Data Mining on large Video Recordings

Hamid Benhadda, Jose Luis Patino Vilchis, Etienne Corvee, François Bremond, Monique Thonnat

Data Mining on large Video Recordings

H. BENHADDA (*), J.L. PATINO (**), E. CORVEE (**), F. BREMOND (**), M. THONNAT (**),

Hamid.Benhadda@fr.thalesgroup.com, {jlpatino, Etienne.Corvee, Francois.Bremond, Monique.Thonnat}@sophia.inria.fr

(*)Thales Comunication, 160 Boulevard de Valmy, 92704 Colombes (FRANCE),

(**) INRIA, 2004 route des Lucioles, 06902 Sophia Antipolis (FRANCE)

Mots clefs :

Fouille de données, gestion des connaissances, management, mémoire d'entreprise, ingénierie des connaissances, modélisation des connaisssances, innovation, collecte d'informations, extraction de connaissances, classification automatique, analyse relationnelle, traitement de la video.

Keywords:

Data mining, knowledge management, management, knowledge engineering, innovation, knowledge modeling, information gathering, knowledge discovery, clustering, relational analysis, video processing.

Palabras clave :

Administración del conocimiento, ingeniería del conocimiento, innovación, formalización del conocimiento, colecta de información, extraccion de conocimientos, clustering, analisis relacional, procesamiento de video.

Résumé

L'exploration de larges données vidéo est une tâche qui devient possible grâce aux avancées techniques dans la détection et suivi d'objets. Les méthodes de fouille d'information comme le clustering sont typiquement employées. Celles-ci ont été principalement appliquées pour la segmentation/indexation vidéo mais l'extraction de connaissances sur l'activité présente dans la vidéo a été seulement partiellement adressée. Dans cet article nous présentons comment ces techniques peuvent être utilisées pour traiter de l'information vidéo pour l'extraction de motifs d'activité. Tout d'abord, les objets d'intérêt sont détectés en temps réel. Ensuite, dans un traitement supplémentaire, l'information liée aux objets détectés est mise dans un format qui correspond a un modèle adapté pour la représentation et l'extraction de connaissances. Nous appliquons ensuite deux types de clustering : 1) Clustering hiérarchique agglomératif pour trouver de motifs des trajectoires principales des personnes dans la video 2) Clustering par analyse relationnelle pour extraire des relations spatio-temporelles entre les personnes et les objets contextuels dans la scène vidéo. Nous présentons des résultats de obtenus sur des vidéos du métro de Turin (Italie).

The exploration of large video data is a task which is now possible because of the advances made on object detection and tracking. Data mining techniques such as clustering are typically employed. Such techniques have mainly been applied for segmentation/indexation of video but knowledge extraction on the activity contained in the video has been only partially addressed. In this paper we present how video information is processed with the ultimate aim to achieve knowledge discovery of people activity in the video. First, objects of interest are detected in real time. Then, in an off-line process, the information related to detected objects is set into a model format suitable for knowledge representation and discovery. We then apply two clustering processes: 1) Agglomerative hierarchical clustering to find the main trajectory patterns of people in the video 2) Relational analysis clustering, which we employ to extract spatio-temporal relations between people and contextual objects in the scene. We present results obtained on real videos of the Torino metro (Italy).

1 Introduction

Nowadays, more than ever, the technical and scientific progress requires human operators to handle more and more quantities of data. To treat this huge amount of data, most of the work can now be performed in the data-mining field to synthesize, analyze and extract valuable information, which is generally hidden in the raw data. Mining of text documents [4,8,15], web-related information [5,7,12], and video [6,13] are some examples. Clustering is one of the most commonly used techniques in data mining to perform knowledge discovery tasks on large amount of data with no prior knowledge of what could be hidden in the data. There exists many clustering techniques in the literature, and the main goal of all these techniques is to obtain a partition of the data by organizing it automatically into separate groups where the objects inside a specific group are more similar to each other (with regards to their extracted and measured attributes, or variables) than to the objects of the other groups. Applying data mining techniques in large video data is now possible also because of the advances made on object detection and tracking [14]. However in most cases data mining techniques have mainly been employed for segmentation/indexation of video but knowledge extraction on the activity contained in the video has been only partially addressed. In this paper we present how clustering techniques can be employed on large video data to to achieve knowledge discovery of people activity in the video and find patterns of interaction between people and contextual objects in the scene. The sequences analysed were captured by the surveillance cameras in the metro sites of Torino (Italy). We have developed specific algorithms [1,14] to detect in real time objects and events of interest present in the video. The procedure is explained in section 2. Then, in an off-line process, we set the information related to the detected objects and events into a suitable knowledge modelling format, which we have conceived. In this model, we calculate a series of descriptors to obtain clear and compact information of human activity in the observed scene. We then apply two clustering processes: 1) Agglomerative hierarchical clustering to find the main trajectory patterns of people in the video 2) Relational analysis clustering [2,3,11], which we employ to extract spatio-temporal relations between people and contextual objects in the scene. The knowledge modelling format and the trajectory characterisation using the agglomerative hierarchical clustering are described in section three. In section four we present the background of relational analysis while section five gives the obtained results.

