SOCIAL FORCE MODEL AIDED ROBUST PARTICLE PHD FILTER FOR MULTIPLE HUMAN TRACKING

Pengming Feng[1], *Wenwu Wang*[2], *Syed Mohsen Naqvi*[1], *Satnam Dlay*[1], *Jonathon A. Chambers*[1]

1. Communications, Sensors, Signal and Information Processing Group, Newcastle University, UK
Emails: {p.feng2, s.m.r.naqvi, satnam.dlay, jonathon.chambers}@newcastle.ac.uk
2. Center for Vision Speech and Signal Processing, University of Surrey, UK
Email: w.wang@surrey.ac.uk

ABSTRACT

In this paper, we propose a novel robust multiple human tracking approach based upon processing a video signal by utilizing a social force model to enhance the particle probability hypothesis density (PHD) filter. In traditional dynamic models, the states of targets are only predicted by their own history; however, in multiple human tracking, the information from interaction between targets and the intentions of each target can be employed to obtain more robust prediction. Furthermore, such information can mitigate the problems of collision and occlusion. The cardinality of variable number of targets can also be estimated by using the PHD filter, hence improving the overall accuracy of the multiple human tracker. In this work, a background subtraction step has also been employed to identify the new born targets and provide the measurement set for the PHD filter. To evaluate tracking performance, sequences from both the CAVIAR and PETS2009 datasets are employed for evaluation, which shows clear improvement of the proposed method over the conventional particle PHD filter.

Index Terms— Social force model, PHD filter, multiple human tracking

1. INTRODUCTION

Multiple human tracking based upon a video signal has potential application in many areas such as monitoring, assistive living and homeland security, where it can be employed to achieve localization and behavioural analysis of human targets. Earlier methods including the Kalman filter [1] and particle filter [2] can perform basic multiple human tracking where the number of targets is assumed to be known and fixed. However, there are still many challenges such as variable number of targets, occlusion, and computational complexity in multiple human tracking [3]. What is more, in fundamental Bayesian filtering methods, the states of the targets are only predicted based upon the history information of the individual targets, and the interactions between targets are not considered.

For a variable number of targets, the random finite set (RFS) [4] based probability hypothesis density (PHD) filter has been recently proposed for multiple human tracking. The advantage of the PHD filter is that it can estimate both the number of targets and their states [5][6][7]. Moreover, it avoids the computational complexity growing exponentially as often occurs in other multiple target tracking approaches such as multiple hypothesis tracking (MHT) [7] by only utilizing the first moment of the posterior distribution rather than the whole distribution. In practice, many researchers have found it beneficial by employing a more effective dynamic model to predict the states of targets [8][9], for example, the social force model [10]. In an ordinary social force model, people are driven by their future destination, taking into account their environment, anticipating collisions, and adjusting their trajectories at an early stage in order to avoid them. In this way, a more accurate dynamic model is built to replace the fixed state model in the Bayesian filtering framework and thereby achieving more accurate tracking results. However, by employing the ordinary social force model [10], the parameters are only used to improve the state model by simply adding them together, in our formulation, we use a Gaussian based social force model to predict the states of the human targets which fits more naturally into the Bayesian framework for multiple human tracking.

In this paper, therefore, a Gaussian based social force model is employed to build a posterior distribution within the prediction stage of the particle PHD filter. In order to identify the new born targets and form the measurement set, a background subtraction step is employed to detect the targets in each frame. To evaluate the performance of our proposed robust PHD filter, two sequences from the CAVIAR [11] and PET2009 [12] datasets are employed which include appearance, occlusion, and disappearance of humans in the field of view of a camera. The results show that our proposed robust particle PHD filter can obtain more accurate results and perform better than a conventional particle PHD filter when tracking a variable number of humans in an enclosed environment.

