# Trajectory Analysis Using Switched Motion Fields: a Parametric Approach

Jorge S. Marques[1,2], João M. Lemos[1,3], Mário A. T. Figueiredo[1,4], Jacinto C. Nascimento[1,2], Miguel Barão[3,5]

[1] Instituto Superior Técnico, Portugal
[2] Instituto de Sistemas e Robótica, Portugal
[3] INESC-ID, Portugal
[4] Instituto de Telecomunicações, Portugal
[5] Universidade de Évora, Portugal

**Abstract.** This paper presents a new model for trajectories in video sequences using mixtures of motion fields. Each field is described by a simple parametric model with only a few parameters. We show that, despite the simplicity of the motion fields, the overall model is able to generate complex trajectories occuring in video analysis.

## 1 Introduction

The analysis of trajectories plays an important role in computer vision [1]-[8]. Consider, *e.g.*, a video surveillance system, tracking moving people or vehicles in a parking lot or in a street. The trajectory of each object is a rich source of information about its behavior. We should therefore be able to learn what are the typical trajectories and how can they be characterized so that we can distinguish typical behaviors from abnormal ones and discriminate different types of common behaviors.

A trajectory model must be rich enough to allow different types of behaviors occurring at the same place. For example, several types of trajectories may occur in a hotel lobby. The same happens if we wish to characterize the traffic in a city or in part of it. A single motion field is not enough to characterize people of vehicle motion in a scene.

A generative model for trajectory analysis based on *switched motion fields* was recently proposed in [1]. The model is equipped with estimation methods that are able to learn a set of motion fields describing typical behaviors of objects in a scene. It is assumed that each trajectory is driven by one of the motion fields at each instant of time, the so-called *active field*. Switchings between active fields are allowed and may occur at any position and any instant of time, according to suitable probabilities. This model is rich enough to describe a variety of behaviors and simple enough to be efficiently learned from experimental data.

This work was supported by FCT (plurianual funding) through the PIDDAC Program funds and by project PTDC/EEACRO/098550/2008.

The work presented in [1] adopts a non-parametric model for the motion fields based on first order splines. In this paper, we extend that work to simple parametric models that depend on a small number of parameters and discuss if it is still possible to obtain flexible overall behaviors and efficient estimation from observed data. The main contribution consists in a parametric approach to the switched motion field model.

## 2 Switched Motion Field Model

Let $x = (\mathbf{x}_1, \ldots, \mathbf{x}_L),\ \mathbf{x}_t \in \mathbb{R}^2$, denote the trajectory of an object in the image. We assume that $x$ is generated by a bank of K vector fields $\mathbf{T}_i : \mathbb{R}^2 \rightarrow \mathbb{R}^2,\ i \in \{1, \ldots, K\}$, according to

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

where $k_t$ is the label of the active field at instant of time $t$ and $\mathbf{w}_1, \ldots, \mathbf{w}_L$ is a white random sequence with normal distribution $\mathbf{w}_t \sim \mathcal{N}(0, \sigma_{k_t}^2 \mathbf{I})$.

Furthermore, we assume that label sequence $k = (k_1, \ldots,\ k_L)$ is a Markov chain with space varying transition probabilities, *i.e.*, the next active field depends on the current active field as well as on the position of the object in the scene. This is an important issue. For example, consider a cross between two streets. Many pedestrians and vehicles change their direction and velocity at the cross. Therefore, the transition probabilities should be higher at the cross than elsewhere.

To be specific, model switching is characterized by the transition probabilities,

$$P(k_t = j | k_{t-1} = i, \mathbf{x}_{t-1}) = B_{ij}(\mathbf{x}_{t-1}) \ , \tag{2}$$

where $B_{ij}(\mathbf{x})$ is the probability of switching from the field $i$ to the field $j$ at position $\mathbf{x}$. Therefore, $\mathbf{B}(\mathbf{x}) = \{B_{ij}(\mathbf{x})\}$ is a field of stochastic matrices which verify the following properties at each position $\mathbf{x}$:

$$B_{ij}(\mathbf{x}) \geq 0, \qquad \sum_{p=1}^{K} B_{ip}(\mathbf{x}) = 1, \qquad \forall i, j \ . \tag{3}$$

Figure 1 shows a set of of trajectories which can be easily generated by this model using three vector fields. This figure suggests the variety of behaviors that can be generated with the proposed approach.