2 Object and Event detection

This section describes the functionalities first of the on-line multiple object tracker algorithm and second the on-line event detector algorithm.

2.1 Multiple object tracking

The multiple object tracker first detects and classifies into semantic classes mobile objects from a given background (reference) image as described in the following section. The latter section then describes how these classified objects are tracked throughout time by the 'long term tracker' algorithm.

2.1.1 Moving object detection

A motion detector algorithm allows the detection of objects before being classified and tracked throughout time. The motion detector segments from the background reference image the foreground pixels which belong to moving objects by a simple thresholding operation. The foreground pixels are then spatially grouped into moving regions represented by the bounding boxes as shown in figure 2.1. This figure shows 9 moving regions whereas there are actually 7 persons in the scene. The

two farthest persons from the camera could not be detected because of the noise present in the image and because of their small dimensions on the image plane. The two persons using the ticket vending machines are too close to each other to be separated into two individual sets of connected pixels. The closest person from the camera, located on the right, was well segmented. The two remaining persons located on the left side of the image are in an dynamic occlusion scenario where one is located in front of the other and although they are far apart they were segmented into one unique moving region. These mobile objects are then classified into semantic model classes according to their 3D sizes. The 3D semantic models used in this application are the person, luggage, group of persons and the crowd model. In figure 2.2, a person and a crowd object are detected as well as an unknown object which 3D dimensions do not fit any given model. In this figure, the hall's ground boundaries are drawn in red and three contextual objects are also drawn: two automatic ticket dispensers on the right side of the image referred to as 'Vending machine 1 and 2' and the 10 validating ticket machines are drawn as one contextual object referred to as 'Gates'.

Figure 2.1 Bounding boxes, representing moving regions, that were found by the motion detector algorithm

Figure 2.2 Person detected (right-most bounding box) and crowd detected (upper-left bounding box)

2.1.2 Long term tracking

A Frame to Frame (F2F) tracker links the list of mobile objects detected in each pair of successive frames. The output of the F2F tracker is a graph of linked mobile objects. This graph provides all the possible trajectories that a mobile object may take. The link between two mobile objects is associated

with a weight, or a 'matching probability', computed from three criteria: the similitude between their semantic classes (e.g Person, Luggage, Crowd), the similarity between their 2D dimensions (bounding boxes width and height) and their 3D distances from each other on the ground plane.

The tracking problem is a central issue in scene interpretation, as erroneous object detection and loss of tracked objects prevents the analysis of their behaviours. For example, trackers have problems coping with detection errors (e.g. caused by shadows) and with occlusions (e.g. two persons crossing each other). For this reason, in order to obtain a more reliable tracking and thus more efficient behaviour recognition, an additional step was added to improve the tracking of objects in the long term by computing the history of the tracked mobile objects evolving in the scene. This method called the 'Long Term Tracker' or LTT, takes as input the results of the F2F (Frame to Frame) tracker which keeps a record of the links between the mobile in a small temporal windows of frames. The LTT uses the explicit geometric model of individual and computes several possible paths for each individual (using a time delay). This is achieved by building the model of paths and individuals as described in the following.

- The path model

A path represents a possible trajectory of one individual in the scene throughout a small period of time interval T. A path is composed of a temporal sequence of mobile objects, fulfilling two conditions. The first condition is introduced in order to limit the number of paths computed in case of over-detection (if several mobile objects represent different parts of the same person we choose the best one). The second condition prevents the creation of paths with large spatial discontinuities. A quality coefficient Q_t qualifies each path at time t, taking into account three coefficients C_1, C_2, and C_3 and the previously computed coefficient Q_{t-1}:

- C_1: the coherency between each mobile of the path and a model of human being (including 3D size and class of mobile objects, like Person or Group).
- C_2: the temporal coherency between the different mobile objects composing the path (e.g. 3D size, motion).
- C_3: the inverse of the 2D distance between two successive mobiles compared with a human movement model.

This quality coefficient is given by: $Q_t = C_1 . C_2 . C_3 . Q_{t-1}$

A new path is created when a mobile object is detected. A path is updated using the new mobile objects detected at current time t and is deleted when its quality coefficient is too low.

- The individual model

An individual is a structure representing a real person moving in the scene at time $t-T$, where T is the delay between the update of paths, performed at time t, and the update of individuals (see Figure 2.3). It represents the output of the Long Term Tracker algorithm. Each individual is associated with a set of paths, representing all its possible trajectories. A path corresponds to a track of a potential person whereas the structure individual corresponds to

a real person. A new individual is created when we are sure it corresponds to a real person i.e. when a path **P**, not yet associated with any individual, fulfils the following four conditions:

- **P** starts in an authorised entering or exiting zone.
- **P** ends in an authorised entering or exiting zone.
- **P** is composed of **T** mobile objects.
- **P** does not overlap with a path already associated with another individual.

