Predicting Social Interactions for Visual Tracking

Xu Yan
xyan6@uh.edu

Ioannis A. Kakadiaris
ioannisk@uh.edu

Shishir K. Shah
sshah@central.uh.edu

Department of Computer Science
University of Houston
Houston, TX 77204-3010
USA

Abstract

Human interaction dynamics are known to play an important role in the development of robust pedestrian trackers that are applicable to a variety of applications in video surveillance. Traditional approaches to pedestrian tracking assume that each pedestrian walks independently and the tracker predicts the location based on an underlying motion model, such as a constant velocity or autoregressive model. Recent approaches have begun to leverage interaction, especially by modeling the repulsion force, among pedestrians to improve motion predictions. However, human interaction is more complex and is influenced by both repulsion and attraction effects. This motivates the use of a more complex human interaction model for pedestrian tracking. In this paper, we propose a novel visual tracking method by leveraging complex social interactions. We present an algorithm that decomposes social interactions into multiple potential interaction modes. We integrate these multiple social interaction modes into an interactive Markov Chain Monte Carlo tracker. We demonstrate how the developed method translates into a more informed motion prediction, resulting in a robust tracking performance. We test our method on videos from unconstrained outdoor environments and compare it against popular multi-object trackers.

1 Introduction

Multiple pedestrian tracking in unconstrained environments remains a challenging task that has received considerable attention from the computer vision community in recent years. Accurate and real-time multiple pedestrian tracking can greatly improve the performance of activity recognition and analysis of high level events through a surveillance system. A number of approaches, that address this issue, have been proposed in the past few years [2, 13]. However, even the best of the existing systems tend to exhibit a loss of accuracy and tracking precision in unconstrained environments and fail to account for complex pedestrian interactions in a scene.

Recently, tracking approaches have explored the social interaction among pedestrians, which provides additional constraints and guides motion prediction. The linear trajectory avoidance model [15] used the repulsion effect among pedestrians to predict local motion

paths for individual pedestrians. Luber *et al*. [12] extended the use of repulsion effect to include scene obstacles to improve motion prediction. Choi *et al*. [5] considered both repulsion effects and group motion dynamics within a joint prediction model. Broadly speaking, all models based on the use of repulsion effects [5, 12, 15] are applicable only in environments where pedestrians are moving in a crowd and do not deviate from their global destination. In contrast, the modeling of more complex interactions among people exhibited through both repulsion and attraction effects has not been adequately explored. Incorporation of the attraction effect into a social interaction model is a logical extension to improve the accuracy of representing pedestrian dynamics. In this paper, we present a visual tracker with a pedestrian dynamic model that combines both attraction and repulsion effects among pedestrians. Without explicit knowledge of the exact social interaction at any instance in time, we consider the motion of each pedestrian to be a combination of all potential social interaction modes. Motivated by the work of Kwon and Lee [11], we propose to decompose a complex human interaction into several deterministic social interaction modes. Specifically, we build a social force model [9] to describe the attraction and repulsion effects among pedestrians. The key contributions of our work are as follows:

1. *A repulsion and attraction model describing pedestrian dynamics*. We model both repulsion and attraction effects in pedestrian dynamics. Such interactions are more common in unconstrained environments and can be leveraged to capture various interaction behaviors such as people meeting, people following, and/or group interactions.

2. *A decomposed social interaction model*. We propose a decomposed motion model that approximates complex social interactions and enables motion prediction without the knowledge of instantaneous interaction modes. The model predicts and tracks multiple pedestrians based on multiple social interaction modes and combines the tracking results together.

The rest of this paper is organized as follows. Section 2 presents the proposed social interaction model and describes its decomposition into multiple models. The incorporation of the proposed model within a Bayesian tracking framework and the design of the compound tracker is presented in Section 3. Section 4 presents the experiments performed and a qualitative and quantitative assessment of the tracker performance. Comparative analysis against multiple existing trackers is also presented. Finally, conclusions are presented in Section 5.