2. BACKGROUND AND PRELIMINARIES

2.1. PHD filter for multiple human tracking

Based upon the Random Finite Set concept, the PHD filter is employed to address multiple human tracking, where only the first-order moment of the multi-target posterior is propagated instead of the posterior itself [7]. Denoting $\mathbf{D}_{k|k}(\mathbf{x})$ as the PHD filter at discrete time k associated with the multi-target posterior density $p_{k|k}(\mathbf{X}_k|\mathbf{Z}_{1:k})$, where $\mathbf{X}_k = \{\mathbf{x}_k^m, m = 1, ..., M\}$ includes the 2D position of all the human targets, $\mathbf{x}_k^m$ denotes the state of the m^{th} target at time k, M is the number of targets and $\mathbf{Z}_{1:k}$ denotes the measurements up to time k. The PHD prediction step is defined as:

$$\mathbf{D}_{k|k-1}(\mathbf{x}_k^m) = \int \phi_{k|k-1}(\mathbf{x}_k^m, \xi)\mathbf{D}_{k-1|k-1}(\xi)d(\xi) + \Upsilon_k \quad (1)$$

where Υ_k is the intensity function of the new target birth RFS, $\phi_{k|k-1}(\mathbf{x}_k^m, \xi)$ is the analogue of the state transition probability in the single target case which is calculated from

$$\phi_{k|k-1}(\mathbf{x}_k^m, \xi) = e_{k|k-1}(\xi)f_{k|k-1}(\mathbf{x}_k^m|\xi) + \beta_{k|k-1}(\mathbf{x}_k^m|\xi) \quad (2)$$

in which $f_{k|k-1}$ is the multi-target transition density, $e_{k|k-1}(\xi)$ is the probability that the target still exists at time k and $\beta_{k|k-1}(\mathbf{x}_k^m|\xi)$ is the intensity of the RFS that a target is spawned from the state ξ. The PHD update step is defined as [13]:

$$\mathbf{D}_{k|k}(\mathbf{x}_k^m) = \left[p_M(\mathbf{x}_k^m) + \sum_{z \in \mathbf{Z}_k} \frac{\psi_{k,z}(\mathbf{x}_k^m)}{\kappa_k + \langle \psi_{k,z}, \mathbf{D}_{k|k-1} \rangle} \right] \mathbf{D}_{k|k-1}(\mathbf{x}_k^m) \quad (3)$$

where p_M is the missing detection probability, $\psi_{k,z}(\mathbf{x}_k^m) = (1 - p_M)g_k(\mathbf{z}|\mathbf{x}_k^m)$ is the single-target likelihood defining the probability that a measurement $\mathbf{z}$ is generated by a target with state $\mathbf{x}_k^m$ and κ_k is the clutter intensity.

2.2. Social force model

Modeling the behaviour of pedestrians has been an important area of research within multiple target tracking [14]. Pedestrian behaviours have been studied from a crowd perspective, with macroscopic models for pedestrian density and velocity [14]. At the other end of the spectrum, microscopic models deal with individual pedestrians, which are called social force models [15], where pedestrians are assumed to react to energy potentials caused by other pedestrians and static obstacles through a repulsive force, while trying to keep a desired speed and motion direction. In the PHD filter, the social force model can be integrated into the prediction step to obtain more accurate prediction for targets.

By employing $\mathbf{X}_k = \{\mathbf{x}_k^m, m = 1, ..., M\}$ to represent the states of a current set of targets at time k based on the information of position, velocity and walking behaviour [14], the position information $\mathbf{p}_k^m = [p_{k,x}^m, p_{k,y}^m]^T$ and the velocity information $\mathbf{v}_k^m = [v_{k,x}^m, v_{k,y}^m]^T$, are employed to describe the state of target m, which can be used to represent the social force between targets, where $(\cdot)^T$ denotes the transpose operator. By assuming each target intends to avoid collisions with other targets, the social force model for target m is calculated between target m and all other targets n $(n \neq m)$ based upon factors such as distance and angular displacement between m and n, which is represented by $d_k^m(n)$ and $A_k^m(n)$ respectively, and the distance factor $d_k^m(n)$ can be calculated as [14]

$$d_k^m(n) = \|\mathbf{p}_k^m + t\mathbf{v}_k^m - \mathbf{p}_k^n - t\mathbf{v}_k^n\| \quad (4)$$

where $\|\cdot\|$ denotes the Euclidean norm and t is the time interval between time frame $k-1$ and k. Then the angular displacement factor $A^m(n)$ is calculated as

$$A_k^m(n) = 1 + \cos(\phi/2) \quad (5)$$

where ϕ denotes the angle displacement between target m and n based upon the original point of the frame.

