Trajectory Modeling Using Mixtures of Vector Fields *

Jacinto C. Nascimento, Mário A. T. Figueiredo, Jorge S. Marques

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

Abstract. Trajectory analysis plays a key role in human activity recognition and video surveillance. This paper proposes a new approach based on modeling trajectories using a bank of vector (velocity) fields. We assume that each trajectory is generated by one of a set of fields or by the concatenation of trajectories produced by different fields. The proposed approach constitutes a space-varying framework for trajectory modeling and is able to discriminate among different types of motion regimes. Furthermore, the vector fields can be efficiently learned from observed trajectories using an expectation-maximization algorithm. An experiment with real data illustrates the promising performance of the method.

1 Introduction

Motion analysis is a central block in many computer vision systems, namely those designed for human activity recognition and video surveillance. When the camera is close to the subject(s) being observed, a wide variety of cues characterizing human activities can be retrieved, *e.g.*, silhouette, head and hands, spatio-temporal templates, color histograms, or articulated models. However, when the camera is far away and has a wide field of view, it is not possible to obtain a detailed description of the subjects. In this case, only trajectory information can be reliably acquired. Motion cues such as velocities and trajectories are therefore a key source of information.

Different trajectory analysis problems (such as classification and clustering) have been addressed using pairwise (dis)similarity measures; these include Euclidean [8] and Hausdorff [11,18] distances, and dynamic time warping [15]).

Another class of methods adopts probabilistic generative models for the trajectories [7,12,14,17], usually of the hidden Markov model (HMM) family. These approaches have the important advantage of not requiring trajectory alignment/registration; moreover, they provide a solid statistical inference framework, based on which model parameters may be estimated from observed data.

* This work was supported by Fundação para a Ciência e a Tecnologia (ISR/IST pluri-anual funding) through the POS Conhecimento Program which includes FEDER funds.

2 Overall Idea

This paper describes a novel approach for modeling object (e.g., pedestrian) trajectories in video sequences. We assume that the object motion is characterized by a set of vector fields, which are learned from observed trajectories. The use of multiple velocity fields aims to describe a variety of behaviors which can be observed in a scene. The system should be able to select the most appropriate velocity field for each sequence.

Two models are considered in this paper. The first model (*M1)* assumes that each object trajectory is generated by one vector field (we don't know which). This leads to a generative model based on a mixture of velocity fields which is flexible enough to represent many types of trajectories. The second model (*M2)* assumes that each trajectory is obtained by the concatenation of segments, each of them generated by one velocity field. Therefore, switching between velocity fields is allowed. Furthermore, it is assumed that the switching mechanism, follows a probabilistic distribution which can be location-dependent.

To illustrate the concept consider two intersecting roads, as depicted in Fig. 1. Given a set of trajectories (Fig. 1 left) we wish to estimate the vector fields which describe the trajectory of cars in the image. Fig. 1 (center and right) shows the expected solution if the problem is addressed using model *M2* (which includes switching). At the intersection, cars have a significant probability of changing direction, whereas far from it, the switching probability is low.

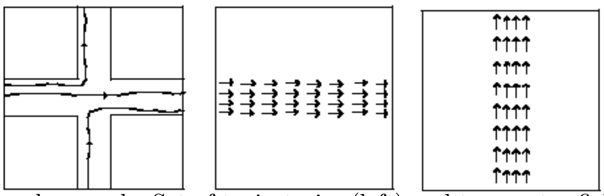

Fig. 1. Cross road example. Set of trajectories (left) and two vector fields modeling to the trajectories (center and right).

Both models are flexible enough to represent a wide variety of trajectories and allow space-varying behaviors without resorting to non-linear dynamical models which are infamously hard to estimate from observed data.

3 Generative Motion Model

For the sake of simplicity, we assume that objects may move freely in the image domain. The object position at time t is represented by a vector $\mathbf{x}_t$ in $\mathbb{R}^2$.

