MULTIPLE MOTION FIELDS FOR MULTIPLE TYPES OF AGENTS

Catarina Barata *Mário A. T. Figueiredo* *Jorge S. Marques*

Instituto de Sistemas e Robótica, Instituto de Telecomunicações,
Instituto Superior Técnico, Universidade de Lisboa, Portugal$^{(a)}$

ABSTRACT

Complex surveillance scenarios comprise different types of agents (*e.g.*, bikers, cars, and pedestrians) that must be efficiently characterized in order to facilitate tasks such as tracking or abnormality detection. This paper proposes an unsupervised hierarchical multiple motion fields model to represent different types of agents, which relies in the combination of hierarchical Markov model and velocity fields. Model parameters are estimated using the expectation-maximization algorithm. The proposed framework was applied to synthetic and real datasets (*Stanford Drone Dataset*), showing the ability to characterize and classify different agents in an unsupervised way.

***Index Terms*—** Surveillance, trajectory analysis, motion fields, hierarchical Markov models

1. INTRODUCTION

The development of motion models that provide reliable descriptions of object trajectories plays an important role in several applications, such as surveillance, robot navigation, and person re-identification [1, 2, 3]. However, most motion models assume that there is only one type of agent in the video, usually pedestrians, and treat other agents (*e.g.*, bikers or cars) as abnormalities [4, 5, 6]. Although this may be valid in indoor scenes, where vehicles are not expected, several applications require the analysis of outdoor videos, where multiple types of agents (*e.g.*, pedestrians, cars, bikers) interact with the environment in different ways and exhibit distinctive motions. Moreover, each type of agent is associated with specific abnormal behaviors. Thus, it is important to develop methods that consider the possibility of multiple agent classes. This task may be solved in a supervised fashion, where we have access to both the trajectories and their respective classes, as reported in [7]. However, fully annotated datasets (see Fig. 1) are difficult to obtain. Hence, it is important to develop a motion model that is able to represent and group the motion patterns of different agent classes in an unsupervised fashion, thus suitable to deal with any kind of data.

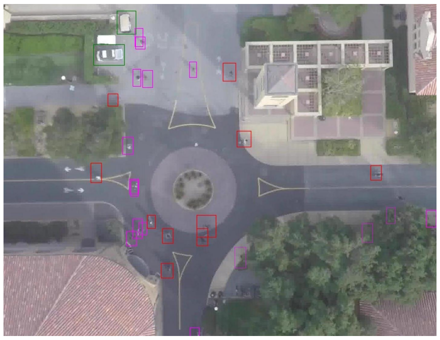

Fig. 1. Image from the *Stanford Drone Dataset* [10]. The bounding box colors identify different agents.

In this work, we propose a framework based on motion fields and a hierarchical hidden-Markov model (HHMM) to group and characterize motion patterns of different agent classes, in an unsupervised way. Recently, the interactions between different types of agents were also modeled in an unsupervised framework, using both the social forces model and recurrent neural networks [8, 9]. However, the goal of those methods was to perform trajectory prediction, while our goal is to identify and characterize class-specific motion patterns, which can be incorporated into a generative motion model for different applications (*e.g.*, tracking, prediction, abnormality detection).

2. RELATED WORK AND EXTENSION TO MULTI-AGENT

Motion fields have been used for trajectory analysis in many applications, such as the characterization of trajectory data from multiple sources, namely GPS positions of pedestrians and vehicles, call detail records of cellphones, and hurricane data [11].

Our work is related with [12], where it was assumed that pedestrian trajectories could be divided into segments, each of them generated by one motion field. The transition between motion fields was possible and defined as a first-order hidden Markov (HMM) model, where the transition probabilities vary across space. The aforementioned method was proposed to deal with pedestrians, and its extension to multiple classes would require either: *i)* separate estimation of motion mod-

This work was supported by the FCT project [PTDC/EEIPRO/0426/2014] and supported by FCT plurianual funding [UID/EEA/50009/2019].

els for each of the agent classes; or *ii)* estimation of a large number of motion fields, as well as the setting of constraints on the transition matrices, to avoid transitions between fields associated with different classes. The former approach would demand a fully annotated dataset, which is not available in most cases, while the latter would require a proper initialization of the model parameters in order to enforce the desired constraints.

