A CLASS OF SPACE-VARYING PARAMETRIC MOTION FIELDS FOR HUMAN ACTIVITY RECOGNITION

Jacinto C. Nascimento[(a)] *Jorge S. Marques*[(a)] *João M. Lemos*[(b)]

Instituto de Sistemas e Robótica[(a)] Instituto de Engenharia de Sistemas e Computadores[(b)]
Instituto Superior Técnico, 1049-001 Lisboa, Portugal

ABSTRACT

Video cameras monitoring human activities in public spaces are commonplace in cities worldwide. Such monitoring task is important for safety and security purposes but is also extremely challenging. In this paper, we propose a class of algorithms for far-field human activity recognition, a central task in video surveillance. More specifically, we explore a class of parametric motion vector fields learned from the trajectories described by people in real-world scenarios. The work proposed herein is a space dependent framework, in sense that the vector fields depend on the pedestrian position. Thus, the model is flexible leading to an expressive description of complex trajectories. Also, a model selection strategy is addressed to automatically choose the appropriate number of underlying motion fields presented in the trajectories. Experimental evaluation is conducted in real settings testifying the usefulness of the proposed approach for human activity recognition.

1. INTRODUCTION AND RELATED WORK

Video cameras monitoring human activities in public spaces are commonplace in cities worldwide. Such monitoring task is important for safety and security purposes [1], [7] but is also difficult to be fulfilled. Human operators are generally employed for this task, but even the most vigilant humans miss important information that could ultimately contribute to unfavorable consequences.

Major research efforts are under way to develop security systems able to report events of interest or even statistics reporting most common activities in surveillance scenarios. Essential are methods by which information can be extracted from video data in order to recognize behaviors or/and detect abnormal events. Such information (*e.g.* features extracted from the image) depend on the surveillance settings, *i.e.* short range, where the camera is close to the observed people, thus detailed information of human gestures, pose, gait can be extracted; or far field where the camera covers a wide area, thus the system is no longer able to acquire information with great detail.

This paper explores cutting-edge technique based on the assumption the majority of people activities observed are described by the performed trajectories. This is particularly relevant if we consider far-field scenario, which is the presented case. Particularly, this work extends our previous work [4]. In [4] it is addressed the problem of estimating (i), the velocity fields $\mathcal{T}$ (ii) noise variances Σ and (iii) switching fields $\mathbf{B}$, thus having the set of model parameters $\boldsymbol{\theta} = (\mathcal{T}, \mathbf{B}, \Sigma)$ which explain well all the trajectories that take place in the scenario. To achieve this, a grid of $N \times N$ is defined over the image domain. The main hindrance is that the model parameters $(\mathcal{T}, \mathbf{B})$ have to be estimated at every node in the grid. With the approach proposed in this paper we can strongly alleviate the computational effort. The vector fields $\mathcal{T}$ do not need to be estimated grid-wise since they belong to a class of parametric models allowing a simpler trajectory model estimation and a drastic reduction of the number of parameters.

Most related work concerning surveillance falls in the category of trajectory-modeling. These methods are essentially deterministic based on trajectory comparison. For instance, trajectory analysis problems have been addressed using similarity measures to measure the mismatch between trajectories; Euclidean [9] and Hausdorff [2] are typical examples. Dynamic time warping [3] has also been proposed.

Another class of approaches assumes that trajectories are produced by probabilistic generative models *e.g.* [6], [8], usually based on the hidden Markov models. These approaches have the important advantage of not requiring alignment/registration of the trajectories being compared; moreover, they provide a solid probabilistic inference framework, based on which model parameters may be obtained from observed data. Our work falls in the latter class of approach. A space dependent HMM is proposed where switching probabilities depend not only on the current active field but also on the pedestrian position. This allows a much better description os non-stationary trajectories observed in far-field surveillance settings.

The paper is organized as follows. Section 2 describes the generative model. Section 3 presents the proposed parametric models, that are the original contribution of the paper. Section 4 explains how the parametric models are learned using the EM algorithm. Section 5 presents experimental validation with real data. Section 6 concludes the paper.