Let $\mathcal{T} = \{\mathbf{T}_1, \ldots, \mathbf{T}_K\}$, with $\mathbf{T}_k : \mathbb{R}^2 \to \mathbb{R}^2$, for $k \in \{1, \ldots, K\}$, be a set of K vector (velocity) fields. The velocity vector at point $\mathbf{x} \in \mathbb{R}^2$ of the k-th field is denoted as $\mathbf{T}_k(\mathbf{x})$. At each time instant, one of these velocity fields is *active*, *i.e.*, is driving the motion. Formally, each object trajectory is generated according to

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{T}_{k_t}(\mathbf{x}_{t-1}) + \mathbf{w}_t, \quad t = 2, ..., L, \tag{1}$$

where $k_t \in \{1, ..., K\}$ is the label of the active field at time t, $\mathbf{w}_t \sim \mathcal{N}(0, \sigma^2_{k_t}\mathbf{I})$ is a realization of white Gaussian noise with zero mean and variance $\sigma^2_{k_t}$ (which may be different for each field), and L is the length (number of points) in the trajectory. The initial position follows some distribution $p(\mathbf{x}_1)$.

The conditional probability density of a trajectory $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_L)$, given the sequence of active models $\mathbf{k} = \{k_1, \ldots, k_L\}$ is

$$p(\mathbf{x}|\mathbf{k}, \mathcal{T}) = p(\mathbf{x}_1) \prod_{t=2}^{L} p(\mathbf{x}_t|\mathbf{x}_{t-1}, k_t),$$

where $p(\mathbf{x}_t|\mathbf{x}_{t-1}, k_t) = \mathcal{N}(\mathbf{x}_t|\mathbf{x}_{t-1} + \mathbf{T}_{k_t}(\mathbf{x}_{t-1}), \sigma^2_{k_t}\mathbf{I})$ is a Gaussian density.

The sequence of active fields $\mathbf{k} = (k_1, \ldots, k_L)$ is modeled as a realization of a first order Markov process, with some initial distribution $P(k_1)$, and a space-varying transition matrix, *i.e.*, $P(k_t = j|k_{t-1} = i, \mathbf{x}_{t-1}) = \mathbf{B}_{ij}(\mathbf{x}_{t-1})$. This model allows the switching probability to depend on the location of the object. The matrix-valued field $\mathbf{B}$ can also be seen as a set of K^2 fields with values in $[0, 1]$, under the constraint that $\sum_j B_{ij}(\mathbf{x}) = 1$, for any $\mathbf{x}$ and any i. If we want to forbid switching between vector fields (as in model *M1*) all we have to do is to assume that $\mathbf{B}(\mathbf{x}) = \mathbf{I}$ (the identity matrix), for any $\mathbf{x}$.

The joint distribution of a trajectory and the underlying sequence of active models, is given by

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

Of course, $P(k_t|k_{t-1}, \mathbf{x}_{t-1})$ is a function of $\mathbf{B}$ and $p(\mathbf{x}_t|\mathbf{x}_{t-1}, k_t)$ is a function of $\mathcal{T}$; to keep the notation lighter, we abstain from explicitly including these dependencies. Finally, a graphical model representation of our generative process is depicted in Fig. 2.

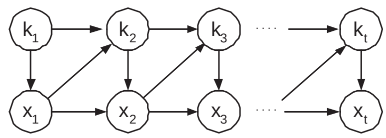

Fig. 2. Graphical model of the trajectory generation process.

4 Learning the Fields

In this section we address the problem of learning, from a set of trajectories, the set of velocity fields $\mathcal{T}$, the field of transition matrices $\mathbf{B}$, and the set of noise variances $\boldsymbol{\sigma} = \{\sigma_1^2, ..., \sigma_K^2\}$. Consider 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's hidden).

In this paper, we focus on model *M1*; *i.e.*, we fix the matrix field $\mathbf{B}$ equal to identity everywhere, $\mathbf{B}(\mathbf{x}) = \mathbf{I}$, thus it does not have to be estimated. In this case, all the elements of each label sequence are identical, *i.e.*, $\mathbf{k}^{(j)} = (k^{(j)}, ..., k^{(j)})$, so we represent the active field of each trajectory simply by $k^{(j)} \in \{1, ..., K\}$. The missing set of labels is thus $\mathcal{K} = \{k^{(1)}, ..., k^{(S)}\} \in \{1, ..., K\}^S$. Finally, we denote the set of the fields and parameters to estimate as $\boldsymbol{\theta} = (\mathcal{T}, \boldsymbol{\sigma})$.