2   Social Interaction Model

The social force model by Helbing [9] is a computational model in which the interactions among pedestrians are described by using the concept of forces between physical entities. Each pedestrian feels a social force from other pedestrians that is proportional to the distance between them. In this model, a pedestrian i makes motion decisions based on the sum of forces $\vec{F}_i$ exerted.

The social force can be divided into repulsive, $\vec{f}^r$, and attractive, $\vec{f}^a$, forces. The repulsive force captures the behavior where people try to avoid collisions with each other and the attractive force captures the behavior when a person approaches another person with an intent to meet. A sum of these forces can be represented by:

$$\vec{F}_i = \sum_{j \neq i} \vec{f}^r_{ij} \; + \sum_{k \neq i, k \neq j} \vec{f}^a_{ik}, \tag{1}$$

where $\vec{f}^r_{ij}$ indicates the repulsive force between pedestrians i and j and $\vec{f}^a_{ik}$ indicates the attractive force between pedestrians i and k. In the social force model, it is assumed that every pedestrian knows the location of every other pedestrian in the scene.

The attraction and repulsion forces share a similar model with different directions and monotonicity. This can be expressed as:

$$\vec{f}^r_{ij} = F^r * e^{(\frac{r_{ij}-d_{ij}}{b})}\vec{u}_{ji}, and \tag{2}$$

$$\vec{f}^a_{ik} = F^a * e^{(\frac{d_{ik}-r_{ik}}{b})}\vec{u}_{ik}, \tag{3}$$

where F^r and F^a are the magnitudes of repulsion and attraction force, respectively, b is the boundary of the influence of the force, d_{ij} is the Euclidean distance between i and j, d_{ik} is the Euclidean distance between i and k, $\vec{u}_{ji}$ is the unit vector from j to i, and $\vec{u}_{ik}$ is the unit vector from i to k. The private sphere of a pedestrian is represented by a circle of radius r with r_i, r_j, and r_k defining the private sphere of pedestrians i, j, and k, respectively. Further, $r_{ij} = r_i + r_j$ and $r_{ik} = r_i + r_k$ define the radius of influence for pedestrians i and j, and i and k, respectively.

Under the social force model, the motion model that predicts the positional information for a tracked pedestrian i is given by:

$$\frac{\vec{F}_i}{m_i} = \frac{\partial \vec{v}_i}{\partial t}, \tag{4}$$

where $\vec{v}_i$ is the instantaneous velocity and m_i is the mass.

2.1 Social Interaction Decomposition

It is hard to define the social interaction mode for every pedestrian pair without an explicit knowledge of their intent. To address this point, we treat pedestrian interaction as linear combination of potential social interaction modes with dynamically adjusted weights at different moments in time. We decompose social interaction into potential social interaction modes and quantify them using a social force model [9] as shown in Eq. 5. For pedestrian i, the total force is given by:

$$\vec{F}_i = \sum_{n=1}^{q_i} w_i^{d_n} \vec{F}_i^{d_n}, d_1, d_2, \ldots, d_{q_i} \in D_i, \tag{5}$$

where d_n denotes the nth social interaction mode, $\vec{F}_i^{d_n}$ represents social forces based on mode d_n, $w_i^{d_n}$ is a weighting coefficient, D_i is the set of potential social interaction modes, and q_i is the number of social interaction modes. For any pedestrian, the decomposition is defined by first building social links with other pedestrians. The attraction and repulsion effects of the link are denoted by $\{+\}$ and $\{-\}$, respectively, and a social interaction mode is denoted by a set of social links' effects. For example, the set with two links would constitute the following interaction modes: $\{\{+,+\},\{+,-\},\{-,+\},\{-,-\}\}$, and the set with one link would include $\{\{+\},\{-\}\}$. Each pedestrian's social interaction set is expanded to have the same size as the maximum number in the social interaction set represented by $q_i = q$. This is done to ensure that the number of decomposed motion models remain the same across all tracked pedestrians.

