Human Behaviour Recognition based on Trajectory Analysis using Neural Networks

Jorge Azorín-López, Marcelo Saval-Calvo, Andrés Fuster-Guilló, José García-Rodríguez

***Abstract*—Automated human behaviour analysis has been, and still remains, a challenging problem. It has been dealt from different points of views: from primitive actions to human interaction recognition. This paper is focused on trajectory analysis which allows a simple high level understanding of complex human behaviour. It is proposed a novel representation method of trajectory data, called Activity Description Vector (ADV) based on the number of occurrences of a person is in a specific point of the scenario and the local movements that perform in it. The ADV is calculated for each cell of the scenario in which it is spatially sampled obtaining a cue for different clustering methods. The ADV representation has been tested as the input of several classic classifiers and compared to other approaches using CAVIAR dataset sequences obtaining great accuracy in the recognition of the behaviour of people in a Shopping Centre.**

I. Introduction

HUMAN behaviour analysis is an old subject in computer science, but it remains a very important and researched topic. It has been named such as behaviour, action/ activity, event, scenario… recognition/analysis, depending on the authors, or the purpose of the research. In this paper we will use human behaviour because the aim is oriented to high level understanding of activities performed by people.

The problem of human behaviour recognition usually takes two things into account: the level of understanding which is going to be analysed, and the method used to do this. Different terminology has been proposed for levels of understanding discrimination, reviewed in diverse papers, including [1], [2] and [3].

Diverse methods can be found in the literature to analyse behaviours from video sequences. They could be grouped in state models (e.g. Bayesian, HMM), pattern recognition (e.g. Neural Networks, SVM), and semantic models (e.g. Petri Nets, grammars), following the scheme in [1]. Both state and semantic models have to predefine a model and rules to evaluate the behaviour. However, pattern recognition methods use the actual data to cluster the space of solutions, being more flexible.

Taking into account the two aforementioned aspects, this paper is focused in the highest level of understanding, the behaviour (also known as activity, event or context), and the use of pattern recognition methods to analyse it. Within this category of methods, approaches including Support Vector Machines, Neural Networks, and Nearest Neighbours have been presented for human behaviour or activity recognition, Particularly, human behaviour study could be dealt in many ways from gestures, movements, trajectories, etc. We are interested in trajectories due to it is simple to calculate, represent and allow high-level understanding of many human behaviours. For example, as the motivation of this paper that is focused in a commercial purpose, where the movements of people along the scenario are relevant in some marketing study tasks. By the analysis of trajectories in a particular scenario, it is possible detect areas of interest and redistribute them to redirect people to different paths.

Normally, trajectories are not studied using pattern analysis due the varying in length of data (same trajectory pattern can be done slower or making small variations of the path). Therefore, a normalization of data has to be done. Hu et al. [4] propose a normalization by using a maximal length component vectors, filling the empty data of shorter paths with no movement. In [5–7] they use PCA to sample the trajectories. Meanwhile, Xi et al. [8] proposed a *Trajectory Directional Histogram* (TDH) to describe the statistic directional distribution of one trajectory. In [9], on the other side, they use a *Discrete Fourier Transform* to reduce the components of the trajectories.

In this paper we propose a novel representation of trajectories that makes use of transformed tracked points on the ground plane where people are moving. The ground plane is spatially sampled in cells containing the proposed Activity Description Vector (ADV). The ADV describes the number of occurrences of a person in a specific location and the movements performed in four directions (up, down, left and right). The ADV has implicit information about the velocity of displacement and the spatial trajectory.

Once the ADV is calculated for the whole scenario, it is used as the input for a clustering method. In this paper, we have used classic technique to show the ability of the representation to recognize behaviours. Specifically, we have used five different methods: *Self-Organizing Map* (SOM), *k-Nearest Neighbour* (kNN), *Neural Gas* (NG), *Supervised SOM* (SSOM) and the *Linear Discriminant Analysis* (LDA).

Self-Organizing Maps have been previously used for trajectory classification, such as in [10] using a flow vector to sample the track information to train a SOM. Martinez-Contreras et al. [11] use SOMs only for motion (trajectory) sampling. A SOM is trained with different motions, and then a new motion is classified and the template is used in a Hidden Markov Model to determinate the action. Schreck et al. [12] developed a framework to classify trajectories using SOMs, scaling the paths into unit square values and sampling them in a predefined number of parts. Hu et al.[4]

This work was supported in part by the University of Alicante under Grant GRE11-01

J. Azorin-Lopez, M. Saval-Calvo, A. Fuster-Guilló and J. Garcia-Rodriguez are with the University of Alicante, E-03080, Alicante, SPAIN (e-mail: {jazorin, msaval, fuster, jgarcia}@dtic.ua.es).

introduce the new fuzzy SOM to classify paths and a previously mentioned sampling. In the case of Nearest Neighbour algorithms, k-Means has been used in trajectory analysis such as in Suzuki et al. [13], where they use HMM to model time-series features of human position and use k-Means to cluster to acquire human motion patterns. Airspace trajectories are analysed in [14] where k-means is used for characteristic turning point clustering of the paths. Blunsden et al. Finally, Blunsden et al. proposed a features vector and a nearest neighbour-based classifier for human interaction analysis based on trajectories [15].