4.1 Estimation Criterion: Marginal MAP

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

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \; p(\boldsymbol{\theta}) \prod_{j=1}^{S} \sum_{k^{(j)}=1}^{K} p(\mathbf{x}^{(j)}, \mathbf{k}^{(j)}|\boldsymbol{\theta}), \tag{3}$$

where each factor $p(\mathbf{x}^{(j)}, \mathbf{k}^{(j)}|\boldsymbol{\theta})$ has the form in (2), with $\mathbf{k}^{(j)} = (k^{(j)}, ..., k^{(j)})$. Next, we derive the E and M steps of the EM algorithm for solving (3). For simplicity, we assume that the initial distributions $p(\mathbf{x}_1)$ and $P(k_1)$ are known.

4.2 The E-step

As is well known, the E-step consists in computing the conditional expectation of the complete log-likelihood, given the current estimates $\widehat{\boldsymbol{\theta}}$ and the observations $\mathcal{X}$. The complete log-likelihood is given by

$$\log p(\mathcal{X}, \mathcal{K}|\boldsymbol{\theta}) = \sum_{j=1}^{S} \log p(\mathbf{x}^{(j)}, \mathbf{k}^{(j)}|\boldsymbol{\theta}) \tag{4}$$

where each $p(\mathbf{x}^{(j)}, \mathbf{k}^{(j)}|\boldsymbol{\theta})$ has the form (2), with $\mathbf{k}^{(j)} = (k^{(j)}, ..., k^{(j)})$. The conditional expectation, usually called the Q-function and denoted as $Q(\boldsymbol{\theta}; \widehat{\boldsymbol{\theta}}) \equiv \mathbb{E}\left[\log p(\mathcal{X}, \mathcal{K}|\boldsymbol{\theta}) \,\middle|\, \mathcal{X}, \widehat{\boldsymbol{\theta}}\right]$, can thus be written as

$$Q(\boldsymbol{\theta}; \widehat{\boldsymbol{\theta}}) = \sum_{j=1}^{S} \sum_{t=2}^{L_j} \sum_{l=1}^{K} \bar{y}_l^{(j)} \log \mathcal{N}(\mathbf{x}_t^{(j)} | \mathbf{x}_{t-1}^{(j)} + \mathbf{T}_l(\mathbf{x}_{t-1}^{(j)}), \sigma_l^2 \mathbf{I}) \tag{5}$$

where $\bar{y}_l^{(j)}$ is the posterior probability that the j-th trajectory was generated by field l, given by

$$\bar{y}_l^{(j)} = P\left[k^{(j)} = l \,\middle|\, \mathbf{x}^{(j)}, \widehat{\boldsymbol{\theta}}\right] = \frac{p(\mathbf{x}^{(j)}, \mathbf{k}^{(j)} = (l, ..., l)|\widehat{\boldsymbol{\theta}})}{\displaystyle\sum_{m=1}^{K} p(\mathbf{x}^{(j)}, \mathbf{k}^{(j)} = (m, ..., m)|\widehat{\boldsymbol{\theta}})}. \tag{6}$$

4.3 The M-step

In the M-step, the field and parameter estimates are updated according to

$$\widehat{\boldsymbol{\theta}}_{\text{new}} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}; \widehat{\boldsymbol{\theta}}) + \log p(\boldsymbol{\theta}). \tag{7}$$

This section describes this maximization in detail, as well as the adopted priors, by looking separately at the maximization with respect to $\mathcal{T}$ and $\boldsymbol{\sigma}$.