Consider the scene illustrated in Fig. 1 comprising of six pedestrians. For pedestrian P_1, the decomposition would be realized by building two social interaction links $\{l_{12}, l_{13}\}$ based on the boundary of the social force's influence. In this case, the set of social interaction modes would be represented as $D_1 = \{\{+,+\}, \{+,-\}, \{-,+\}, \{-,-\}\}$ and the social interaction mode of pedestrian P_1 at that moment would be one mode from the set. The social interaction force model is leveraged to compute the repulsion and attraction forces, which are $\vec{f}_{21}^r$, $\vec{f}_{31}^r$, $\vec{f}_{12}^a$, and $\vec{f}_{13}^a$. As an example, the interaction force for mode $d_1 = \{+,+\}$ would be given by $\vec{F}_1^{d_1} = \vec{f}_{12}^a + \vec{f}_{13}^a$. In this example, considering pedestrians $P_1 - P_4$, the maximum size of the set of social interaction modes is four due to D_1. Although pedestrian P_4 has only one link $\{l_{43}\}$, the set D_4 is expanded to the same size. That is, $D_4 = \{\{+\}, \{-\}, \{+\}, \{-\}\}$ and accordingly $\vec{F}_4^{d_1} = \vec{f}_{43}^a$. Pedestrian P_5 has no social interaction link indicating that the social force under any interaction mode is zero. The set of interaction modes is expanded for D_5 to include $D_5 = \{\emptyset, \emptyset, \emptyset, \emptyset\}$ and accordingly $\vec{F}_5^{d_1} = \vec{0}$. Similarly, the forces for each of the pedestrians are computed and finally the force vector resulting from a single social interaction mode is given as $\vec{F}^{d_n} = [\vec{F}_1^{d_n}, \vec{F}_2^{d_n}, \dots, \vec{F}_P^{d_n}]^T$ where P is the number of pedestrians.
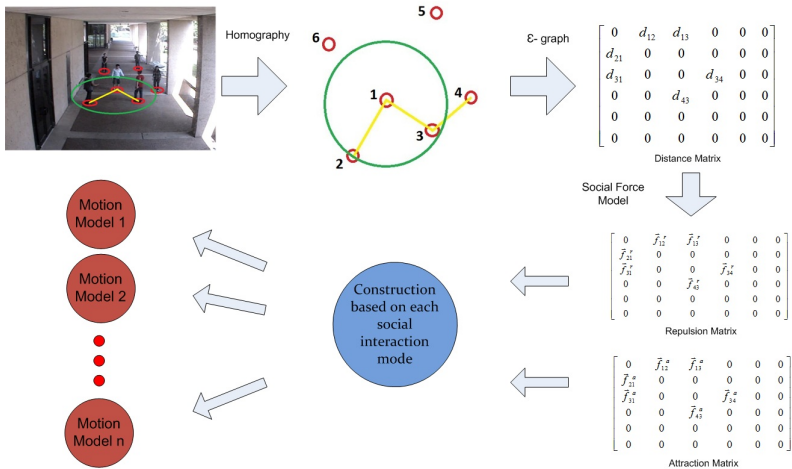

Figure 1: Depiction of the social interaction decomposition framework

The proposed decomposition is computationally unsustainable. Within the context of a Bayesian tracking framework, the posterior distribution for each interaction mode d_n, given a motion prior, can be approximated by constructing a Markov chain for sampling. Hence, an increase in the number of social links leads to an increase in the number of social interaction modes within a set and a corresponding increase in the number of sampling chains. Considering P pedestrians in a scene, if every pedestrian has a social interaction link with each other, then, the number of chains is given by $Q = 2^L = 2^{\frac{P(P-1)}{2}}$, where L is the number of links. There is an exponential growth in the number of chains with respect to the number of pedestrians to be tracked. We address this issue by limiting the number of social links between pedestrians based on the distance between them. The construction of social links is based on the ε-graph [4]. A link is established between pedestrian i and j if $E(i, j) < \varepsilon$ where E is the Euclidean distance in real world coordinates. By adjusting the value of ε, we build sparse links among pedestrians. In this paper, ε is set to be equal to the value of the

forces' boundary b.

