SIViP
DOI 10.1007/s11760-014-0620-0

ORIGINAL PAPER

Moving horizon estimation of pedestrian interactions using multiple velocity fields

Ana Portelo · Mário A. T. Figueiredo · João M. Lemos · Jorge S. Marques

Received: 9 July 2013 / Revised: 5 January 2014 / Accepted: 23 January 2014

Abstract This paper describes a model, and algorithms based on it, for the analysis of pedestrian interactions in outdoor scenes. Pedestrian activities are described by their trajectories in the scene, and we wish to know if they were independently generated or if they are correlated. Two models are considered: (i) a model based on multiple velocity fields recently proposed in Nascimento et al. (IEEE Trans Image Process 22(5):1712–1725, 2013) and (ii) an interaction model based on the attraction/repulsion between pairs of pedestrians. Several combinations of these models are studied and evaluated. An estimation method based on a moving horizon optimization of a quadratic cost functional is proposed. Experimental results with synthetic data and real video data are presented to assess the performance of the algorithms.

Keywords Human activity recognition · Trajectory analysis · Probabilistic models

1 Introduction

The analysis of pedestrian interactions in outdoor scenes is a challenging problem with applications in video surveillance [1,2]. The goal is to detect the presence of interacting pedestrians and to recognize the type of interaction that they are performing. Examples of pedestrian activities include the following: meeting, walking together, stopping and starting, or pursuit.

A typical approach to tackle this problem consists of extracting the pedestrian trajectories from the video sequence. The trajectories are then analyzed using a probabilistic motion model learned from the video data [3–6]. That model should be able to discriminate noninteracting pedestrians from interacting ones and should also provide information about the type of interaction.

A generative model for the pedestrian motion in the scene, based on multiple velocity fields, was recently proposed in Nascimento et al. [7,8]. Each pedestrian is represented by his/her center of mass in the image, and the pedestrian motion is driven by one of the velocity fields at each frame: the active field. Switching among velocity fields is allowed at each position. The velocity fields represent typical motion patterns in the scene, and the switching mechanism allows abrupt changes of direction and increases the flexibility of the model.

The multiple motion field model accounts for the motion of isolated pedestrians in the scene, but does not consider interactions among pedestrians. To account for interactions, the model is extended with an interaction term that describes attractive/repulsive behaviors between pairs of pedestrians. This term is inspired by the social force model proposed by Helbing and Molnar [9,10], which has been used for crowd simulation and, more recently, for crowd behavior analysis (e.g., see [11,12]).

This paper has two main contributions:

– Extension of the multiple velocity fields model to deal with pedestrian interactions, by describing these interactions using attractive/repulsive terms;

A. Portelo · J. M. Lemos
INESC-ID, Instituto Superior Tecnico, Lisbon, Portugal

M. A. T. Figueiredo
Instituto de Telecomunicações, Instituto Superior Tecnico, Lisbon, Portugal

J. S. Marques (✉)
Institute for Systems and Robotics, Instituto Superior Tecnico, Lisbon, Portugal
e-mail: jsm@isr.ist.utl.pt

Published online: 21 March 2014

– Application of a moving horizon algorithm to estimate the active field that drives each trajectory and the interaction parameters at each instant of time.

The paper is organized as follows. Section 2 describes the interaction model. Section 3 describes the parameter estimation method. Section 4 presents experimental results, and Sect. 5 draws conclusions.

2 State of the art

Several models have been proposed to describe pedestrian interactions in video sequences. Some of those models consider a small number of interacting pedestrians (e.g., two or three), while others are focused on large groups (crowds). Several models are inspired by physical concepts and consider pedestrians as a set of particles obeying physical laws. Henderson [13] used gas-kinetic and fluid mechanics to describe the movement of crowds. The magnetic force model proposed in Okazaki [14] describes pedestrians as particles acted by electric forces as if the pedestrians and the scene were described by charged particles. The best-known model inspired by physics is perhaps the social force model proposed by Helbing and Molnar [9,10]; in that model, the movement of each pedestrian is described by the Newton law of mechanics, under the action of several forces that represent the interaction with other pedestrians, with the scene, and with the desired goal, each of which has an attractive or repulsive action on the pedestrian. The social force model has been used for simulation purposes, and it was shown that several human behaviors can be modeled by it. In addition, it has also been used for video analysis, and several works propose inference methods to estimate the attraction/repulsion parameters from object trajectories [11] or optical flow [15].

Another class of methods extracts spatiotemporal patterns from the video stream. Those patterns include the object trajectories or the evolution of feature vectors (e.g., relative position, velocity, bounding box) [2]. The evolution of the pedestrian features has been represented by using Bayesian networks, hidden Markov models, and several variants of these modeling tools. The interaction among pedestrians is tackled in [3] using coupled hidden Markov models (CHMM). However, only two pedestrians are considered in this model, and the extension to a higher (and varying) number of pedestrians is not considered. Activity recognition with a varying number of pedestrians has been tackled using observation decomposed hidden Markov models (ODHMMs) [16]. The number of agents associated with pedestrians is kept equal to three in the experiments reported.

Yet, another set of methods represents complex interactions using sub-events, each of which characterized by a different model. The interaction among multiple pedestrians has been addressed by different methods, such as context-free grammars (CFG) [1], and networks of dynamic probabilistic models [4]. That approach also includes switched hidden Markov models [8], and switched motion field models [7], which express complex motion patterns using simple models. However, those models have only been used to describe the activity of isolated pedestrians.

