ESTIMATION OF SPACE-VARYING COVARIANCE MATRICES

Catarina Barata *João M. Lemos* *Jorge S. Marques*

Institute for Systems and Robotics, INESC-ID,
Instituto Superior Técnico, Universidade de Lisboa, Portugal

ABSTRACT

This paper considers the representation of human trajectories in video signals. These trajectories are modeled by switched dynamical models, based on motion fields that drive the pedestrian during consecutive time intervals. This paper addresses the estimation of uncertainty in trajectory generation by using space-varying covariance matrices estimated from the video data. Experimental results show that the proposed model outperforms previous methods, based on static and isotropic covariance matrices.

***Index Terms*—** Human motion, multiple dynamical models, space-varying covariance matrices.

1. INTRODUCTION

Understanding and predicting human motion in a scene is of major importance to many applications, such as surveillance (*e.g.*, identification of abnormal behaviors), sport performance analysis, and human-robot interactions [15]. The images can be acquired using two different settings: i) short range, where the camera is close to the observed person and we have access to detailed information of the human pose; and ii) far field, where the camera covers a wide area (see Fig. 1) and is no longer able to provide detailed information. In this work, we are interested in the latter setting, in which most methods rely on the characterization of the trajectories performed by the pedestrians.

Two classes of approaches have been used to characterize trajectories. The first one is based on human-human interactions, basically assuming that the pedestrian motion is governed by attractive and repulsive forces, created by neighbor pedestrians (social forces models) [10, 13]. Alternatively, one can assume that the motion is defined by the interaction between the pedestrian and the geometry of the scene. In this case, the number of typical motions is finite, making it possible to estimate the motion patterns [17]. The methodologies proposed in this context may be divided into two groups: discriminative and generative. Discriminative methods include splines [16, 2], clustering [2, 18], and classification algorithms (*e.g.*, reinforcement learning [12] or recurrent neural networks [1, 14]) strategies. Despite their promising results, these methods may be too rigid and, thus, unable to model the inherently stochastic behavior of the trajectories, which depends on factors such as the presence of other agents in the scene or the time of the day. Some discriminative methods have tried to tackle this issue by combining motion models with social forces approaches (*e.g.*, [1, 8]). On the other hand, generative methods, such as the ones based on Gaussian processes (GP) [7, 11, 6] or switching motion vector fields [20], allow us to be explicit regarding the uncertainty induced by the aforementioned factors.

In this work we will focus on modeling trajectories using vector fields, which have been shown to efficiently describe not only pedestrian trajectories [20], but also GPS data from hurricanes, vehicles, and cellphones [9]. Our starting point will be the method proposed in [20], where trajectories are defined as a sequence of points, each generated by one vector field. Although only one motion model is active at each time instant, it is assumed that switching between them may occur at any point in the domain, according to a space-varying probability. This makes the model flexible enough to characterize a wide variety of motion patterns. The drawback of the previous approach is the formulation of the noise covariance of the trajectory segments, *i.e.*, their uncertainty. This covariance is assumed to be isotropic and the same across all the domain, which is rigid and unrealistic. Moreover, it is incorrect to estimate covariances in regions of the domain where no data is available. These limitations are addressed in this paper, where we propose a framework to estimate space-varying covariance matrices that also depend on the type of motion (see Fig. 1). Covariance matrices will only be estimated in regions where observations are available and set to a high value in all of the others. Models based on GP also allow the estimation of space-varying covariances. However, GP do not consider switching between models and the covariances are a combination of the trajectory noise and model uncertainties [6, 3].

Fig. 1: Pedestrian trajectory and corresponding space-varying noise covariance.

To sum-up, the main contributions of this paper are the probabilistic reformulation of [20] to include space-varying noise covariances, as well as the description of an optimization strategy to estimate them. The remaining of the paper is organized as follows. Section 2 describes the trajectory model and Section 3 explains the formulation of the space-varying covariance matrices. Section 4 addresses the estimation of the covariance matrices and the optimization method. Section 5 presents the experimental results and Section 6 concludes the paper.