The pair $(\mathbf{x}_t, k_t)$ can be considered as a hybrid state since it summarizes all the past information needed to generate the future samples of the process. The joint probability function associated to the pair of sequences $\{x, k\}$ is given by

$$p(x, k) = p(\mathbf{x}_1, k_1) \prod_{t=2}^{L} p(\mathbf{x}_t | k_t, \mathbf{x}_{t-1}) p(k_t | \mathbf{x}_{t-1}, k_{t-1}), \tag{4}$$

where

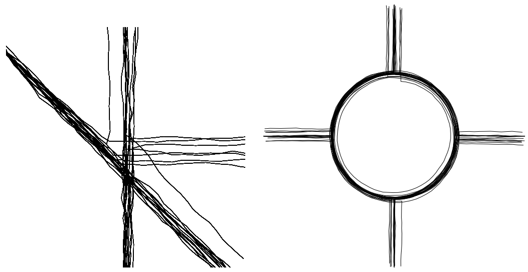

**Fig. 1.** Synthetic trajectories for the cross and roundabout problems

$$p(\mathbf{x}_t|k_t, \mathbf{x}_{t-1}) = \frac{1}{2\pi\sigma_{k_t}^2} e^{-\frac{1}{2\sigma_{k_t}^2}\|\mathbf{x}_t - \mathbf{x}_{t-1} - T_{k_t}(\mathbf{x}_{t-1})\|^2} , \quad (5)$$

and $p(k_t|\mathbf{x}_{t-1}, k_{t-1}) = B_{k_{t-1},k_t}(\mathbf{x}_{t-1})$.

The parameters to be learned from the video data are: i) the number of models $K$; ii) the motion fields $\mathbf{T}_1, \ldots, \mathbf{T}_K$; iii) the field of transition matrices $\mathbf{B}$; and iv) the noise variances $\sigma_1^2, \ldots, \sigma_K^2$.

In [1] the fields $\mathbf{T}_1, ..., \mathbf{T}_K$ and $\mathbf{B}$ are modeled in a non-parametric way. They are specified at the nodes of a regular grid and interpolated using first order splines

$$\mathbf{T}_k(\mathbf{x}) = \sum_{i=1}^{N} \mathbf{t}_k^i \phi_i(\mathbf{x}), \qquad \mathbf{B}(\mathbf{x}) = \sum_{i=1}^{N} \mathbf{b}^i \phi_i(\mathbf{x}) , \quad (6)$$

where $\mathbf{t}_k^i, \mathbf{b}^i$ are the velocity vector and the transition matrix associated to the $i - th$ node of the grid and $\phi_i(\mathbf{x})$ is the corresponding spline function, centered at the i-th node. As a consequence, that approach can be classified as non-parametric since we are not imposing any kind of structure, and each field depends on a large number of parameters (typically a few hundreds) which have to be estimated from the data. Some kind of regularization (Gaussian field priors, in [1]) is required to obtain meaningful estimates for these parameters.

In this paper we follow a different approach by adopting parametric models for the motion fields. This results in a much smaller number of parameters to be estimated. Although we are making strong assumptions about each motion field, a flexible trajectory model is expected at the end because trajectories are decomposed into segments, each of them generated by a different motion field. Parametric field may be tuned to a specific space region, if necessary.

## 3 Parametric Motion Fields

We consider several parametric motion models, which are often used in image alignment and registration, namely [9]: translation (T), Euclidean (E), similarity (S) and affine (A) transforms. All these models are expressed by

$$\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{t} , \quad (7)$$

| Name | Transformation | Motion field |
|---|---|---|
| T | $\mathbf{z} = \mathbf{x} + \mathbf{t}$ | $\mathbf{T}(\mathbf{x}) = \mathbf{t}$ |
| E | $\mathbf{z} = \mathbf{R}\mathbf{x} + \mathbf{t}$ | $\mathbf{T}(\mathbf{x}) = (\mathbf{R} - \mathbf{I})\mathbf{x} + \mathbf{t}$ |
| S | $\mathbf{z} = \mathbf{sR}\mathbf{x} + \mathbf{t}$ | $\mathbf{T}(\mathbf{x}) = (s\mathbf{R} - \mathbf{I})\mathbf{x} + \mathbf{t}$ |
| A | $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{t}$ | $\mathbf{T}(\mathbf{x}) = (\mathbf{A} - \mathbf{I})\mathbf{x} + \mathbf{t}$ |