3 Motion model

3.1 Isolated pedestrians

The pedestrian model proposed in [7,8] assumes that each pedestrian walks under the influence of the scene i.e., different scenes lead to different pedestrian trajectories and behaviors. Furthermore, it assumes that the pedestrian trajectories are described by K velocity fields $\mathbf{T}_k : [0,1]^2 \rightarrow \mathbb{R}^2$, for $k \in \{1,\ldots,K\}$ that represent typical motion regimes. For the sake of simplicity, the image domain is assumed to be $[0,1]^2$.

The proposed model considers that one velocity field is driving the pedestrian motion at each instant of time (active field), and the pedestrian trajectory is generated by Nascimento [7]

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

where $\mathbf{x}(t) \in \mathbb{R}^2$ denotes the pedestrian position at the discrete time t, $k(t) \in \{1,\ldots,K\}$ is the label of the active field, and $\mathbf{w}(t)$ is a sequence of random and uncorrelated displacements, with normal distribution, $\mathbf{w}(t) \sim N(\mathbf{0}, \sigma^2_{k(t)}\mathbf{I})$; σ^2_k denotes the variance associated with the kth velocity field.

The label sequence of the active field $k(1),\ldots,k(t)$ is assumed to be a Markov chain, characterized by the conditional probabilities

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

where $B_{ij}(\mathbf{x})$ is the transition probability from velocity field i to j, when the pedestrian is located at the position $\mathbf{x}$. This means that the transition probabilities depend on the position of the pedestrian in the scene. For example, if the pedestrian is at the middle of a cross between two streets, there is a high probability of changing the walking direction.

The velocity fields and the space-varying transition matrix are estimated on a uniform grid of nodes and interpolated when the pedestrian is located at an arbitrary position that does not coincide with a grid node. The model parameters are as follows: (i) the velocity fields; (ii) the space-varying transition matrix; and (iii) the noise variances. In practice, these parameters are estimated from a set of typical trajectories extracted from the video signal, during the learning phase (see [7] for details).

3.2 Switched interaction model (SIM)

Consider now pairs of pedestrians that may interact, or not, during the observation interval. If the pedestrians do not interact (hypothesis H0), we will assume that their trajectories are described as before, i.e.,

$$
\begin{aligned}
&H0: \\
&\mathbf{x}^{(1)}(t) = \mathbf{x}^{(1)}(t-1) + \mathbf{T}_{k^{(1)}(t)}(\mathbf{x}^{(1)}(t-1)) + \mathbf{w}^{(1)}(t) \\
&\mathbf{x}^{(2)}(t) = \mathbf{x}^{(2)}(t-1) + \mathbf{T}_{k^{(2)}(t)}(\mathbf{x}^{(2)}(t-1)) + \mathbf{w}^{(2)}(t),
\end{aligned} \tag{3}
$$

where $\mathbf{x}^{(i)}, i = 1, 2$ denotes the position of pedestrian i; $k^{(i)}, i = 1, 2$ is the label of the active field for the $i-th$ pedestrian, and $\mathbf{w}^{(i)}(t), i = 1, 2$, is a random displacement.

Equation (3) describes the motion of isolated pedestrians in the scene but does not consider pedestrian interactions. To account for the interactions (hypothesis H1), we will assume that each pedestrian may be attracted or repulsed by the other, leading to the following interaction model

$$
\begin{aligned}
&H1: \\
&\mathbf{x}^{(1)}(t) = \mathbf{x}^{(1)}(t-1) + \alpha^{(1)}(t-1)\boldsymbol{\phi}^{(1,2)}(t-1) + \mathbf{w}^{(1)}(t) \\
&\mathbf{x}^{(2)}(t) = \mathbf{x}^{(2)}(t-1) + \alpha^{(2)}(t-1)\boldsymbol{\phi}^{(2,1)}(t-1) + \mathbf{w}^{(2)}(t),
\end{aligned} \tag{4}
$$

where

$$
\boldsymbol{\phi}^{(i,j)}(t-1) = \frac{\mathbf{x}^{(j)}(t-1) - \mathbf{x}^{(i)}(t-1)}{\|\mathbf{x}^{(j)}(t-1) - \mathbf{x}^{(i)}(t-1)\|}, \tag{5}
$$

is a unit vector pointing from pedestrian i toward pedestrian j; and $\alpha^{(1)}, \alpha^{(2)}$ are the interaction parameters: a positive $\alpha^{(i)}$ represents attraction, while negative $\alpha^{(i)}$ represents repulsion; $\mathbf{w}^{(i)}(t)$ is a random perturbation as before with Gaussian distribution $N(\mathbf{0}, \sigma_0^2\mathbf{I})$ where σ_0^2 is the noise variance in the presence of interaction.

Both models are used in the generation of the pedestrians trajectories. During the noninteraction intervals, pedestrians trajectories are generated by (3), and during interaction intervals, the interaction model (4) is used instead. This involves a binary switching between both generation mechanisms (H0, H1). Therefore, the overall model is denoted by *Switched Interaction Model (SIM)*.