The rest of the paper is organized as follows. Section II presents the ADV representation model proposed in this research that contains the proper information that allows classic classifiers to recognize human behaviour. Experiments are discussed and compared to other approaches in Section III. Finally, conclusions about the research are presented in Section IV.

II. Representation Model of Trajectories to Recognize Behaviour

In the literature, different approaches have been developed to sample trajectories into same-size useful values to extract later the behaviour. In this paper, we propose a new representation of trajectories based on sampling the scenario taking into account simple extracted features from the trajectory that correspond to a specific sample of the scenario instead of sampling only the trajectory values.

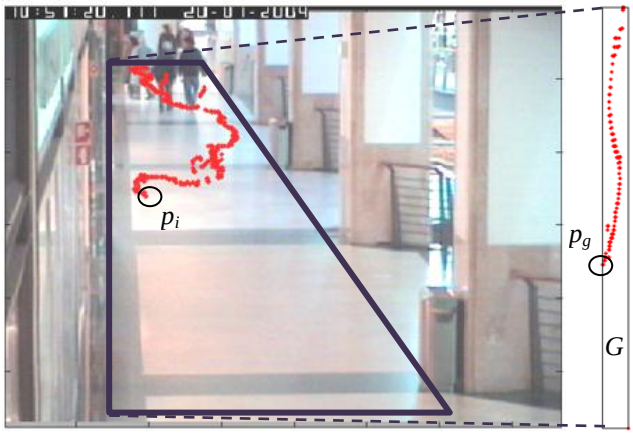

Fig. 1. Projective transformation to obtain the basic geometric model able to represent the trajectory of a person in the scenario.

The novel representation method takes the ground where people are moving as the basic geometric model to describe the trajectory of the individuals. We consider that data values of the scenario have to be without perspective. Therefore, the space of values has to be perpendicular to the point of view of the camera. If the camera is not on the roof, any information contained on the image plane captured from a static camera has to be transformed to the corresponding plane that fits the ground by means of a Homography, H (1). The projective transformation allows us to consider the whole space of movements of the people in the Euclidean space (see Figure 1). Then, any point p_i on the image is transformed to a point p_g on the ground plane G.

$$p_g = H \cdot p_i \quad (1)$$

Since we are only interested in the spatial trajectory information, to obtain a simple representation to analyse the behaviour, the information needed to track the objects in the scene are the positions of an individual in the scene. They set a list of tracked points *LTP* on *G*.

$$LTP = \{p_1, p_2, p_3, \ldots, p_n\} \quad (2)$$

Typically, surveillance cameras have a frame rate of about 25 frames per second, and due to segmentation and tracking errors, the blobs that surround the object analysed could vary in their shape. This can make little noisy motions that have to be avoided. Then we propose a sampling of the *LTP* by taking only values of each *t* frames and modelling the trajectory with a spline curve, recovering a smoothed trajectory of LTP.

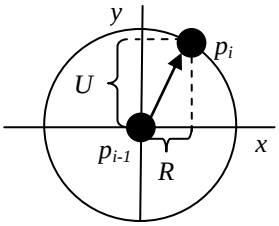

Fig. 2. Representation over axis *x* an *y* of movements Up (U) and Right (R) in a particular displacement between the point p_{i-1} and p_i.

From the smoothed tracked positions, we are able to calculate the movements of a person. Instead of calculate the global positions from an origin; we consider the displacements occurred with a particular trajectory in each axis considering a local origin for each tracked point in the trajectory of the person. Therefore, one particular movement from one tracked point to another will be calculated per each axis considering the displacement and the direction. In order to calculate it, we consider four directions for each point on *G*: Up, U, (3), Down, D, (4), Left, L, (5) and Right, R (6). The displacement is calculated as the dot product of the displacement vector between two consecutive tracked points on *LTP*, p_i and p_{i-1}, and the corresponding normal vector for each axis (see Figure 2). Therefore, for a displacement of a person, movements will be:

$$U(p_i) = \begin{cases} (p_i - p_{i-1}) \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix} & \text{if } \dfrac{(p_i - p_{i-1}) \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix}}{\|(p_i - p_{i-1})\|} > 0 \\ 0 & \textit{otherwise} \end{cases} \quad (3)$$

$$D(p_i) = \begin{cases} (p_i - p_{i-1}) \cdot \begin{bmatrix} 0 \\ -1 \end{bmatrix} & \text{if } \dfrac{(p_i - p_{i-1}) \cdot \begin{bmatrix} 0 \\ -1 \end{bmatrix}}{\|(p_i - p_{i-1})\|} > 0 \\ 0 & \textit{otherwise} \end{cases} \quad (4)$$