The goal of these four conditions is to avoid creating a new individual using a wrong path due to detection errors (e.g. shadows). Time delay **T** is to ensure that the individual chooses the correct path. An individual is updated with the most compatible path (see below). It is deleted when its last mobile objects are inside/outside a zone and when there is no longer a mobile object newly detected (occurs when the real person has left the field of view).
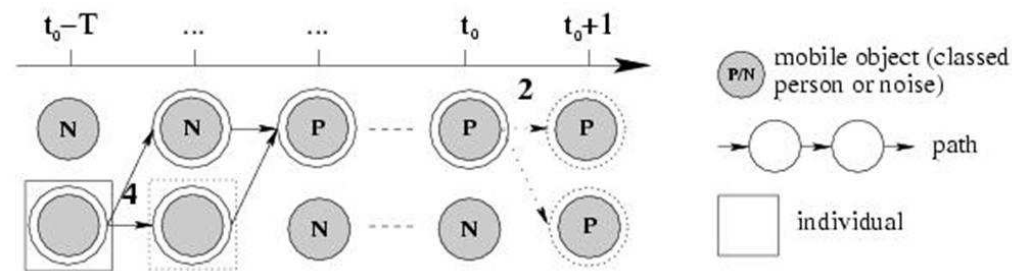
Figure 2.3 Long term tracker Individual structure

Having an already existing set of individuals, the Long Term Tracking algorithm computes all the possible paths that these individuals can undertake. The individuals are associated with the most probable path and their tracks are then updated before the same procedure is launched for the newly grabbed frame. Example of tracked individuals are shown in figure 2.4. These mobiles objects were well classified and tracked although the segmentation results were poor (see figure 2.1 as example) and so did the classification.

Figure 2.4 Tracking results for images corresponding to frames indexed 978, 999, 1008 and 1059. Three major characters are evolving in this scene: two persons and one group of persons.

2.2 Simple events detection

The trajectories of each detected objects given by the tracking algorithm aims to build a relationship between the objects with the contextual content of the scene. In this application, 10 'gates' (i.e. machines enabling the access to the platform), 2 ticket vending machines and one platform (or central hall) compose the scene. The platform delimits the ground floor where all mobiles are allowed to evolve and the machines describe the contextual objects the mobiles interact with. The object trajectories give temporal information about the interaction of the objects with the scene components. The detected events are the following:

- '*inside_zone(o, z)*': when an object 'o' is inside the zone 'z'.
- '*stays_inside_zone(o,z,T1)*': when the event 'inside_ zone(o,z)' is being detected successively for at least T1 seconds
- '*close_to(o, eq, D)*': when the 3D distance of an object location on the ground plane is less than the maximum distance allowed, D, from an equipment object 'eq'
- '*stays_at(p,eq,T2)*': when the event e '*close(o, eq, Dmax, T2)*' is being consecutively detected for at least T2 seconds.
- '*crowding_in_zone(c,z)*': when the event e '*stays_inside_zone (c, z, T3)*' is detected for at least T3 seconds.

where,

* object o={p, g, c, l, u} with p=person, g=group, c=crowd, l=luggage, and u=unknown.
* zone z={platform, validating_zone, vending_zone}
* equipment eq={g1, …, g10, vm1, vm2 } where 'gi' is the ith gate and vmi is the ith vending machine.

T1=60 s, D=1m50, T2=5 s, T3=120 s

Figure 2.5 and 2.8 show one person staying at the Gates and one staying at gate number 7 respectively. The corresponding event is defined by '*stays_at(p, gates, 5 seconds)*' and '*stays_at(p, gate7, 5 seconds)*'. Figure 2.5 also shows a detected crowd. Figure 2.6 and 2.9 show a group of person staying at the vending machine number 2 and at the gates respectively. Figure 2.7 and 2.10 show a person and a group of persons staying inside the hall referred to as platform. The corresponding events are '*stays_inside_zone(p,platform)*' and '*stays_inside_zone(g,platform)*'.

These figures also show erroneously detected objects. For example, the legs of one person in figure 2.10 were detected as a luggage and various unknown objects were detected in the other frames. Although noise often occurs, the Long Term tracker and event detection algorithms allowed meaningful trajectories to be obtained for relatively reliable simple event detection.

Figure 2.5 Person detected staying at the gates (smaller bounding box) and a crowd.

Figure 2.6 Group of persons staying at the vending machine (biggest bounding box).

Figure 2.7 Person detected staying in the hall.

Figure 2.8 Person detected staying at gate 7 (biggest bounding box).

Figure 2.9 Group of persons staying at the gates (biggest bounding box).