**Table 1.** Parametric motion models

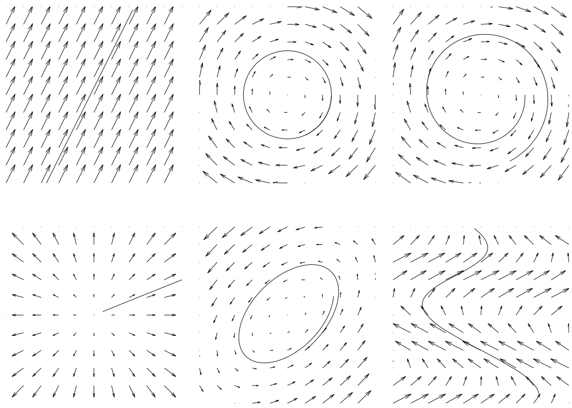

**Fig. 2.** Examples of motion fields and trajectories generated by the T, E, S models (top row) and A, A, non parametric models (bottom row).

where $\mathbf{z}$ is the transformed position of the point $\mathbf{x}$, $\mathbf{A}$ is a $2 \times 2$ matrix and $\mathbf{t}$ is a $2 \times 1$ translation vector. The only difference between these models lies in the structure of matrix $\mathbf{A}$ as shown in Table 1. In this table, $\mathbf{R}$ (a rotation matrix) and $s\mathbf{R}, s \in \mathbb{R}$, have the following structure

$$\mathbf{R} = \begin{bmatrix} \cos\theta \; \sin\theta \\ -\sin\theta \; \cos\theta \end{bmatrix} \qquad s\mathbf{R} = \begin{bmatrix} a \; b \\ -b \; a \end{bmatrix} , \tag{8}$$

and $\mathbf{A}$ is an arbitrary $2 \times 2$ matrix.

Figure 2 shows examples of the motion fields and trajectories generated by these models and by a non-parametric one. Only the translation and the Euclidean transform generate trajectories in which the object moves with constant speed since the eigenvalues of the matrix $\mathbf{A}$ lie on the unit circle. In the other cases, velocity has an exponential growth or decay. This is not a major problem since each model is only used for a short period of time. We stress that, at this point, no switching was allowed in the generation of these trajectories. More complex trajectories can be generated if we allow model switching.

## 4 Model Estimation

In an ideal setting, we would like to turn on the camera and ask the system to learn the behavior of all the pedestrians and vehicles in the scene. Assuming that the tracking task is solved (although this is by no means a trivial task) we would like to estimate the number of fields and the field parameters from a set of observed trajectories $\mathcal{X} = \{x^{(1)}, \ldots, x^{(S)}\}$ where $x^{(s)} = (\mathbf{x}_1^{(s)}, \ldots, \mathbf{x}_{L_s}^{(s)})$ is the s-th trajectory.

The maximum likelihood (ML) estimates of all the model parameters, collectively denoted as $\theta$, can be obtained by solving the following optimization problem

$$\hat{\theta} = \arg\max_{\theta} \ \log p(\mathcal{X}|\theta). \tag{9}$$

However, the likelihood function cannot be directly computed. Since we do not know the sequence of active models underlying each trajectory, we should marginalize the complete likelihood function $p(\mathcal{X}, \mathcal{K}|\theta)$, *i.e.*,

$$p(\mathcal{X}|\theta) = \sum_{\mathcal{K}} p(\mathcal{X}, \mathcal{K}|\theta) = \sum_{\mathcal{K}} \prod_{s=1}^{S} p(x^{(s)}, k^{(s)}|\theta), \tag{10}$$

where $\mathcal{K} = \{k^{(1)}, \ldots, k^{(1)}\}$ are the (unobserved) label sequences (active models) and $p(x^{(s)}, k^{(s)}|\theta)$ is the joint density defined in (4). The marginalization involves a sum for all sequences of labels $\mathcal{K}$ which is unfeasible since it involves a huge number of operations.

This difficulty can be circumvented by using the Expectation-Maximization (EM) method. The EM method is based on the optimization of an auxiliary function: the conditional expectation of the complete log-likelihood

$$U(\theta, \hat{\theta}) = \mathbb{E}\left\{\log p(\mathcal{X}, \mathcal{K})|\mathcal{X}, \hat{\theta}\right\}, \tag{11}$$

where $\hat{\theta}$ is the currently available estimate of the model parameters. The EM method generates a sequence of estimates by iteratively optimizing $U(\theta, \hat{\theta})$ with respect to $\theta$, to update the parameter estimates:

$$\hat{\theta}(t+1) = \arg\max_{\theta} \ U(\theta, \hat{\theta}(t)). \tag{12}$$

The expected value of the complete log-likelihood with respect to these variables can be written as

$$U(\theta, \hat{\theta}) = \bar{\mathcal{A}}(\mathcal{X}, \mathcal{K}) + \bar{\mathcal{B}}(\mathcal{X}, \mathcal{K}), \tag{13}$$

with

$$\begin{aligned} \bar{\mathcal{A}}(\mathcal{X}, \mathcal{K}) = C - \textstyle\sum_{s=1}^{S} \sum_{t=2}^{L_s} \sum_{i=1}^{K} w_i^{(s)} \Big[ &\log(2\pi\sigma_i^2) \\ &+ \tfrac{1}{2\sigma_i^2} \|\mathbf{x}_t^{(s)} - \mathbf{x}_{t-1}^{(s)} - \mathbf{T}_i(\mathbf{x}_{t-1}^{(s)})\|^2 \Big], \end{aligned} \tag{14}$$

$$\bar{\mathcal{B}}(\mathcal{X}, \mathcal{K}) = \sum_{s=1}^{S} \sum_{t=2}^{L_s} \sum_{i,j=1}^{K} w_{i,j}^{(s)} \log B_{ij}(\mathbf{x}_{t-1}^{(s)}), \tag{15}$$

where $w_i^{(s)}(t) = P(k_t^{(s)} = i|\mathbf{x}^{(s)}, \hat{\theta})$ is the probability of label $i$ at time $t$ and $w_{i,j}^{(s)}(t) = P(k_{t-1}^{(s)} = i, k_t^{(s)} = j|\mathbf{x}^{(s)}, \hat{\theta})$ the probability of the pair of labels $i, j$ at

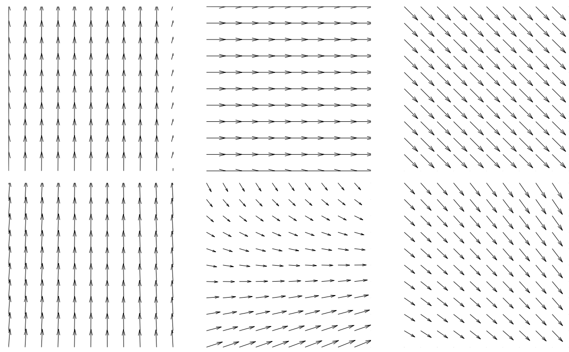

**Fig. 3.** Estimated fields for the cross problem with translation (1st row) and affine (2nd row) models ($K = 3$)

consecutive instants of time. The probabilities $w_i^{(s)}(t), w_{i,j}^{(s)}(t)$ are computed in the E-step of the EM method using the forward-backward algorithm [10].

The M-step maximizes $U$ with respect to the model parameters. The maximization with respect to the noise variances and switching matrix field is done as in [11]. The optimization with respect to the motion parameters depends on the motion model adopted but this is straight forward.

## 5 Results

The proposed model was applied to synthetic trajectories. Figure 1 shows two sets of 30 trajectories which are denoted as cross and roundabout. The first case simulates a cross between three streets with two entries. The second case simulates a roundabout with four entries and combines linear and circular segments. The field of transition matrices was modeled in a non-parametric way, using a $11 \times 11$ regular grid of points and first order interpolation splines as in [1].

Figure 3 shows the estimates obtained by the EM method assuming affine motion fields and translation fields. These results were obtained after 5 iterations. Both models are able to correctly extract the correct motion fields. The second affine motion field is not uniform but it is approximately uniform in the region of interest. We stress that we do not have any *a priori* knowledge about the active model at each instant of time. This information must be guessed by the EM algorithm using the soft assignment variables $(w_i^{(s)}(t), w_{i,j}^{(s)}(t))$.

The second problem is more complex since it involves a larger number of fields (five) and a mixture of circular and uniform fields. Figure 4 (1st row) shows the estimates obtained using Euclidean motion fields showing that the EM method is able to correctly estimate the field parameters. It should be mentioned that the output of the EM method depends on the initialization as shown in the Figure 4 (bottom): two motion fields are associated to the circular motion and the third motion field in the Figure tries to represent two different motion directions.