$$L(p_i) = \begin{cases} (p_i - p_{i-1}) \cdot \begin{bmatrix} -1 \\ 0 \end{bmatrix} & \text{if } \dfrac{(p_i - p_{i-1}) \cdot \begin{bmatrix} -11 \\ 0 \end{bmatrix}}{\|(p_i - p_{i-1})\|} > 0 \\ 0 & \textit{otherwise} \end{cases} \quad (5)$$

$$R(p_i)=\begin{cases}(p_i-p_{i-1})\cdot\begin{bmatrix}1\\0\end{bmatrix} & \text{if } \dfrac{(p_i-p_{i-1})\cdot\begin{bmatrix}1\\0\end{bmatrix}}{\|(p_i-p_{i-1})\|}>0\\ 0 & \textit{otherwise}\end{cases} \quad (6)$$

These four particular movements have information about the direction of the trajectory and the velocity of a person in a specific point on G.

Additionally, we consider the frequency, F, as the number of occurrences of a person that is in a specific point of G. That is, the number of frames that a person has been in a specific location. F contains information about the spatial trajectory of a person but not considering the movements itself.

Finally, the ground plane G is spatially sampled in a matrix C of *m* x *n* cells, so that the transformed points p_g and the functions of frequency and movements of it are in one of the cells of the matrix C. Each cell will describe the activity happened in that region of the scene considering the vector of relevant values, called *Activity Description Vector* (ADV). This vector will be composed by the frequency and the U, D, L and R movements of all points of the ground plane inside a cell:

$$\text{ADV}=\langle \text{F}, \text{U}, \text{D}, \text{L}, \text{R}\rangle. \quad (7)$$

Therefore, within a particular cell, the accumulative histograms of the movements U, D, L, R and frequency F for the points on G of the cell $C_{i,j}$ of C are calculated. Let u $_x$ v the actual size of the scenario, m $_x$ n the cells it has been split and $p_{k,l}$ the point located in the position *k* and *l* of the G space, each ADV in a cell is:

$$\forall c_{i,j}\in C \wedge \forall p_{k,l}\in G \,/\, i=\left\lfloor \frac{k\times m}{u}\right\rfloor \wedge j=\left\lfloor \frac{k\times n}{v}\right\rfloor$$
$$ADV_{i,j}=\begin{pmatrix}\sum F(p_{k,l}),U(p_{k,l}),\sum D(p_{k,l}),\\ \sum L(p_{k,l}),\sum R(p_{k,l})\end{pmatrix} \quad (3)$$

Hence, for a scenario space of $u_x v$, split in $m_x n$ cells, each data in the ADV will have 5 $_x$ *m* $_x$ *n* values divided in five meaningful parts with size *m* $_x$ *n*. Figure 3 shows an example of a trajectory and the ADV representation.

III. Experiments

A. Experimental setup

Experiments have been carried out using the CAVIAR database [16]. Specifically, validation of the representation of trajectories to recognize human behaviours makes use of the 26 clips from the Shopping Centre in Portugal recorded from frontal view of the scenario. This set of sequences contains 1500 frames on average of 384x288 pixels, capturing 235 individuals at 25 frames per second. Each sequence was labelled frame-by-frame by hand and each individual is tracked using a unique identifier in the sequence. Therefore, each frame has a set of tracked individuals visible in that frame that are surrounded by a bounding box and labelled according to the situation in which the individual is involved.

Each tracked individual have a set of labels that describes the context, the situations, the movement and the role. The context (*shop enter*, *windowshop*, *shop exit*, *shop reenter*, *browsing*, *immobile and walking*) is unique for each tracked person and involve the person in a sequence of situations (*browsing*, *inactive*, *moving*, *shop enter*, *and shop exit*). The individual also have been labelled according how much he or she is moving (*inactive*, *active* and *walking*) and the role that takes in the sequence (*browser* and *walker*).

The objective of experiments is to validate the ability of the simple representation model to recognize the behaviour, which is the use of simple representation to recognize complex situations. In consequence, we only take into