3.3 Combined interaction model (CIM)

Until now, we have considered separate motion models for the pedestrians motion with and without interaction. In practice, these two models may simultaneously influence the pedestrians motion, i.e., we may assume that two pedestrians may wish to combine scene-driven motion with some kind of interaction. This can be achieved by using a combined model in which both motion mechanisms are added

$$
\begin{aligned}
\mathbf{x}^{(1)}(t) &= \mathbf{x}^{(1)}(t-1) + \mathbf{T}_{k^{(1)}(t)}(\mathbf{x}^{(1)}(t-1)) \\
&\quad + \alpha^{(1)}(t-1)\boldsymbol{\phi}^{(1,2)}(t-1) + \mathbf{w}^{(1)}(t) \\
\mathbf{x}^{(2)}(t) &= \mathbf{x}^{(2)}(t-1) + \mathbf{T}_{k^{(2)}(t)}(\mathbf{x}^{(2)}(t-1)) \\
&\quad + \alpha^{(2)}(t-1)\boldsymbol{\phi}^{(2,1)}(t-1) + \mathbf{w}^{(2)}(t)
\end{aligned} \tag{6}
$$

where $k^{(1)}(t), k^{(2)}(t)$ are the labels of the active fields. This generative model is denoted *Combined Interaction Model (CIM)*.

There is a simple relationship between combined and switched interaction models. The two motion mechanisms in SIM are special cases of CIM:

- No-interaction: the SIM model is equal to CIM for $\alpha^{(1)} = \alpha^{(2)} = 0$;
- Interaction: the SIM model is equal to CIM using null fields $\mathbf{T}_0 = \mathbf{0}$.

For the sake of symmetry, we have created an additional velocity field, the null field $\mathbf{T}_0$, which is not trained from the video data.

3.4 Model identifiability

Two questions can be asked about the two proposed models: (i) do they accurately represent interactions between pairs of pedestrians in a scene? (ii) Can we always distinguish the two motion regimes described above (isolated motion and pedestrian interaction)?

The first question will be answered through experimental evaluation of the models in Sect. 5. Concerning the second question, the answer is "not always." Indeed, the two regimes use different equations [Eqs. (3, 4)] to update the pedestrian position. In the first case, the update is based on the active motion field $\mathbf{T}_{k(t)}(\mathbf{x}(t-1))$, while in the second case, the update is based on $\boldsymbol{\phi}(t-1)$. If these two vectors are colinear at a given instant of time, there is no way to distinguish scene-driven motion from interaction at that instant of time and an infinite number of trade-offs are allowed.

This difficulty can be attenuated by considering multiple time instants and some regularization condition on the evolution of the model parameters, e.g., by penalizing changes in the attraction/repulsion coefficients or changes in the label sequence.

4 Parameter estimation

Given a pair of trajectories $\mathbf{x}^{(1)}(t), \mathbf{x}^{(2)}(t), t = 1, \ldots, T$, we wish to know if interaction exists and what is the evolution of the attraction/repulsion coefficients. The answer depends on the model adopted (either SIM, CIM). Let us consider the

estimation of the CIM model parameters first, since it is the most general one.

Four parameters have to be estimated at each instant of time ($\alpha^{(i)}(t), k^{(i)}(t)$, for $i = 1, 2$). However, we can simplify the problem and separately estimate the parameters of each pedestrian, provided $\boldsymbol{\phi}$ is known.

We will assume that $\alpha^{(i)}(t)$ changes sparsely along time and the label $k^{(i)}(t)$ has a small number of transitions. Therefore, we will adopt a sliding interval (moving horizon) with length H and assume that these two variables are constant within this time interval. For each pedestrian, the energy of the residue is computed in a sliding time window of length H extending to the past between the current time t_0 and $t_0 - H + 1$

$$E_{t_0}(\alpha, k) = \frac{1}{\sigma_k^2} \sum_{t=t_0-H+1}^{t_0} \|\mathbf{y}(t) - \alpha\phi(t-1)\|^2, \tag{7}$$

where $\mathbf{y}(t) = \mathbf{x}(t) - \mathbf{x}(t-1) - \mathbf{T}_k(\mathbf{x}(t-1))$. The index $i = 1, 2$ was dropped for the sake of simplicity.

The minimization of (7) with respect to α leads to a close form expression

$$\hat{\alpha}(k) = R^{-1} r \tag{8}$$

where

$$R = \sum_{t=t_0-H+1}^{t_0} \phi(t-1)^T \phi(t-1), \tag{9}$$

$$r = \sum_{t=t_0-H+1}^{t_0} \phi(t-1)^T \mathbf{y}(t). \tag{10}$$

The estimation of the index of the active field, k, will be discussed in the sequel.

4.1 Joint estimation (JE)

Two optimization strategies are considered for the estimation of CIM parameters. The first method is based on the optimization of $E_{t_0}(\alpha, k)$ with respect to k and α, simultaneously,

$$(\hat{\alpha}, \hat{k}) = \arg\min_{\alpha,k} E_{t_0}(\alpha, k). \tag{11}$$

Since a closed form expression for α is available, we can replace it by its optimal estimate $\hat{\alpha}$ defined in (8) and minimize $E_{t_0}(\hat{\alpha}(k), k)$ with respect to k. This is a straightforward step since we only need to compute the energy associated with each motion field $k \in \{1, \ldots, K\}$ and choose the smallest.

Although this approach leads to the global minimum of E_{t_0}, it does not always lead to meaningful estimates of the unknown parameters. Indeed, since the model is too flexible, there are several ways to explain the data, i.e., sometimes, we can explain the observations $\mathbf{x}(t)$ using different velocity fields and use the interaction term to compensate the mismatch (see Sect. 3.4). Therefore, we adopted the alternative approach described in the sequel.