In this paper, we address the aforementioned limitation, and propose a new probabilistic formulation for the motion model. In particular, we propose to replace the HMM by a HHMM, which will allow us to condition both the motion fields and the transition matrices on a specific agent class, as described in Section 3. We will show that this new probabilistic formulation allows the grouping of motion patterns of a specific class, without the need for a fully annotated dataset, and still provides reliable representations.

3. HIERARCHICAL SWITCHED MOTION MODEL

Let us assume that the various agents in a scene (*e.g.*, skaters, pedestrians) exhibit a finite number of motion patterns, which are specific of their class $c \in \{1, ..., C\}$. Each agent will be associated with a trajectory $x = (x_1, x_2, ..., x_L)$, where L is the length of the trajectory and $x_t \in [0,1]^2$ is the position at time instant t. The motion patterns that characterize the trajectories may be summarized into a set of K^c motion fields, where $T^c_k : [0,1]^2 \rightarrow \mathbb{R}^2$ is the $k-th$ motion field belonging to class c. Thus, we generate the position x_t as follows

$$x_t = x_{t-1} + T^{c_t}_{k_t}(x_{t-1}) + w_{k_t}, \tag{1}$$

where $T^{c_t}_{k_t}$ is the active motion field, conditioned on class c_t, and $w_{k_t} \sim \mathcal{N}(0, \Sigma^{c_t}_{k_t}(x_{t-1}))$ is the class-specific space-varying white Gausian noise perturbation, associated with the uncertainty of the position.

Only one motion field may be active at each time instant. However, we assume that it is possible to switch between motion fields of the same class at specific positions. Additionally, we postulate that is also possible for an agent to change class, although with a lower probability, at certain positions of the space (*e.g.*, a skater picks up the skate and starts walking). These transitions are modeled as a HHMM, as explained in the following sub-section.

3.1. Hierarchical Motion Model

HHMM have been introduced by Fine et al. [13] as an extension of the standard HMM to problems that exhibit a hierarchical structure. The main idea of this model is that the hidden states are organized in hierarchical levels, such that the hidden states at the uppers levels, called "internal" states, are responsible for activating states at the lower levels. Each internal state is only able to activate some of the states of the level below it, and these lower states are not shared. The process of activation is carried out until a state at the lowest level is reached. This level, usually called "production", is responsible for generating the observations, similarly to a traditional HMM [14].

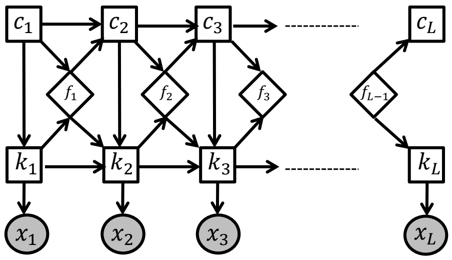

Fig. 2. Graphical representation of the proposed switched model.

Our model can be defined as a two-level HHMM (see Fig.2), where the upper level models the class and the lower level the active motion field. A binary variable f_t is used to identify the end of the production level; this allows the model to decide whether to stay in the production level ($f_t = 0$), or to leave the production level and return the control to the upper level ($f_t = 1$). This makes it possible for the model to generate the next position by either: i) maintaining the same class; or ii) changing class.

Based on the aforementioned formulation, we define the following probabilities for our motion model [15]:

$$p(k_t = j | k_{t-1} = l, f_{t-1}, c_t = u, x_{t-1}) = \begin{cases} \tilde{B}^u_{lj}(x_{t-1}), & \text{if } f_{t-1} = 0 \\ \pi^u_j(x_{t-1}), & \text{if } f_{t-1} = 1 \end{cases}, \tag{2}$$

where $\tilde{B}^u_{lj}(x)$ is the element (l, j) of the stochastic matrix $\tilde{B}^u(x)$ associated with class u, and $\pi^u_j(x)$ is the initial distribution of motion field j, given the class u. Both variables are evaluated at position x. Similarly to the traditional HMM, $\tilde{B}^u(x)$ comprises the probabilities of transition between states l and j. However, it is also necessary to account for the ending probabilities, *i.e.*, the probability of transition to $f_t = 1$, which we will loosely refer to *end*

$$p(f_t = 1 | k_t = j, c_t = u, x_{t-1}) = B^u_{jend}(x_{t-1}). \tag{3}$$

Thus, we define $B^u_{lj} = (1 - B^u_{lend}(x_{t-1}))\tilde{B}^u_{lj}(x_{t-1})$ as a rescaled version of $\tilde{B}^u_{lj}(x_{t-1})$. At the upper level, the transition between classes is also governed by a stochastic matrix $A(x)$ evaluated at position x, such that

$$p(c_t = u | c_{t-1} = v, f_{t-1}, x_{t-1}) = \begin{cases} \delta(v, u), & \text{if } f_{t-1} = 0 \\ A_{vu}(x_{t-1}), & \text{if } f_{t-1} = 1 \end{cases}. \tag{4}$$