2. TRAJECTORY MODEL

Different types of trajectories can be observed in an outdoor scene. These can be aggregated into different motion regimes, where each regime is represented by a motion field, as suggested in [20]. Let $T_k : [0,1]^2 \rightarrow \mathbb{R}^2$, be the $k-th$ motion field, where $[0,1]^2$ denotes

This work was supported by the FCT project and plurianual funding: [PTDC/EEIPRO/0426/2014], [UID/EEA/50009/2013].

the image plane and $k \in \{1, \ldots, K\}$ is the motion regime (label). We assume that only one motion field drives the pedestrian trajectory at each instant of time, the so-called *active field*, but switching between different motion fields is allowed, during the pedestrian evolution in the scene.

Let $(x_1, \ldots, x_L)$, with $x_t \in [0,1]^2$, be the trajectory of a pedestrian in the image plane and $(k_1, \ldots, k_L)$, with $k_t \in \{1, \ldots, K\}$, be the sequence of active fields. We will assume that x_t is generated according to the dynamical equation

$$x_t = x_{t-1} + T_{k_t}(x_{t-1}) + w_t, \tag{1}$$

where $T_{k_t}(x_{t-1})$ is the active motion field driving the pedestrian at time t, and w_t is a white noise perturbation with Gaussian distribution $w_t \sim N(0, \sigma^2_{k_t} I)$. We also assume that switching between motion fields is described by a first order Markov chain with space-varying transition probabilities

$$P(k_t = j | k_{t-1} = i, x_{t-1}) = b_{ij}(x_{t-1}), \tag{2}$$

where $b_{ij}(x)$ is the element ij of the transition matrix $B(x)$ (stochastic matrix), evaluated at position x.

The space-varying motion fields, T_k, and transition matrix, B, are defined using a regular grid of $\sqrt{n} \times \sqrt{n}$ nodes. The motion fields and transition matrix are estimated at the grid nodes, and defined at the other points by bilinear interpolation.

To be specific, the $k-th$ motion field is obtained by multiplying a dictionary of node velocities, $\mathcal{T}_k \in \mathbb{R}^{2\times n}$, by the vector of interpolation coefficients $\Phi(x) \in \mathbb{R}^{n\times 1}$, *i.e.*,

$$T_k(x) = \mathcal{T}_k \Phi(x)\ . \tag{3}$$

The space-varying transition matrix, $B(x)$, is represented in a similar way using the same regular grid.

3. TRAJECTORY UNCERTAINTY

The previous model, adopts flexible descriptions for the motion fields and transition matrix. However, it uses a rigid formulation for the noise covariance matrix, $\text{cov}\{w_t\} = \mathbf{\Sigma}_{k_t} \in \mathbb{R}^{2\times 2}$. This matrix is assumed to be isotropic and the same across all the pedestrian positions, *i.e.*, $\mathbf{\Sigma}_k = \sigma_k^2 I$. These assumptions are unrealistic, since: i) pedestrian motion is conditioned by the geometry of the scene; and ii) there is no guarantee that the covariance of the noise is the same in all the directions.

Therefore, in this work we postulate that the noise covariance depends not only on type of motion, but also on the pedestrian position, *i.e.*, $\text{cov}\{w_t\} = \mathbf{\Sigma}_{k_t}(x) \in \mathbb{R}^{2\times 2}$. The noise covariance associated with the $k-th$ motion field and position x is obtained by interpolating the covariance matrices associated to a regular grid of nodes, similar to the one used in previous section

$$\mathbf{\Sigma}_k(x) = \sum_{i=1}^{n} \mathbf{\Sigma}_k^i \phi^i(x), \tag{4}$$

where $\mathbf{\Sigma}_k^i$ and $\phi^i(x)$ are, respectively, the covariance matrix and the interpolation coefficient of the $i-th$ node. Since $0 \le \phi^i(x) \le 1, \forall i$, the interpolated matrix is guaranteed to be semidefinite positive.

4. SPACE-VARYING COVARIANCE ESTIMATION

We wish to estimate the model parameters $\theta = (\mathcal{T}, \mathcal{B}, \mathbf{\Sigma})$ where $\mathcal{T}$ is the dictionary of the motion vectors, $\mathcal{B}$ is the dictionary of the transition matrices (stochastic) and $\mathbf{\Sigma}$ is the dictionary of the noise covariance matrices associated to the nodes. These parameters should be retrieved from pedestrian trajectories detected in the video signal.