2. TRAJECTORY GENERATION

Trajectory generation follows the same generative model as in [4], that is

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{T}_{k_t}(\mathbf{x}_{t-1}) + \mathbf{w}_t, \tag{1}$$

where $\mathbf{x}_t$ is the target position at instant t, $k_t \in \{1, ..., K\}$ is the label of the active field at time t, $\mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, \sigma^2_{k_t}\mathbf{I})$ is white Gaussian noise with zero mean and variance $\sigma^2_{k_t}$ which may be different for each field. At each time instant, one of these velocity fields is *active*, that is, is driving the motion.

Throughout the paper we will adopt the following notation: $\mathbf{T}_k(\mathbf{x})$ denotes the velocity vector at point $\mathbf{x} \in \mathbb{R}^2$ of the k-th field, $\mathbf{T}_k : \mathbb{R}^2 \to \mathbb{R}^2$, for $k \in \{1, \ldots, K\}$; $\mathcal{T} = \{\mathbf{T}_1, \ldots, \mathbf{T}_K\}$, denotes the set of K vector fields and $\boldsymbol{\sigma} = \{\sigma^2_1, ..., \sigma^2_K\}$ the set of noise variances.

According to the dynamic model (1), we can write the joint distribution of a trajectory $\mathbf{x} = (\mathbf{x}_1, .., \mathbf{x}_L)$ and the underlying

This work was supported by the FCT project [PEst-OE/EEI/LA0009/2011], by the ARGUS project PTDC/EEA-CRO/098550/2008, and by the European Project FP6-2005-IST-6-045062-URUS.

Table 1. Parametric motion models ($\mathbf{R}$ denotes a rotation matrix).

Model	*Transformation matrix* $\mathbf{Ax}$
Translation	$\mathbf{I}$ ($\mathbf{A} \rightarrow$ *identity matrix*)
Rigid Body	$\mathbf{R}$ ($\mathbf{A} \rightarrow$ *rotation matrix*)
Euclidean similarity	s$\mathbf{R}$ ($\mathbf{A} \rightarrow$ *Euclidean similarity*)
Affine	$\mathbf{A}$ ($\mathbf{A} \rightarrow$ *arbitrary*)

sequence of active fields $\mathbf{k} = (k_1, ..., k_L)$ as

$$\begin{aligned} p(\mathbf{x}, \mathbf{k}|\mathcal{T}, \mathbf{B}, \Sigma) &= p(\mathbf{x}_1)P(k_1)\prod_{t=2}^{L} p(\mathbf{x}_t, k_t|\mathbf{x}_{t-1}, k_{t-1}) \\ &= p(\mathbf{x}_1)P(k_1)\prod_{t=2}^{L} p(\mathbf{x}_t|\mathbf{x}_{t-1}, k_t)P(k_t|k_{t-1}, \mathbf{x}_{t-1}). \end{aligned} \tag{2}$$

where L is the length of the trajectory $\mathbf{x}$. The first term (in the 2^{nd} line of (2)) is modeled as a normal distribution, *i.e.* $\mathcal{N}(\mathbf{x}_t|\mathbf{x}_{t-1} + \mathbf{T}_{k_t}(\mathbf{x}_{t-1}), \sigma_{k_t}^2\mathbf{I})$; the second term is a first order space dependent HMM, that is, $P(k_t = j|k_{t-1} = i, \mathbf{x}_{t-1}) = \mathbf{B}_{ij}(\mathbf{x}_{t-1})$; and $p(\mathbf{x}_1)$ is the distribution of the initial position.

Here, $P(k_t|k_{t-1}, \mathbf{x}_{t-1})$ is a function of $\mathbf{B}$, $p(\mathbf{x}_t|\mathbf{x}_{t-1}, k_t)$ is a function of $\mathcal{T}$ and σ_{k_t}, and $p(\mathbf{x}_t, k_t|\mathbf{x}_{t-1}, k_{t-1})$ is a function of $\mathcal{T}$, $\mathbf{B}$, and σ_{k_t}.