Figure 2.10 Group of persons staying in the hall (biggest bounding box) and a luggage (smallest bounding box) erroneously detected.

3 Knowledge Modelling

In order to have a clear and compact representation of the human activity evolving on the video and with the aim to achieve further knowledge discovery, we have divided all related information to objects and events detected on the video into three different semantic tables. Mobile objects table, events table and contextual objects table. Some structured knowledge representations have been presented before [9,10], but in this contribution we propose a model which takes into account interactions between tracked objects in the video and their environment.

3.1 Mobile Objects Table

The mobile objects table allows us to characterise the detected objects in the scene. The following fields describe the mobile objects table:

m_id: the identifier label for the object.

m_type: the class the object belongs to: Person, Group, Crowd or Luggage.

m_start: time the object is first seen.

m_end: time the object is last seen.

m_shape: the label describing the object's shape depending on the object's ratio height/width.

m_involved_events_id: all occurring Events related to the identified object.

m_significant_event: the most significant event among all events. This is calculated as the most frequent event related to the mobile object.

m_trajectory_type: the trajectory pattern characterising the object.

3.1.1 Agglomerative Hierarchical Clustering for Mobile Objects

To describe this last field in the mobile objects table, we implemented a hierarchical clustering algorithm to group similar trajectories after a given observation time. The input feature vector of the clustering algorithm is composed of two key points defining each trajectory, namely the first and last point [xi(1),yi(1),xi(end),yi(end)]. The dendrogram resulting after applying the algorithm is unique but the final number of clusters in which the data set is to be divided is subjective. In our case, the end-user can interactively choose the final number of clusters. As an example, Figure 3.1 shows the dendrogram's results after the clustering of about 100 trajectories. Figure 3.2 shows a cluster representative trajectory. The clustering of trajectories was validated on the Caviar dataset, an EC founded project that has made available a dataset of video clips with hand-labelled ground truth (http://homepages.inf.ed.ac.uk/rbf/CAVIAR/).

3.2 Events Table

The Events table allows us to learn the normal activity in the scene observed by the camera. The events table contains the following fields:

e_id: The identifier label for the detected Event.

e_type: The class the Event belongs to ('close_to', 'stays_at', …)

e_start: first moment on which the Event is detected.

e_end: last moment on which the Event is seen.

e_involved_mobile_object_id: the identifier label of the object involved in that event.

e_involved_contextual_object_i: the name of the contextual object involved in that event.

Figure 3.2 Trajectory type corresponding to the third cluster (left to right) in figure 3.1.
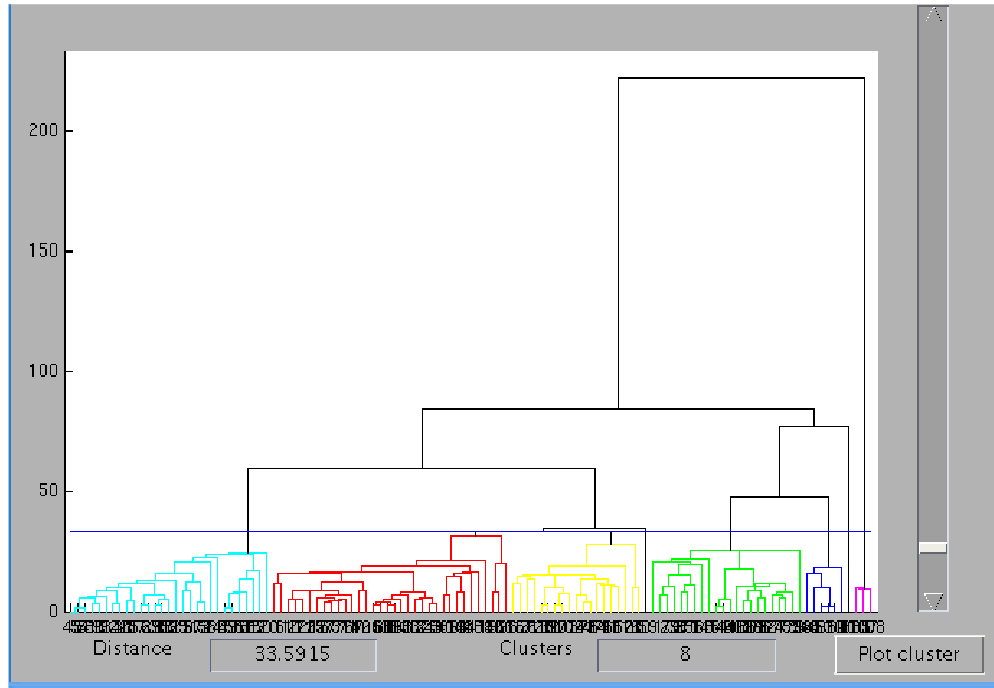
Figure 3.1 Dendrogram formed with the agglomerative hierarchical clustering algorithm. The y-axis represents the distance between records in the dataset. The horizontal line is the threshold at which the partition groups are formed. Here the distance threshold is 33.59, which leads to 8 clusters.