Since the intention and ordinary velocity of each target are also considered, the change of velocity U_k^m and the cosine between velocity and destination path $\mathbf{des}^m$ are also calculated as the parameters for the social force model [10]. We assume that each pedestrian m walks towards a destination $\mathbf{des}^m = [des_x^m, des_y^m]^T$, and in doing so tries to maintain a desired speed $\mathbf{u}^m = [u_x^m, u_y^m]^T$, these two components can be described as two energy functions U_k^m and D_k^m, which denote the change of velocity and cosine between current velocity and destination path for target m respectively

$$U_k^m = \|(\mathbf{v}_k^m - \mathbf{u}^m)\| \quad (6)$$

$$D_k^m = \frac{(\mathbf{des}^m - \mathbf{p}_k^m) \cdot \mathbf{v}_k^m}{\|\mathbf{des}^m - \mathbf{p}_k^m\| \cdot \|\mathbf{v}_k^m\|} \quad (7)$$

where $\mathbf{v}_k^m$ denotes the velocity of target m at time k, where k is equal to the frame number.

After calculating the above parameters for the social force model, the overall social force for target m at time k can be written as [14]

$$S_k^m = \sum_{n \neq m} d_k^m(n)A_k^m(n) + \lambda_1 U_k^m + \lambda_2 D_k^m \quad (8)$$

where λ_1 and λ_2 control the influence of the two regularizers. When the social forces for each target are obtained, they can be adopted within the prediction part of the particle filter, in this way, the interactions between the targets are considered. However, the results from the social force model for particle prediction can be further improved by building a posterior distribution within the prediction stage as in Section 3.

2.3. Background subtraction

We detect the target in the video using the background subtraction method [16]. The new-born target for the PHD filter can thereby be obtained more accurately than by selecting particles in the whole frame randomly. What is more, the results from the background subtraction can also be used as the measurement set for likelihood calculation for each target. In this paper, we used a codebook method for background subtraction [17]. Some results from the background subtraction are shown in Fig. 1

Fig. 1: Background subtraction results for three frames in the sequence from CAVIAR dataset, where the green part in the first figure shows the occlusion of two human targets; the red part in the second figure shows the appearance of another human target and the yellow part in the third figure shows the disappearance of a human target.

After obtaining the results from the background subtraction, the center of each foreground region is calculated, which is used to establish an RFS of measurement set $\mathbf{Z}_k$ for the PHD filter, which contains the localization information of each measurement [16].

3. SOCIAL FORCE MODEL AIDED PARTICLE PHD FILTER FOR MULTIPLE HUMAN TRACKING

There are many solutions for (1) and (3), in this work, one of which is obtained using a sequential Monte Carlo method that approximates the PHD with a set of weighted random samples, which is called the particle PHD filter and is the focus of this paper. We use this method because it performs well in the non-Gaussian noise and non-linear model framework, besides, it is straightforward to be adopted with a social force model. After identifying the new born targets from the background subtraction model as described in Section 2.3, the particles are predicted with the state model and a set of random particles associated with the weights, $\mathbf{X}_k = \{\mathbf{x}_k^i, i = 0, ..., N + J_k\}$ is obtained, where N is the number of particles we employed in the particle PHD filter and J_k is the number of particles used to represent the new born targets.

The social force model can then be used to achieve more accurate prediction for the particle PHD filter. Based upon the traditional social force model described in Section 2.2, as introduced in [14], a Gaussian distribution based energy function has been established to describe the social force model for prediction. When a particle $\mathbf{x}_k^{m,i}$ is predicted to represent the state of target m, $\mathbf{x}_k^m$, at time k, its weight is predicted by the social force model with other existing targets. The distance for social force model between $\mathbf{x}_k^{m,i}$ and target n can be then represented as (9), so the larger the distance between the predicted particle and the selected target, the lower energy they have from the distance aspect

$$w_{k,d}^{m,i}(n) = e^{\frac{d_k^{m,i}(n)}{2\sigma_d^2}} \tag{9}$$

where σ_d controls the influence from distance factor on the social force models. $w_{k,d}^{m,i}(n)$ becomes minimum if the linear trajectories collide with each other. Then the angular displacement can be represented as