3. A CLASS OF PARAMETRIC MOTION FIELDS

As defined in Section 1, the parameters to be estimated are denoted as $\boldsymbol{\theta} = (\mathcal{T}, \mathbf{B}, \Sigma)$, where $\mathcal{T}$ is the set of vector fields, $\mathbf{B}$ is the field of transition matrices, and Σ, is the set of noise variances (we assume a common diagonal covariance).

In [4] the motion fields $\mathcal{T}$ and the matrix $\mathbf{B}$ are estimated using a non-parametric model, *i.e.*, no structure is imposed to the model variables, instead they are defined at the nodes of a regular grid and interpolated in other image points using $1st$ order spline interpolation. The consequence is that each field depends on a large number (hundreds) of parameters. Here, we assume that the motion fields are described by parametric models. The number of parameters to be estimated, is thus much smaller (< 10).

Four parametric motion models are considered. All the models are expressed by

$$\mathbf{z} = \mathbf{Ax} + \mathbf{t}, \tag{3}$$

where $\mathbf{z}$ is the transformed position of $\mathbf{x}$, $\mathbf{A}$ is a 2×2 matrix and $\mathbf{t}$ is a 2×1 translation vector. The difference between these models are shown in Table 1.

The motion fields are given by

$$\mathbf{T}_k(\mathbf{x}) = \mathbf{A}_k\mathbf{x} + \mathbf{t}_k \quad k = 1, ..., K \tag{4}$$

4. LEARNING THE VECTOR FIELDS

We now address the estimation of the models parameters. More specifically, how to learn the set of velocity fields $\mathcal{T}$, the field of transition matrices $\mathbf{B}$, and the set of noise variances $\Sigma = \{\sigma_1^2, ..., \sigma_K^2\}$, from a set of observed trajectories. To accomplish this, we will assume to have a training set of S independent trajectories $\mathcal{X} = \{\mathbf{x}^{(1)}, ..., \mathbf{x}^{(S)}\}$, where $\mathbf{x}^{(j)} = (\mathbf{x}_1^{(j)}, ..., \mathbf{x}_{L_j}^{(j)})$ is the j-th observed trajectory, assumed to have length L_j. Naturally, we assume that the corresponding set of sequences of active fields, $\mathcal{K} = \{\mathbf{k}^{(1)}, ..., \mathbf{k}^{(S)}\}$, is not observed (it is hidden).

4.1. MMAP criterion

Since the active field labels $\mathcal{K}$ are missing, this fact suggests the use of an EM algorithm to find a *marginal maximum a posteriori* (MMAP) estimate of $\boldsymbol{\theta}$ under some prior $p(\boldsymbol{\theta}) = p(\mathcal{T})p(\mathbf{B})p(\Sigma)$; formally,

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \Big[\sum_{\mathcal{K}} \prod_{j=1}^{S} p(\mathbf{x}^{(j)}, \mathbf{k}^{(j)}|\boldsymbol{\theta})\Big] p(\boldsymbol{\theta}) \tag{5}$$

where $p(\boldsymbol{\theta})$ is the prior on the parameters. Recall that the sum over all possible sequences of labels leads to an exponential growth which can not be directly computed. Instead the EM method will be used.

The EM algorithm is based on the conditional expectation of the complete log-likelihood

$$\log p(\mathcal{X}, \mathcal{K}|\boldsymbol{\theta}) = \sum_{j=1}^{S} \log p(\mathbf{x}^{(j)}, \mathbf{k}^{(j)}|\,\boldsymbol{\theta}) \tag{6}$$

where $p(\mathbf{x}^{(j)}, \mathbf{k}^{(j)}|\,\boldsymbol{\theta})$ is given by (2). As is common in dealing with missing labels, we use binary indicator variables, defined as follows: each label $k_t^{(j)} \in \{1, ..., K\}$ (the active field at time t of trajectory j) is represented by a binary vector $\mathbf{y}_t^{(j)} = (y_{t,1}^{(j)}, ..., y_{t,K}^{(j)}) \in \{0, 1\}^K$, where $y_{t,l}^{(j)} = 1 \Leftrightarrow k_t^{(j)} = l$. With this notation, the complete log-likelihood becomes

