Abnormal behavior detection in video protection systems

Luis Patino[1], Hamid Benhadda[2], Nedra Nefzi[1,2], Bernard Boulay[1], François Bremond[1], and Monique Thonnat[1]

[1] INRIA Sophia Antipolis - Méditerranée
2004 route des Lucioles - BP 93 - 06902 Sophia Antipolis, France
[2] Thales Services, Campus Polytechnique
1 Avenue Augustin Fresnel - 91767 Palaiseau Cedex, France

Abstract. In this work we present a system to extract in an unsupervised manner the main activities that can be observed by a camera monitoring a scene on the long-term and with the ultimate aim to discover abnormal events. To allow for semantically interpretable results, the activities are characterised by referring them to contextual elements of the observed scene. By contextual elements, we understand activity areas such as building entrances, people meeting areas, road areas,...etc. The system thus stars in a first step by the unsupervised learning of the main activity areas of the scene. We employ trajectory-based analysis of mobiles in the video to discover the points of entry and exit of mobiles appearing in the scene and ultimately deduce the different areas of activity. In a second step, mobile objects are then characterised in relation to the learned activity areas. Two kind of behaviours can then be defined either 'staying in a given activity zone' or 'transferring from an activity zone to another' or a sequence of the previous two behaviours. In a third step we employ a high-level clustering algorithm to group mobiles according to their behaviours and discover both, frequent/normal behaviours and unusual/abnormal events. We employ soft computing techniques in the first two steps to handle the uncertainty inherently present in low-level trajectory data. The high-level clustering algorithm is based on relational analysis methodology. We have applied our system in two domains, the monitoring of activities in the hall entrance of an underground station and to traffic control by monitoring a bus reserved street lane. We show that our current results are encouraging.

1 Introduction

Nowadays with the increasingly growing needs of protection of people and personal properties, video surveillance has become a big concern of everyday life. A consequence of these needs has led to the proliferation of cameras almost everywhere. Combining these facts with the ease and cheapness of data collection and storage, the video surveillance systems can collect huge amounts of video data that can be very useful for investigation purposes of events occurrence or for the needs of predictions in order to improve these systems themselves.

Paradoxically, these data become virtually unusable for analysis because of their hugeness, hence the need of intelligent analysis systems able to help the end user by, at least semi-automating the data processing analysis. In this paper, we will use clustering techniques to gather similar behaviours and/or trajectories in the same clusters to help the end user finding more rapidly the most important events of interest.

The remainder of this paper is as follows. In the following section, we give a short overview of the related work. Some preliminary steps for data processing are given in section 3. The methodology for activity extraction is presented in section 4; the experimental results are to be found in section 5. Finally, Section 6 draws the main conclusions and describes some perspectives.

2 Related Work

Extraction of the activity contained in the video by applying data-mining techniques represents a field that has only started to be addressed. Although the general problem of unsupervised learning has been broadly studied in the last couple of decades, there are only a few systems which apply them in the domain of behaviour analysis. A few systems employ soft computing techniques to characterize video activity patterns [7, 10] but the methodology to self-discover new activities is still missing. Because of the complexity to tune parameters or to acquire knowledge, most systems limit themselves to object recognition [6]. For behaviour recognition, three main categories of learning techniques have been investigated.

• The first class of techniques learns the parameters of a video understanding program. These techniques have been widely used in case of event recognition methods based on neural networks [8], Bayesian classifiers and HMMs [11, 16].

• The second class consists in using unsupervised learning techniques to deduce abnormalities from the occurring events [17].

• The third class of methods focuses on learning behaviour based on trajectory analysis. This class is the most popular learning approach due to its effectiveness in detecting normal/abnormal behaviours; for instance, on abnormal trajectory detection on roads [13, 15] or pedestrian trajectory characterisation [2]. Hidden Markov Models (HMM) have also been employed to detect different states of pre-defined normal behaviour [3, 14].

Our main contribution to the state of the art is to employ trajectory information to automatically learn activity zones and generate from them behaviour/event information. Trajectoty and event information analysed together can make visible important activity patterns difficult to see in the raw data.