3.3 Contextual Objects Table

This last semantic table allow us to learn what the interactions between persons and contextual objects of the scene are. The descriptive fields are the following:

c_id; c_type; c_involved_events_id; c_significant_event: are defined in the same way as for the mobile objects but referring to contextual objects.
c_start and c_end: refer to the first and last instant the mobile object interacts with the contextual object respectively.
c_rare_event: the rarest event.

c_event_histogram: gives the frequency of occurrence of all involved events.

c_involved_mobile_objects_id: all detected mobile objects interacting with the contextual object of interest.

c_histogram_mobile_objects: gives the frequency of appearance for all involved mobile objects.

c_use_duration: percentage of occupancy (or use of a contextual object). For instance, the Ticket Machine has a 10% of use over the observation time.

c_mean_time_of_use: average time of interactions between the mobile object and the contextual object.

Having defined these three semantic tables, relational analysis for the discovery of hidden information can now be performed as described in the next section.

4 Relational analysis theory

The relational analysis theory is a data analysis technique that has been initiated and developed at IBM in the 1970s, by F. Marcotorchino and P. Michaud [11]. This technique is used to resolve many problems that occur in fields like: preferences, voting systems, clustering, …etc. It uses the "pairwise comparison principle". If the data is made up of N objects $(O_1, O_2, \ldots, O_N)$ on which M attributes (or variables) $(V_1, V_2, \ldots, V_M)$ have been measured then the "pairwise comparison principle" consists in transforming the data, which is usually, represented by a $N \times M$ rectangular matrix into two squared $N \times N$ matrices S and $\overline{S}$. The matrix S, which is called the global relational Condorcet's matrix, of general term s_{ij} representing the global similarity measure between the two objects O_i and O_j over all the M attributes and matrix $\overline{S}$ of general term $\overline{s}_{ij}$ representing the global dissimilarity measure of these two objects. To get matrix S, each V_k attribute is transformed into a squared $N \times N$ matrix S^k of general term s_{ij}^k representing the similarity measure between the two objects O_i and O_j with regards to attribute V_k. To get matrix $\overline{S}$, a dissimilarity measure $\overline{s}_{ij}^k$ of objects O_i and O_j with regards to attribute V_k is then computed as the complement to the maximum similarity measure possible between these two objects. As the similarity between two different objects is less or equal to their self-similarities: that is if $s_{ij}^k \leq \min\left(s_{ii}^k, s_{jj}^k\right)$ then $\overline{s}_{ij}^k = \min\left(s_{ii}^k, s_{jj}^k\right) - s_{ij}^k$. This leads to a dissimilarity measure matrix $\overline{S}^k$. The matrices S and $\overline{S}$ are then obtained by summing, respectively, all the matrices S^k and $\overline{S}^k$, that is $S = \sum_{k=1}^{M} S^k$ and $\overline{S} = \sum_{k=1}^{M} \overline{S}^k$. The global similarity between each two objects O_i and O_j is thus $s_{ij} = \sum_{k=1}^{M} s_{ij}^k$ and their global dissimilarity is $\overline{s}_{ij} = \sum_{k=1}^{M} \overline{s}_{ij}^k$.

4.1 Condorcet's criterion

To cluster a population of N objects described by M variables, the relational analysis theory maximises the Condorcet's criterion $C(X) = \sum_{i=1}^{N} \sum_{j=1}^{N} \left(s_{ij} x_{ij} + \overline{s}_{ij} \overline{x}_{ij} \right)$

where X is a binary $N \times N$ matrix of general term x_{ij} representing the partition to discover in the data, that is:

$$x_{ij} = \begin{cases} 1 & \textit{if } i \textit{ and } j \textit{ are in the same cluster} \\ 0 & \textit{otherwise} \end{cases}$$

and $\overline{x}_{ij}$ is defined as : $\overline{x}_{ij} = 1 - x_{ij}$.

The mathematical formulation of the associated clustering problem based upon the relational analysis theory is:

$$Max(C(X))$$

$$\textit{where } X \textit{ satisfies}:$$

$$\begin{cases} x_{ii} = 1 & \textit{reflexivity} \\ x_{ij} = x_{ji} & \textit{symetry} \\ x_{ij} + x_{jk} - x_{ik} \leq 1 & \textit{transitivity} \end{cases}$$

One of the advantages that relational analysis brings with regards to most other existing clustering techniques is that there is no need to fix arbitrarily the number of clusters to be discovered, the clusters and their number are found automatically by the algorithm (in a really unsupervised way). Even if exact solutions of this binary linear programming problem has been obtained for $N \leq 120$, for bigger sizes of data we use ad-hoc heuristics

Using the definition of $\overline{x}_{ij}$, the Condorcet's criterion can be written as $C(X) = \sum_{i=1}^{N} \sum_{j=1}^{N} \left(s_{ij} - \overline{s}_{ij} \right) x_{ij} + \sum_{i=1}^{N} \sum_{j=1}^{N} \overline{s}_{ij}$. This formulation of the Condorcet's criterion shows

that to get the criterion maximization it is necessary to, a priori, put two objects in the same cluster of the resulting partition whenever their global similarity is greater than their global dissimilarity.