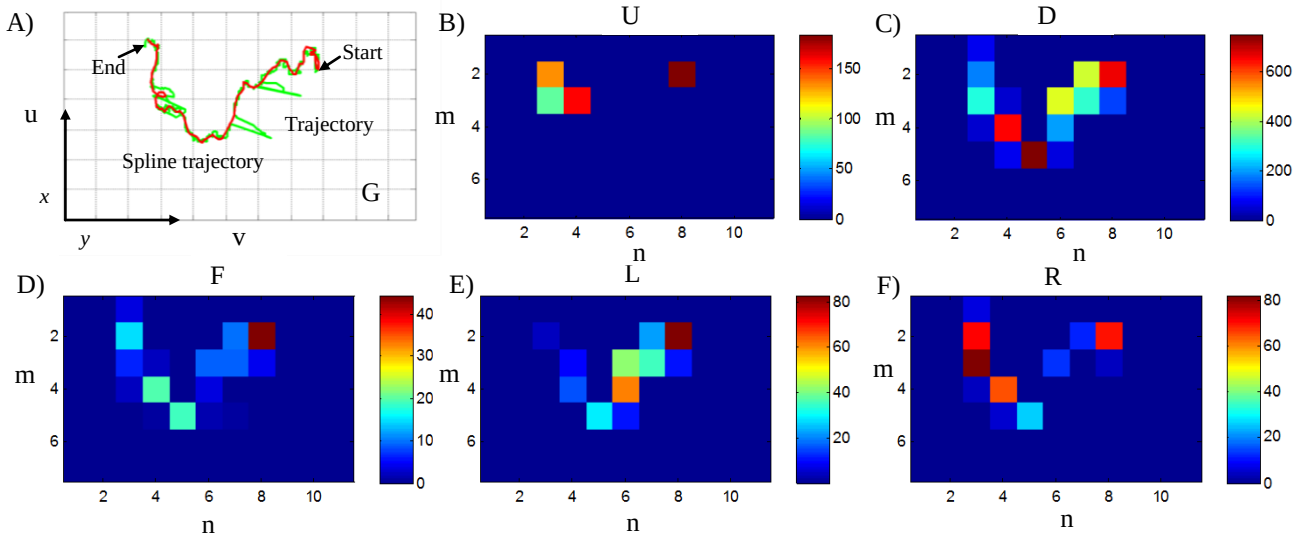

Fig. 3. Example of a trajectory of a person mainly going down. A) Trajectory and smoothed spline trajectory on G from start point to end point. The spatial sampling of the $u_x v$ space G into 7x11 cells ($m_x n$) of *C* is also represented. B) C) E) and F) represent the accumulative displacement in each cell of C for movements Up, Down, Left and Right respectively. Finally D) represents the accumulative histograms of frequency that conforms the ADV representation using the sampling space.

account the context label of the CAVIAR sequences as the high-level interpretation of the behaviour of a person in the scene. This information is subjective and depends on the observer. Additionally, we use the bounding box positions as the low-level data to describe the trajectory of a person. In this case, the information is objective but there is some variation in it due to the labelling was done by humans.

The 235 persons in the 26 clips labelled from the Shopping Centre perform 255 different trajectories (some persons have different contexts for the sequence) that are classified into the 7 contexts. As we can see in the Table I, the samples are imbalanced. Thus, the Synthetic Minority Over-Sampling Technique (SMOTE) [17] has been applied to obtain the same number of samples for each context. Also, for the *Walking* context, samples are undersampled randomly getting, finally, 70 samples per context.

TABLE I
SAMPLES USED IN EXPERIMENTS

Context	Samples	Average frames
Shop Enter	55	344
WindowShoping	18	1119
Shop Exit	63	405
Shop Reenter	5	151
Browsing	10	750
Inmobile	22	573
Walking	82	575

As we mentioned before, the bounding box positions used as the tracking points for individuals have some variations in pixels positions (and consequently to the transformed positions on the plane). In order to avoid the variations, a data sampling have been carried at a sampling frequency of 1 Hz (i.e. we take into account the position data each 25 frames). Finally, a SPLINE curve is calculated from the sampled data to obtain the trajectories included in each context.

B. Results and discussion

In order to validate the ability of the simple representation of trajectories to analyse the person behaviour, we have selected classic classifiers: Self-Organizing Map (SOM) [18], Supervised Self-Organizing Map (SSOM) [19], Neural GAS (NGAS) [20], Linear Discriminant Analysis (LDA) and k-Nearest Neighbour (kNN) [21]. Moreover, a multiclassifier (MC) designed from the above classifiers has been designed. The MC calculates from an input the most frequent class classified by the mentioned classic techniques.

Experiments have been performed for different grid sizes: 1x1, 3x5, 5x7 and 7x11 in order to evaluate the ability of the representation to synthesize the information extracted from the scene (see Figure 5 for an example of 3 behaviours with an ADV of 3x5 grid) f. Additionally, inputs for each classifier have been normalized to the range (0 1) dividing each component of the ADV vector by the maximum value for each component.

For each grid selection, a 10-fold cross validation has been performed obtaining Sensitivity and Specificity values, and the ROC curves is presented to analyse the performance of classifiers with ADV of different scenario sampling.

