Data Mining in a Video Data Base

Jose Luis Patino Vilchis, Hamid Benhadda, Francois Bremond

Chapter 14

Data Mining in a Video Database

14.1. Introduction

The term "data mining" denotes both a process and a set of techniques used to automatically discover the knowledge contained in very large databases. Its main goal is to extract non-trivial information that may be hidden among the mass of data. Data mining techniques have been used in various domains such as marketing, medicine and finance. These techniques can be divided into two main methods: "supervised" methods, which are based on existing models to recognize new objects or events, and "unsupervised" methods, which discover hidden and useful models in the data. Regarding the application of data mining in the domain of video, we can cite the classification of videos into "categories" for quick searching (video on demand), detection of unusual or suspect activities in prerecorded video-surveillance sequences and selection of traffic models to optimize and best manage traffic.

Comment [sk1]: AQ : Perhaps it should be replaced with [events]. Please check and confirm.

Comment [sk2]: AQ : We have replaced the word [detection] with [selection]. Please check and confirm.

This chapter deals with the issue of unsupervised extraction of activities observed in a video scene recorded by a video-surveillance camera. To do so, we use clustering techniques, which facilitate automatic discovery of typical behaviors and trajectories observed in the scenes. The aim is to automatically learn the zones of activity and use these zones to generate information about the behaviors and events of the trajectories being analyzed. By analyzing both trajectories and events, we can discover significant models of activity which are difficult to apprehend from the raw data.

To obtain semantically interpretable results, we find that the activities we wish to discover are characterized on the basis of contextual elements of the scene under

---

Chapter written by Luis PATINO, Hamid BENHADDA and François BRÉMOND.

observation. These contextual elements may be zones of activity such as building entrances, areas where people come together, road areas, etc.

The global process of analysis presented here is divided into three main stages. In the first stage, the trajectories of the mobile objects evolving in the recorded video scenes are extracted and analyzed to identify the points at which those mobile objects enter and exit the scene and, with them, deduce the various activity zones. In the second stage, the moving objects are characterized in relation to learned zones of activity. Three types of behaviors are considered:

– the moving object stays in a given activity zone;

– the moving object passes from one activity zone to another;

– a combination of the above two behaviors (the moving object stays for a certain amount of time in a given zone and *then* goes into another).

In the third stage, a high-level "clustering" algorithm, based on relational analysis [BEN 07], is used to classify the moving objects identified in the scenes and group them together in clusters, based on the similarity of their behaviors. This enables us to reveal both frequent (habitual) behaviors and unusual (uncommon) behaviors.

These three stages are detailed in sections 14.3–14.5 of this chapter. "Soft computing" techniques are used in the first two stages to deal with the uncertainty inherent in low-level trajectory data. This family, for which a more formal definition was given by L.A. Zadeh [ZAD 94], encompasses techniques imitating human reasoning and based on fuzzy logic, neural networks, genetic algorithms and probabilistic techniques.

14.2. State-of-the-art

Although the general issue of unsupervised learning has been widely studied over the course of the past two decades [MAS 08, PON 08], the use of data mining tools to extract activities from videos is a relatively new discipline, and its application in the field of video behavioral analysis is as yet rather limited. At present, only a few systems use soft computing techniques to determine the characteristic traits of the activities observed in video scenes [DOU 00, LEE 02]. In view of the complexity of configuring the parameters or acquisition of knowledge, most systems are limited to object recognition [DAL 06].

Comment [sk3]: AQ : Please provide full reference detail for references [DAL 06 ; [FOR 04; POR 04].

Three main categories of learning techniques have been studied for behavior recognition. The first category deals with the parameters of a video interpretation program. The techniques under this category have been widely used in the case of

event-recognition methods based on neural networks [FOR 04] or Bayesian classifiers [LV 06, WIL 01]. The second category consists of using unsupervised learning techniques for the detection of abnormal events on the basis of observed events [XIA 05]. The third category is centered on deducing the behavior of objects from their trajectories. This category is the most popular because of its efficiency in detecting normal/abnormal behaviors, e.g. for detecting abnormal trajectories on the roads [PIC 05, STA 05], categorizing pedestrians' trajectories [ANJ 07] or characterizing daily domestic activities [PUS 12, ZUN 12]. Hidden Markov models have also been used to detect different predefined states of normal behaviors [BAS 07, POR 04]. The approach presented below belongs to the third category.