4.2 Illustrative example

To illustrate the construction of the different matrices quoted above, let us consider a population made up of five objects noted $\{O_1, O_2, O_3, O_4, \text{ and } O_5\}$ on

which three qualitative (or categorical) attributes $V^1 =$ "Sex", $V^2 =$ "Socio-professional category "having two modalities {Executive, Worker} and $V^3 =$ "nationality" having three modalities {French, Spanish and English} had been measured. Let's suppose that:
- the first three individuals are women and the two last are men,
- the first two individuals are executives and the three last are workers and
- the first two individuals are French, the two following individuals are Spanish and the last one English

As all the attributes are of categorical type, we can define the similarity s_{ij}^k and dissimilarity $\bar{s}_{ij}^k$ between each two objects O_i and O_j with regards to each attribute V_k $(k=1,2,3)$ as:

$$s_{ij}^k = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are of the same category} \\ 0 & \text{otherwise} \end{cases}$$

and $\bar{s}_{ij}^k = 1 - s_{ij}^k$. The raw data and the three relational matrices are given bellow:

Raw data representation				Relational representation																	
				S^1						S^2						S^3				
V^1	V^2	V^3	Individuals	O_1	O_2	O_3	O_4	O_5		O_1	O_2	O_3	O_4	O_5		O_1	O_2	O_3	O_4	O_5
Woman	Executive	French	O_1	1	1	1	0	0		1	1	0	0	0		1	1	0	0	0
Woman	Executive	French	O_2	1	1	1	0	0		1	1	0	0	0		1	1	0	0	0
Woman	Worker	Spanish	O_3	1	1	1	0	0		0	0	1	1	1		0	0	1	1	0
Man	Worker	Spanish	O_4	0	0	0	1	1		0	0	1	1	1		0	0	1	1	0
Man	Worker	English	O_5	0	0	0	1	1		0	0	1	1	1		0	0	0	0	1

The global similarity and dissimilarity matrices are then:

	Global matrices									
	S						$\bar{S}$			
Individuals	O_1	O_2	O_3	O_4	O_5	O_1	O_2	O_3	O_4	O_5
O_1	3	3	1	0	0	0	0	2	3	3
O_2	3	3	1	0	0	0	0	2	3	3
O_3	1	1	3	2	1	2	2	0	1	2
O_4	0	0	2	3	2	3	3	1	0	1
O_5	0	0	1	2	3	3	3	2	1	0

5 Results

We analysed two video sequences of the Torino metro. One lasting 45 minutes and the second one 2 h 20 min. On the first case we detected 2052 mobile objects and clustered their associated trajectories into 21 clusters employing the hierarchical clustering algorithm. Each cluster represents then a trajectory type. The characterisation of trajectories gives important information on behaviour and flows of people. For instance, trajectory cluster 6 shows people coming from north doors and going to use the vending machines. Trajectory cluster 12 shows people coming from north doors / gates and exiting through south doors. (see Figure 5.1 and 5.2 respectively). Once the trajectories of mobile objects were characterised, all information was formatted according to the semantic tables given in section 3. From the contextual objects table, we were able to follow the evolution of contextual objects. Figure 5.3 shows the number of people occupying the platform during the observation time. For this example, it exceeds 200 at 6 h 45 min. Figure 5.4 shows the evolution on the usage of a vending machine. Interestingly, a user spends more time in the vending machine when the platform hall is not too much busy.

One portion of the mobile objects table is presented in Figure 5.5., this table was then employed as input of the relational algorithm presented in section 4. Some of the clusters found using relational analysis are presented below. Figure 5.6 (Cluster 9) represents people in platform hall but associated with trajectories of type 6 'entering north doors – going to the vending machine (shown in Figure 5.1). Figure 5.7 presents another cluster (11) including only 28 persons but they all have in common a trajectory of type 12 'exiting through south doors' (shown in Figure 5.2) and all persons were detected inside the platform.
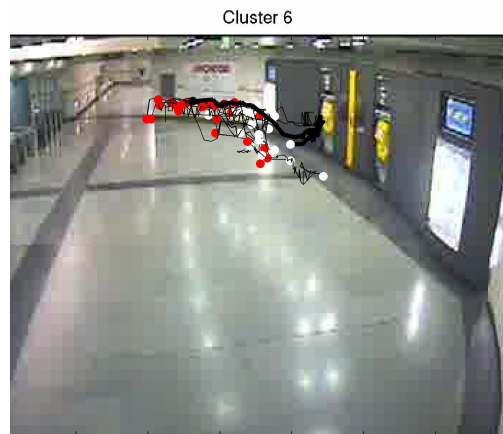