Updating $\widehat{\boldsymbol{\sigma}}$ Adopting flat priors, *i.e.*, looking for the usual maximum likelihood noise variance estimates, computing the partial derivative of $Q(\boldsymbol{\theta}; \widehat{\boldsymbol{\theta}})$ with respect to each component σ_k^2 of $\boldsymbol{\sigma}$, and equating to zero, yields

$$(\widehat{\sigma}_k^2)_{\text{new}} = \left(\sum_{j=1}^{S} \sum_{t=2}^{L_j} \bar{y}_k^{(j)} \, \| \mathbf{x}_t^{(s)} - \mathbf{x}_{t-1}^{(s)} - \mathbf{T}_k(\mathbf{x}_{t-1}^{(s)}) \|^2 \right) \left(\sum_{j=1}^{S} \sum_{t=2}^{L_j} \bar{y}_k^{(j)} \right)^{-1},$$

for $k = 1, ..., K$.

Updating $\widehat{\mathcal{T}}$ Estimating the velocity fields requires some sort of regularization. Moreover, these fields live in infinite dimensional spaces (if we ignore the discrete nature of digital images), thus optimization with respect to them constitute a difficult variational problems. We sidestep this difficulty by adopting a finite dimensional parametrization, in which each velocity field is written as a linear combination of basis functions, *i.e.*,

$$\mathbf{T}_k(\mathbf{x}) = \sum_{n=1}^{N} \mathbf{t}_k^{(n)} \, \phi_n(\mathbf{x}), \tag{8}$$

where each $\mathbf{t}_k^{(n)} \in \mathbb{R}^2$ and $\phi_n(\mathbf{x}) : \mathbb{R}^2 \to \mathbb{R}$, for $n = 1, \ldots, N$, is a set of basis functions (scalar basis fields). Collecting all these vector coefficients in $\boldsymbol{\tau}_k \in \mathbb{R}^{N \times 2}$, defined according to $\boldsymbol{\tau}_k^T = [\mathbf{t}_k^{(1)}, ..., \mathbf{t}_k^{(N)}]$ and letting $\Phi(\mathbf{x}) = [\phi_1(\mathbf{x}), ..., \phi_N(\mathbf{x})] \in \mathbb{R}^N$, we can write

$$\mathbf{T}_k(\mathbf{x}) = \Phi(\mathbf{x}) \, \boldsymbol{\tau}_k, \tag{9}$$

thus estimating $\mathbf{T}_k$ becomes equivalent to estimating the coefficient vector $\boldsymbol{\tau}_k$.

To encourage smoothness of each velocity field $\mathbf{T}_k$, we adopt a zero mean Gaussian prior, with a covariance function chosen to assign low probability to large velocity differences between nearby locations

$$p(\boldsymbol{\tau}_k) \propto \exp\{-\frac{1}{2\,\alpha^2} \, \boldsymbol{\tau}_k^T \, \boldsymbol{\Gamma}^{-1} \boldsymbol{\tau}_k\}, \tag{10}$$

where α^2 is a global variance factor that allows controlling the "strength" of the prior. The covariance $\boldsymbol{\Gamma}$ and the basis functions ϕ_i determine the covariance function of $\mathbf{T}_k$; for details, see [16].

The term of $Q(\boldsymbol{\theta};\widehat{\boldsymbol{\theta}}) + \log p(\boldsymbol{\tau}_k)$ which is a function of $\boldsymbol{\tau}_k$ is (apart from a 1/2 factor) equal to

$$\sum_{j=1}^{S}\sum_{t=2}^{L_j} \bar{y}_k^{(j)} \|\mathbf{x}_t^{(j)} - \mathbf{x}_{t-1}^{(j)} - \Phi(\mathbf{x}_{t-1}^{(j)})\boldsymbol{\tau}_k\|^2 + \boldsymbol{\tau}_k^T\, \boldsymbol{\Gamma}^{-1}\boldsymbol{\tau}_k \tag{11}$$

Computing the gradient with respect to $\boldsymbol{\tau}_k$ and equating to zero, leads to a linear system of equations,

$$\left(\mathbf{R}_k + \frac{\boldsymbol{\Gamma}^{-1}}{\alpha^2}\right)\, \boldsymbol{\tau}_k = \mathbf{r}_k \tag{12}$$

where

$$\mathbf{R}_k = \sum_{j=1}^{S}\sum_{t=2}^{L_j} \bar{y}_k^{(j)} \left(\Phi(\mathbf{x}_{t-1}^{(j)})\right)^T \Phi(\mathbf{x}_{t-1}^{(j)}) \tag{13}$$

and

$$\mathbf{r}_k = \sum_{j=1}^{S}\sum_{t=2}^{L_j} \bar{y}_k^{(j)} \left(\Phi(\mathbf{x}_{t-1}^{(j)})\right)^T (\mathbf{x}_t^{(j)} - \mathbf{x}_{t-1}^{(j)}). \tag{14}$$