2.2 Motion Model with Social Interaction Modes

In a Bayesian context, the tracking problem is to quantify the posterior density $p(x_t|y_{1:t})$, where the observations are specified by $y_{1:t} = y_1, y_2, ..., y_t$. Given the state x_t at time t, the posterior probability is estimated by:

$$p(x_t|y_{1:t}) = cp(y_t|x_t) \int p(x_t|x_{t-1})p(x_{t-1}|y_{1:t-1})dx_{t-1}, \qquad (6)$$

where c is a normalization constant, $p(y_t|x_t)$ specifies the likelihood function of the current observation given the current state, $p(x_t|x_{t-1})$ specifies the probability of the current state given the previous state and the prior probability, and $p(x_{t-1}|y_{1:t-1})$ specifies the previous posterior probability.

The goal of our proposed method is to find the best state $\hat{x}_t$ at time t given the observation $y_{1:t}$. It can be obtained by using the Maximum a Posteriori (MAP) estimate over the M samples at each time t, denoted by:

$$\hat{x}_t = \arg\max_{x_t^\ell} p(x_t^\ell|y_{1:t}) \text{ for } l = 1, \ldots, M \quad , \qquad (7)$$

where x_t^ℓ indicates the ℓth sample of the state x_t. Our method estimates an accurate value of posterior probability by designing a sophisticated motion model $p(x_t|x_{t-1})$. Following the decomposition idea by Kwon and Lee [11], the motion model is designed as the weighted linear combination of its basic components:

$$p(x_t|x_{t-1}) = \sum_{n=1}^{N} w_t^n p_n(x_t|x_{t-1}), \text{ and } \sum_{n=1}^{N} w_t^n = 1, \qquad (8)$$

where $p_n(x_t|x_{t-1})$ denotes the nth basic motion model, w_t^n is the weighting variable at time t, and N is the number of decomposed motion models, which is equal to cardinality of the social interaction set. Hence, $N = q$.

Let $\vec{c}_t = [\vec{c}_{1,t}, \vec{c}_{2,t}, \ldots, \vec{c}_{P,t}]^T$ and $\vec{v}_t = [\vec{v}_{1,t}, \vec{v}_{2,t}, \ldots, \vec{v}_{P,t}]^T$ denote the positions and velocities of the pedestrians, respectively, at time t. Given $\vec{F}^{d_n} = [\vec{F}_1^{d_n}, \vec{F}_2^{d_n}, \ldots, \vec{F}_P^{d_n}]^T$, the position information for pedestrians in the next frame can be predicted. Using a constant velocity motion model, the motion prediction is defined as $\vec{c}_t = \vec{c}_{t-1} + \vec{v}_{t-1}\Delta t$ in which Δt is the time interval between two frames. Incorporating the social interaction force for prediction, the state update at a fixed interval of time Δt is given as:

$$\begin{bmatrix} \vec{c}_t^{d_n} \\ \vec{v}_t^{d_n} \end{bmatrix} = \begin{bmatrix} \vec{c}_{t-1} + \vec{v}_{t-1}\Delta t + \frac{1}{2}\frac{\vec{F}^{d_n}}{m}\Delta t^2 \\ \vec{v}_{t-1} + \frac{\vec{F}^{d_n}}{m}\Delta t \end{bmatrix}, \qquad (9)$$

where $\vec{c}_t^{d_n}$ and $\vec{v}_t^{d_n}$ denote the predicted positions and velocity of pedestrians under the interaction mode d_n, respectively. For a single pedestrian i, the motion prediction $\vec{c}_{i,t}$ is given by a set of locations predicted by each of its social interaction modes, $\vec{c}_{i,t} = \{\vec{c}_{i,t}^{d_1}, \vec{c}_{i,t}^{d_2}, \ldots, \vec{c}_{i,t}^{d_q}\}$. In this work, each basic motion model $p_n(x_t|x_{t-1})$ is modeled as a Gaussian distribution and is given by:

$$p_n(x_t|x_{t-1}) \propto \mathcal{N}(\vec{c}_t; \vec{c}_t^{d_n}, \sigma^2). \qquad (10)$$

2.3   Observation Model

Following Perez *et al.* [16], we use a color observation model based on the Hue-Saturation-Value (HSV) space. Given the initialization for any pedestrian to be tracked (bounding box), we perform a kernel density estimate, $g^* = g(x_0)$, of the color distribution in frame 0. The data likelihood is derived based on the Bhattacharyya similarity coefficient and is defined as $S[g^*, g(x_t)]$. The likelihood of observation, based on Gibbs distribution, is given by:

$$p(y_t|x_t) \propto \exp(-\lambda S^2[g^*, g(x_t)]),\qquad(11)$$

where $\lambda = 20$ as suggested in [16].