$$w_{k,\phi}^{m,i}(n) = (A_k^{m,i}(n))^{\beta} \tag{10}$$

where β controls the influence from the direction of the velocity. Based on (9), the influence of multiple subjects can now be modeled as a weighted product, where particle $\mathbf{x}_k^{m,i}$ gets assigned a weight with each target n $(n \neq m)$, namely weight $w_k^{m,i}(n)$ depending on its current distance and angular displacement ϕ [14]

$$w_k^{m,i}(n) = w_{k,d}^{m,i}(n) w_{k,\phi}^{m,i}(n) \tag{11}$$

then the two energy functions $U_k^m(\cdot)$ and $D_k^m(\cdot)$, which denote the change of velocity and cosine between the current velocity and destination path of target m for particle $\mathbf{x}_k^{m,i}$ respectively can also be replaced by two energy functions:

$$E_{k,U}(\mathbf{x}_k^{m,i}) = e^{-\frac{U_k^{m,i}}{2\sigma_v^2}} \quad E_{k,D}(\mathbf{x}_k^{m,i}) = e^{-\frac{D_k^{m,i}}{2\sigma_D^2}} \tag{12}$$

where σ_v and σ_D control the influence of changing the velocity and destination on the social force of the target respectively.

The overall interaction energy for particle $\mathbf{x}_k^{m,i}$ which is predicted to represent the state of target m can be described as

$$S_k(\mathbf{x}_k^{m,i}) = \prod_{n \neq m} w_k^{m,i}(n) E_{k,U}(\mathbf{x}_k^{m,i}) E_{k,D}(\mathbf{x}_k^{m,i}) \tag{13}$$

By utilizing the above equations (9) to (13), the social force weight function is built, which can be used for establishing a prior distribution for sampling the particles. By calculating the social force from other targets, the prior weight for all particles from prediction s_k^i can be given by normalizing S_k. In this way, a posterior distribution for the prediction of all particles is established, where s_k^i is employed as the predicted weight for particle $\mathbf{x}_k^i$, so we can have

$$\{\tilde{\mathbf{x}}_k^i, \tilde{w}_k^i\}_{i=1}^{N+J_k} \tag{14}$$

where $\tilde{w}_k^i = s_k^i$ is obtained from the social force model and '˜' denotes the value from estimation.

After achieving the random measurement set from background subtraction, the PHD filter updating step is employed to update the particles

$$\widetilde{w}_k^i = \left[p_M(\widetilde{\mathbf{x}}_k^i) + \sum_{\forall z \in \mathbf{Z}_k} \frac{(1 - p_M(\widetilde{\mathbf{x}}_k^i))\psi_{k,z}(\widetilde{\mathbf{x}}_k^i)}{\kappa_k(z) + C_k(z)} \right] \widetilde{w}_k^i \tag{15}$$

where

$$C_k(z) = \sum_{i=1}^{N+J_k} \psi_{k,z}(\widetilde{\mathbf{x}}_k^i)\widetilde{w}_k^i \tag{16}$$

and $\psi_{k,z}(\widetilde{\mathbf{x}}_k^i)$ is the likelihood given by the Euclidean distance between the state position and measurement position $p(\mathbf{z}_k^i|\mathbf{x}_k^i)$

$$\psi_{k,z}(\widetilde{\mathbf{x}}_k^i) = p(\mathbf{z}_k^i|\mathbf{x}_k^i) = e^{-\frac{(\tilde{\mathbf{x}}_k^i-\mathbf{z}_k^i)^T(\tilde{\mathbf{x}}_k^i-\mathbf{z}_k^i)}{\sigma_R^2}} \tag{17}$$

and the number of targets is calculated by the sum of the weights for all particles. However, since at each iteration k, J_k new particles are added to the old N particles for the new born targets; to limit the growth of the number of particles, a resampling step is performed after the update step, which is the same as the traditional particle filter introduced in [2].

By the method described above, a novel social force model is employed to establish a posterior distribution for the particle PHD filter, hence improving the prediction for the targets; in the next section, some simulation results will be given to show the accuracy of the proposed system.