Here $\delta(v, u)$ is the Kronecker delta and $A_{vu}(x)$ is the (v, u) element of $A(x)$. Based on this formulation, the joint probability $p(x, k, f, c)$ of a trajectory x associated with a sequence of motion fields k, classes c, and binary variables f, is defined as follows:

$$
\begin{aligned}
p(x,k,f,c) &= p(x_1,k_1,f_1,c_1)\prod_{t=2}^{L} p(x_t,k_t,f_t,c_t|x_{t-1},k_{t-1},f_{t-1},c_{t-1}) \\
&= p(x_1,k_1,f_1,c_1)\prod_{t=2}^{L} p(x_t|x_{t-1},k_t,c_t)p(c_t|x_{t-1},c_{t-1},f_{t-1}) \\
&.p(k_t|x_{t-1},c_t,f_{t-1},k_{t-1})p(f_t|k_t,c_t,x_{t-1}). \qquad (5)
\end{aligned}
$$

3.2. Model Estimation

All the model parameters $\theta = (\mathcal{T},\mathcal{B},\mathcal{A},\Pi,\mathbf{\Sigma})$ are defined on a regular grid of $\sqrt{n} \times \sqrt{n}$ nodes, where $\mathcal{T}$ is a dictionary of motion fields, $\mathcal{B}$ and $\mathcal{A}$ are dictionaries of fields and classes transition matrices, Π is the dictionary of motion fields probabilities, and $\mathbf{\Sigma}$ is a dictionary of covariance matrices. The parameters are estimated at the grid nodes and computed elsewhere using bilinear interpolation [16]. These values may be estimated using a set of S observed trajectories $\mathcal{X} = \left\{x^{(1)},...,x^{(S)}\right\}$, with variable lengths:

$$
\hat{\theta} = \arg\max_{\theta} \left[\log p(\mathcal{X}|\theta) + \log p(\theta)\right]. \qquad (6)
$$

Since there are several hidden variables in the model (the sequences $k^{(s)}$, $c^{(s)}$, and $f^{(s)}$), for $s = 1,...,S$, we will naturally resort to the expectation-maximization (EM) algorithm (see details in [17]). The auxiliary function to be maximized w.r.t. θ, given the previous estimate θ', is

$$
\begin{aligned}
U(\theta,\theta') &= E\left\{\log p(\mathcal{X},\mathcal{K}|\theta)|\mathcal{X},\theta'\right\} + \log p(\theta) \\
&= U_1(\theta,\theta') + U_2(\theta,\theta') + U_3(\theta,\theta') + U_4(\theta,\theta') + U_5(\theta,\theta'), \qquad (7)
\end{aligned}
$$

where

$$
U_1(\theta,\theta') = \sum_{s=1}^{S}\sum_{t=2}^{L_s}\sum_{c=1}^{C}\sum_{k=1}^{K^c}\sum_{f=0}^{1} \gamma_{ckf}^{(s)}(t)\log\det\left(\mathbf{\Sigma}_k^c(x_{t-1}^{(s)})\right),
$$

$$
U_2(\theta,\theta') = \sum_{s=1}^{S}\sum_{t=2}^{L_s}\sum_{c=1}^{C}\sum_{k=1}^{K^c}\sum_{f=0}^{1} \gamma_{ckf}^{(s)}(t)\|v_t^{(s)} - T_k^c(x_{t-1}^{(s)})\|^2_{\mathbf{\Sigma}_k^c(x_{t-1}^{(s)})},
$$