Let $\mathcal{X} = \left\{x^{(1)}, ..., x^{(S)}\right\}$ be a set of S observed trajectories. The MAP estimate of θ is obtained by

$$\hat{\theta} = \arg\max_{\theta} \left[\log p(\mathcal{X}|\theta) + \log p(\theta)\right]. \tag{5}$$

Since the model depends on hidden parameters (label sequences, $k^{(s)}, s = 1, \ldots, S$), the likelihood function cannot be easily obtained. Therefore, we will adopt the EM algorithm [20] to estimate θ. This amounts to maximizing the auxiliary function

$$\begin{aligned} U(\theta, \theta') &= E\left\{\log p(\mathcal{X}, \mathcal{K}|\theta) | \mathcal{X}, \theta'\right\} + \log p(\theta) \\ &= U_1(\theta, \theta') + U_2(\theta, \theta') + U_3(\theta, \theta'), \end{aligned} \tag{6}$$

where

$$U_1(\theta, \theta') = -\frac{1}{2} \sum_{s=1}^{S} \sum_{t=2}^{L_s} \sum_{k=1}^{K} w_k^{(s)}(t) \log\left(\det\left(\sum_{i=1}^{n} \mathbf{\Sigma}_k^i \phi^i(x_{t-1}^{(s)})\right)\right),$$

$$U_2(\theta, \theta') = -\frac{1}{2} \sum_{s=1}^{S} \sum_{t=2}^{L_s} \sum_{k=1}^{K} w_k^{(s)}(t) \|x_t^{(s)} - x_{t-1}^{(s)} - T_k(x_{t-1}^{(s)}\|^2_{\mathbf{\Sigma}_k(x_{t-1}^{(s)})},$$

$$U_3(\theta, \theta') = \sum_{s=1}^{S} \sum_{t=2}^{L_s} \sum_{p,q=1}^{K} w_{p,q}^{(s)}(t) \log B_{pq}(x_{t-1}^{(s)}). \tag{7}$$

where $w_p^{(s)}(t) = P(k_t^{(s)} = p | x^{(s)}, \theta')$ and $w_{p,q}^{(s)}(t) = P(k_{t-1}^{(s)} = p, k_t^{(s)} = q || x^{(s)}, \theta')$. These weights are obtained in the E step.

The optimization with respect to $\mathcal{T}, \mathcal{B}$ is described in [20, 19, 4]. The optimization with respect to $\mathbf{\Sigma}$ is new and requires the computation of the derivatives $\frac{\partial(U)}{\partial \mathbf{\Sigma}_\eta^\gamma}$, for each node γ and type of motion η. After cumbersome calculations, we obtain

$$\begin{aligned} \frac{\partial U_1}{\partial \mathbf{\Sigma}_\eta^\gamma} &= -\frac{1}{2} \frac{\partial}{\partial \mathbf{\Sigma}_\eta^\gamma} \sum_{s=1}^{S} \sum_{t=1}^{L_s} \sum_{k=1}^{K} w_k^{(s)}(t) \log\left(\det\left(\sum_{i=1}^{n} \mathbf{\Sigma}_k^i \phi^i\left(x_{t-1}^{(s)}\right)\right)\right) \\ &= -\frac{1}{2} \sum_{s=1}^{S} \sum_{t=1}^{L_s} w_\eta^{(s)}(t) \frac{\partial}{\partial \mathbf{\Sigma}_\eta^\gamma} \log\left(\det\left(\sum_{i=1}^{n} \mathbf{\Sigma}_\eta^i \phi^i\left(x_{t-1}^{(s)}\right)\right)\right) \\ &= -\frac{1}{2} \sum_{s=1}^{S} \sum_{t=1}^{L_s} w_\eta^{(s)}(t) \phi^\gamma(x_{t-1}^{(s)}) \left(\sum_{i=1}^{n} \mathbf{\Sigma}_\eta^i \phi^i(x_{t-1}^{(s)})\right)^{-1}, \end{aligned}$$