3 Data preprocessing

In order to discover meaningful activity, it is of prime importance to have available detailed information allowing to detect the different possible interactions between mobiles. As our system is based on trajectory analysis, the first step to prepare the data for the activity clustering methodology is to extract tracklets of fairly constant speed allowing to characterise the displacements of the mobile or its stationary state. Describing the detection and tracking algorithm is out of the scope of the paper. Suffices to say background substraction and a Kalman filter-based tracker (KLT) are employed to extract mobile trajectories.

If the dataset is made up of N objects, the trajectory tr_j for object O_j in this dataset is defined as the set of points $[x_j(t), y_j(t)]$ corresponding to their position points; x and y are time series vectors whose length is not equal for all objects as the time they spend in the scene is variable. The instantaneous speed for that mobile at point $[x_j(t), y_j(t)]$ is then $v(t) = \left(\dot{x}(t)^2 + \dot{y}(t)^2\right)^{\frac{1}{2}}$. The objective is then to detect those points of changing speed allowing to segment the trajectory into tracklets of fairly constant speed so that the trajectory can be summarised as a series of displacements at constant speed or in stationary state.

The mobile object time series speed vector is analysed in the frame of a multiresolution analysis of a time series function $v(k)$ with a smoothing function, $\rho_{2^s}(k) = \rho(2^s k)$, to be dilated at different scales s. In this frame, the approximation A of $v(k)$ by ρ is such that $A_{2^{s-1}}v$ is a broader approximation of $A_{2^s}v$. By analyzing the time series v at coarse resolutions, it is possible to smooth out small details and select those points associated with important changes.

The speed change points are then employed to segment the original trajectory tr_j into a series of i tracklets tk. Each tracklet is defined by two key points, these are the beginning and the end of the tracklet, $[x_j^i(1), y_j^i(1)]$ and $[x_j^i(end), y_j^i(end)]$. By globally reindexing all tracklets, let m be the number of total tracklets extracted, we obtain the following tracklet feature vector :

$$tk_m = [x_m(1), y_m(1), x_m(end), y_m(end)] \tag{1}$$

4 Activity Analysis and Clustering

We understand activity as the interactions occurring between mobile objects themselves and those between mobiles and the environment. We employ trajectory-based analysis of mobiles in the video to discover the points of entry and exit of mobiles appearing in the scene and ultimately deduce the different areas of activity. In a second step, mobile objects are then characterised in relation to the learned activity areas. In a third step we employ a high-level clustering algorithm to group mobiles according principally to their behaviours and discover both, frequent/normal behaviours and unusual/abnormal events. We employ soft computing techniques in the first two steps to handle the uncertainty inherently present in low-level trajectory data. The high-level clustering algorithm is based on relational analysis methodology.

4.1 Unsupervised learning of activity areas

Modeling the spatial context of the scene is essential for recognition and interpretation of activity. Because it is not possible to define a-priori all activity zones, manually defined Contextual zones do not suffice to describe all possible situations or evolving actions in the monitored scene. We thus propose to learn the complementary activity zones.

The feature vector defined in equation 1 constitute a set of simple descriptors that have proven experimentally to be enough to describe activities in a large variety of domains (such as traffic monitoring, subway control, monitoring smart environments), mainly because they define where the object is coming from and where it is going to and also with approximative constant speed.

In our system, mobiles starting (or ending up) their displacement at nearby positions are seen to share a common activity area. Finding thus the activity areas on the scene could be seen as an equivalent task of clustering entry/exit tracklet points. Different common clustering algorithms such as k-means, Self Organising Maps, etc... can be used but the drawback is, first, the number of clusters must be known in advance, and second, the spatial morphology of the resulting clusters is always circular, which may not correspond to the real scene topology. In our approach, we propose to find first high resoluted (small in size) entry/exit activity zones, which in a second step could be further regrouped to obtain broader activity zones (following an hierarchical agglomerative algorithm).

Clustering of tracklet entry/exit points For this first step, we employ the well-known clustering Leader algorithm [9]. It has the advantage to work on-line without needing to specify the number of clusters in advance. In this method, it is assumed that a threshold T is given. The algorithm constructs a partition of the input space (defining a set of clusters) and a leading representative for each cluster, so that every object in a cluster is within a distance T of the leading object. The threshold T is thus a measure of the diameter of each cluster. The algorithm makes one pass through the dataset, assigning each object to the cluster whose leader is the closest and making a new cluster, and a new leader, for objects that are not close enough to any existing leaders. Defining T is application dependent. In this work we have employed the threshold setting suggested by [1].