Notice that since $\Phi(\mathbf{x}_{t-1}^{(j)})$ is $1 \times N$, matrix $\mathbf{R}_k$ is $N \times N$ and $\mathbf{r}_k$ is $N \times 2$ (as is $\boldsymbol{\tau}_k$). Solving (12), yields $(\widehat{\boldsymbol{\tau}}_k)_{\text{new}}$, for $k = 1, ..., K$, which in turn define $\widehat{\mathcal{T}}_{\text{new}} = (\widehat{\mathbf{T}}_1, ..., \widehat{\mathbf{T}}_K)_{\text{new}}$.

5 Experimental Results

The proposed algorithm was applied to several synthetic and real data sets. We present here only an example showing the ability of the proposed method to separate different types of trajectories, according to their structure, and to estimate a space-varying model for the trajectories of each group.

Fig. 3 (left, center) shows the trajectories of points of interest in two video sequences often used in multi-point tracking (rotating disk and golf ball). We applied the proposed method to a data set containing all the trajectories extracted from both video sequences. It was assumed that the number of vector fields is known to be equal to 2. Fig. 4 shows the velocity fields estimates obtained by the EM algorithm. It is clear that the proposed algorithm successfully separated the two sets of trajectories and estimated the motion fields explaining each of them. This approach has also been applied to surveillance videos obtained in a university campus with success. Those results will be presented in a forthcoming paper due to lack of space.

6 Conclusions

This paper described probabilistic models for trajectories in video sequences using multiple velocity fields. Two models were considered: (i) each trajectory is

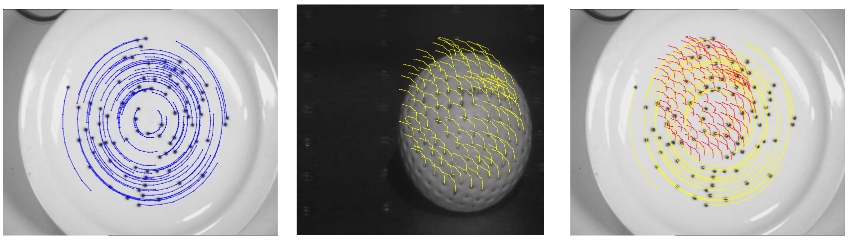

Fig. 3. Data set: original velocity fields (left, center) and their superposition (right)

generated by a single velocity field, randomly selected (M1); (ii) switching among vector fields during the trajectory is allowed (M2). We have described in detail an EM algorithm to estimate the fields under model M1. A proof-of-concept experiment with real data was presented, giving evidence of the adequacy of the proposed approach.

The implementation of the second scenario, which involves the estimation of a field of stochastic matrices, as well as a comprehensive experimental evaluation in video surveillance tasks will be presented in a forthcoming paper.

Acknowledgement The image sequences used in this paper were kindly provided by K. Shafique.