4. SIMULATION

In order to evaluate the performance of the proposed social force model aided particle PHD filter, sequences from the CAVIAR and PET2009 datasets are employed. In this work, 300 particles are employed to represent each target and each particle for a target contains $\mathbf{x}_k^m = [p_x^m, p_y^m, v_x^m, v_y^m, h^m, w^m]^T$ which includes the position, velocity and the size information for the targets. The zero-mean noise vector $\mathbf{w}_k$ for prediction in the state model has covariance structure $cov\{\mathbf{w}_k\} = Diag\{25, 25, 16, 16, 4, 4\}$ and for $\mathbf{v}_k$ $cov\{\mathbf{v}_k\} = Diag\{25, 25\}$. The parameters for the social force model are chosen as: $\sigma_d = 0.361$, $\sigma_v = 0.02$, $\sigma_D = 0.9$, $\beta = 2$, $\lambda_1 = 2.33$ and $\lambda_2 = 2.073$.

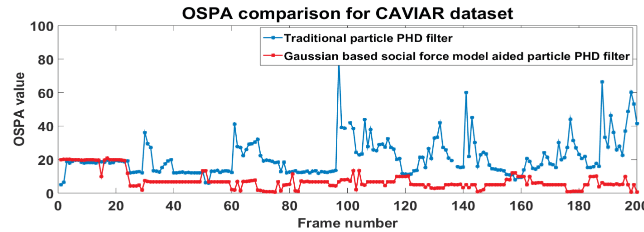
In order to evaluate the proposed tracking system, the sum of error in each frame is compared and shown in Table 1.

Table 1: Comparisons of different tracking results

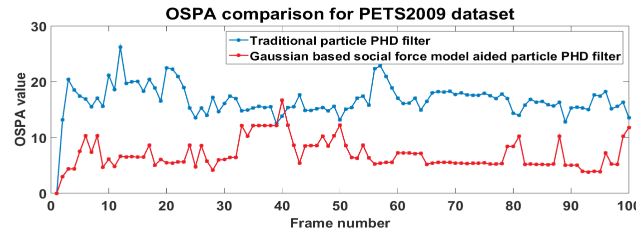
	CAVIAR			PETS2009		
	PHD	SFM-PHD	G-SFM-PHD	PHD	SFM-PHD	G-SFM-PHD
ME(pixel)	38.25	31.81	14.28	72.78	68.25	40.76
Improvement(%)	-	17.84%	55.11%	-	6.22%	40.28%

Here ME is the mean of the sum of error on each frame, PHD denotes the results from the traditional particle PHD filter, SFM-PHD denotes the results from traditional social force model aided particle PHD filter and G-SFM-PHD denotes our proposed Gaussian based social force model aided particle PHD filter. It can be observed that the accuracy is improved by employing our proposed tracking system.

The optimal subpattern assignment (OSPA) metric [18], which is popularly used by researchers is also employed



(a) OSPA comparison for CAVIAR dataset



(b) OSPA comparison for PETS2009 dataset

Fig. 2: Comparison of OSPA value between the social force model aided particle PHD filter and the traditional particle PHD filter. Subfigure (a) is the comparison for the CAVIAR dataset and (b) is for the PETS2009 dataset. The blue line denotes the traditional particle PHD filter from [16], the red line denotes the algorithm proposed in our paper.

to evaluate our tracking system, where the error from both localization and cardinality are considered to evaluate the tracking system. The results are shown in Fig. 2, where the mean of OSPA value is reduced from 21.11 to 7.32 in the CAVIAR dataset and from 16.71 to 6.94 in the PETS2009 dataset, which means our proposed Gaussian based social force model particle PHD filter performs better in both localization and cardinality. However, in the first few frames of the CAVIAR dataset, the targets are moving together towards the same destination, hence the social force model is not applicable, and the results are not improved in those early frames. To make more thorough evaluations, additional sequences have been used in [21] and more measures such as OSPAMT [19] and CLEAR MOT [20] have also been employed for evaluation, the results not shown for space limitation also confirm the improvement and robustness of our proposed tracking system.