Comment [sk4]: AQ : Please check if the sentence [This ... activities] edited retains the intended sense.

14.3. Pre-processing of the data

To discover significant activities, it is of crucial importance that we have detailed information that can be used to detect the different possible interactions between the objects. Given that the approach presented here is based on trajectory analysis, the first stage in preparing the data that will be used as input for the method of activity clustering is to break down each trajectory into subtrajectories with near-constant velocity, called "tracklets", which characterize the state of an object. This pre-processing stage is illustrated in Figure 14.1.

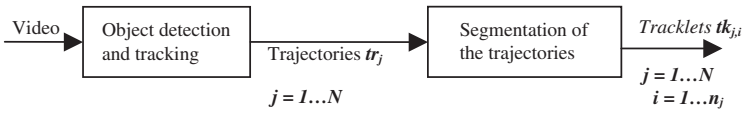


Figure 14.1. *Stages in pre-processing of the data*

The description of object detection and tracking algorithms is beyond the scope of the work presented in this chapter. We can refer to Chapters 7 and 8 of this book, or to [CHA 11].

NOTATION. – The data set is made up of N objects O_j $(j = 1, \ldots, N)$, each object being characterized by its trajectory tr_j, which is, itself, defined as a series of positions $[x_j(t), y_j(t)]$, taken at regular intervals over the course of the observation. In addition, a velocity measurement is defined at each moment by:

$$v_j(t) = \sqrt{(x_j'(t)^2 + y_j'(t)^2)} \qquad [14.1]$$

where $x_j'(t)$ and $y_j'(t)$ denote the (approximate) time derivatives of the functions $x_j(t)$ and $y_j(t)$, respectively.

The aim is to detect the points at which the velocity changes, in order to segment the trajectory into tracklets of near-constant velocity, so that the trajectory is summarized as a series of stationary states or movements at constant velocity. The velocity of the moving objects is studied as part of a multiresolution analysis for a series of discrete temporal data $v(k)$ with a Haar wavelet smoothing function, $\rho_s(k) = \rho(2^s.k)$, dilated to different scales s. In this context, the approximation A of $v(k)$ by ρ is such that $A_{s-1}\ v$ is an approximation whose lowest resolution is $A_s\ v$:

$$\rho_s(k) = \rho(2^s.k),\ A_0\ v = v,\ A_s\ v = \rho_s * v = \rho * A_{s-1}\ v \qquad [14.2]$$

where * denotes the convolution operator.

By analyzing the temporal series v at coarse resolutions, it is possible to take out the insignificant details and choose points associated with significant changes; the location in time of these points is then refined by using finer scales. The points at which the velocity changes are then used to segment the original trajectory tr_j into a series of n_j tracklets $tk_{j,i}$, $i = 1, \ldots, n_j$, each tracklet being defined by its start and end positions:

Comment [sk5]: AQ: Perhaps you mean [take out], so we have replaced [iron out] with [take out].

$$[x_{j,i}(1), y_{j,i}(1)] \text{ and } [x_{j,i}(\text{end}), y_{j,i}(\text{end})] \qquad [14.3]$$

The number M of tracklets extracted from the set of trajectories detected in a scene is equal to the sum of the numbers of tracklets for all the trajectories:

$$M = \sum_{j=1}^{N} n_j$$

By globally re-indexing all the tracklets (let m represent the subscript that denotes the m^{th} tracklet tk_m in any trajectory tr_j from those detected in the scene and $m = 1, \ldots, M$), we can get the following tracklet feature vector tk_m:

$$tk_m = [x_m(1), y_m(1), x_m(\text{end}), y_m(\text{end})] \qquad [14.4]$$

14.4. Activity analysis and automatic classification

The general approach is illustrated in Figure 14.2. We define an activity as being the set of interactions that occur between the moving objects themselves and between those objects and the environment. We use an analysis based on the trajectory of the moving objects in the video to discover the points at which they enter and exit the scene and, finally, deduce the different zones of activity from these observations. In the second stage, the moving objects are characterized in relation to the learned zones of activity. In the third stage, a clustering algorithm is

applied to the moving objects, grouping them together based on their behaviors and thereby revealing frequent/normal and inhabitual/abnormal behaviors and events. The high-level clustering algorithm used to do this is based on the theory of relational analysis.