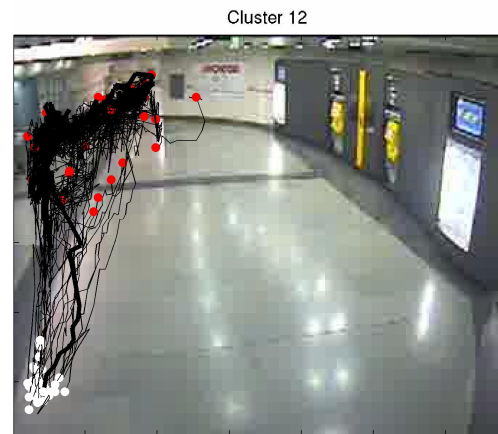Figure 5.1 Trajectory cluster 6: people coming from north doors and going to use the vending machines

Figure 5.2 Trajectory cluster 12: people coming from north doors / gates and exiting through south doors
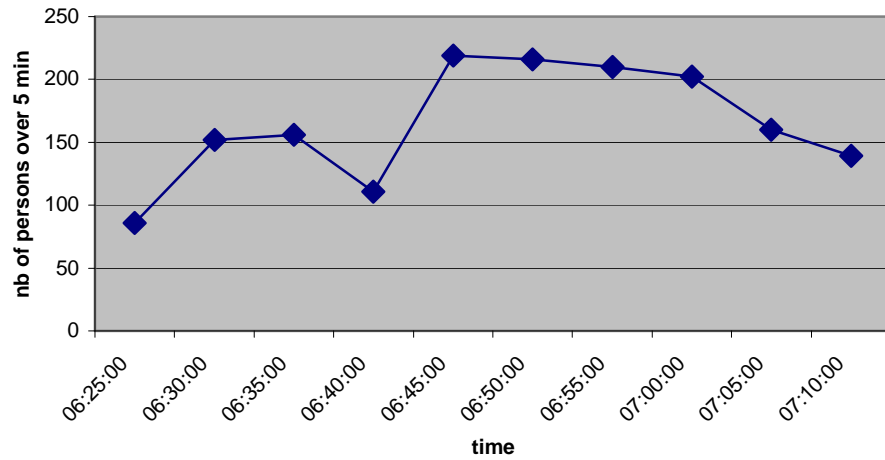
Figure 5.3 Temporal evolution in the number of people occupying the station hall.
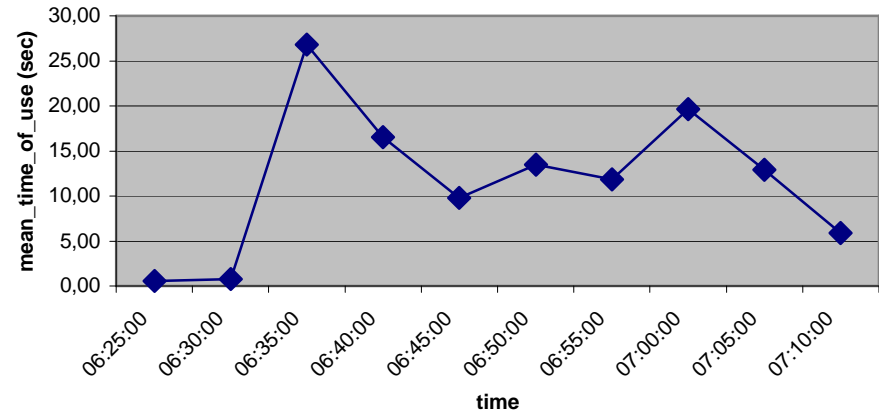
Figure 5.4 Temporal evolution in the mean time a user spends on a vending machine in the station hall.

mob_obj_id	mob_obj_type	startframe	endframe	traj_type	shape_type	sig_evnt_id
6409	Person	81305	81365	2	small	inside_zone_Platform
6412	Person	81335	81335	1	tall	inside_zone_Platform
6392	PersonGroup	81130	81280	2	small	group_inside_zone_Platform
6400	PersonGroup	81210	81210	1	small	group_inside_zone_Platform
6399	Person	81200	81205	1	tall	inside_zone_Platform
6385	Person	81090	81150	2	tall	inside_zone_Platform
6381	PersonGroup	81030	81075	2	small	group_inside_zone_Platform
6374	Luggage	80950	81025	2	big	inside_zone_Platform
6370	Person	80900	80945	1	tall	inside_zone_Platform
6369	Luggage	80855	80935	2	large	inside_zone_Platform
6366	Person	80775	80775	1	tall	inside_zone_Platform
6353	Person	80205	80390	2	small	inside_zone_Platform
6361	Person	80310	80325	1	tall	inside_zone_Platform
5942	Person	76930	80265	2	tall	inside_zone_Platform
6342	Person	79780	80195	2	small	inside_zone_Platform
6322	PersonGroup	79565	79875	1	large	group_inside_zone_Platform