$$
\begin{aligned}
U_3(\theta,\theta') &= 2\sum_{s=1}^{S}\sum_{t=2}^{L_s}\sum_{c=1}^{C}\sum_{k=1}^{K^c} \xi_{ck1}^{(s)}(t)\log B_{kend}^c(x_{t-1}^{(s)}) \\
&+ \xi_{ck0}^{(s)}(t)\log(1 - B_{kend}^c(x_{t-1}^{(s)})),
\end{aligned}
$$

$$
\begin{aligned}
U_4(\theta,\theta') &= 2\sum_{s=1}^{S}\sum_{t=2}^{L_s}\sum_{c=1}^{C}\left[\sum_{k=1}^{K^c} \eta_{ck}^{(s)}(t)\log \pi_k^c(x_{t-1}^{(s)})\right] \\
&+ \left[\sum_{p,q}^{K^c} w_{cpq}^{(s)}(t)\log \tilde{B}_{pq}^c(x_{t-1}^{(s)})\right],
\end{aligned}
$$

$$
U_5(\theta,\theta') = 2\sum_{s=1}^{S}\sum_{t=2}^{L_s}\sum_{j,c}^{C} \chi_{jc}^{(s)}(t)\log A_{jc}(x_{t-1}^{(s)}),
$$

where, $v_t^{(s)} = x_t^{(s)} - x_{t-1}^{(s)}$, $\gamma_{ujf}^{(s)}(t) = p(k_t^{(s)} = j, c_t^{(s)} = u, f_{t-1}^{(s)}|x^{(s)},\theta')$ is the smooth state probability, $\xi_{ujf}^{(s)}(t) = p(k_t^{(s)} = j, c_t^{(s)} = u, f_t^{(s)}|x^{(s)},\theta')$ gives us the ending and non-ending probabilities, $\eta_{uj}^{(s)}(t) = p(k_t^{(s)} = j, c_t^{(s)} = u, f_{t-1}^{(s)} = 1|x^{(s)},\theta')$ is the probability of a vertical transition from the class level to the motion model one, $w_{uij}^{(s)}(t) = p(k_{t-1}^{(s)} = i, k_t^{(s)} = j, c_t^{(s)} = u, f_{t-1}^{(s)} = 0|x^{(s)},\theta')$ is the probability of an horizontal transition at the motion model level, and $\chi_{vu}^{(s)}(t) = p(c_{t-1}^{(s)} = v, c_t^{(s)} = u, f_{t-1}^{(s)} = 1|x^{(s)},\theta')$ is the probability of an horizontal transition at the class level. All of these values are computed in the E-step, using the generalized Baum-Welch algorithm [13].

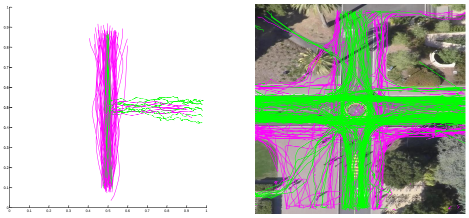

Fig. 3. Datasets: Synthetic (left) and *Stanford Drone Dataset - Little* (right). The colors represent the two classes.

The M-step consists in maximizing $U(\theta,\theta')$. While $\mathcal{T}$ and $\mathbf{\Sigma}$ are updated as proposed in [18, 19], the estimation of $\mathcal{B}$, $\mathcal{A}$, and Π pose a new problem. We refer to [17] for details. Unfortunately, these updates do not have an explicit solution, thus we resort to the gradient descent algorithm, followed by a projection on the simplex [20], to ensure that the obtained parameters are probabilities.

4. EXPERIMENTAL RESULTS

We test the proposed approach on: synthetic trajectories and real trajectories from the *Stanford Drone Dataset* [10], which were recorded using a quadcopter with a 4k camera. Each type of dataset comprises trajectories of two classes of agents, as we will detail below. The hierarchical motion models are trained as described in Section 3.2. In both experiments, we have set the grid size to 23 × 23 nodes.

To evaluate the estimated models, we check if trajectories of the same type of agent are grouped together. The ground-truth class labels are available for both synthetic and real datasets, and it is possible to predict the class of each trajectory segment using the smooth state probability as follows

$$
\hat{c}_t^{(s)} = \arg\max_c \sum_{k=1}^{K^c}\sum_f \gamma_{ckf}^{(s)}(t), \qquad (8)
$$

where $\gamma_{ckf}^{(s)}(t)$ is computed using the generalized Baum-Welch algorithm [13].