$$\begin{aligned} \mathcal{L} &= \log p(\mathcal{X}, \mathcal{K}|\boldsymbol{\theta}) \\ &= C + \sum_{j=1}^{S}\sum_{t=2}^{L_j}\sum_{l=1}^{K} y_{t,l}^{(j)} \log \mathcal{N}(\mathbf{x}_t^{(j)}|\mathbf{x}_{t-1}^{(j)} + \mathbf{T}_l(\mathbf{x}_{t-1}^{(j)}), \sigma_l^2\mathbf{I}) \\ &+ \sum_{j=1}^{S}\sum_{t=2}^{L_j}\sum_{l=1}^{K}\sum_{g=1}^{K} y_{t-1,g}^{(j)}\, y_{t,l}^{(j)} \log B_{g,l}(\mathbf{x}_{t-1}^{(j)}), \end{aligned} \tag{7}$$

where $C = \sum_{j=1}^{S} \log p(\mathbf{x}_1^{(j)}) + \log P(k_1^{(s)})$ is a constant.

4.2. The EM algorithm

The E-step aims at computing the conditional expectation of the complete log-likelihood given in (7), given the current estimates of the parameters $\widehat{\boldsymbol{\theta}}$, *i.e.*, $Q(\boldsymbol{\theta}; \widehat{\boldsymbol{\theta}}) \equiv \mathbb{E}\Big[\mathcal{L}|\mathcal{X}, \widehat{\boldsymbol{\theta}}\Big]$. This leads to the computation of the conditional expectations with respect of missing binary indicators $y_{t,l}^{(j)}$ and product switching indicators $y_{t-1,g}^{(j)}\, y_{t,l}^{(j)}$. These probabilities are obtained by a modified forward-backward procedure, recall that the transition matrix is not constant, but depends on the trajectories.

In the M-step, the estimates are updated according to

$$\widehat{\boldsymbol{\theta}}_{\text{new}} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}; \widehat{\boldsymbol{\theta}}) + \log p(\boldsymbol{\theta}). \tag{8}$$

The maximization, as well as the adopted priors, are obtained by looking separately at the maximization with respect to each component of $\boldsymbol{\theta} = (\mathcal{T}, \mathbf{B}, \Sigma)$. This study is provided in our previous work, where we study this maximization in more detail, as well as the adopted priors.

The update of the $\mathbf{B}$ and Σ are the same as in [4]. What is different is the update of the vector fields $\mathcal{T}$ that depend on the

model being used. After straightforward manipulation, *i.e.*, deriving in order to $\mathbf{A}$ and zeroing the first term in (7), (see also (4)), the M-step provides the following updates:

1. *Translation*: $\mathbf{A} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix
2. *Rigid body*: $\mathbf{A}$ is a rotation matrix, *i.e.* a unitary matrix where $\mathbf{A}^\top = \mathbf{A}^{-1}$ with the constraint, $\det(\mathbf{A}) = 1$. The estimation is $\mathbf{A} = \mathbf{U}\mathbf{V}^\top$ obtained with through SVD, *i.e.* $\mathbf{R}_{01} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$
3. *Euclidean similarity*: $\mathbf{A} = \begin{bmatrix} \theta_1 & \theta_2 \\ -\theta_2 & \theta_1 \end{bmatrix}$ where θ is a 2×1 vector given by $\theta = (\mathrm{Tr}(\mathbf{R_{11}}))^{-1}\mathbf{R}_{01}^\top \mathbf{1}$, where $\mathbf{1}$ is the unity vector, and $\mathbf{R}_{ij} \triangleq \frac{1}{L-1}\sum_{t=2}^{L} \omega_t \widetilde{\mathbf{x}}_{t-i}\mathbf{x}_{t-j}^\top$
4. *Affine model*: $\mathbf{A} = \mathbf{R}_{01}(\mathbf{R}_{11})^{-1}$.