Let's consider the position of a mobile is *L(x,y)*, its influential zone, *Zn*, is defined by a radial basis function (RBF) centered at the position L; and the belonginess of a new point p(x,y) to that zone is given by:

$$Zn(L,p) = \phi(L,p) = exp(-\|p-L\|^2 T^2) \tag{2}$$

The RBF function has a maximum of 1 when its input is $p = L$ and thus acts as a similarity detector with decreasing values outputted whenever p strides away from L. An object element will be included into a cluster Zn if $Zn(L,p) \geq 0.5$, which is a natural choice. The cluster receptive field (hyper-sphere) is controlled by the learnt parameter T

Merging tracklet entry/exit zones We find the final activity areas by aggregating similar entry/exit tracklet zones. We look to establish a similarity relation between the different zones defined by the tracklets. On the end, new zones are given by the fulfillment of two relations: 1) zone Zn_i overlaps zone Zn_j, and 2) zone Zn_j does not overlap any a-priori user-defined contextual Zone $Zctx_q$. These relations are defined:

$R1_{ij}$: Zone Zn_i overlaps Zone Zn_j

$$R1_{ij} = \sum_{k=1}^{3} \left[\sum_{(x,y)\in(X_{ik},Y_{ik})} Zn_j(x,y) \right] \tag{3}$$

and $X_{ik} = \left\{ \frac{(k+1)}{3} T\cos(\theta) + L_i(1) \right\}$,
$Y_{ik} = \left\{ \frac{(k+1)}{3} T\sin(\theta) + L_i(2) \right\}$ with $\theta = 0, ..., \frac{\pi}{8}, ..., 2\pi$
That is, points $(x, y) \in (X_{ik}, Y_{ik})$ belonging to concentric circles to L_i are tested to verify how they fit Zn_j in order to calculate the overlap/similarity between Zn_i and Zn_j. This allows avoiding equity problems with clusters defined on sparse regions (some clusters may be defined with a much larger number of points than others).

$R2_{iq}$: Zone Zn_i overlaps Zone $Zctx_q$

$$R2_{iq} = \sum_{k=1}^{3} \left[\sum_{(x,y)\in(X_{ik},Y_{ik})} Zctx_q(x,y) \right] \tag{4}$$

It is possible to transform R2 into a new relation, R3, which links Zn_i and Zn_j if both clusters are related to the same Zone $Zctx_q$ through the fulfillment of R2. The relation between Zn_i and Zn_j is then given by

$$R3_{ij} = \max_q \min\left[R2_{iq}, R2_{qj}\right] \tag{5}$$

Remark that $\overline{R3}$, the complement to $R3$ given by $\overline{R3} = -R3$, represents the relation linking Zn_i and Zn_j if both clusters are not related to any contextual Zone ($Zctx_q$). $R1$ and $\overline{R3}$ can be aggregated employing a soft computing aggregation operator such as $R = R1 \cap \overline{R3} = \max\left(0, R1 + \overline{R3} - 1\right)$ and made transitive with:

$$R \circ R\,(x,y) = \max_z \min\left[R\,(x,z), R\,(z,y)\right] \tag{6}$$

R is now a transitive similarity relation with R indicating the strength of the similarity. If we define a discrimination level α in the closed interval [0,1], an $\alpha - cut$ can be defined such that

$$R^\alpha\,(x,y) = 1 \Leftrightarrow R\,(x,y) \geqslant \alpha \tag{7}$$

It is thus implicit that $\alpha_1 > \alpha_2 \Leftrightarrow R^{\alpha_1} \subset R^{\alpha_2}$; thus, the R^α form a nested sequence of equivalence relations, or from the classification point of view, R^α

induces a partition $\pi^\alpha = \{Z_i^\alpha\}$ of $X \times Y$ (or X^2) such that $\alpha_1 > \alpha_2$ implies π^{α_1} is a refinement of π^{α_2}.

At this point, the difficulty comes down to select the appropriate $\alpha - cut$ such that π^α from R^α represents the best partition of the data. This is still a difficult and open issue that we choose to approach by selecting the alpha-values, which induce a significant change from π^{α_k} to $\pi^{\alpha_{k+1}}$.