$$\begin{aligned} \frac{\partial U_2}{\partial \mathbf{\Sigma}_\eta^\gamma} &= -\frac{1}{2} \frac{\partial}{\partial \mathbf{\Sigma}_\eta^\gamma} \sum_{s=1}^{S} \sum_{t=1}^{L_s} \sum_{k=1}^{K} w_k^{(s)}(t) \|x_t^{(s)} - x_{t-1}^{(s)} - T_k(x_{t-1}^{(s)})\|^2_{\mathbf{\Sigma}_k(x_{t-1}^{(s)})} \\ &= -\frac{1}{2} \frac{\partial}{\partial \mathbf{\Sigma}_\alpha^\gamma} \sum_{s=1}^{S} \sum_{t=1}^{L_s} \sum_{k=1}^{K} w_k^{(s)}(t) \text{tr}\left(\left(\sum_{i=1}^{n} \mathbf{\Sigma}_k^i \phi^i(x_{t-1}^{(s)})\right)^{-1} \mathbf{G}_t^{(s)}\right) \\ &= -\frac{1}{2} \sum_{s=1}^{S} \sum_{t=1}^{L_s} w_\eta^{(s)}(t) \frac{\partial}{\partial \mathbf{\Sigma}_\eta^\gamma} \text{tr}\left(\left(\sum_{i=1}^{n} \mathbf{\Sigma}_\eta^i \phi^i(x_{t-1}^{s})\right)^{-1} \mathbf{G}_t^{(s)}\right) \\ &= -\frac{1}{2} \sum_{s=1}^{S} \sum_{t=1}^{L_s} w_\eta^{(s)}(t) \phi^\gamma(x_{t-1}^{(s)}). \\ &\cdot \left(\left(\sum_{i=1}^{n} \mathbf{\Sigma}_\eta^i \phi^i(x_{t-1}^{(s)})\right)^{-1} \mathbf{G}_t^{(s)} \left(\sum_{i=1}^{n} \mathbf{\Sigma}_\eta^i \phi^i(x_{t-1}^{(s)})\right)^{-1}\right)^T, \\ \frac{\partial U_3}{\partial \mathbf{\Sigma}_\alpha^\gamma} &= 0, \end{aligned} \tag{8}$$

where

$$\mathbf{G}_t^{(s)} = (x_t^{(s)} - x_{t-1}^{(s)} - T_\eta(x_{t-1}^{(s)}))(x_t^{(s)} - x_{t-1}^{(s)} - T_\eta(x_{t-1}^{(s)}))^T. \tag{9}$$

The stationary conditions for node γ and motion regime η

$$\frac{\partial U}{\partial \mathbf{\Sigma}_\eta^\gamma} = 0\ , \tag{10}$$

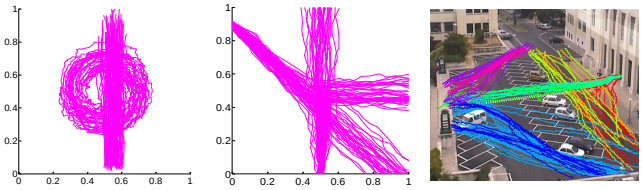

Fig. 2: Experimental dataset: syntethic case 1 (left), synthetic case 2 (mid), and Campus dataset (right).

cannot be analytically solved. Therefore, we will resort to numerical optimization methods (gradient algorithm)

$$\boldsymbol{\Sigma}_\eta^\gamma \leftarrow \boldsymbol{\Sigma}_\eta^\gamma + \mu \frac{\partial U}{\partial \boldsymbol{\Sigma}_\eta^\gamma}, \tag{11}$$

where μ is the step. This recursion should be applied for each grid node of each motion regime until a stop condition is met. The choice of the step η is done using the Barzilai-Borwein (BB) method [5].

The covariance estimates should be symmetric and positive definite matrices. The first condition is automatically guaranteed by the gradient algorithm but the second is not. To overcome this difficulty, we project the covariance estimates into the set of positive definite matrices in each iteration of the algorithm. This is accomplished by performing eigendecomposition of the covariance estimates and replacing negative and null eigenvalues by small positive ones. We stress that null singular values are not acceptable since they lead to instabilities in the computation of U.

5. EXPERIMENTAL RESULTS

We evaluate the proposed methodology in three datasets: i) synthetic trajectories generated using a linear or circular motion field, without switching (case 1); ii) synthetic trajectories generated using motion fields with switching probabilities (case 2); and iii) real trajectories from the Campus data set [20]. The experiments with the synthetic data allow us to compare the estimates of the model parameters $\theta = (\mathcal{T}, \mathcal{B}, \boldsymbol{\Sigma})$ against a known ground truth, while the experiment with the Campus dataset gives us a qualitative evaluation of the proposed method in the real world. In both of the synthetic experiments, three subsets of equal size are generated, each using a different configuration for $\boldsymbol{\Sigma}$: i) isotropic and equal for each node of a motion field; ii) anisotropic in the direction of the motion and equal for each node of a motion field; and iii) anisotropic in the direction of the motion and space varying (random). All of the noise covariances are in the range $\left[1 \times 10^{-6}, 1 \times 10^{-3}\right]$.