5. EXPERIMENTAL EVALUATION

This section considers the application of the proposed model in a real scenario shown in Fig. 1, with trajectories superimposed and taken at UPC campus (Barcelona)[1]. The images were obtained from a remote fixed network camera located at UPC campus. Thousands of images ($\approx 1.3 \times 10^5$) were acquired corresponding to 4 hours of recording at approximately frame rate 10 images/sec.[2]

Before applying the proposed model to estimate a set of motion fields, we need to extract the trajectories from the video sequences by tracking the pedestrians. For that purpose, we used the Lehigh omnidirectional tracking system to detect regions, followed by region association, as in [5]. The trajectories are then projected onto a view orthogonal to the ground plane to enforce viewpoint invariance. The total number of the trajectories considered is 270.

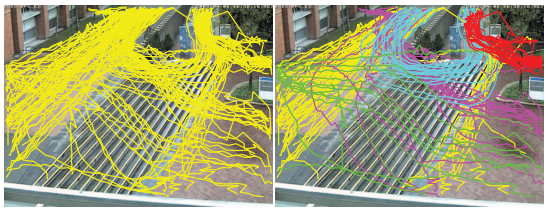

Fig. 1. (Left) Detected trajectories in UPC campus, (right) the same set of trajectories, but each trajectory has a different color label according to the trajectory-class (activity). In this scenario the following classes are considered (see right image): $a_1 \rightarrow$ *walking and stepping down the stairs*, (magenta); $a_2 \rightarrow$ *walking along*, (yellow); $a_3 \rightarrow$ *crossing and stepping down the stairs*, (green); $a_4 \rightarrow$ *pass diagonally*, (red); $a_5 \rightarrow$ *turning the campus* (cyan).

To classify human activities, the generative model in (1) can be cast into a maximum a posteriori classifier. Therefore, we consider a training set $\{\mathcal{X}^{(a)}, a = 1, .., A\}$, where A is the number of different activities and $\mathcal{X}^{(a)}$ is the set of trajectories belonging to the a-th activity. Thus, we estimate A generative models $\{\widehat{\boldsymbol{\theta}}^{(a)}, a = 1, .., A\}$ using he EM algorithm described in Section 4.2. Finally, given a new trajectory $\mathbf{x}$, the MAP classifier is given by

$$\widehat{a}(\mathbf{x}) = \arg \max_{a \in \{1,\ldots,A\}} p(\mathbf{x}|\widehat{\boldsymbol{\theta}}^{(a)})P(a) \qquad (9)$$

where $P(a)$ is the prior probability for activity a (we assume equiprobable classes), and $p(\mathbf{x}|\boldsymbol{\theta}^{(a)})$ is the probability density function of the trajectory $\mathbf{x}$ under the model with parameters $\widehat{\boldsymbol{\theta}}^{(a)}$.

[1]UPC images were acquired in the scope of the EU project FP6-EU-IST-045062 URUS - *Ubiquitous Networking Robotics in Urban Settings*.

[2]This low frame rate is due to the limitations of the cameras network and due to the fact that it is not intended to store much data.

Table 2. Parameters initialization for the models.

parameter	initialization
number of motion fields K	$\{1, 2, ..., 8\}$
diagonal noise variances Σ	$\{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}\}$
translation $\mathbf{t}$ of the vector field $\mathbf{T}$	random in $[-0.1, 0.11]$
transition matrix $\mathbf{B}$	$b_{ii} = 1 - \frac{K-1}{10}$, $b_{ij} = 1 - b_{ii}$

5.1. Discriminative parameters selection

We now describe the procedure to automatically determine the number of motion fields. As mentioned above, the framework is fully described by the triplet $\boldsymbol{\theta} = (\mathcal{T}, \mathbf{B}, \Sigma)$. Besides, we also have to specify the number of motion models K. To accomplish this, the underlying assumption is that, K is estimated assuming that the model is going to be used for a specific task: activity recognition. Particularly, pedestrian motion is characterized by trajectories which are modeled by vector fields. The idea, is thus, to select the generative model that achieves the best classification performance. This reasoning is also applied for the remaining parameters in $\boldsymbol{\theta}$. Thus, the parameter selection is discriminative since their choice will lead to the best classification accuracy.