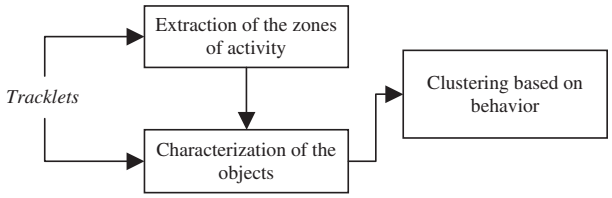


Figure 14.2. *Stages of processing for activity analysis and automatic classification*

14.4.1. *Unsupervised learning of zones of activity*

Modeling the spatial context of a given scene is essential for the recognition and interpretation of the activities observed in that scene. Because it is difficult to manually define all the contextual zones of a given scene, it is not possible to describe all the situations or actions evolving in the scene under observation; hence, it is necessary to learn additional zones for an exhaustive description of that scene.

The feature vector defined in equation [14.4] constitutes a set of descriptors that we can use to describe the activities in a great variety of domains (e.g. traffic surveillance, subway monitoring, surveillance of smart environments), because they primarily define the origin and destination of an object.

In our system, the moving objects that begin (or end) their displacement at positions near one another are considered to share a common zone of activity. Thus, the task of finding the zones of activity in the scene may be considered tantamount to clustering the entry/exit points of the tracklets. Therefore, a number of classification algorithms – e.g. *k*-means clustering or self-organizing maps (SOMs) – can be used, but these algorithms present two drawbacks: on the one hand, we have to fix the number of clusters that must be known in advance and, on the other hand, the morphology of the zones resulting from these techniques is always circular in form (the result of the very definition of the cluster obtained: a center of the cluster and a radius of influence), which cannot correspond exactly to the actual topology of the scene. In our approach, we attempt to find zones of activity with high spatial resolution (small surface), which are then clustered to obtain larger zones of activity, using a hierarchical agglomerative clustering algorithm.

14.4.1.1. *Clustering of the start/end points of the tracklets*

Clustering of the start/end points of the tracklets is performed by a well-known clustering algorithm called "Leader" [HAR 75], which offers the advantage of being able to be executed without having to specify the number of clusters in advance. In this method, it is supposed that a distance threshold T from the center of each cluster is given. The algorithm constructs a partition of the entry space (defining a set of clusters) and a representative or "leader" for each cluster, so that each object in a cluster is at a distance less than T from the leader. The threshold T is thus a measure of the diameter of each cluster. The algorithm processes the entire data set, assigning each object to the cluster whose leader is closest, and constructs a new cluster and a new leader for objects whose distance from all the existing leaders is greater than the threshold. The value of T depends on the application. To set T, we have used the configuration suggested by [PAT 11].

Consider the location of an object to be L; its zone of influence Zn is defined by a radial basis function (RBF) centered on the location L; whether or not a new point p belongs to that zone is measured by equation [14.5]:

$$Zn(L, p) = \phi(L, p) = \exp(-\|p - L\|^2 \,.\, T^2) \qquad [14.5]$$

The RBF has a maximum value of 1 when its input is $p = L$, and therefore acts as a similarity detector; its values decrease as p gets further away from L. An object will be included in a cluster Zn if $Zn(L, p) \geq 0.5$. The receptive surface (hypersphere) of the cluster is controlled by the learned parameter T.

14.4.1.2. *Merging of the tracklets' entry/exit zones*

We find the definitive zones of activity by merging similar entry/exit zones of the tracklets. The new zones are obtained by satisfying two relations:

– $R1$: the zone Zn_i must overlap zone Zn_j;

– $R2$: the zone Zn_i must not overlap a contextual zone $Z\text{ctx}_q$, defined beforehand by the user.

The process is illustrated in Figure 14.3. $R2$ is transformed into a new relation ($R3$) of similarity between zones Zn_i, obtained by transitivity in relation to the overlap of the zones Zn_i with the contextual zones $Z\text{ctx}_q$. Then, by merging relations $R1$ and $R3$ into a final relation R, we can establish the similarity/overlap between zones Zn that do not overlap the zones $Z\text{ctx}_q$ defined by the user. Definitive zones of activity are obtained by thresholding this relation matrix R.