The following metrics are used to evaluate the estimated model parameters:

$$e_{\mathcal{T}_k} = \frac{1}{\#\Omega} \sum_{i \in \Omega} \|T_k^i - T_k^{o,i}\|_2^2, \tag{12}$$

$$e_{B^i} = \|\text{vec}(B^i) - \text{vec}(B^{o,i})\|_2^2, \tag{13}$$

$$e_{\boldsymbol{\Sigma}_k} = \frac{1}{\#\Omega} \sum_{i \in \Omega} \sum_{s=1}^{S} \sum_{t=2}^{L_s} w_k^s(t) \phi^i(x_{t-1}^s) \|\text{vec}(\boldsymbol{\Sigma}_k^i) - \text{vec}(\boldsymbol{\Sigma}_k^{o,i})\|_2^2, \tag{14}$$

where $T_k^{o,i}$, $B^{o,i}$, and $\boldsymbol{\Sigma}_k^{o,i}$ are the ground truth values of the $i-th$ node and $k-th$ motion type, and Ω is the set of nodes that is supported by, at least, 30% of the observations.

5.1. Synthetic Data Without Switching

In this example, which we will call synthetic case 1, the trajectories are generated according to $K = 2$ motion regimes: counterclockwise and linear, as shown in Fig. 2 (left). Linear trajectories start at a region around point $[0.55, 0.05]^T$, while circular trajectories are centered on point $[0.5, 0.5]^T$ and exhibit varying radius. We define the motion fields in a regular grid of 11×11, being impossible to switch between fields in any of the nodes.

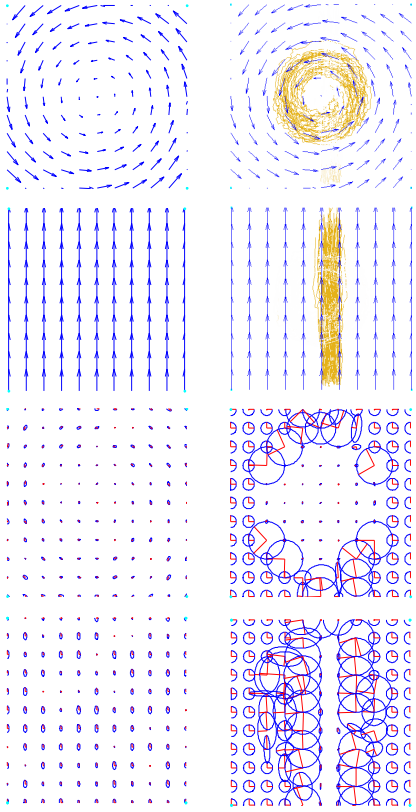

Fig. 3: Ground truth (left) and estimated velocity fields $\mathcal{T}$ and space-varying covariance matrices $\boldsymbol{\Sigma}$ (right) for synthetic case 1. In yellow we show the trajectories associated to each motion regime.

Table 1: Fields and covariances estimation errors for synthetic case 1. "Fixed" identifies the isotropic covariance $\sigma_k^2 I$ and "Space-Varying" identifies the proposed strategy. In **bold** we highlight the best results.

Noise type	Covariance type	$e_{\mathcal{T}_k}$	$e_{\boldsymbol{\Sigma}_k}$
Isotropic	Fixed [20]	$2.80 \times 10^{-3}/2.30 \times 10^{-3}$	$5.19 \times 10^{-5}/5.19 \times 10^{-5}$
	Space-Varying	$\mathbf{2.50 \times 10^{-3}/1.50 \times 10^{-3}}$	$\mathbf{2.63 \times 10^{-5}/1.12 \times 10^{-5}}$
Anisotropic	Fixed [20]	$\mathbf{1.10 \times 10^{-3}}/2.70 \times 10^{-3}$	$2.67 \times 10^{-5}/1.14 \times 10^{-4}$
	Space-Varying	$1.30 \times 10^{-3}/\mathbf{2.60 \times 10^{-3}}$	$\mathbf{9.20 \times 10^{-6}/5.87 \times 10^{-5}}$
Random	Fixed [20]	$\mathbf{1.40 \times 10^{-3}}/1.30 \times 10^{-3}$	$1.89 \times 10^{-5}/8.82 \times 10^{-5}$
	Space-Varying	$1.60 \times 10^{-3}/\mathbf{3.00 \times 10^{-4}}$	$\mathbf{6.20 \times 10^{-6}/3.49 \times 10^{-5}}$