To automatically detect those significant partition changes we choose to study the cluster area and number of clusters induced at each partition π^α. We achieve this in the frame of a multiresolution analysis. By analysing induced partitions at coarse resolutions, it is possible to smooth out small details and select the $\alpha - cut$ levels associated with important changes. From the monitored scene, it would be useful to distinguish among different information levels: (i) grouped activity on large spaces, (ii) very detailed individual activity, (iii) somewhere meaningful in-between the last two. For this reason, when performing activity zone discovery, we automatically select the three highest change-inducers $\alpha - cut$ levels from the previous analysis. The result is then that we end up with a three levels hierarchy of activity zones.

4.2 Behaviour definition

We aim at creating a system for the recognition and interpretation of human activity and behaviour, and extract new information of interest for end-users. Low-level tracking information is thus expected to be transformed into high-level semantic descriptions conveying useful and novel information. In our application, we establish a semantic meaning from the scene model built as decribed in the previous section (4.1). The behaviour knowledge can be thus expressed with semantic concepts, instead of using quantitative data, thanks to the learned zones and any user-defined contextual zone. Let us assume, we have in total $k = 1, ..., K$ contextual zones on the scene defined either a-priori, or after the zone learning procedure. Two different kinds of behaviours can then be identified:

- Mobile moving from Zone $Zctx_k$ to Zone $Zctx_{k'}$
- Mobile Inside Zone $Zctx_k$

4.3 High level activity analysis

To perform deep high level activities analysis that are not easy to highlight from the raw video data we will try to exhibit complex and hidden correlations between the mobile objects in the video scenes on one hand and between the mobile objects and contextual objects on the other hand. To reach this goal we will use relational analysis clustering methodology.

Relational analysis clustering methodology The foundations of relational analysis clustering methodology dates from the former works of Condorcet related to the voting theory in 1785 [5]. The concept underlying these works is based on the "consensus" or "majority" rule, which means that a decision can be considered valid as soon as it is supported by a majority of voters. This methodology is very well adapted to categorical data where the variables measured on the objects of interest can be considered as judges deciding if any two objects of the data are or are not of the same category. This methodology has been formalized for the first time as a linear programming problem $\mathcal{P}$ in the late eighties by Marcotorchino F. and Michaud P. [12]. If X is a squared matrix representing the unknown final partition to be discovered and $\mathcal{I}$ represents the whole data set, then the mathematical formulation of the problem is :

$$(\mathcal{P}) \quad \begin{cases} \max_X \mathcal{C}(X) & & \\ x_{ii} = 1 & \forall\, i \in \mathcal{I} & \text{(reflexivity)} \\ x_{ii'} - x_{i'i} = 0 & \forall\, (i,i') \in \mathcal{I}^2 & \text{(symmetry)} \\ x_{ii'} + x_{i'i''} - x_{ii''} \leq 1 & \forall\, (i,i',i'') \in \mathcal{I}^3 & \text{(transitivity)} \\ x_{ii'} \in \{0,1\} & \forall\, (i,i') \in \mathcal{I}^2 & \text{(binarity)} \end{cases}$$

where $\mathcal{C}(X)$ is the Condorcet's criterion defined by the relation:

$$\mathcal{C}(X) = \sum_{i=1}^{n}\sum_{i'=1}^{n}(s_{ii'} - \overline{s}_{ii'})x_{ii'}$$

$x_{ii'}$ is the general term of matrix X, defined as:

$$x_{ii'} = \begin{cases} 1 & \text{if } i \text{ and } i' \text{ are in the same cluster of the final partition} \\ 0 & \text{otherwise} \end{cases}$$

$s_{ii'}$ is the number of variables (or voters) supporting "object i and object i' are in the same cluster of the final parition" and $\overline{s}_{ii'}$ is the number of variables (or voters) supporting the opposite decision. Hence to maximise the Condorcet's criterion, two objects i and i' are likely to be in the same cluster of the final obtained partition as soon as $(s_{ii'} - \overline{s}_{ii'})$ is is strictly positive and are likely to be separated otherwise. More details about relational analysis clustering methodology can be found in [4].