4.2 Hierarchical estimation (HE)

In order to tackle the identifiability problem mentioned in Sect. 4.1, we assume that the motion field is the main responsible for explaining the observations. The interaction term is used only when the motion field model is not enough to describe the trajectories. Therefore, we estimate the motion field first, assuming no interaction

$$\hat{k} = \arg\min_{k} E_{t_0}(\alpha = 0, k). \tag{12}$$

After obtaining $\hat{k}$, we estimate α by solving the following optimization problem

$$\hat{\alpha} = \arg\min_{\alpha} E_{t_0}(\alpha, \hat{k}). \tag{13}$$

This procedure can be done analytically using (8) with $\hat{k}$ estimated from (12). The whole algorithm is very fast and can be easily speed up by recursively computing R and r. The above hierarchical approach proved to be much more robust than the joint minimization procedure described in Sect. 4.1.

4.3 Mutually exclusive estimation (MEE)

The SIM model uses different motion models for representing trajectories without and with interaction. In the first case (H0: no interaction), the model only depends on the label k with $\alpha = 0$, while in the second case (H1: interaction), the model only depends on the α coefficient with $k = 0$. Therefore, the minimization of both parameters is decoupled and can be stated as follows

$$H0: \ \hat{k} = \arg\min_{k} E_{t_0}(\alpha = 0, k) \tag{14}$$

$$H1: \ \hat{\alpha} = \arg\min_{\alpha} E_{t_0}(\alpha, k = 0) \tag{15}$$

We choose the hypothesis $H0$ or $H1$ that has the smallest energy.

In all the three estimation methods, after computing $\hat{\alpha}, \hat{k}$ for the instant t_0, we shift the analysis window by one sample (or more) and repeat the procedure. This technique is inspired in the *Moving horizon estimation (MHE)* method used in state estimation problems [17].

5 Experimental results

The interaction models (SIM, CIM) and the estimation methods (JE, HE, MEE) proposed in this paper were tested with synthetic and video data. A few selected experiments are described in this section to illustrate the performance of the algorithms.

5.1 Synthetic data

First, the model is evaluated with synthetic data. In each experiment, a pair of trajectories is generated using the combined interaction model with two velocity fields: a vertical (up) field and an horizontal (right) field. The evolution of the active field labels and interaction parameters, $k(t)^{(i)}, \alpha^{(i)}, i = 1, 2$, are specified by the user. Then, we applied three estimation methods (JE, HE, MEE) to estimate these parameters and the range of interaction between the pedestrians. We stress that only the first two methods (JE, HE) are compatible with the generative model used to create the data. The third method (MEE) is based on a different model, and therefore, it is expected to perform worse in these tests.

Figure 1 shows one of these experiments. The 1st row shows the synthetic trajectories generated by the CIM model with interaction between the pedestrians in the time interval $\{71, \ldots, 200\}$. A cross is drawn every 10 frames to include time information on the trajectories. The following rows present the results obtained with JE, HE, and MEE, respectively. For the sake of compactness, we will only present the estimates for the 1st pedestrian (blue line). The estimates for second pedestrian (red line) are similar. The best results are obtained by the HE method which is able to detect the interaction interval and provides accurate estimates for $k^{(1)}(t)$ and $\alpha^{(1)}(t)$ parameters. The other two methods are significantly worse. The JE method leads to a bad estimate of the interaction interval due to excess of flexibility in the estimation process. The vector field displacement can be compensated by an appropriate choice of the $\alpha^{(i)}(t)$ coefficient. The MEE method also fails to detect the interaction range since it is not consistent with the model used to generate the data.

To assess the performance of these methods, we have generated 100 trajectory pairs and built receiver operating characteristics (ROC curves) for the detection of interaction in each frame. In order to detect pedestrian interaction in the same way in all these methods, $|\alpha^{(i)}(t)|$ is compared with a threshold λ and interaction is detected if $|\alpha^{(i)}(t)| > \lambda$. The ROC curve is obtained by varying the value of λ. The ROC curve describes the trade-off between false positives and false negatives. The performance of the detector is often assessed by the area under the curve (AUC) that would be equal to 1 if the detector was perfect and had no errors. Figure 2 shows the three ROC curves and the corresponding AUC measurements. As expected, the best results are achieved by the HE method with AUC = 93 %.

5.2 Video data

The proposed models were also applied in the analysis of pedestrian trajectories in outdoor scenes. The video signal was acquired using a video camera Sony HDR-CX260 with a resolution of 8.9 megapixels per frame and working at a frame rate of 30 frames per second. A tracking algorithm was used to extract the pedestrians trajectories. This procedure was done using a background subtraction algorithm for the detection of active regions, followed by matching active regions at consecutive frames using the assignment algorithm described in [18]. The trajectories were then subsampled at a frame rate of 1 frame per second, and the association errors were corrected.

Figures 3 and 4 show two examples of pedestrian interactions (meet and go) in an university campus and the results obtained with the HE and MEE methods. The JE method is not considered since it performs worse than the HE. Five motion fields are considered in these examples. Four of them represent pedestrian motion in four directions: up (U), down (D), left(L), right (R). The last field is the null field, which represents stopped (S) pedestrians.