We randomly generate 100 trajectories according to (1), for each type of noise covariances (isotropic, anisotropic, and random). Due to space constraints, we only show the trajectories generated using the random anisotropic covariances (Fig. 2 left) . Noise covariances are shown in Fig. 3 (left). Please note that the displayed covariances are anisotropic and depend on the motion. Each set of trajectories is used to estimate the model parameters $\theta = (\mathcal{T}, \boldsymbol{\Sigma})$ (here $\mathcal{B}$ is not considered), as described in Section 4. Furthermore, for comparison purposes, we also estimate the parameters using the formulation of [20], where $\boldsymbol{\Sigma}_k$ is set to $\sigma_k^2 I$ in all of the grid nodes. The motion fields $\mathcal{T}$ are randomly initialized and the covariance matrices $\boldsymbol{\Sigma}$ are initialized as $\sigma_{init}^2 I$, where $\sigma_{init}^2 = \{10^{-3}, 5 \times 10^{-3}, 10^{-2}\}$.

Table 1 reports the best error values $e_{\mathcal{T}_k}$ and $e_{\boldsymbol{\Sigma}_k}$ for each motion model and $\sigma_{init} = 10^{-3}$, using the proposed method and [20]. Fig. 3 (right) shows the estimated fields and covariances. These results demonstrate that the proposed method is able to successfully estimate the model parameters $\theta = (\mathcal{T}, \boldsymbol{\Sigma})$ in the nodes where tra-

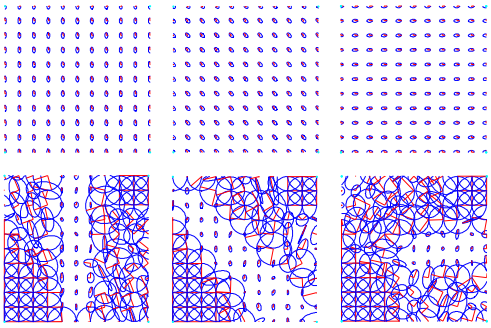

Fig. 4: Ground truth (1st row) and estimated space-varying covariance matrices $\mathbf{\Sigma}$ (2nd row) for synthetic case 2.

jectories are observed, while assigning high covariances to nodes that have few observations. Moreover, the error values prove that the proposed method outperforms [20] not only estimation of the covariance matrices, which was expected, but also in the estimation of the fields.

5.2. Synthetic Data With Switching

The synthetic case 2 comprise $K = 3$ motion regimes, as exemplified in Fig. 2 (mid). These trajectories are generated as follows. First, we randomly choose the starting point between a small region around $[0.5, 0]^T$ and $[0, 0.9]^T$. In the first case, the motion regime is an up motion from the bottom to the top of the image, while in the second case it is a down motion with approximately 45° of orientation. The trajectories may switch between these motions and a third motion regime, orientated from left to right, at the center of the image. This switching is governed by the following transition matrix $B^c = \begin{bmatrix} 0.7 & 0.1 & 0.2 \\ 0.2 & 0.2 & 0.6 \\ 0 & 0 & 1 \end{bmatrix}$. As before, we define the model in a regular grid of 11×11 nodes.

We conduct the experiments described in the previous section, using three datasets of 100 randomly generated trajectories with varying types of noise covariance. Here we are also interested in estimating $\mathcal{B}$, as well as in evaluating the error e_{B^i} in the transition node. The experimental results are summarized in Table 2 and the estimated covariances for the anisotropic noise are shown in Fig. 4. Once more, the proposed method is able to successfully estimate the model parameters and outperform the previous approach is most of the results. The estimated switching matrix for the example is $\hat{B}^c = \begin{bmatrix} 0.8 & 0.05 & 0.15 \\ 0.23 & 0.17 & 0.6 \\ 0 & 0 & 1 \end{bmatrix}$, which is similar to the real one.

Table 2: Fields and covariances estimation errors for synthetic case 2. "Fixed" identifies the isotropic covariance $\sigma_k^2 I$ and "Space-Varying" identifies the proposed strategy. In **bold** we highlight the best results.