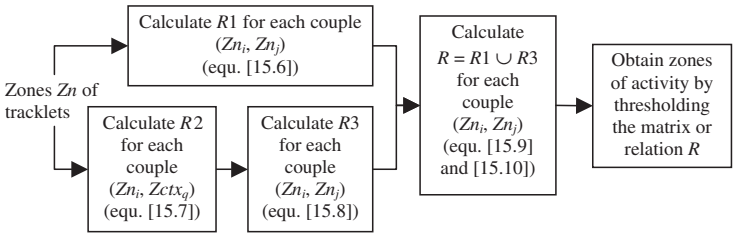


Figure 14.3. *Stages of processing for obtaining zones of activity*

These relations are defined and evaluated by:

– $R1_{ij}$: Zn_i overlaps Zn_j:

$$R1_{ij} = \sum_{k=1}^{3} \left[\sum_{p(x,y) \in (X_{ik}, Y_{ik})} Zn_j \left(L_j, p(x,y) \right) \right] \quad [14.6]$$

where

$$\left(X_{ik}, Y_{ik} \right) = \left\{ \frac{k+1}{3} \cdot T \cos\left(\theta\right) + L_i \right\}, \theta = 0,..., \frac{\pi}{8}, ...2\pi$$

In other words, the summation is performed for a set of points belonging to three concentric circles centered around L_i (thus inside Zn_i). These points are evaluated in order to verify whether they also belong to Zn_j to calculate the overlap/similarity between Zn_i and Zn_j. This helps avoid problems of homogeneity with the clusters defined on scattered regions (certain clusters can be defined with a larger number of points than others);

– $R2_{iq}$: Zn_i overlaps $Zctx_q$:

$$R2_{iq} = \sum_{k=1}^{3} \left[\sum_{p(x,y) \in (X_{ik}, Y_{ik})} Zctx_q \left(p(x,y) \right) \right] \quad [14.7]$$

– $R3_{ij}$: Zn_i and Zn_j are linked to the same contextual zone:

$$R3_{ij} = \max_q \min \left(R2_{iq}, R2_{qj} \right) \quad [14.8]$$

– R_{ij}: Zn_i overlaps Zn_j and the two zones are not linked to the same contextual zone.

Note that $\overline{R3}$, the counterpart of $R3$ given by $\overline{R3} = -R3$, represents the relation between Zn_i and Zn_j if neither cluster is linked to a contextual zone. $R1$ and $\overline{R3}$ can be aggregated by using a soft computing aggregation operator such that:

$$R = R1 \cap \overline{R3} = \max\left(0, R1 + \overline{R3} - 1\right) \quad [14.9]$$

Next, R is rendered transitive with:

$$R \circ R = \max_k \left\{ \min\left[R(Zn_i, Zn_k), R(Zn_k, Zn_j) \right] \right\} \quad [14.10]$$

R is now a transitive relation matrix. R_{ij} gives the extent of similarity between zones Zn_i and Zn_j. If we define a level of discrimination α in the closed interval [0,1], the value of the α-cut can be defined such that:

$$R^{\alpha}(Zn_i, Zn_j) = 1 \Leftrightarrow R(Zn_i, Zn_j) \geq \alpha \quad [14.11]$$

Thus, implicitly, $\alpha_1 > \alpha_2 \Leftrightarrow R^{\alpha_1} \subset R^{\alpha_1}$. Thus, R^{α} induces a grouping of zones $\pi^{\alpha} = \left\{ Zn_i \right\}$ such that $\alpha_1 > \alpha_2$ implies that π^{α_1} is a refined version of π^{α_2}.

At this point, the difficulty lies in selecting an appropriate α-cut value such that π^{α} of R^{α} represents the best distribution of the data. This is always a difficult and open-ended question that we have chosen to deal with by selecting the value of the α-cut, which causes a significant change from π^{α_k} to $\pi^{\alpha_{k+1}}$. To automatically detect significant changes, we have chosen to examine the surface of the cluster and the number of clusters caused at each partition π^{α}. We achieve this objective as part of a multiresolution analysis. By analyzing the partitions caused at coarse resolutions, it is possible to take out the small details and select the α-cut values associated with significant changes.

Comment [sk6]: AQ: Perhaps you mean [take out], so we have replaced [iron out] with [take out].