Figure 5.5 Mobile objects table

Typical Individual		
Variable	Modality	Percent
TRAJECTORY_TYPE	6	100.0%
ENDFRAME	45845-84375	60.0%
SIGN_EVENT_ID	inside_platform	60.0%
STARTFRAME	48975-84220	56.0%
MOB_OBJ_TYPE	person	46.0%
SHAPE_TYPE	small	30.0%

Figure 5.6 Cluster 9 in the Relational Analysis Result

Typical Individual		
Variable	Modality	Percent
TRAJECTORY_TYPE	12	100.0%
SIGN_EVENT_ID	inside_platform	82.0%
MOB_OBJ_TYPE	person	60.0%
ENDFRAME	45845-84375	53.0%
STARTFRAME	48975-84220	50.0%
SHAPE_TYPE	small	32.0%

Figure 5.7 Cluster 11 in the Relational Analysis Result

6 Conclusion

In this paper it has been shown how clustering techniques can be applied on video data for the extraction of meaningful information. First hierarchical clustering was applied in order to obtain the most common trajectories undertaken by underground users. Then, we apply the relational analysis with the aim to achieve knowledge discovery of higher complex events. For this purpose we created a specific knowledge modelling format that gathers all information from tracked objects of interest in the scene. This kind of representation allows the end-user to explore the interactions between people and contextual objects of the scene. In this way it is possible to obtain statistics on the underground activity and thus optimise available resources. The relational analysis works directly on all features characterising mobile objects as it has the enormous advantage of being able to analyse heterogeneous variables. This let us find relationships between people, their trajectories and their occurrences. In our future work we will refine features such as object's trajectory and shape by adding sub-categories and look for new relationships employing relational analysis.

7 Bibliographie

[1]   AVANZI A., BREMOND F., TORNIERI C., THONNAT M., *Design and assessment of an intelligent activity monitoring platform*, *EURASIP*, pp. 2359-2374, 2005

[2]   BENHADDA H., *La similarité régularisée et ses applications en classification automatique*, doctorat de l'université paris VI, 1997

[3]   BENHADDA H., MARCOTORCHINO F., *Introduction à la similarité régularisée en analyse relationnelle*, *revue de statistique*, Vol. 46, N°1, pp. 45-69, 1998

[4]   BLATAK J., *First-order Frequent Patterns in Text Mining*, *Proceedings of EPIA,* pp. 344 – 350, 2005

[5]   CHIA-HUI CH., KAYED, M., GIRGIS, M.R., SHAALAN, K.F., *A Survey of Web Information Extraction Systems, IEEE Transactions on Knowledge and Data Engineering*, Vol. 18, pp. 1411-1428, 2006

[6]   DAI K., ZHANG J., LI G., *Video mining: concepts, approaches and applications, Proceedings of the 12th International Multi-Media Modelling Conference*, 4 pp., 2006

[7]   FACCA FM., LANZI PL., *Mining interesting knowledge from weblogs : a survey, DataKnowledge Engineering,* Vol. 53, pp. 225-241, 2005

[8]   LEMOINE J., BENHADDA H., AH-PINE J., *Classification non supervisee de documents heterogenes : Application au corpus '20 Newsgroups', The 11th IPMU International Conference,* Vol. 3, pp. 2538-2544, 2006

[9]   LIN H., CHEN A. L. P., *Motion event derivation and query language for video databases*, *Proceedings of SPIE*, Vol. 4315, pp. 208-218, 2001

[10] LIU D., HUGHES C. E., *Deducing Behaviors from Primitive Movement Attributes*, *Defense and Security Symposium, Proceedings of the SPIE*, Vol. 5812, pp. 180-189, 2005

[11] MARCOTORCHINO F., MICHAUD P., *Optimisation en analyse ordinale des données*, Masson, 1978

[12] MCCURLEY K.S., TOMKINS, A., *Mining and knowledge discovery from the Web*, *Proceedings of the 7th International Symposium on Parallel Architectures, Algorithms and Networks*, pp. 4-9, 2004.

[13] OH J., BANDI B., *Multimedia data mining framework for raw video sequences, Multimedia data mining MDM/KDD*, pp. 1-10, 2002

[14] VU VT., BREMOND F., THONNAT M., *Automatic video interpretation: A novel algorithm for temporal scenario recognition, Proceedings of the IJCAI'03,* pp. 1295-1302, 2003

[15] XING J., AH-HWEE T., *Mining ontological knowledge from domain-specific text documents, Fifth IEEE International Conference on Data Mining*, 4 pp., 2005