4.1. Synthetic Data

A synthetic dataset composed of $S = 100$ samples, was created (see Fig. 3-left). We assume $C = 2$ classes of agents, each with $K^c = 2$ possible motion fields. One of the classes

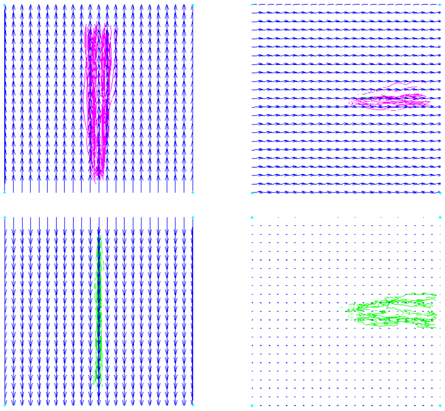

Fig. 4. Estimated motion fields per class (1st and 2nd rows) and associated trajectories for synthetic example. The arrows are colored according to the most probable class and their length is proportional to the velocity module.

exhibits an upward motion pattern, with the possibility of switching to a ”left-to-right” motion (magenta). The second class exhibits a downward motion pattern, also with the possibility of switching to a ”left-to-right” pattern (green). However, in this case, the velocity is $\frac{1}{10}$ of that observed for the first class. Moreover, we define the possibility of switching between classes at a specific node that corresponds to approximately the position $(0.75, 0.5)$ in the image, such that the agent switches from moving from ”left-to-right” according to the velocity of class 2 and starts moving at the velocity of class 1. The corresponding transition matrix is $= \left[\begin{smallmatrix} 1.0 & 0 \\ 0.3 & 0.7 \end{smallmatrix}\right]$.

The estimated fields in Fig. 4 show that the proposed model is able to correctly separate most of the trajectories and motion patterns of a given class. In particular, we observe that 99.5% of the examples from ground-truth class were correctly classified as $c = 1$ (top row), and 99.8% of the elements from class two were classified as $c = 2$ (bottom row). Additionally, we are able to compare the estimated class transition matrix $\hat{A} = \left[\begin{smallmatrix} 0.99 & 0.01 \\ 0.28 & 0.72 \end{smallmatrix}\right]$ with the ground-truth one, and observe that they are very similar.

4.2. Real Data

The experiments with real data were carried out on the *Stanford Drone Dataset* [10], using the videos from the *Little* scenario (see Fig. 3-right). In these videos, the authors have tracked and manually labeled the agents according to several possible classes. Here, we focus on the trajectories of *pedestrians* (#200, magenta) and *bikers* (#385, green), due to the large number of trajectories for each of these classes. The videos were not recorded from the same position, had to be aligned through a spatial transformation.

Three of the *Little* videos (IDs 1,2,3) were used for training the motion model, considering $C = 2$ classes, each with $K^c = 4$ fields. These fields were roughly initialized to represent the directions *North-South*, *South-North*, *East-West*, and *West-East*. During training, the model did not have access to the ground truth labels of the trajectories. The test video (Id 0) was used to evaluate the estimated fields and classes, using a moving window strategy. Consecutive portions of the trajectory $(x_{t_o}, ..., x_{t_{o+\Delta}})$ with an overlap of $\frac{\Delta}{2}$ $(\Delta = 10)$ were analyzed and associated with a class by using (8) to label consecutive time instants. Majority voting gave a label for the whole portion. Finally, these labels were compared with the ground-truth and a confusion matrix was built (Table 1).

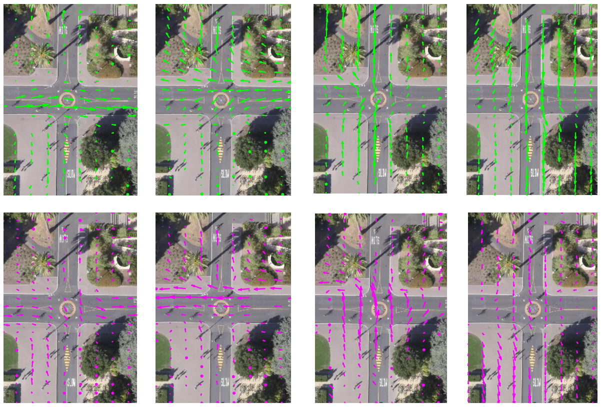

Fig. 5. Estimated motion fields per class (1st and 2nd rows) for the *Little* scenario. The arrows are colored according to the most probable class and their length is proportional to the velocity module.