14.4.2. *Definition of behaviors*

Our aim is to create a system to recognize and interpret human activity and behavior, and to extract new information that could be beneficial to end users. The low-level information should thus be transformed into useful descriptions of a high semantic level. In our application, we establish semantic meaning based on the model of the scene constructed as described in section 14.4.1. The behavior can thus be expressed with semantic concepts, instead of using quantitative data, with the learned zones (contextual zones). Suppose we have K contextual zones $Z\text{ctx}_k$ in the

scene in total, defined *a priori* or after the zone-learning procedure. Two types of behaviors can then be identified:

– objects moving from zone $Z\text{ctx}_{k1}$ to $Z\text{ctx}_{k2}$;

– objects inside zone $Z\text{ctx}_{k}$.

14.4.3. *Relational analysis*

To carry out an in-depth, high-level analysis of activities that are not easy to highlight using raw video data, we use an automated classification technique, "relational analysis", which will enable us to extract complex and hidden correlations between the moving objects in the scenes of the video, on the one hand, and between the moving objects and the contextual objects, on the other hand.

14.4.3.1. *General principles of relational analysis*

The foundation of relational analysis goes back to the work of the Marquis de Condorcet on the voting system [CON 85] in 1785. This work was based on the principle of "consensus" or "majority rule", which stipulates that when a number of judges give their respective opinions about the "validity" or "non-validity" of a decision, the final result must be in line with the majority of opinions given by these judges. In other words, the final decision will be that in favor of which the suffrage represents at least 50% of the judges.

Relational analysis is a clustering technique that lends itself very well to the processing of categorial (or qualitative) data. For this type of data, each of the variables measured on the objects to be clustered may be considered to be a judge, which decides, for each pair of objects, whether or not those two objects belong to the same category. The mathematical formulation in the form of a linear programming problem was first given in [MAR 78]. This formulation was defined as follows: if X is a square matrix, representing the final distribution to be found, in which each term $x_{i,j}$ is equal to 1 if i and j are of the same category in the final distribution and equal to 0 if not, and if $c_{ii'}$ represents the number of variables (or voters) in favor of the decision that the objects i and i' belong to the same category and $\overline{c}_{ii'}$ is the number of variables (or voters) supporting the opposite decision (that the two objects are not in the same category), then the Condorcet criterion $C(X)$ is defined by:

$$C(X) = \sum_{i=1}^{n} \sum_{i'=1}^{n} (c_{ii'} - \overline{c}_{ii'}) x_{ii'} \qquad [14.12]$$

The mathematical formulation in the form of a linear programming problem, which we can use to find the "consensus" distribution, depending on the different variables measured on the objects, is given by equation:

$$(\Pi)\ \begin{cases} \underset{X}{\text{Max}}\left(C(X)\right) & & \\ x_{ii'} \in \{0,1\} & \forall (i,i') \in \Omega^2 & (\text{binarity}) \\ x_{ii} = 1 & \forall i \in \Omega & (\text{reflexivity}) \\ x_{ii'} - x_{i'i} = 0 & \forall (i,i') \in \Omega^2 & (\text{symmetry}) \\ x_{ii'} + x_{i'i''} - x_{ii''} \leq 1 & \forall (i,i',i'') \in \Omega^3 & (\text{transitivity}) \end{cases} \quad [14.13]$$

The classes (or clusters) obtained are described by the modalities that have played the greatest part in their formation. Two indicators are used to this effect: the "characteristic ratio" (*CR*) and the "discriminating ratio" (*DR*). For a given class C with cardinal $|C|$ and a modality j, if n_j^C is the number of objects in the class exhibiting the modality j and n_j is the total number of objects in Ω with that modality, the two indicators are defined by:

$$CR(j) = \frac{n_j^C}{|C|} \qquad DR(j) = \frac{n_j^C}{n_j} \quad [14.14]$$

14.4.3.2. *Application of relational analysis to video data*

The main two types of concepts on which we perform high-level clustering are the moving objects (or "mobiles") observed and the events that have occurred in a video scene. To analyze these concepts, and with the aim of putting the data into a format appropriate for data mining, we have extracted the following specific characteristics from the raw data:

– Mobile ID: the label identifying the object;

– Start: the instant when the object is seen for the first time in the scene;

– End: the instant when the object is seen for the last time in the scene;

– Duration: the total duration of the period when the object is observed in the scene (expressed in seconds);

– Dist_org_dest: the total distance covered between the origin and the destination of a trajectory (expressed in meters);

– Average_velocity: the average of the instantaneous velocities calculated at all the points on the trajectory of the mobile object;

– Significant_event: the main event deduced for the object, with the learned zones taken into consideration.