At the end of the process, the obtained clusters will be described by the variables modalities that have participated strongly to their constitution thanks to two indicators named $ratioC$ and $ratioD$. for a cluster C and a modality j, $ratioC(j)$ is the proportion of objects belonging to the cluster that have the modality j, and $ratioD(j)$ is the proportion of cluster objects having modality j with regard to the whole data set objects having this modality,that is :

$$ratioC(j) = \frac{nj\_C}{|C|} \quad (8) \qquad ratioD(j) = \frac{nj\_C}{nj} \quad (9)$$

where :
- nj_C is the number of objects of cluster C having modality j
- $|C|$ is the cardinal of cluster C
- nj is the total number of objects of the whole data set having modality j

Relational analysis clustering applied to video data There are two main types of concepts to be represented from the video and on which we perform high-level clustering: mobile objects of the observed scene and video events occurring in the scene. For the off-line analysis of both types of concepts, and with the aim of setting the data in a suitable format to achieve knowledge discovery, this information is characterized by a set of specific features:

1. **mobileid**: the identifier label of the object
2. **start**: the time when the object is first seen.
3. **end**: the time when the object is last seen.
4. **duration**: the total time the object has been observed. (formated in seconds)
5. **dist_org_dest**: the total distance walked from origin to destination. (formated in meters)
6. **mean_speed**: the average of the instantaneous speed calculated at all points of the detected mobile object trajectory.
7. **Significant_evnt**: inferred event for the object considering the learned zones.

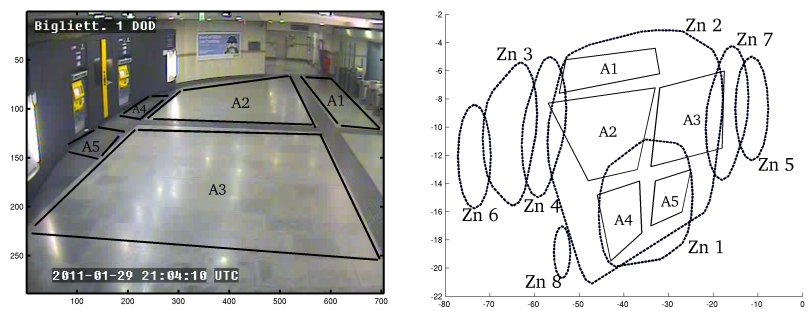

Fig. 1. Left panel: Original underground scene observed by the camera. A1 to A5 correspond to user defined areas delimitating the scene. Right panel: Learned zones. Zn1 to Zn8 correspond to activity areas as discovered with our algorithm.

5 Results and Evaluation

The algorithm for unsupervised learning of activity areas (as described in section 4.1) was applied to a one hour-duration video recorded at one entrance hall from the Torino underground system. The final relation R given in equation (6), which verifies the transitive closure, is thresholded for different $\alpha - cut$ values going from 0 to 0.9 and with a step value of 0.05. The algorithm automatically selects the best $\alpha - cut$ value giving a well detailed composition of the activity areas of the scene as mentioned in section 4.1. Figure 1 presents those learned zones corresponding to the analysed video.

As mentioned in section 4.2, we achieve behaviour carachterisation by linking low-level tracking to the learned zones. The whole activity observed from the scene can then be reported following the behaviours inferred from the learned zones. For instance, for the ticket machine area (zone Zn1), the obtained report is given in table 1. Zone Zn1 is actually the second most employed one of all zones just after Zn2, wich corresponds to the main hall area (occupancy of 39.85%; not shown in table 1). Most people going to the vending machines depart from the main hall (Zn2); and all people at the vending machines go to the main hall. People directly going to the vending machines from any of the station entrances do this from entrance zones Zn5 or Zn7.

Table 1. Activity reporting (not exhaustive) related to the vending machines area.

Proportion	Number of mobiles	Description
25.92%	106	at Zone 1
3.91%	16	Zone 1 to Zone 2
2.20%	9	Zone 2 to Zone 1
0.49%	2	Zone 7 to Zone 1
0.24%	1	Zone 5 to Zone 1

In the last step towards a knowledge discovery task, we have applied the relational analysis process explained before in section 4.3. The input variables are those corresponding to the mobile object semantic table described above also in section 4.3, which in this case for the analysed video contains in total 409 detected objects. Variables are weighted so to favour the formation of as many clusters as Significant_events in the table are. The interest of such procedure is to explain through other variables what characterizes each of the Significant_events.

Table 2. Example of activity clusters obtained after application of the Relational Analysis procedure.