Figures 3 and 4 display the warped trajectories, the energy associated with the admissible hypotheses the interaction detection, and the estimates of $k^{(i)}(t), \alpha^{(i)}(t)$. All these variables are represented as a function of time (center of the analysis window). To assess the binary decision (interaction/no-interaction) provided by the model, we compare it with a manual decision provided by the user (dashed line). As we did before, we only represent the variables associated with the first pedestrian (blue trajectory) for the sake of compactness.

In the first example (Fig. 3), both methods manage to extract meaningful estimates of the active field for the blue trajectory. Furthermore, they both agree that the pedestrian is moving to the right. The α curves are, however, quite different. In the case of MEE, the α parameter is not estimated when there is no interaction and the α coefficient is specially important in the initial phase of attraction when the two pedestrians meet. This also makes sense. Concerning the interaction detection, the two methods adopt different strategies. The HE compares the α coefficient with a threshold, and the MEE takes the decision by considering an additional hypothesis (no-interaction) and computing the energy function for this hypothesis. The best results are achieved by the MEE method in this example since the detection output (solid line) is much closer to the user decision (dashed line) in the case of the MEE method.

The second example (Fig. 4) considers two pedestrians that move in different directions and suddenly see each other

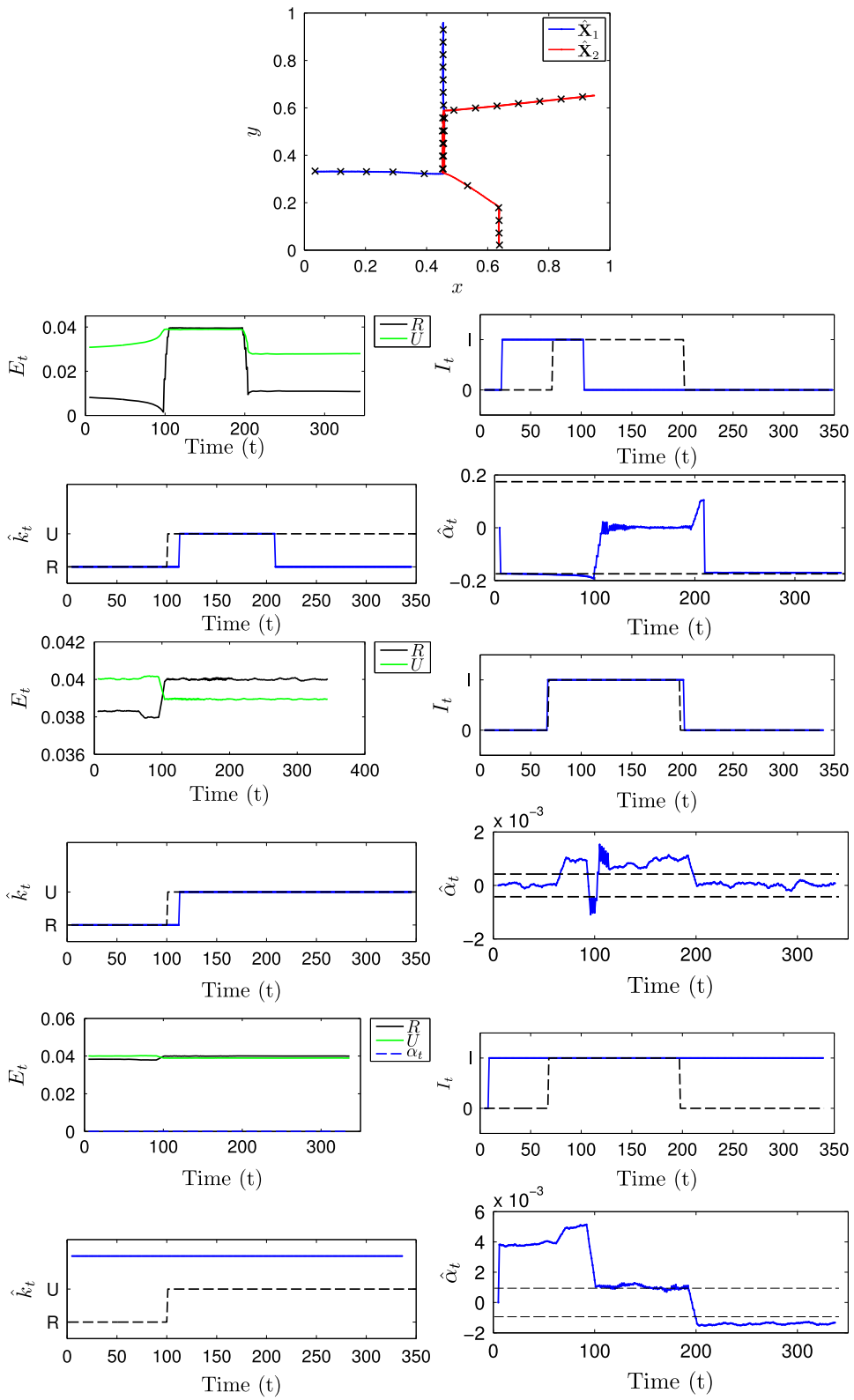

Fig. 1 Synthetic example (noise variance $\sigma^2 = 0.01$)—interaction analysis with JE, HE, and MEE methods: synthetic trajectories (*1st row*); JE results: energy and interaction detection (*2nd row*); active field and interaction coefficient estimates, $\hat{k}_t^{(1)}, \hat{\alpha}_t^{(1)}$ (*3rd row*); HE results: energy and interaction detection (*4th row*); active field and interaction coefficient estimates, $\hat{k}_t^{(1)}, \hat{\alpha}_t^{(1)}$ (*5th row*); MEE results: energy and interaction detection (*6th row*); active field and interaction coefficient estimates, $\hat{k}_t^{(1)}, \hat{\alpha}_t^{(1)}$ (7th row)

and meet. After a while, they separate. Both algorithms agree that the blue pedestrian moves right and then up. The MEE considers that the pedestrians do not interact when they meet, and the HE method considers that they are stopped which also makes sense. However, the interaction detection provided by the MEE is close to the user decision (dashed line),