14.5. Results and evaluations

The algorithm for unsupervised learning of zones of activity (described in section 14.4.1) was applied to an hour-long video scene, recorded in one of the entrance halls of the metro station in Turin. The final relation R given in equation [14.10], which verifies transitive closure, is thresholded for different α-cut values ranging from 0 to 0.9 and with a "step" value of 0.05. The algorithm automatically selects the best α-cut value, attributing a precise composition of activity sectors in the scene, as mentioned in section 14.4.1. Figure 14.4 presents the learned zones corresponding to the video scene being analyzed.

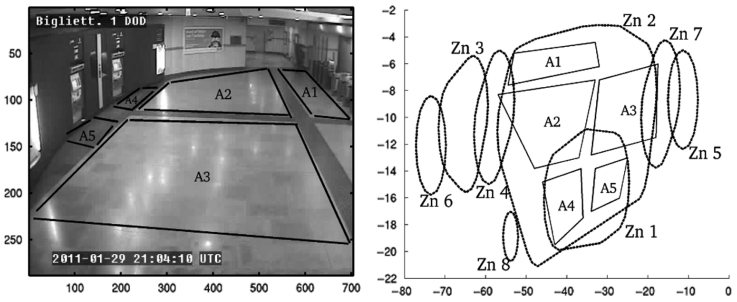


Figure 14.4. *(a) Original scene observed by the camera. A1 to A5 are the zones defined by the user delimiting the scene; (b) the learned zones. Zn1 to Zn8 correspond to the zones of activity discovered by our algorithm*

As mentioned in section 14.4.2, we characterize the behaviors by matching the low-level detections with the learned zones. The collection of activities observed in the scene can then be spoken of as a set of behaviors referring to the learned zones. For instance, for the ticket machine area (zone *Zn*1), the activity report obtained is given in Table 14.1. Zone *Zn*1 is in fact the second busiest of all the zones, just after *Zn*2, which corresponds to the main concourse (39.85% occupancy; not shown in Table 14.2). Most people who go toward the vending machines come from the main concourse (*Zn*2), and all who move away from the machines go toward the main concourse. Those people who go directly to the automatic ticket machines from one of the entrances to the station do so from the entrance zones, *Zn*5 or *Zn*7.

Proportion (%)	Number of mobiles	Description
25.92	106	Inside 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

Table 14.1. *Non-exhaustive report of activity in relation to the ticket machine zone*

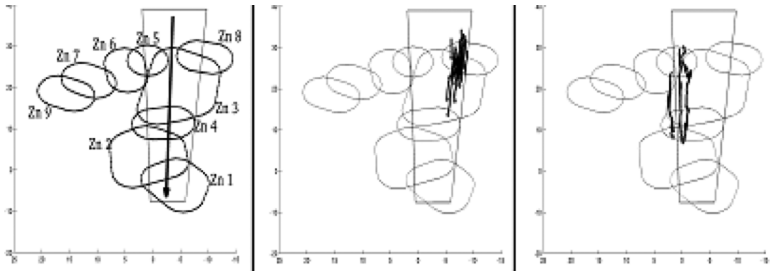
In the first stage of discovery of new information, we apply the process of relational analysis (see section 14.4.3). The input variables are those corresponding to the table of moving objects described above and in section 14.4.3.2, which, in the case of the particular video being analyzed, contains 409 detected objects in total. The variables are weighted so as to favor the formation of as many clusters as there are significant events in the table. The advantage of such a procedure is to use other variables to explain what characterizes each of the significant events.

Cluster 2 (106 elements)			Cluster 4 (16 elements)		
Modality	C ratio	D ratio	Modality	C ratio	D ratio
Event in Zone 1	100	100	Event Zone 1 to Zone 2	100	100
Dist_org_dest 0–3.84	89	92	Dist_org_dest 3.84–7.68	43	12
Duration0–4.2	75	26	Duration 0–4.2	37	2
Velocity 0–1.44	57	44	Velocity 2.88–4.32	37	20
Cluster 19 (2 elements)					
Modality	C ratio	D ratio			
Event Zone 7 to Zone 1	100	100			
Start 22:16:20	50	100			
End 22:16:24.4	50	100			