2.4   Compound Tracker: Integrating Decomposed Motion Models

Given each decomposed motion model, a basic tracker is composed of a pair of observation and motion models as illustrated in Fig. 2. This generates a basic tracker n that uses the observation model $p(y_t|x_t)$ and the motion model $p_n(x_t|x_{t-1})$, describing a specific social interaction mode d_n. The total number of basic trackers for a pedestrian i is the same as the total number of social interaction modes in the set D_i. A Markov chain is constructed for each of the trackers. The state space is updated according to the MAP estimate obtained via the Metropolis Hasting algorithm [8]. The Interactive Markov Chain Monte Carlo (IM-CMC) algorithm is leveraged to sample across the basic trackers and combine their sampling results. The IMCMC algorithm consists of two main parts: parallel process and interacting process. The Metropolis Hasting algorithm is executed in parallel across all chains during the parallel process while the optimal state value is determined during the interacting process through communication between all chains. The weight w_t^n is implicitly estimated during the interacting process. The implementation used in this paper follows the approaches proposed in Kwon and Lee [11] and Corander *et al.* [6].
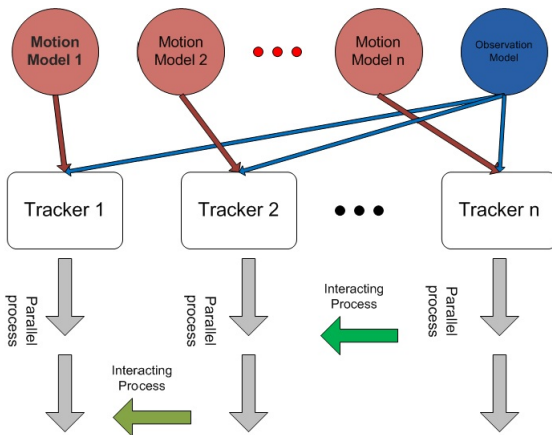

Figure 2: Overview of the proposed tracking framework

Figure 3: The tracking results for selected frames from the BEHAVE dataset: our proposed tracker (row1), BPF (row2), MCMC (row3), and VTD (row4).

3 Experiments

The proposed tracker is evaluated across multiple datasets and compared against several popular visual trackers. Specifically, we compare the tracking results of our method with those of boosted particle filter (BPF) tracker [14], a standard MCMC particle filter tracker [10, 16], and the Visual Tracking Decomposition (VTD) tracker [11]. Since the standard MCMC particle filter only has a single Markov Chain, we perform N iterations of the MCMC tracker where N is the maximum number of Markov Chains used in our method. All the trackers are initialized manually by specifying a bounding box in the first frame and data association is entirely based on the generative observation model without any dynamic update. To initialize the social interaction model, we assume that a pedestrian has a private sphere of radius equal to 0.1 m and a mass of 80 kg. The magnitude of social force is 500 N and the boundary is set to be 3 m. The relaxation time used for all video sequences is 0.2 s. The tracking cycle is equal to the discrete time interval Δt according to every video sequence's frame rate. The variance of motion model, σ^2, is set to 2.

3.1 Datasets

Two video sequences were included from the "**BEHAVE**" Interactions Test Case Scenarios [1]. The videos were acquired at 25 frames/sec and the tracking cycle used is 0.04 s. Two

Figure 4: The tracking results for selected frames from our dataset: our proposed tracker (row1), BPF (row2), MCMC (row3), and VTD (row4).

campus pedestrian sequences from the "**EPFL**" dataset [∎] were also included in our analysis. These videos were acquired at 25 frames/sec and the tracking cycle used is 0.04 *s*. Finally, we also included two video sequences acquired by our team in an outdoor passageway with six pedestrians in the scene. The videos were acquired at 30 frames/sec and the tracking cycle is 0.034 *s*. The resolution of each frame in the video is 704×480 pixels.