Noise type	Covariance type	$e_{\mathcal{T}_k}$	$e_{\mathbf{\Sigma}_k}$	e_{B^i}
Isotropic	Fixed [20]	$\mathbf{3.20 \times 10^{-3}}$	5.28×10^{-5}	0.61
		$\mathbf{1.50 \times 10^{-3}}$	4.59×10^{-5}	
		$\mathbf{1.20 \times 10^{-3}}$	5.77×10^{-5}	
	Space-Varying	3.40×10^{-3}	$\mathbf{2.44 \times 10^{-5}}$	**0.31**
		1.90×10^{-3}	$\mathbf{2.24 \times 10^{-5}}$	
		1.40×10^{-3}	$\mathbf{3.27 \times 10^{-5}}$	
Anisotropic	Fixed [20]	5.10×10^{-3}	1.28×10^{-4}	0.72
		8.00×10^{-4}	1.93×10^{-4}	
		3.20×10^{-3}	$\mathbf{1.43 \times 10^{-4}}$	
	Space-Varying	$\mathbf{4.30 \times 10^{-3}}$	$\mathbf{3.80 \times 10^{-5}}$	**0.19**
		$\mathbf{4.00 \times 10^{-4}}$	$\mathbf{1.33 \times 10^{-4}}$	
		2.70×10^{-3}	1.57×10^{-4}	
Random	Fixed [20]	2.20×10^{-3}	8.61×10^{-5}	0.75
		$\mathbf{1.00 \times 10^{-3}}$	1.33×10^{-4}	
		2.50×10^{-3}	1.12×10^{-4}	
	Space-Varying	$\mathbf{1.10 \times 10^{-3}}$	$\mathbf{4.00 \times 10^{-5}}$	**0.32**
		1.10×10^{-3}	$\mathbf{1.20 \times 10^{-4}}$	
		$\mathbf{1.80 \times 10^{-3}}$	$\mathbf{1.03 \times 10^{-4}}$	

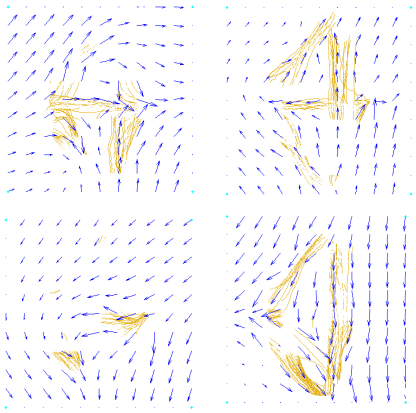

Fig. 5: Estimated velocity fields $\mathcal{T}$ for the Campus dataset. In yellow we show the trajectories associated to each motion regime.

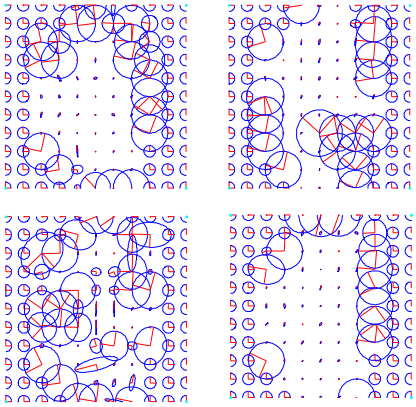

Fig. 6: Estimated node covariances $\mathbf{\Sigma}$ for the Campus dataset.

5.3. Real Data

This experiment was carried on the Campus data set [20], which comprises 134 pedestrian trajectories acquired by a static far field camera at the IST university campus in Lisbon (see Fig. 2, right). An homography was performed to correct the distortion cause by the perspective projection, and the transformed trajectories were used to estimate the model parameters $\theta = (\mathcal{T}, \mathcal{B}, \mathbf{\Sigma})$.

The number of motion regimes was set to $K = 4$ (north, south, east, and west) and the regular grid was defined to be 11×11. Similarly to the previous experiments, the fields were randomly initialized, while the covariance matrices were set to be $\mathbf{\Sigma}_k = \sigma_{init}^2 I$. Since we do not know the ground truth of any of the parameters, we are only able to perform a qualitative assessment of the estimated fields $\mathcal{T}$ and space varying noise covariances $\mathbf{\Sigma}$. Figures 5 and 6 respectively display each of these estimates.

6. CONCLUSIONS

This paper proposes a new methodology for the estimation of the uncertainty in pedestrian trajectory generation, applied to models that are based on switching motion fields. It is assumed that the covariance of the trajectory noise depends both on the type of motion and pedestrian position, *i.e.*, it is considered to be space-varying. The approach used to estimate the space-varying covariances is introduced and several experiments are carried out to evaluated the method. Synthetic and real datasets have been used to conduct the experiments, and it was demonstrated that our assumption is reasonable and that it outperforms the previous method based on global and isotropic noise covariances.