Table 1. Confusion matrix: model vs ground truth labels.

	Ground-Truth	
Model Labels	*Biker*	*Pedestrian*
$c = 1$	69.3%	30.7%
$c = 2$	29.3%	70.7%

These results yield a balanced accuracy of 70% for the two classes on the test set, which is very promising and testifies for the potential of the proposed framework. Fig. 5 shows the estimated fields for each of the classes. We believe that this interesting performance may be further improved, because currently the model is too flexible. This may be tackled using appropriate priors to constraint the ending and class transition probabilities $\mathcal{B}$.

5. CONCLUSIONS

This paper proposes a hierarchical model for unsupervised trajectory analysis. The model is able to successfully describe and group trajectories from different kinds of agents, both synthetic and real (*bikers* and *pedestrians*). Moreover, the experimental results show the ability of the proposed method to estimate the model parameters and to identify agents of the same class. Forthcoming steps will include the use of priors, extension to other types of agents (*e.g*, vehicles), and an extensive comparison with other approaches, in particular supervised ones.

6. REFERENCES

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

[2] M. S. Ibrahim, S. Muralidharan, Z. Deng, A. Vahdat, and G. Mori, "A hierarchical deep temporal model for group activity recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1971–1980.

[3] H. Yao, A. Cavallaro, T. Bouwmans, and Z. Zhang, "Guest editorial introduction to the special issue on group and crowd behavior analysis for intelligent multi-camera video surveillance," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 3, pp. 405–408, 2017.

[4] W. Li, V. Mahadevan, and N. Vasconcelos, "Anomaly detection and localization in crowded scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 1, pp. 18–32, 2014.

[5] S. Coşar, G. Donatiello, V.a Bogorny, C. Garate, L. O. Alvares, and F. Brémond, "Toward abnormal trajectory and event detection in video surveillance," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 3, pp. 683–695, 2017.

[6] M. Ravanbakhsh, M. Nabi, E. Sangineto, L. Marcenaro, C. Regazzoni, and N. Sebe, "Abnormal event detection in videos using generative adversarial nets," in *IEEE International Conference on Image Processing*, 2017, pp. 1577–1581.

[7] P. Coscia, F. Castaldo, F. A. N. Palmieri, A. Alahi, S. Savarese, and L. Ballan, "Long-term path prediction in urban scenarios using circular distributions," *Image and Vision Computing*, vol. 69, pp. 81–91, 2018.

[8] 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.

[9] 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," in *CVPR*, 2017, pp. 336–345.

[10] A. Robicquet, A. Sadeghian, A. Alahi, and S. Savarese, "Learning social etiquette: Human trajectory understanding in crowded scenes," in *European conference on Computer Vision*, 2016, pp. 549–565.

[11] 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*, 2013, vol. 32, pp. 201–210.

[12] 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.

[13] S. Fine, Y. Singer, and N. Tishby, "The hierarchical hidden Markov model: Analysis and Applications," *Machine learning*, vol. 32, no. 1, pp. 41–62, 1998.

[14] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, pp. 257286, 1989.

[15] W. Kong, Z. Y. Dong, D. J. Hill, J. Ma, J. H. Zhao, and F. J. Luo, "A hierarchical hidden Markov model framework for home appliance modeling," *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 3079–3090, 2018.

[16] R. Szeliski, *Computer vision: algorithms and applications*, 2010.

[17] C. Barata, M. A. T. Figueiredo, and J. S. Marques, "Hierarchical motion fields," Tech. Rep., SPARSIS Project, 2019, `http://users.isr.ist.utl.pt/~jsm/SPARSIS/2019_Catarina_report.pdf`.

[18] C. Barata, J. C. Nascimento, and J. S. Marques, "A sparse approach to pedestrian trajectory modeling using multiple motion fields," in *IEEE International Conference on Image Processing*, 2017, pp. 2538–2542.

[19] C. Barata, J. M. Lemos, and J. S. Marques, "Estimation of space-varying covariance matrices," in *IEEE International Conference on Image Processing*, 2018, pp. 4003–4007.

[20] L. Condat, "Fast projection onto the simplex and the $l1$ ball," vol. 158, pp. 575–585, 2016.