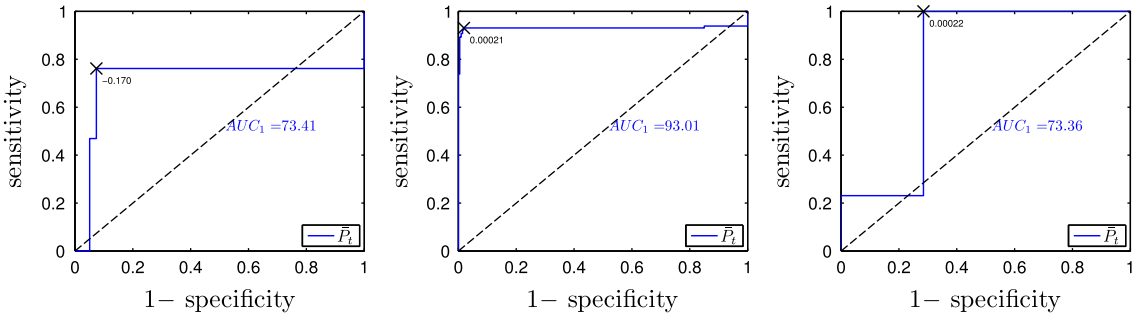

Fig. 2 ROC curves for JE (*left*) HE (*center*) and MEE (right), for $\sigma^2 = 0.01$