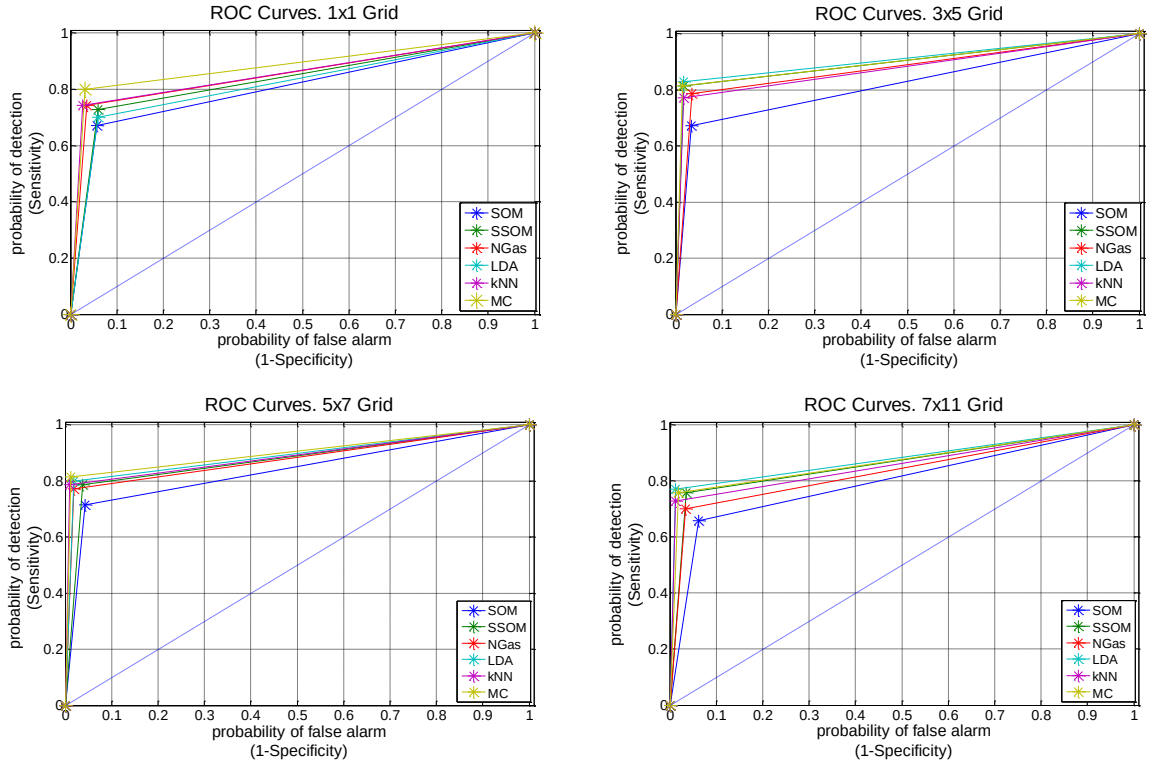

Fig. 4. ROC curves for 1x1 (a), 3x5 (b), 5x7 (c) and 7x11 (d) grid sizes.