3.2   Qualitative Comparison

Across all six video sequences, exhibiting varying environments, our method successfully tracked all pedestrians. An illustrative example from one of the "**BEHAVE**" sequences is shown in Fig. 3. This scene shows two pedestrians approaching a third and eventually forming a group. Four representative frames from each algorithm's tracking results are shown. The proposed tracker and BPF track all pedestrians well before the three pedestrians start exhibiting group interaction. However, only our tracker continues tracking successfully through the group interaction even though there is significant occlusion. The other two trackers fail due to misleading background and poor image quality. Similar observations can be made for the results from the video sequences from the "**EPFL**" dataset. Once again the MCMC and the VTD tracker failed early due to complex background and occlusions. BPF tracked well before the occurrence of partial occlusions. In contrast, our tracker tracks the occluded pedestrians even through abrupt motion changes due to robust prediction based on accountable social interaction modes. Figure 4 shows four representative frames of each algorithm's tracking results from a video sequence acquired by our team in an outdoor pas-

sageway. Due to poor image quality, similar appearance among pedestrians, and partial occlusions, MCMC and VTD trackers lose track of several pedestrians. The proposed tracker and the BPF tracker successfully localize all pedestrians correctly through the video. However, overall, the proposed algorithm exhibits better tracking performance compared to the BPF tracker.

3.3 Quantitative Comparison

To quantitatively compare the result under different scenarios, we manually labeled the ground truth in the six video sequences. The tracking performance is measured based on CLEAR MOT metrics [4]. We report multiple object tracking precision (MOTP), miss rate, false positive rate and number of ID switches. The threshold of building a matched pair between a tracking result and the ground truth is half of the rectangles' diagonal in the ground truth. MOTP measures the ability of the tracker to estimate precise pedestrian positions, which is independent of an algorithm's tracking accuracy. In this paper, MOTP is computed as the average error of matched pairs' center over all frames, measured in pixels.

Tables 1-3 present the results of all four algorithms for each of the video sequences from the "BEHAVE", "EPFL", and our datasets, respectively. Our tracker outperforms all the other trackers in terms of miss rate, false positives rate and ID switches. In terms of MOTP, our tracker outperforms the other trackers in four video sequences and achieves the second best in two other video sequences. Overall, the proposed algorithm tracks multiple pedestrians more robustly by leveraging individual social interaction modes. The results clearly indicate that the improved motion model also contributes to better localization accuracy.

	Seq#1				Seq#2			
Method	MOTP	Misses	False Positives	ID Switches	MOTP	Misses	False Positives	ID Switches
OURS	**3.62**	**0.00 %**	**0.00 %**	**0**	3.21	**0.00 %**	**0.00 %**	**0**
BPF	6.48	54.49 %	54.49 %	0	3.14	18.21 %	18.21 %	0
MCMC	3.86	29.28 %	29.28 %	0	4.99	26.65 %	26.65 %	43
VTD	5.46	44.64 %	44.64 %	12	**2.60**	47.93 %	47.93 %	17

Table 1: BEHAVE dataset results.

	Seq#1				Seq#2			
Method	MOTP	Misses	False Positives	ID Switches	MOTP	Misses	False Positives	ID Switches
OURS	**2.71**	**0.00 %**	**0.00 %**	**0**	**1.22**	**0.00 %**	**0.00 %**	**0**
BPF	4.64	8.92 %	8.92 %	0	3.58	9.57 %	9.57 %	0
MCMC	3.97	27.48 %	27.48 %	0	14.53	50.43 %	50.43 %	5
VTD	4.93	37.12 %	37.12 %	17	3.37	38.55 %	38.55 %	2

Table 2: EPFL dataset results.