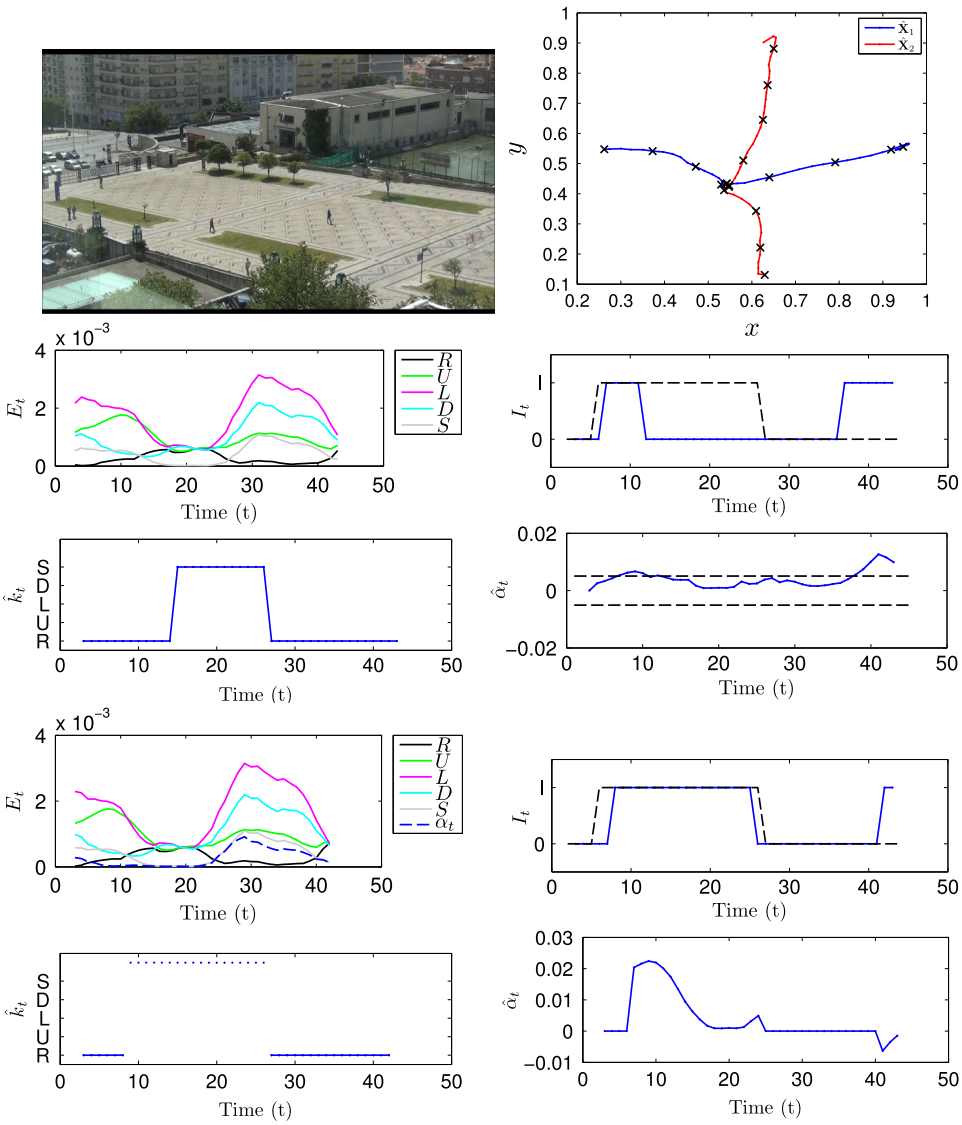

Fig. 3 Example 1—interaction analysis with HE and MEE methods: sample image and warped trajectories (*1st row*); HE results: energy and interaction detection (*2nd row*); active field and interaction coefficient estimates, $\hat{k}_t^{(1)}, \hat{\alpha}_t^{(1)}$ (*3rd row*); MEE results: energy and interaction detection (*4th row*); active field and interaction coefficient estimates, $\hat{k}_t^{(1)}, \hat{\alpha}_t^{(1)}$ (*5th row*)

while the interaction detection provided by the HE tends to shorten the interaction intervals. Both algorithms provide a spurious detection of interaction in the first frames of the sequence.

We can conclude from these two examples (and others, we have considered) that both algorithms (HE, MEE) perform well in the estimation of pedestrian interactions although spurious decisions may occur. Trajectory artifacts may be

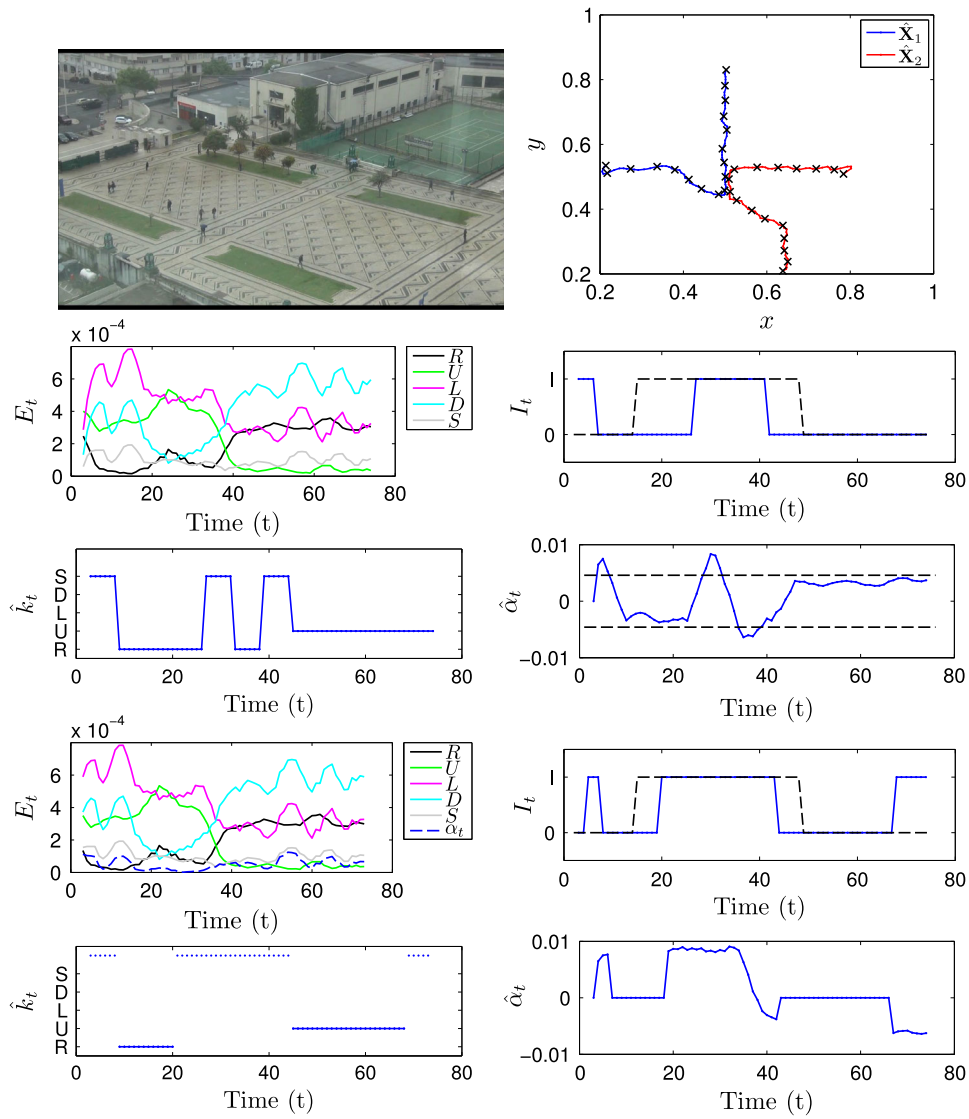

Fig. 4 Example 2—interaction analysis with HE and MEE methods: sample image and warped trajectories (*1st row*); HE results: energy and interaction detection (*2nd row*); active field and interaction coefficient estimates, $\hat{k}_t^{(1)}$, $\hat{\alpha}_t^{(1)}$ (*3rd row*); MEE results: energy and interaction detection (*4th row*); active field and interaction coefficient estimates, $\hat{k}_t^{(1)}$, $\hat{\alpha}_t^{(1)}$ (*5th row*)

erroneously confused with interactions. The MEE method provides more robust estimates of the interaction interval. This suggests that pedestrian motion is not driven by external velocity fields, during interactions. Therefore, the switched interaction model (SIM) performs better than the combined interaction model (CIM) in these examples.