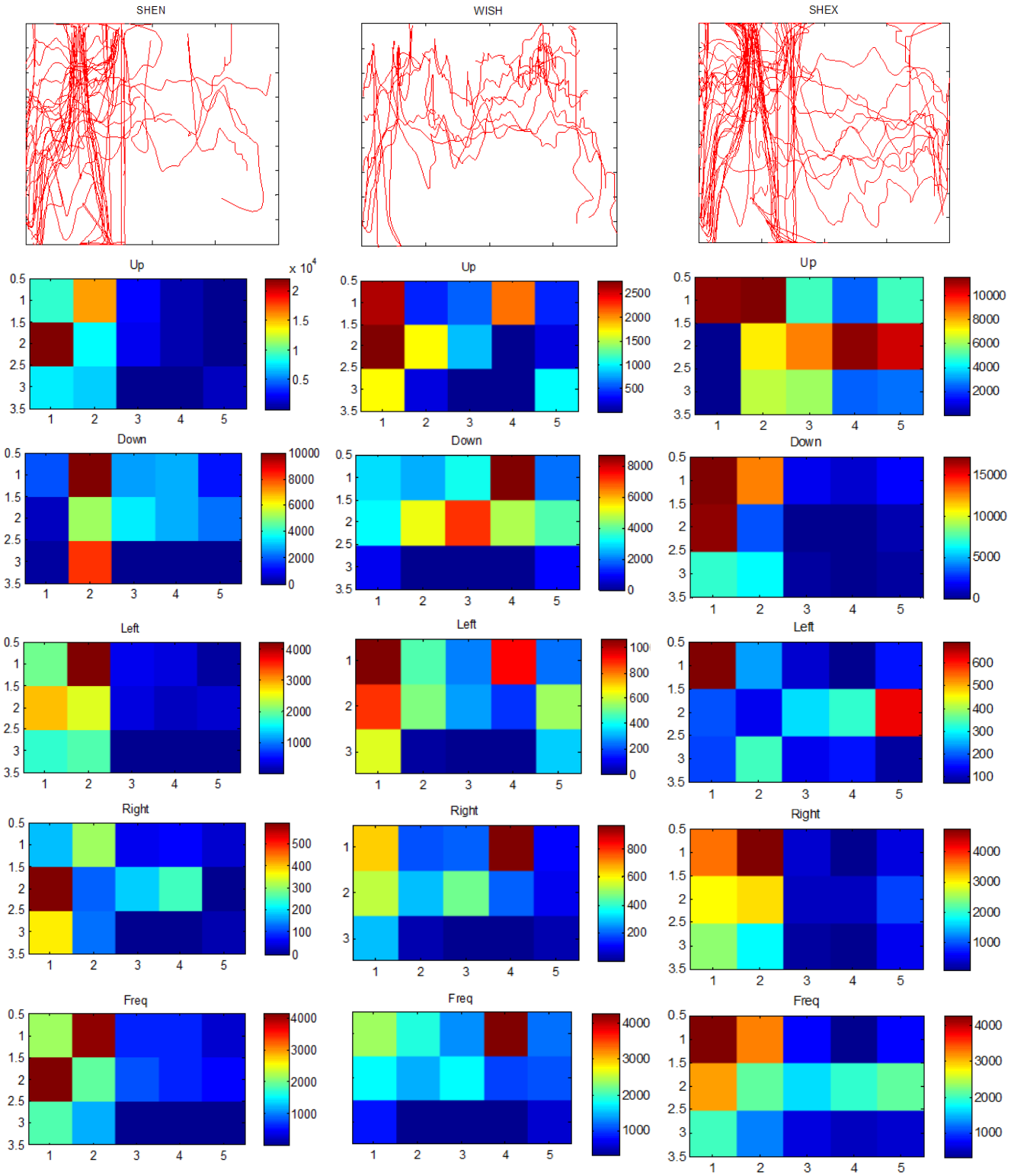

Fig. 5. Smoothed trajectories for all samples of *Shop enter* (SHEN), *Windowshopping* (WISH) and *Shop exit* (SHEX) behaviour (first row). The rest of rows show the accumulative Up, Down, Left and Right movements and Frequency that set the ADV representation.

Table II shows the results of classification accuracy for each classifier with different grid size. Columns present the values of *Sensitivity* (correctly classified positive samples / true positive samples), *Specificity* (correctly classified negative samples / true negative samples), *Correct Rate* (correctly classified samples / classified samples), and *Error Rate* (incorrectly classified samples / classified samples). Bolded values represent the best for each classifier. Best

results are achieved in 3x5 and 5x7 sampling for Sensitivity. For the rest of values, 3x5 and 5x7 continue been the best results except for LDA classifier, where 7x11 is the best size for sampling due the PCA prior application. These results show that, for human behaviour analysis, 7x11 could produce oversampling, producing error classification. On the other hand, 1x1 (no sampling) does not achieve best results in any case, proving that the proposed sampling method enhance the classification process. Hence, our representation is able to recognize the behaviour of the persons in the Shopping Centre with great accuracy.

TABLE II
CLASSIFICATION PERFORMANCE FOR DIFFERENT CLASSIFICATION METHODS

Classification method	Grid	Sens.	Spec.	Correct Rate	False Rate
SOM	1x1	0.6714	0.9429	0.7490	0.2510
	3x5	0.6714	**0.9667**	**0.7980**	**0.2020**
	5x7	**0.7143**	0.9571	0.7878	0.2122
	7x11	0.6571	0.9381	0.7755	0.2245
SSOM	1x1	0.7286	0.9405	0.6612	0.3388
	3x5	**0.8143**	**0.9833**	0.7592	0.2408
	5x7	0.7857	0.9619	**0.7735**	**0.2265**
	7x11	0.7571	0.9643	0.7429	0.2571
NGAS	1x1	0.7429	0.9643	0.7653	0.2347
	3x5	**0.7857**	0.9643	0.8367	0.1633
	5x7	0.7714	**0.9810**	**0.8469**	**0.1531**
	7x11	0.7000	0.9667	0.8122	0.1878
LDA	1x1	0.7000	0.9405	0.6286	0.3714
	3x5	**0.8286**	0.9833	0.8633	0.1367
	5x7	0.8000	0.9810	0.8612	0.1388
	7x11*	0.7714	**0.9881**	**0.8857**	**0.1143**
kNN	1x1	0.7429	0.9738	0.8224	0.1776
	3x5	0.7714	0.9833	**0.8592**	**0.1408**
	5x7	**0.7857**	**0.9905**	0.8510	0.1490
	7x11	0.7286	0.9881	0.8245	0.1755
MC	1x1	0.8000	0.9690	0.7980	0.2020
	3x5	**0.8143**	0.9857	**0.8776**	**0.1224**
	5x7	**0.8143**	**0.9881**	0.8694	0.1306
	7x11	0.7571	0.9810	0.8735	0.1265

* Due to the size of the ADV, a prior PCA of 200 components has been calculated.
Sens: Sensitivity, Spec.: Specificity

We have used the ROC (receiver operating characteristic) to illustrate the performance of classifiers in Figure 4. ROC curve represents the relation between the Sensitivity and 1-Specificity.

From ROC curves it is seen that for 1x1 sampling MC is the best option with a high different respect the rest. In 3x5 and 7x11 LDA classifies better than the others, and in 5x7 MC is the best, but in both cases they achieve similar results. In the four cases, SOM is the worst classifier. The dotted line represents the middle value where a classifier will have the same percentage of success and failure. All classifier curves are over this value, which means that success rate is always higher than failure rate.