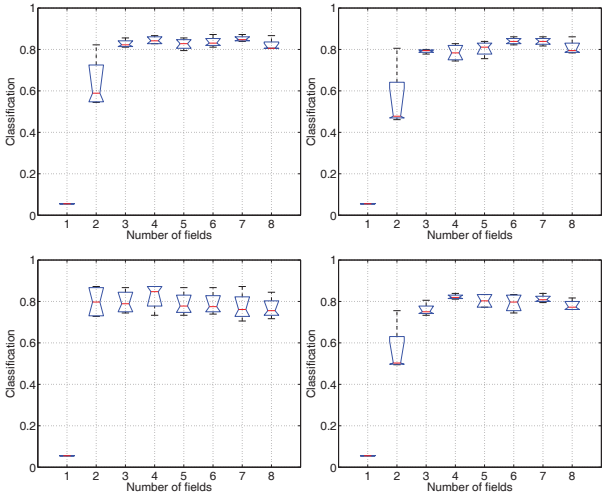

Fig. 2. (a) Classification accuracy for different number of motion fields. Top row: Translation (left) Rigid Body (right) models. Bottom row: Euclidean similarities (left), affine (right) models.

We used a training set that contains approximately 35% of the trajectories for training, and 65% for test. From the training set, 10% was used to estimate the model parameters $\boldsymbol{\theta}$ and the remaining to estimate the model order by cross validation.

Table 2 details how the model parameters are varied to perform the model selection. For each value of K, we consider sev-

Table 3. Classification results of the parametric models for the five activities: $a_1 \rightarrow$ *walking and stepping down the stairs*, $a_2 \rightarrow$ *walking along*, $a_3 \rightarrow$ *crossing and stepping down the stairs*, $a_4 \rightarrow$ *pass diagonally*, $a_5 \rightarrow$ *turning the campus*.

	Translation				
	a_1	a_2	a_3	a_4	a_5
a_1	**70.59**%	13.23%	16.18%	0%	0%
a_2	10.42%	**82.3**%	0%	7.28%	0%
a_3	9.62%	3.85%	**80.77**%	3.85%	1.91%
a_4	1.19%	3.57%	2.38%	**89.29**%	3.57%
a_5	0%	0%	0%	0%	**100**%

	Rigid Body				
	a_1	a_2	a_3	a_4	a_5
a_1	**77.94**%	16.18%	5.88%	0%	0%
a_2	7.29%	**89.58**	0%	3.13%	0%
a_3	19.23%	1.92%	**73.08**%	5.77	0%
a_4	1.19%	0%	0%	**97.62**%	1.19%
a_5	0%	0%	0%	0%	**100**%

	Euclidean Similarities				
	a_1	a_2	a_3	a_4	a_5
a_1	**92.65**%	2.94%	4.41%	0%	0%
a_2	23.96%	**65.63**%	0%	1.04%	9.37%
a_3	13.46	0%	**84.62**%	1.92%	0%
a_4	0%	0%	0%	**97.62**%	2.38%
a_5	0%	0%	0%	0%	**100**%

	Affine model				
	a_1	a_2	a_3	a_4	a_5
a_1	**75**%	17.65%	7.35%	0%	0%
a_2	5.21%	**92.71**%	1.04%	0%	1.04%
a_3	23.08%	3.85%	**67.31**%	1.91%	3.85%
a_4	2.38%	4.76%	1.19%	**89.29**%	2.38
a_5	0%	0%	0%	0%	**100**%

eral noise variances (see 2^{nd} line). This table also shows how we set the initial values for the diagonal (b_{ii}) and non-diagonal (b_{ij}) entries of the transition matrix allowing all kinds of transitions between vector fields at all space positions. The vector field $\mathbf{T}$ is randomly initialized in the interval $[-0.1, 0.1]$. Therefore, a velocity is randomly chosen and it is equal for every nodes in the grid.