6 Conclusions

This paper proposed two models for the analysis of interactions between pairs of pedestrians in outdoor scenes. The models assume that each of the pedestrians may interact, or not, with the other. Noninteracting pedestrians move according to a multiple velocity field model that was recently proposed in [7,8], while interacting pedestrians are subject to attraction/repulsion effects. Two models were considered to account for the pedestrian motion: a switched interaction model (SIM), in which these two effects are mutually exclusive, and a combined interaction model (CIM), in which the two types of motion are combined, according to weights to be estimated. Estimation methods were provided to estimate the model parameters (label of the active field, attraction/repulsion coefficient) using a moving horizon strategy. This means that a sliding window is used to estimate the parameters. We assumed that the unknown parameters are constant inside the window and are estimated by minimizing the energy of the residue inside the window.

The proposed models were applied to synthetic and video data, in order to illustrate their performance. It was shown that they are flexible enough to represent pedestrian motion and interactions in outdoor scenes. It was also concluded that it is not always simple to separate the interaction effects from the noninteraction motion, since pedestrians do not always follow the velocity fields directions, and this mismatch can be interpreted as an interaction.

Several work directions are worth following in the future. First, we should characterize and improve the trade-off between interaction and velocity field motion, by considering additional constraints to the type of motion assumed in both cases. Second, the model should be extended to the case of multiple pedestrian interactions. The attraction/repulsion terms used in this paper to account for human interactions can be easily extended to multiple interacting pedestrians; however, the reliable estimation of these parameter is a challenging task for which additional constrains or regularization terms have to be considered. As a last comment, when evaluating the algorithms, more difficult interactions can be considered in future work.

Acknowledgments This work was supported by FCT in the framework of contract PTDC/EEA-CRO/098550/2008, PEst-OE/ EEI/ LA0009/2013, and PEst-OE/EEI/LA0021/2013.

References

1. Aggarwal, J.K., Ryoo, M.S.: Human activity analysis: a review. ACM Comput. Surv. (CSUR) Surv. Homepage Arch. **43**(3) (2011)
2. Turaga, P., Subrahmanian, R., Subrahmanian, V., Udrea, O.: Machine recognition of human activities: a survey. IEEE Trans. Circuits Syst. Video Technol. **18**(11), 1473–1488 (2008)
3. Oliver, N., Rosario, B., Pentland, A.: A bayesian computer vision system for modeling human interactions. IEEE Trans. Pattern Anal. Mach. Intell. **22**(8), 831–843 (2000)
4. Suk, H.-I., Jain, A., Lee, S.-W.: A network of dynamic probabilistic models for human interaction analysis. IEEE Trans. Circuits Syst. Video Technol. **21**(7), 932–945 (2011)
5. Ntalampiras, S., Arsic, D., Hofmann, M., Andersson, M., Ganchev, T.: PROMETHEUS: Heterogeneous Sensor Database in Support of Human Behavioral Patterns in Unrestricted Environments. Signal, Image and Video Processing. Springer (2012)
6. Mahbub, U., Imtiaz, H., Ahad, A.R.: Action recognition based on statistical analysis from clustered flow vectors. Signal, Image and Video Processing. Springer (2013)
7. Nascimento, J., Figueiredo, M.A.T., Marques, J.: Activity recognition using mixture of vector fields. IEEE Trans. Image Process. **22**(5), 1712–1725 (2013)
8. Nascimento, J., Marques, J., Figueiredo, M.A.T.: Classification of complex pedestrian activities from trajectories. IEEE International Conference Image Processing, pp. 3481–3484 (2010)
9. Helbing, D.: A mathematical model for the behavior of pedestrians. Behav. Sci. **36**, 298–310 (1991)
10. Helbing, D., Molnar, P.: Social force model for pedestrian dynamics. Phys. Rev. E **51**, 4282–4286 (1995)
11. Mehran, R., Oyama, A., Shah, M.: Abnormal crowd behavior detection using social force model. In: IEEE International Conf. on Computer Vision and Pattern Recognition, pp. 935–942 (2009)
12. Zhang, Y., Qin, L., Yao, H., Huang, Q.: Abnormal crowd behavior detection based on social attribute-aware force model. In: IEEE Internatinal Conference Image Processing, pp. 2689–2692 (2012)
13. Henderson, L.F.: On the fluid mechanic of human crowd motions. Transp. Res. **8**, 509–515 (1974)
14. Okazaki, S.: A study of pedestrian movement in architectural space, part 1: pedestrian movement by the application on of magnetic models. Trans. A.I.J **283**, 111–119 (1979)
15. Luber, M., Stork, J., Tipaldi, G., Arras, K.O.: People tracking with human motion predictions from social forces. In: IEEE Conference on ICRA, pp. 464–469 (2010)
16. Liu, X., Chua, C.: Multi-agent activity recognition using observation decomposed hidden markov models. Image Vis. Comput. **24**, 166–175 (2006)
17. Alessandri, A., Baglietto, M., Battistelli, G.: Moving-horizon state estimation for non-linear discrete-time systems: new stability results and approximation schemes. Automatica **44**, 1753–1765 (2008)
18. Veenman, C.J., Reinders, M.J.T., Backer, E.: Resolving motion correspondence for densely moving points. IEEE Trans. Pattern Anal. Mach. Intell. **23**, 54–72 (2001)