Next, confusion matrixes are presented in Table III for the classifiers with best ROC curve to study in depth the classification. Matrix columns represent the true classes, and rows represent the classifier prediction. The ideal classifiers will have only non-zero numbers in the main diagonal.

Classification has been done with 70 samples of each class. Table III shows a high accuracy in classifying for each pattern, being the SHRE (shop reenter) the best classified because it is the most different trajectory among the whole possible tested paths. On the contrary, WALK (walking) is the most failure sample classified. This is because all trajectories, except immobile, have walking component, then the classifiers cannot distinguish between the generic walk and a specific walk for another action.

TABLE III
CONFUSION MATRIXES

SHEN	WISH	SHEX	SHRE	BROW	IMMO	WALK
Sample size 1x1, MC classifier						
56	2	1	0	1	0	9
2	59	1	0	3	7	0
0	0	61	0	2	0	13
0	0	0	70	1	0	1
0	3	2	0	57	3	14
3	6	0	0	4	60	5
9	0	5	0	2	0	28
Sample size 3x5, LDA classifier						
58	1	1	0	0	0	5
3	69	0	0	0	0	4
0	0	62	0	0	0	3
0	0	4	70	0	0	3
1	0	0	0	60	0	7
3	0	0	0	0	59	3
5	0	3	0	10	11	45
Sample size 5x7, MC classifier						
57	0	0	0	0	1	4
7	70	0	0	0	0	13
0	0	60	0	0	0	3
1	0	1	70	0	0	1
0	0	1	0	69	1	4
3	0	2	0	1	67	12
2	0	6	0	0	1	33
Sample size 7x11, LDA Classifier						
57	0	0	0	0	1	4
7	70	0	0	0	0	13
0	0	60	0	0	0	3
1	0	1	70	0	0	1
0	0	1	0	69	1	4
3	0	2	0	1	67	12
2	0	6	0	0	1	33

' where SHEN = shop enter; WISH = windows shop; SHEX = Shop exit; SHRE = Shop reenter; BROW = Browsing; IMMO = immobile; WALK = walking

In order to show the accuracy of the proposed representation to include behaviour information, the MC classifier using the ADV has been compared to other contemporary methods. Sensitivity and specificity results of context classification have been calculated from reported success rates in [16] and [22] of comparable experiments on the same dataset. These methods are grouped as state and semantic models using predefined models and rules to evaluate behaviours.

In [16], two approaches were presented. The first, a rule-based approach, used semantic rules on both the role and movement classifications to evaluate the context from video sequences. The second, used an extension of the HMM. Specifically, to interpret the context, hidden semi-Markov model (HSMM) [23]. HSMMs extend the standard Hidden

Markov model with an explicit duration model for each state [24]. Finally, in [22] Lavee et al. proposed the use of Petri Nets (PN) for recognition of event occurrences in video. The Petri Net was used to express semantic knowledge about the event domain as well as for recognizing events as they occur in a particular video sequence.

Table IV shows results for the above three methods (Rule-based, HSMM, PN) and the proposed multiclassifier (MC) for the ADV representation using a 5x7 grid.

TABLE IV
CLASSIFICATION PERFORMANCE COMPARISON

	Rule-based	HSMM	PN	**MC (5x7)**
Sensitivity	0.57	0.6508	0.8085	**0.8143**
Specificity	N/A	0.9866	0.9680	**0.9881**

As is shown in the table, the ADV approach achieves a significant improvement over both the Rule-based and the HSMM results for sensitivity and specificity. Although, the improvement over the PN model is smaller, around 0.6% and 2% for sensitivity and specificity respectively (better improvement can be achieved comparing PN to LDA results about 2% for sensitivity and 1.5% for specificity, see Table II), the ADV representation outperform the results without having semantic knowledge about behaviour.

IV. CONCLUSIONS

In this paper a human behaviour recognition method based on trajectory analysis using neural networks is proposed. The method is based on sampling trajectory data as a cue for different classifiers. The proposed model uses the "Activity Description Vector" (ADV) to describe the activity happened in each region of the scene (cells). Different clustering models have been used to test the ADV sampling method (SOM, Supervised SOM, NGAS, LDA, kNN, MC as a combination of the others). Experiments have been carried out using the CAVIAR database. The experimental results show the capacity of the ADV representation to organize the context situations of persons with clearly separated clusters. Moreover, predefined models and rules to evaluate behaviours are not needed in this method, as occurs in state and semantic models (Bayesian, HMM, Petri Nets, Grammars,…) [1]. The proposed scheme is able to recognize behaviour by only using global information and data from tracking to generate the ADV. Experimental results shows how classic classifiers are able to cluster the input vectors allowing the system to correctly recognize human behaviour in complex situations with great accuracy. We are currently exploring the feasibility of ADV sampling method in other contexts of human behaviour to analyse the generality of the representation.

V. REFERENCES

[1] G. Lavee, E. Rivlin, and M. Rudzsky, "Understanding Video Events : A Survey of Methods for Automatic Interpretation of Semantic Occurrences in Video," vol. 39, no. 5, pp. 489–504, 2009.