Using the above setup we performed, for each parametric model, 128 configurations of the EM (each of them with 4 different EM initializations) to find the most appropriate number of motion fields. Thus, a total of $128 \times 4 = 512$ runs of the EM are performed.

Fig. 2 shows the performance of each parametric model with the variation of the number of motion models K. From this figure, we conclude that for larger values of K the performance of the translation and rigid body models improves. More specifically, to achieve the highest performance, it is required to have $K = 7$ (translation) and $K = 6, 7$ (rigid body). Concerning the Euclidean similarity and Affine models, we conclude that, now the best performances are reached for lower values, *i.e.*, $K = 3, K = 4$. This is expected, since the simpler is the parametric model, the higher is the number of motion fields required to represent the pedestrian's trajectories. Although not shown, the noise variances (see 2^{nd} line in Table 2) that lead to the best accuracy for all models is $\Sigma = 0.1$ for all the parametric models.

5.2. Human activity classification

This section presents statistical evaluation of the method concerning the classification of the activities. A independent test set is used for this purpose. The model parameters as well as the model order are set to the values that provided best discriminative classification, *i.e.* ($K = 6$ for the translation and rigid body models, and $K = 4$ for Euclidean and affine). The best value for the diagonal noise variances is $\Sigma = 0.1$ for all models. Also, we run 4 different initializations of the EM varying the parameter $\mathbf{T}$.

The classification results are reported in Table 3, where the mean values (over the 4 EM estimates) are shown. From this table we see that all the parametric models provide quite remarkable accuracy for surveillance application. More than, trying to select the best model, the most important conclusion is that the class of parametric models herein proposed is a useful framework for classifying pedestrian's activities performed in far-field settings.

6. CONCLUSIONS

This paper presented a novel approach to recognize activities from pedestrian's trajectories. The framework is based on a class of parametric motion models based on space-varying vector fields. This approach allows representing a wide variety of trajectories exhibiting space dependent behaviors. Also, it avoids a non-linear dynamical model based representation, making the proposal simpler. Experiments testify the usefulness of the approach, where the main idea to bare in mind is that complex trajectories can be modeled/classified using simpler parametric models depending on small set of parameters (< 10).

7. REFERENCES

[1] O. Boiman and M. Irani, "Detecting irregularities in images and in video," in *IEEE Int. Conf. on Computer Vision*, 2005, pp. 462–469.

[2] O. J. I. Junejo and M. Shah., "Multi feature path modeling for video surveillance," in *ICPR*, 2004.

[3] A. S. S. T. M. Pierobon, M. Marcon, "Clustering of human actions using invariant body shape descriptor and dynamic time warping," in *IEEE Conf. on Advanced Video and Signal Based Surveillance*, 2005.

[4] J. C. Nascimento, M. A. T. Figueiredo, and J. S. Marques, "Classification of complex pedestrian activities from trajectories," in *ICIP*, 2010, pp. 3481–3484.

[5] X. G. T. Boult, R. Micheals and M. Eckmann, "Into the woods: Visual surveillance of non-cooperative camouflaged targets in complex outdoor settings," in *Proc. of the IEEE*, vol. 89, no. 10, 2001, pp. 1382–1402.

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

[7] L. W. W. Hu, T. Tan and S. Maybank, "A survey on visual surveillance of object motion and behaviors," *IEEE Trans. on Systems and Cybernetics–Part C: Applications and Reviews*, vol. 34, no. 3, pp. 334–352, 2004.

[8] G. N. X. Wang, K. Ma and E. Grimson., "Trajectory analysis and semantic region modeling using a nonparametric Bayesian model," in *CVPR*, 2008.

[9] W. H. Z. Fu and T. Tan, "Similarity based vehicle trajectory clustering and anomaly detection," in *ICIP*, 2005.