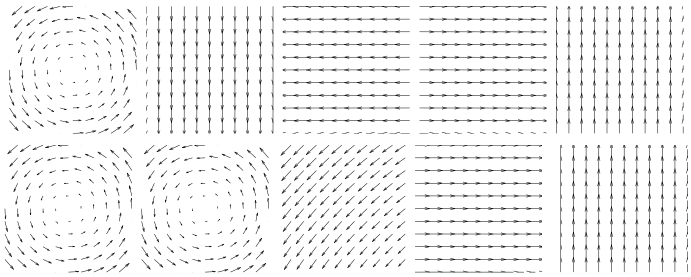

**Fig. 4.** Estimated fields for the roundabout problem with Euclidean transform model ($K = 5$) and two different initializations

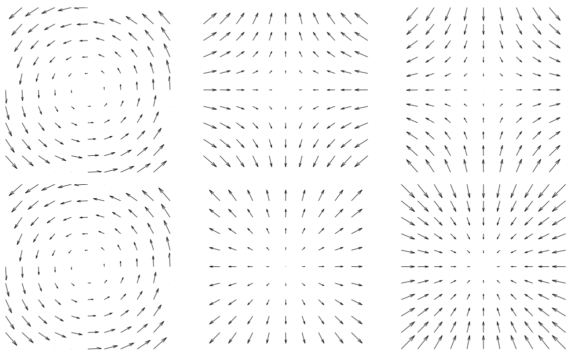

**Fig. 5.** Estimated fields for the roundabout problem with the affine model ($K = 3$) and two different initializations

This is shown in Figure 5 which displays the output of the EM method, assuming an affine model with a smaller number of fields (three) and two different initializations.

Although excellent results were achieved for both problems, the algorithm may converge to local maxima of the likelihood function, leading to poor estimates of the motion fields. This is a consequence of the EM estimation method which does not guarantee the convergence towards the global maxima [12]. However, this effect is stronger in the estimation of parametric vector fields depending on a small number of parameters ($< 10$) then in the case of non-parametric models which depend on hundreds of parameters. The presence of global restrictions increase the attraction towards local maxima.

## 6 Conclusions

This paper presents an extension of the trajectory model with multiple motion fields presented in [1]. It shows that complex trajectories can be generated using a set of simple motion fields (parametric fields) depending on a small number of parameters. The key point is the ability to switch between motion fields at any position in space. In addition we use space dependent switching probabilities which allow different switching behaviors in different regions of space.

Despite the good results obtained there are several open question to be addressed in the future: how to initialize the EM method in an efficient way? how to determine the best number of motion fields for a given problem? how to select the most appropriate model for each field? application of this model to real data and the comparison with the non-parametric model [1]. These questions will be addressed in a forthcoming paper which will include extensive results and a comparison with non-parametric techniques.

## References


1. J. C. Nascimento, M. A. T. Figueiredo, J. S. Marques, Trajectory Analysis in Natural Images Using Mixtures of Vector Fields, IEEE International Conference of Image Processing, 2009.
2. W. Hu, T. Tan, L. Wang, S. Maybank. A survey on visual surveillance of object motion and behaviours, IEEE Trans. on Systems and Cybernetics, Part C, 34, 334-352, 2004.
3. T. Duong, H. Bui, D. Phung, S. Venkatesh. Activity recognition and abnormality detection with the switching hidden semi-Markov model. IEEE Conference on Computer Vision and Pattern Recognition, 2005.
4. Z. Fu, W. Hu, T. Tan. Similarity based vehicle trajectory clustering and anomaly detection. IEEE International Conference on Image Processing, Genoa, Italy, 2005.
5. X. Wang, K. Ma, G. Ng, E. Grimson. Trajectory analysis and semantic region modeling using a nonparametric Bayesian model. IEEE Conference on Computer Vision and Pattern Recognition, 2008.
6. N. Oliver, B. Rosario, A. Pentland. A Bayesian computer vision system for modeling human interactions. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(8):831-843, 2000.
7. I. Junejo, O. Javed, M. Shah, Multi feature path modeling for video surveillance, International Conference on Pattern Recognition, 2004.
8. to be included
9. R. Szeliski, Image Alignment and Stiching: a Tutorial, Foundations and Trends in Computer Graphics and Computer Vision, Vol. 2, 1-104, December 2006.
10. L. R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, Proceedings of the IEEE, vol. 77, no. 2, 257-286, 1989.
11. M. Barão, J. S. Marques, J. M. Lemos, An Improved EM-method for the Estimation of Transition Probabilities in Multiple Model Switching Systems, IFAC Symposium on Nonlinear Control Systems, Bolonha, 2010.
12. C. Wu, On the convergence properties of the EM algorithm, the Annals of Statistics, Vol. 11, 95-103, 1983.