Cluster 2 (106 elements)			Cluster 4 (16 elements)			Cluster 19 (2 elements)		
modality	ratioC	ratioD	modality	ratioC	ratioD	modality	ratioC	ratioD
Evnt.atZone1	100	100	Evnt.Zone1toZone2	100	100	Evnt.Zone7toZone1	100	100
distorgdest.0-3.84	89	32	distorgdest.3.84-7.68	43	12	Start. 22:16:20	50	100
duration.0-4.2	75	26	duration.0-4.2	37	2	End. 22:16:24.4	50	100
meanspeed.0-1.44	57	44	meanspeed.2.88-4.32	37	20			

After application of the relational analysis, 23 activity clusters are obtained; table 2 shows some of the obtained clusters. In this table, for instance, cluster2 corresponds to the activity of people staying in zone Zn1, that is, by the Vending machines zone. It can be observed that people can be described by having a very small displacement (less than 4m) and being almost immobile (speed less than 1.5m/s) for a short period of time ([0-4sec]).This is what characterises people bying tickets. This activity is the second most common activity in the station just after 'at Zone 2', corresponding to people staying in the main hall (cluster1; not shown) . Cluster4 corresponds to people going from the vending machines to the main hall ('Zone1 to Zone2'); as observed in the table, people walks a longer distance ([3.84-7.68m]) at a high walking speed ([2.88-4.32m/s]). The last cluster, Cluster 19 represents a rare activity (only two elements contained in this activity); people going from 'Zone7 to Zone1'. These persons are particulary characterised by the time at which they are detected (no other mobile detected at a nearby time). Thus, this way the relational analysis can help us to group together people having similar behaviour. This is of particular interest to the end-users because the activities in the metro station can be better quantified.

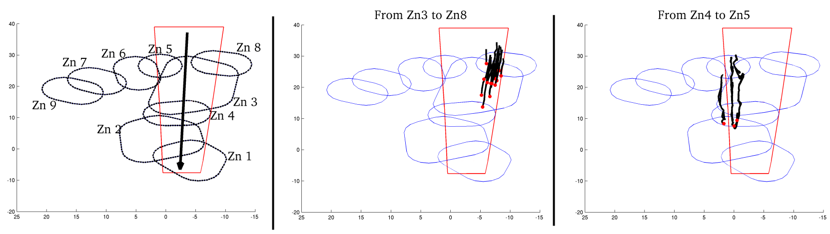

Fig. 2. Left panel: Top view of the bus lane (marked in red) with the zones learned by our system. The arrow indicates the bus flow. Central and Right panels: Two activity examples with mobiles moving against the normal traffic flow. Red points indicate the mobile departure.

We have also applied the proposed approach to another domain, namely, traffic control monitoring and specifically detecting abnormal events on a bus reserved street lane. The interest to work on this dataset is that event ground-truth is available for 45 minutes of video. We applied in this dataset the same processing chain as before, that is, we learned first the activity areas of the scene in an unsupervised manner; then we extracted the behaviour from detected mobile objects by relating them to the learned zones. Finally, we applied again the relational clustering methodology to extract high-level information. Figure 2 shows those learned activity zones of the scene. Zn6, Zn7 and Zn9 correspond to people activity areas, whereas the remainig ones are vehicle activity areas as they overlap the bus lane. Behaviour was then characterised either as staying in a given zone or transferring between two zones. We marked as abnormal those behaviours when transferring between zones was in a south-north direction. Table 3 shows the abnormal behaviours established from the learned zones. When applying the high-level relational analysis methodology, all abnormal events were gathered as expected in a single cluster. In order to test the correctness of the whole processing chain, we compared the abnormal events obtained by our approach with those signaled by the available ground-truth. The results of this comparison can be observed in table 4. The recall measure is fairly high while the precision is acceptable. This is principally because our system has included a relative high amount of false positive detections, which might possibly appear because of detections at the bus lane border, which in the ground-truth were not included.