[2] B. T. Morris and M. M. Trivedi, "A Survey of Vision-Based Trajectory Learning and Analysis for Surveillance," vol. 18, no. 8, pp. 1114–1127, 2008.

[3] R. B. Fisher, "The PETS04 Surveillance Ground-Truth Data Sets," in *Sixth IEEE Int. Work. on Performance Evaluation of Tracking and Surveillance (PETS04)*, 2004, pp. 1 – 5.

[4] W. Hu, D. Xie, T. Tan, and S. Maybank, "Learning Activity Patterns Using Fuzzy Self-Organizing Neural Network," vol. 34, no. 3, pp. 1618–1626, 2004.

[5] N. Anjum and A. Cavallaro, "Single camera calibration for trajectory-based behavior analysis 2 Ground-plane calibration," pp. 147–152, 2007.

[6] N. Anjum and A. Cavallaro, "Multifeature Object Trajectory Clustering for Video Analysis," vol. 18, no. 11, pp. 1555–1564, 2008.

[7] N. Anjum and A. Cavallaro, "Trajectory Clustering for Scene Context Learning and Outlier Detection," in *Video Search and Mining*, vol. 287, D. Schonfeld, C. Shan, D. Tao, and L. Wang, Eds. Springer Berlin Heidelberg, 2010, pp. 33–51.

[8] X. Li, W. Hu, and W. Hu, "A Coarse-to-Fine Strategy for Vehicle Motion Trajectory Clustering," pp. 18–21, 2006.

[9] A. Naftel and S. Khalid, "Classifying spatiotemporal object trajectories using unsupervised learning in the coefficient feature space," *Multimedia Systems*, vol. 12, no. 3, pp. 227–238, Sep. 2006.

[10] J. Owens and A. Hunter, "Application of the Self-Organising Map to Trajectory Classification," pp. 1–7, 2000.

[11] F. Martinez-Contreras, C. Orrite-Urunuela, E. Herrero-Jaraba, H. Ragheb, and S. a. Velastin, "Recognizing Human Actions Using Silhouette-based HMM," *2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*, pp. 43–48, Sep. 2009.

[12] T. Schreck, J. Bernard, T. von Landesberger, and J. Kohlhammer, "Visual cluster analysis of trajectory data with interactive Kohonen maps," *Information Visualization*, vol. 8, no. 1, pp. 14–29, Feb. 2009.

[13] N. Suzuki, K. Hirasawa, K. Tanaka, Y. Kobayashi, Y. Sato, and Y. Fujino, "Learning motion patterns and anomaly detection by Human trajectory analysis," *2007 IEEE International Conference on Systems, Man and Cybernetics*, pp. 498–503, 2007.

[14] M. Gariel, A. N. Srivastava, S. Member, and E. Feron, "Trajectory Clustering and an Application to Airspace Monitoring," vol. 12, no. 4, pp. 1511–1524, 2011.

[15] S. Blunsden, E. Andrade, and R. Fisher, "Non Parametric Classification of Human Interaction," pp. 347–354, 2007.

[16] R. Fisher, J. Santos-Victor, and J. Crowley, "CAVIAR: Context Aware Vision Using Image-Based Active Recognition Project." [Online]. Available: http://homepages.inf.ed.ac.uk/rbf/CAVIAR/. [Accessed: 01-May-2013].

[17] N. V Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE : Synthetic Minority Over-sampling Technique," vol. 16, pp. 321–357, 2002.

[18] T. Kohonen, "The Self-organizing Map," vol. 78, no. 9, pp. 1464–1480, 1990.

[19] S. Papadimitriou, S. Mavroudi, L. Vladutu, G. Pavlides, and A. Bezerianos, "The Supervised Network Self-Organizing Map for Classification," pp. 185–203, 2002.

[20] T. M. Martinetz, S. G. Berkovich, and K. J. Schulten, ";Neural-gas' network for vector quantization and its application to time-series prediction.," *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council*, vol. 4, no. 4, pp. 558–69, Jan. 1993.

[21] T. M. Cover and P. E. Hart, "Nearest Neighbor," IEEE Transactions on Information Theory, vol. 13, no. 1, pp. 21 – 27, 1967.

[22] G. Lavee, M. Rudzsky, E. Rivlin, and A. Borzin, "Video Event Modeling and Recognition in Generalized Stochastic Petri Nets," IEEE Transactions on Circuits and Systems for Video Technology, vol. 20, no. 1, pp. 102–118, 2010.

[23] D. Tweed and R. Fisher, "Efficient Hidden Semi-Markov Model Inference for Structured Video Sequences," in Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 2005. 2nd Joint IEEE International Workshop on, 2005, pp. 247 – 254.

[24] R. Fisher, J. Santos-Victor, and J. Crowley, "CAVIAR Hidden Semi-Markov Model Behaviour Recognition." [Online]. Available: http://homepages.inf.ed.ac.uk/rbf/CAVIAR/hsmm.htm. [Accessed: 01-May-2013].