7. REFERENCES

[1] A. Alahi, V. Ramanathan, K. Goel, A. Robicquet, A. A. Sadeghian, L. Fei-Fei, and S. Savarese, "Learning to predict human behaviour in crowded scenes," *Group and Crowd Behavior for Computer Vision*, pp. 183–207, 2017.

[2] P. Baiget, E. Sommerlade, I. Reid, and J. Gonzalez, "Finding prototypes to estimate trajectory development in outdoor scenarios," in *Proceedings of the 1st THEMIS Workshop*, 2008, pp. 27–34.

[3] M. Barão and J. S. Marques, "Gaussian random vector fields in trajectory modelling," in *IMVIP*, 2017.

[4] C. Barata, J. C. Nascimento, and J. S. Marques, "A sparse approach to pedestrian trajectory modeling using multiple motion fields," in *ICIP*, 2017.

[5] J. Barzilai and J. M. Borwein, "Two-point step size gradient methods," *IMA journal of numerical analysis*, vol. 8, no. 1, pp. 141–148, 1988.

[6] V. Bastani, L. Marcenaro, and C. S. Regazzoni, "Online nonparametric bayesian activity mining and analysis from surveillance video," *IEEE Transactions on Image Processing*, vol. 25, no. 5, pp. 2089–2102, 2016.

[7] D. Ellis, E. Sommerlade, and I. Reid, "Modelling pedestrian trajectory patterns with gaussian processes," in *ICCV Workshops*, 2009, pp. 1229–1234.

[8] T. Fernando, S. Denman, S. Sridharan, and C. Fookes, "Soft+ hardwired attention: An lstm framework for human trajectory prediction and abnormal event detection," *arXiv preprint arXiv:1702.05552*, 2017.

[9] N. Ferreira, J. T. Klosowski, C. E. Scheidegger, and C. T. Silva, "Vector field k-means: Clustering trajectories by fitting multiple vector fields," in *Computer Graphics Forum*, vol. 32, no. 3pt2, 2013, pp. 201–210.

[10] D. Helbing and P. Molnar, "Social force model for pedestrian dynamics," *Physical review E*, vol. 51, p. 42824286, 1995.

[11] K. Kim, D. Lee, and I. Essa, "Gaussian process regression flow for analysis of motion trajectories," in *ICCV*, 2011, pp. 1164–1171.

[12] K. M. Kitani, B. D. Ziebart, J. A. Bagnell, and M. Hebert, "Activity forecasting," in *European Conference on Computer Vision*, 2012, pp. 201–214.

[13] L. Leal-Taixé, M. Fenzi, A. Kuznetsova, B. Rosenhahn, and S. Savarese, "Learning an image-based motion context for multiple people tracking," in *CVPR*, 2014, pp. 3542–3549.

[14] N. Lee, W. Choi, P. Vernaza, C. B. Choy, P. H. S. Torr, and M. Chandraker, "Desire: Distant future prediction in dynamic scenes with interacting agents," *arXiv preprint arXiv:1704.04394*, 2017.

[15] E. Maggio and A. Cavallaro, *Video tracking: theory and practice*. John Wiley & Sons, 2011.

[16] D. Makris and T. Ellis, "Path detection in video surveillance," *Image and Vision Computing*, vol. 20, no. 12, pp. 895–903, 2002.

[17] B. T. Morris and M. M. Trivedi, "A survey of vision-based trajectory learning and analysis for surveillance," *IEEE transactions on circuits and systems for video technology*, vol. 18, no. 8, pp. 1114–1127, 2008.

[18] ——, "Trajectory learning for activity understanding: Unsupervised, multilevel, and long-term adaptive approach," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 11, pp. 2287–2301, 2011.

[19] J. C. Nascimento, M. Barão, J. S. Marques, and J. M. Lemos, "Information geometric algorithm for estimating switching probabilities in space-varying hmm," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5263–5273, 2014.

[20] J. C. Nascimento, M. A. T. Figueiredo, and J. S. Marques, "Activity recognition using a mixture of vector fields," *IEEE Transactions on Image Processing*, vol. 22, no. 5, pp. 1712–1725, 2013.