Table 3. Discovered abnormal events

Abnormal Event	Proportion
Zn1 to Zn8	34
Zn1 to Zn5	3
Zn1 to Zn3	45
Zn1 to Zn4	5
Zn1 to Zn2	34
Zn2 to Zn4	8
Zn2 to Zn3	12
Zn4 to Zn3	5
Zn4 to Zn5	3
Zn4 to Zn8	10
Zn3 to Zn5	2
Zn3 to Zn8	16

Table 4. Evaluation measures

TP	116
FP	61
TN	2174
FN	8
Precision	0.65
Recall	0.93

6 Conclusions

In this paper we have presented a system to extract from video and in an unsupervised manner the main activities that can be observed from the scene. We have setup a processing chain broadly working on three steps: The system stars in a first step by the unsupervised learning of the main activity areas of the scene. In a second step, mobile objects are then characterised in relation to the learned activity areas: either as 'staying in a given activity zone' or 'transferring from an activity zone to another' or a sequence of the previous two behaviours if the tracking persits long enough. In a third step we employ a high-level relational clustering algorithm to group mobiles according to their behaviours and discover what other characteristics from mobile objects are strongly correlated. We have applied our algorithm in two domains. First, monitoring of activities in the hall entrance of an underground station and showed what are the most active areas of the scene and how rare/abnormal (going to low occupated activity zones) and frequent activities (e.g. buying tickets) are characterised. In the second application, monitoring a bus reserved street lane, we were able again to learn the topology of the scene and separate normal from abnormal activities. When comparing with available ground-truth for this application, we obtained a high recall measure while the precision is acceptable but can be improved, which we aim to do in our future work by including a larger series of variables to analyse into the last relational clustering step.

References

1. Incremental learning on trajectory clustering. In: Intelligent Paradigms in Safety and Security, chap. 3. Springer-Verlag (2011)
2. Anjum, N., Cavallaro, A.: Single camera calibration for trajectory-based behavior analysis. In: AVSS 2007, IEEE Conference on Advanced Video and Signal Based Surveillance. pp. 147–152 (2007)
3. Bashir, F., Khokhar, A., Schonfeld, D.: Object Trajectory-Based Activity Classification and Recognition Using Hidden Markov Models. IEEE Transactions on Image Processing 16, 1912–1919 (2007)
4. Benhadda, H., Marcotorchino, F.: L'analyse relationnelle pour la fouille de grandes bases de données. Revue des nouvelles technologies de l'information RNTI-A-2, 149–167 (2007)
5. Condorcet, M.: Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix. De l'imprimerie royale, Paris (1785)
6. Dalal, N., Triggs, B., Schmid, C.: Human detection using oriented histograms of flow and appearance. Lecture Notes in Computer Science 3952, 428 (2006)
7. Doulamis, A.: A fuzzy video content representation for video summarization and content-based retrieval. Signal Processing 80(6), 1049–1067 (juin 2000)
8. Foresti, G., Micheloni, C., Snidaro, L.: Event classification for automatic visual-based surveillance of parking lots. In: Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004. vol. 3, pp. 314–317. IEEE (2004)
9. Hartigan, J.A.: Clustering algorithms. John Wiley & Sons, Inc., New York (1975)
10. Lee, S.W., Mase, K.: Activity and Location Recognition Using Wearable Sensors. IEEE pervasive computing 1(03), 24–32 (2002)
11. Lv, F., Song, X., Wu, B., Singh, V., Nevatia, R.: Left luggage detection using bayesian inference. Proceedings of the 9th IEEE International Workshop (2006)
12. Marcotorchino, F., Michaud, P.: Optimisation en analyse relationnelle des données. Masson, Paris (1978)
13. Piciarelli, C., Foresti, G., Snidaro, L.: Trajectory clustering and its applications for video surveillance. In: Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance, AVSS 2005. vol. 18, pp. 40–45. IEEE (2005)
14. Porikli, F.: Learning object trajectory patterns by spectral clustering. In: 2004 IEEE International Conference on Multimedia and Expo (ICME). vol. 2, pp. 1171–1174. IEEE (2004)
15. Stauffer, C., Grimson, W.E.L.: Learning patterns of activity using real-time tracking. IEEE Trans. Pattern Anal. Mach. Intell. 22(8), 747–757 (2000)
16. Wilson, A.D., Bobick, A.F.: Hidden Markov models for modeling and recognizing gesture under variation. International Journal of Pattern Recognition and Artificial Intelligence 15, 123–160 (2001)
17. Xiang, T., Gong, S.: Video behaviour profiling and abnormality detection without manual labelling. Tenth IEEE International Conference on Computer Vision, 2005. ICCV 2005 2 (2005)