References

1. A. Agarwal, B. Triggs. Tracking articulated motion with piecewise learned dynamical models. *Eur. Conf. on Comp. Vision*, Prague, 2004.
2. S. Ali, M. Shah. A Lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. *IEEE Conf. Comp. Vision and Patt. Rec.*, Minneapolis, 2007.
3. O. Boiman, M. Irani. Detecting irregularities in images and in video. *IEEE Int. Conf. on Comp. Vision*, Beijing, China, 2005.
4. T. Boult, R. Micheals, X. Gao, M. Eckmann. Into the woods: Visual surveillance of non-cooperative camouflaged targets in complex outdoor settings. In *Proc. of the IEEE*, 89(10):1382–1402, 2001.
5. P. Dierchx. *Curve and Surface Fitting with Splines.* Oxford Univ. Press, 1993.
6. J. Duchi, S. Shalev-Shwartz, Y. Singer, T. Chandra. Efficient projections onto the ℓ_1-ball for learning in high dimensions. *Int. Conf. on Machine Learning*, Helsinki, Finland, 2008.
7. T. Duong, H. Bui, D. Phung, S. Venkatesh. Activity recognition and abnormality detection with the switching hidden semi-Markov model. *IEEE Conf. Comp. Vision and Patt. Rec.*, San Diego, CA, 2005.
8. Z. Fu, W. Hu, T. Tan. Similarity based vehicle trajectory clustering and anomaly detection. In *IEEE Int. Conf. on Image Proc.*, Genoa, Italy, 2005.
9. W. Hu, T. Tan, L. Wang, S. Maybank. A survey on visual surveillance of object motion and behaviors. *IEEE Trans. Systems, Man, and Cybern. (Part C)*, 34(3):334–352, 2004.
10. N. Johnson and D. C. Hogg. Learning the distribution of object trajectories for event recognition. *Image and Vision Computing*, 14:583–592, 1996.

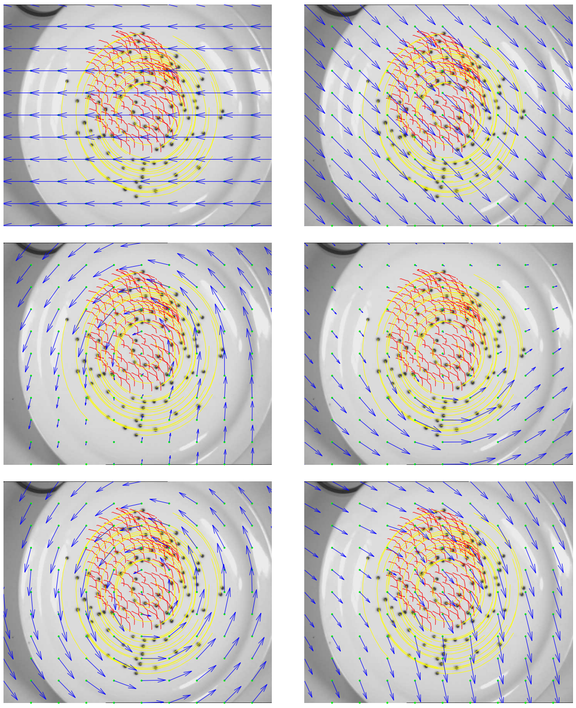

Fig. 4. Left and right columns: the two velocity field estimates, shown in blue. Top row: initialization. Second row: intermediate estimates. Bottom row: final estimates.

11. I. Junejo, O. Javed, M. Shah. Multi feature path modeling for video surveillance. In *Int. Conf. on Patt. Rec.*, Cambridge, UK, 2004.
12. J. Nascimento, M. Figueiredo, J. Marques. Independent increment processes for human motion recognition. *Comp. Vision & Image Underst.*, 109(2):126–138, 2008.
13. J. Nocedal, S. Wright. *Numerical Optimization,* Springer, New York, 2006.
14. N. Oliver, B. Rosario, A. Pentland. A Bayesian computer vision system for modeling human interactions. *IEEE Trans. Patt. Anal. Mach. Intell.,* 22(8):831–843, 2000.
15. M. Pierobon, M. Marcon, A. Sarti, S. Tubaro. Clustering of human actions using invariant body shape descriptor and dynamic time warping. *IEEE Conf. on Adv. Video and Sig. Based Surv.,* pages 22–27, 2005.
16. C. Rasmussen, C. Williams. *Gaussian Processes for Machine Learning.* The MIT Press, Cambridge, MA, 2006.
17. X. Wang, K. Ma, G. Ng, E. Grimson. Trajectory analysis and semantic region modeling using a nonparametric Bayesian model. *IEEE Conf. on Comp. Vision and Patt. Rec.*, Anchorage, 2008.
18. X. Wang, K. Tieu, E. Grimson. Learning semantic scene models by trajectory analysis. *Eur. Conf. on Comp. Vision*, Graz, Austria, 2006.