Table 14.2. *Example of clusters of activity obtained following the application of the procedure of relational analysis*

After the application of relational analysis, 23 clusters of activity are obtained; Table 14.2 shows some examples of the clusters obtained. In this table, for instance, Cluster 2 corresponds to the activity of people in zone *Zn*1, which is the ticket machine area. We can observe that people move little (less than 4 m) and slowly (velocity less than 1.5 m/s) for a short period of time (0–4 s): characteristic behaviors of people buying their tickets. This activity is the second most common activity in the station, just after "inside zone *Zn*2", corresponding to people

remaining in the main concourse (Cluster 1; not shown). Cluster 4 corresponds to people moving away from the ticket machines into the main concourse ("zone 1 to zone 2"); as we can see in the table, these people move a greater distance (between 3.84 and 7.68 m), walking faster (2.88 to 4.32 m/s). The last cluster, Cluster 19, represents a rare activity (only two elements are contained in this activity): the people go from "zone 7 to zone 1". These people are characterized, in particular, by the time at which they are seen (no mobiles detected around the same time). Thus, relational analysis can help us cluster people who exhibit similar behavior. This is of particular interest for the users, because the activities in the metro station can be better quantified.



a) b) c)

Figure 14.5. *(a) Bird's-eye view of the bus lane with the zones learned by our system. The arrow indicates the direction of the flow of buses; (b) and (c) two examples of activities with objects moving against the normal flow of traffic. The ellipses indicate the departure point of these moving objects*

We also apply the proposed method to another domain: the monitoring of traffic control and specific detection of abnormal events in a dedicated bus lane. The advantage of working on this data set is that the ground truth events are available for 45 min of video. We apply the processing chain explained above to this data set; to begin with, our system learns, in an unsupervised manner, the zones of activity in the scene; then we extract the behaviors of the detected mobile objects, associating them with the learned zones. Finally, we apply relational analysis to extract high-level information. Figure 14.5 shows the zones of activity learned from the scene. *Zn*6, *Zn*7 and *Zn*9 correspond to people's zones of activity, whereas the other zones of activity are vehicular zones (see the overlap with the bus lane). The behaviors are then characterized either as remaining in a given zone or moving from one zone to another. We mark south-to-north movement between zones as abnormal activity. Table 14.3 shows the abnormal behaviors established from the learned zones. By applying high-level relational analysis, all the abnormal events – as expected – are gathered together in a single cluster. To test the correctness of the proposed

processing chain, we compare the abnormal events detected by our approach against those indicated by the available ground truth. The results of this comparison can be seen in Table 14.4. The measure of recall is fairly high, and the precision is acceptable. This is mainly because our system produces a relatively high number of false detections, which can occur because of detections on the boundary of the bus lane, which are not included in the ground truths. This is a general problem of human perception, where mobiles only partially observed by the camera are not considered by the expert annotating the video, but those objects are still detected by the system.

Abnormal event	Proportion
*Zn*1 to *Zn*8	34
*Zn*1 to *Zn*5	3
*Zn*1 to *Zn*3	45
*Zn*1 to *Zn*4	5
*Zn*1 to *Zn*2	34
*Zn*2 to *Zn*4	8
*Zn*2 to *Zn*3	12
*Zn*4 to *Zn*3	5
*Zn*4 to *Zn*5	3
*Zn*4 to *Zn*8	10
*Zn*3 to *Zn*5	2
*Zn*3 to *Zn*8	16

Table 14.3. *Abnormal events discovered*

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

Table 14.4. *Measures of evaluation*

14.6. Conclusion

In this chapter, we have presented a system for unsupervised extraction of the main activities observed in a video scene. We have proposed a general processing chain comprising three main stages, beginning with unsupervised learning of the

main zones of activity in the scene. The moving objects are then characterized by employing the learned zones of activity, such as "staying in a given zone of activity", "moving from one zone of activity to another" or a combination of the two activities if the monitoring of the moving objects in the scene takes place over a long enough period of time. Finally, an automatic classification algorithm based on the theory of relational analysis is used to cluster the objects observed based on the similarity of their behavior and to discover correlations that are not at first evident simply from the raw data.

We apply the process to two scenarios. The first scenario relates to the monitoring of the activities in the entrance concourse of a metro station. The results obtained show which zones have the most intense activity in the scene under observation, and rare/uncommon behaviors such as "entering a zone with low occupancy activity" or frequent behaviors such as "buying tickets".