5. CONCLUSION

In this paper, we proposed a social force model aided particle PHD filter for multiple human tracking, in which the social force model is represented by a Gaussian model to establish a posterior model in the prediction stage for the particles. The results show the improvement by our proposed method over the conventional particle PHD filter on both the localization and cardinality from the mean of error on each frame and OSPA measure. In future work, an MCMC step will be employed in order to resample the predicted particles and in measurement updating step, a classifier will be utilized to mitigate the measurement noise.

6. REFERENCES

[1] S. M. Kay, *Fundamentals of Statistical Signal Processing*, vol. I, Prentice Hall PTR, 1993.

[2] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman filter: Particle Filters for Tracking Applications*, Artech House, 2004.

[3] D. E. Clark, "Multiple target tracking with the probability hypothesis density filter," PhD thesis, Heriot-Watt University, 2006.

[4] E. Maggio, M. Taj, and A. Cavallaro, "Efficient multitarget visual tracking using random finite sets," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 8, pp. 1016–1027, 2008.

[5] D.E. Clark and J. Bell, "Data association for the phd filter," in *Proceedings of the 2005 International Conference on Intelligent Sensors, Sensor Networks and Information Processing Conference*. IEEE, pp. 217–222, 2005.

[6] K. Panta, B.N. Vo, and S. Singh, "Novel data association schemes for the probability hypothesis density filter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 2, pp. 556–570, 2007.

[7] E. Maggio and A. Cavallaro, *Video Tracking*, John Wiley and Sons, Ltd, 2011.

[8] C. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer Science & Business Media, 2013.

[9] E. Mazor, A. Averbuch, Y. Bar-Shalom, and J.Dayan. Interacting multiple model methods in target tracking: a survey. *IEEE Transactions on Aerospace and Electronic Systems*, 34(1):103–123, 1998.

[10] M. Luber, J. Stork, G.D. Tipaldi, and K. Arras, "People tracking with human motion predictions from social forces," in *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 464–469, 2010.

[11] Caviar dataset. http://homepages.inf.ed.ac.uk/rbf/CAVIAR/.

[12] I. Goldberg and A. Mikhail., Privacy Enhancing Technologies PETS 2009 in *Proceddings of 9th International Symposium , Seattle, WA, USA, August 5-7, 2009. Proceedings, vol. 5672 of Lecture Notes in Computer Science*. Springer, 2009.

[13] A. Doucet, B.N. Vo, C. Andrieu, and M. Davy, "Particle filtering for multi-target tracking and sensor management," *The International Conference on Informantion Fusion, Annaplis, MD*, pp. 474–481, 2002.

[14] S. Pellegrini, A. Ess, K. Schindler, and L.V. Gool, "You'll never walk alone: Modeling social behavior for multi-target tracking," in *IEEE 12th International Conference on Computer Vision*, pp. 261–268, 2009.

[15] D. Helbing, I. Farkas, and T. Vicsek, "Simulating dynamical features of escape panic," *Nature*, vol. 407, no. 6803, pp. 487–490, 2000.

[16] Y. Wang, J. Wu, A. A. Kassim, and W. Huang, "Tracking a variable number of human groups in video using probability hypothesis density," in *Proc. IEEE 18th of the International Conference on Pattern Recognition.*, pp. 1127–1130, 2006.

[17] M. Yu, A. Rhuma, S. M. Naqvi, and J. Chambers, "Fall detection for the elderly in a smart room by using an enhanced one class support vector machine," in *Proc. IEEE 17th of the International Conference on Digital Signal Processing*, pp. 1–6, 2011.

[18] B. Ristic, B. N. Vo, and D. Clark. "Performance evaluation of multi-target tracking using the OSPA metric," in *Proc. IEEE 13th Conference on Information Fusion*, pp. 1–7, 2010.

[19] T. Vu and R. Evans. "A new performance metric for multiple target tracking based on optimal subpattern assignment," in *17th IEEE International Conference on Information Fusion (FUSION)*, pp. 1–8, 2014.

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

[21] P. Feng, W. Wang, S. M. Naqvi, S. Dlay and J. Chambers, "Social force model-aided MCMC-OCSVM particle PHD filter for multiple human tracking," *in preparation for IEEE Transactions on Signal Processing*.