "Towards generic detection of unusual events in video surveillance," in *AVSS*, 2009.

[5] X. Wang, K. Tieu, and E. Grimson, "Learning semantic scene models by trajectory analysis," in *ECCV*, 2006.

[6] A. Basharat, A. Gritai, and M. Shah, "Learning object motion patterns for anomaly detection and improved object detection," in *CVPR*, 2008.

[7] I. Saleemi, K. Shafique, and M. Shah, "Probabilistic modeling of scene dynamics for applications in visual surveillance," in *ECCV*, 2009.

[8] T. Duong, H. Bui, D. Phung, and S. Venkatesh, "Activity recognition and abnormality detection with the switching hidden semi-markov model," in *CVPR*, 2005.

[9] J. Varadarajan and J.-M. Odobez, "Topic models for scene analysis and abnormality detection," in *Int. Conf. on Computer Vision*, 2009.

[10] C. Loy, T. Xiang, and S. Gong, "From local temporal correlation to global anomaly detection," in *European Conf. on Computer Vision*, 2008.

[11] C. Carincotte, X. Naturel, M. Hick, J.-M. Odobez, J. Yao, A. Bastide, and B. Corbucci, "Understanding metro station usage using closed circuit television cameras analysis," in *ITSC*, 2008.

[12] M.J. Beal, Z. Ghahramani, and C.E. Rasmussen, "The infinite hidden markov model," in *NIPS*, 2002.

[13] D.M. Blei, A.Y. Ng, and M.I. Jordan, "Latent dirichlet allocation," in *JMLR*, 2003.