4 Conclusion

In this paper, we have proposed a new dynamic model for tracking multiple pedestrians. The method leverages the social interaction decomposition to approximate a broader set of human

	Seq#1				Seq#2			
			False	ID			False	ID
Method	MOTP	Misses	Positives	Switches	MOTP	Misses	Positives	Switches
OURS	**4.67**	**0.32** %	**0.32** %	**0**	4.43	**0.00** %	**0.00** %	**0**
BPF	5.89	1.64 %	1.64 %	1	6.20	**0.00** %	**0.00** %	**0**
MCMC	12.85	26.39 %	26.39 %	122	**3.95**	22.25 %	22.25 %	**0**
VTD	8.57	43.01 %	43.01 %	17	10.18	30.47 %	30.47 %	353

Table 3: Our dataset results.

interaction behaviors in unconstrained environments. To the best of our knowledge, this is the first time the social force model has been extended to include the notion of attraction in human tracking. The proposed dynamic model is decomposed through the construction of multiple basic trackers, each representative of a motion model defined by the specific social interaction mode. An IMCMC framework is used to combine the predictions from the basic trackers to find the best state at each time step. The experimental results demonstrate that the proposed method enables tracking of pedestrians in complex scenes with occlusions and varying interaction behaviors.

Acknowledgements

References

[1] Behave interactions test case scenarios, October 2007. URL http://groups.inf.ed.ac.uk/vision/BEHAVEDATA/INTERACTIONS/.

[2] M. Andriluka, S. Roth, and B. Schiele. People-tracking-by-detection and people-detection-by-tracking. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1–8, Alaska, June 2008.

[3] K. Bernardin and R. Stiefelhagen. Evaluating multiple object tracking performance: the CLEAR MOT metrics. *Journal on Image and Video Processing*, 2008:1–10, 2008.

[4] C.W. Chen, R.C. Ugarte, C. Wu, and H. Aghajan. Discovering social interaction in real work environments. In *Proc. International Workshop on Social Behavior Analysis*, pages 1–6, Santa Barbara, March 2011.

[5] W. Choi and S. Savarese. Multiple target tracking in world coordinate with single, minimally calibrated camera. In *Proc. European Conference on Computer Vision*, pages 553–567, Crete, Greece, September 2010.

[6] J. Corander, M. Ekdahl, and T. Koski. Parallel interacting MCMC for learning of topologies of graphical models. *Data Mining and Knowledge Discovery*, 17(3):431–456, 2008.

[7] F. Fleuret, J. Berclaz, R. Lengagne, and P. Fua. Multi-camera people tracking with a probabilistic occupancy map. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):267–282, February 2008.

[8] W.K. Hastings. Monte Carlo sampling method using Markov Chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[9] D. Helbing and P. Molnár. Social force model for pedestrian dynamics. *Physical Review E*, 51(5):4282–4286, 1995.

[10] Z. Khan, T. Balch, and F. Dellaert. MCMC-based particle filtering for tracking a variable number of interacting targets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(11):1805–1818, 2005.

[11] J. Kwon and K.M. Lee. Visual tracking decomposition. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1269–1272, San Francisco, June 2010.

[12] M. Luber, J.A. Stork, G.D. Tipaldi, and K.O. Arras. People tracking with human motion prediction from social forces. In *Proc. International Conference on Robotics and Automation*, pages 464–469, Alaska, May 2010.

[13] D. Mitzel, E. Horbert, A. Ess, and B. Leibe. Multi-person tracking with sparse detection and continuous segmentation. In *Proc. European Conference on Computer Vision*, pages 397–410, Crete, Greece, September 2010.

[14] K. Okuma, A. Taleghani, and N.D. Freitas. A boosted particle filter: multitarget detection and tracking. In *Proc. European Conference on Computer Vision*, pages 28–39, Prague, Czech Republic, May 2004.

[15] S. Pellegrini, A. Ess, K.Schindler, and L. van Gool. You'll never walk alone: modeling social behavior for multi-target tracking. In *Proc. International Conference on Computer Vision*, pages 261–268, Kyoto, Japan, 2009.

[16] P. Perez, C. Hue, J. Vermaak, and M. Gangnet. Color-based probabilistic tracking. In *Proc. European Conference on Computer Vision*, pages 661–675, Copenhagen, Denmark, May 2002.