The second scenario relates to the surveillance of a road lane reserved for buses. Again, we are able – on the one hand – to learn the topology of the scene and – on the other hand – to reveal the normal activities (the passage of a bus into the zone) and abnormal activities (the passage of other vehicles into the reserved lane). By comparing the results obtained with the ground truth, a high level of recall and an acceptable degree of precision are obtained. We will attempt to improve these results in our future work by enriching the data and adding new informative variables to the process.

14.7. Bibliography

[ANJ 07] ANJUM N., CAVALLARO A., "Single camera calibration for trajectory-based behavior analysis", *AVSS*, London, United Kingdom, 2007.

[BAS 07] BASHIR F., KHOKHAR A., SCHONFELD D., "Object trajectory-based activity classification and recognition using hidden Markov models", *IEEE Transactions on Image Processing*, 16, pp. 1912–1919, 2007.

[BEN 07] BENHADDA H., MARCOTORCHINO J.F., "L'analyse relationnelle pour la fouille de grandes bases de données", *Revue des Nouvelles Technologies de l'Information (RNTI)*, vol. A-2, pp. 149–167, 2007.

[CHA 11] CHAU D.P., BREMOND F., THONNAT M., CORVEE E., "Robust mobile object tracking based on multiple feature similarity and trajectory filtering", *VISAPP*, Vilamoura, Portugal, 2011.

[CON 85] 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.

[DOU 00] DOULAMIS A., "A fuzzy video content representation for video summarization and content-based retrieval", *Signal Processing*, vol. 80, no. 6, pp. 1049–1067, June 2000.

Comment [sk7]: AQ: Please provide complete conference titles and month-dates in references [ANJ 07; CHA 11; LV 06; PIC 05; XIA 05].

[HAR 75] HARTIGAN J.A., *Clustering Algorithms*, John Wiley & Sons, New York, 1975.

[LEE 02] LEE S.W., MASE K., "Activity and location recognition using wearable sensors", *IEEE Pervasive Computing*, vol. 1, no. 3, pp. 24–32, 2002.

[LV 06] LV F., SONG X., WU B., SINGH V.K., NEVATIA R., "Left luggage detection using Bayesian inference", *PETS*, New York, NY, 2006.

[MAR 78] MARCOTORCHINO F., MICHAUD P., *Optimisation en analyse relationnelle des données*, Masson, Paris, 1978.

[PON 08] PONCELET P., TEISSEIRE M., MASSEGLIA F., *Data Mining Patterns: New Methods and Applications*, Information Science Reference, Hershey, PA, 2008.

[MAS 08] MASSEGLIA F., PONCELET P., TEISSEIRE M., *Successes and New Directions in Data Mining*, Information Science Reference, Hershey, PA, 2008.

[PAT 11] PATINO L., BREMOND F., THONNAT M., "Incremental learning on trajectory clustering", in REMAGNINO P. (ed.), *Intelligent Paradigms in Safety and Security*, Springer-Verlag, Berlin, 2011.

[PIC 05] PICIARELLI C., FORESTI G., SNIDARO L., "Trajectory clustering and its applications for video surveillance", *AVSS*, Como, Italy, 2005.

[PUS 12] PUSIOL G., Event learning based on trajectory clustering, PhD Thesis, Nice Sophia Antipolis University, France, 2012.

[STA 05] STAUFFER C., GRIMSON W.E.L., "Learning patterns of activity using real-time tracking", *IEEE Transactions on PAMI.*, vol. 22, no. 8, pp. 747–757, 2005.

[WIL 01] 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, no. 1, pp. 123–160, 2001.

[XIA 05] XIANG T., GONG S., "Video behaviour profiling and abnormality detection without manual labelling", *ICCV*, Beijing, China 2005.

[ZAD 94] ZADEH L.A., "Soft computing and fuzzy logic", *IEEE Software*, vol. 11, no. 6, 1994.

[ZUN 12] ZUNIGA M., BREMOND F., THONNAT M., "Hierarchical and incremental event learning approach based on concept formation models", *Neurocomputing*, 2012.

Comment [sk8]: AQ: Please check the insertions (issue no. and page range) in reference [WIL 01].

Comment [sk9]: AQ: Please provide page range in reference [ZAD 94].