Multi Target Tracking Using Determinantal Point Processes

Felipe Jorquera, Sergio Hernández(✉), and Diego Vergara

Laboratorio Geoespacial, Universidad Católica del Maule, Talca, Chile
f.jorquera.uribe@gmail.com, shernandez@ucm.cl,
diego.vergara.letelier@gmail.com
http://www.ucm.cl

Abstract. Multi Target Tracking has many applications such as video surveillance and event recognition among others. In this paper, we present a multi object tracking (MOT) method based on point processes and random finite sets theory. The Probability Hypothesis Density (PHD) filter is a MOT algorithm that deals with missed, false and redundant detections. However, the PHD filter, as well as other conventional tracking-by-detection approaches, requires some sort of pre-processing technique such as non-maximum suppression (NMS) to eliminate redundant detections. In this paper, we show that using NMS is sub-optimal and therefore propose Determinantal Point Processes (DPP) to select the final set of detections based on quality and similarity terms. We conclude that PHD filter-DPP method outperforms PHD filter-NMS.

Keywords: Multi object tracking · Tracking-by-detection
Determinantal Point Processes

1 Introduction

Multi Object Tracking (MOT) is a challenging computer vision task with applications in video surveillance, event recognition and crowd monitoring among others. Conventional MOT methods based on tracking-by-detection, require an object detection step which locates target detections within an image and associate those detections with trajectories.

The general framework for the object detection step is to test image patches using a sliding window and a trained classifier. Then, redundant responses are usually suppressed by non-maximum suppression (NMS) [10]. Otherwise, object proposals entail a large number of raw detection responses around a true object. However, it is hard to detect occluded objects by this way. The detector is trained to distinguish between target classes and not to differentiate between intra-class variations. Therefore, when multiple objects are occluded, the detector assigns high confidence scores to foreground detections and low confidence scores to other objects. Therefore, the NMS scheme generates false negatives when multiple detections occur closely. On the other hand, if object detection is performed using strong appearance cues, this issue can be alleviated to some extent.

M. Mendoza and S. Velastín (Eds.): CIARP 2017, LNCS 10657, pp. 323–330, 2018.
https://doi.org/10.1007/978-3-319-75193-1_39

Currently, some of the top-performing trackers rely on strong affinity models [7]. Sparse appearance models are used in LINF1 [3], online appearance updates in MHT_DAM [6], integral channel feature appearance models in oICF [5] and aggregated local flow of long-term interest point trajectories in NOMT [1]. Recently, Deep Learning approaches have been proposed for tracking applications, e.g., MDPNN16 [13] uses Recurrent Neural Networks to encode appearance, motion, and interaction, and JMC [15] uses deep matching to improve the affinity measure.

Otherwise, whenever strong or weak detections are to be used, the data association step must handle multiple or missed detections. Current MOT algorithms tackle the data association step into an optimization problem, where the cost function is built upon pairwise similarity costs. While still being able to deliver consistent trajectories, most MOT algorithms based on optimization techniques underestimate the true number of tracks. The Probability Hypothesis Density (PHD) filter is a MOT algorithm that propagates the first-order moment of a multi-object density, using a moment-matching approximation of the true posterior [9]. Therefore, the method has the appealing property of being able to recursively estimate the number of tracks and their locations [11].

In this paper, we propose a tracking-by-detection approach based on the Probability (PHD) filter. Due to the strong dependency of the estimated number of tracks and the number of detections, Determinantal Point Processes (DPP) are used as an alternative to NMS.

2 Determinantal Point Processes

In order to solve the aforementioned problems of NMS, DPPS have been proposed to pedestrian detections [8]. DPPs can select a subset of detections by merging their objectness (detection scores) and individualness (similarity between detections candidates) into a single objective function.

Let N be the number of items and $\mathcal{Z} = \{1, 2, \ldots, N\}$ be the corresponding index set and $Z \subset \mathcal{Z}$ be a subset of indices for the selected items. Each item i is represented by its quality q_i and similarity S_{ij} to another item j. Using their qualities and similarities, we can compute a positive definite kernel matrix $L_Z = [q_i q_j S_{ij}]_{i,j \in Z}$. In this way, the DPP likelihood is a measure for the joint probability $\mathcal{P}_L(Z)$ of the selected indices. Then, we can find the most probable subset by solving the following optimization problem:

$$Z^* = \arg\max_{Z \subset \mathcal{Z}} (det(L_Z)) = \arg\max_{Z \subset \mathcal{Z}} (\prod_{i \in Z} q_i^2) \det([S_{ij}]_{i,j \in Z}). \tag{1}$$

Quality Term. The quality term is defined based on the assumption that wrong detections with high confidence scores degrade the accuracy detection and that a bounding box with a small number of raw detections inside it is more likely to contain ground truth detections [8]. Then, let $s = \{s_1, s_2, \ldots, s_N\}$ be set of scores for N raw detections, s_i^o the detection score, n_i the number of

raw detections inside the current bounding box and λ a penalty constant, the detection score can be computed as $s_i = s_i^o \exp(-\lambda n_i)$. Thus, the quality score can be represented as follows:

$$q = \alpha s + \beta, \tag{2}$$

where α and β are weights for the quality term needed to balance the detection scores of different detectors.

Similarity Term. The similarity term combines appearance and spatial information of detections. Then, appearance is determined by the correlation of feature descriptors of bounding boxes, i.e. if we denote y_i as detection features vector for the i-th detection, the appearance term S^c can be computed as $S^c_{ij} = y_i^T y_j$. Also, the spatial information term is designed to give high correlation to multiple boxes around a single detection. Let S^s be the spatial individualness term, it is defined as $S^s_{ij} = \frac{|\sigma_i \cap \sigma_j|}{\sqrt{|\sigma_i||\sigma_j|}} \in [0, 1]$, where σ_i denote a set of pixel indices belonging to the i-th detection box. Thus, the similarity term is defined merging S^c and S^s into a single diversity feature as follows:

$$S = \mu S^c + (1 - \mu) S^s, \tag{3}$$

where $0 \leq \mu \leq 1$ determines the relative importance of each feature term (Fig. 1).

2.1 Optimization Step

At the eliminating redundant detections step, the DPP mode finding (i.e. optimization for Eq. (1)) can be tackled using the following greedy algorithm [8]:

Algorithm 1. Greedy algorithm for solving (1)

Input : q, S, $\mathcal{Z}$, ϵ
Output: Z^*
$Z^* = \emptyset$
while $\mathcal{Z} \neq \emptyset$ **do**
  $j^* = \arg\max_{j \in \mathcal{Z}} (\prod_{i \in Z^* \cup \{j\}} q^2) \det(S_{Z^* \cup \{j\}})$
  $Z = Z^* \cup \{j^*\}$
  if $P_L(Z)/P_L(Z^*) > 1 + \epsilon$ **then**
    $Z^* \leftarrow Z$
    delete j^* from $\mathcal{Z}$
  else
    break
  end
end

3 PHD Filter - DPP Algorithm Description

The PHD filter is an MOT algorithm that is based on random finite sets and point processes theory [12]. Although we introduce DPPs as a novel preprocessing technique, our implementation follows the standard PHD filter algorithm [16]. In order to compute the quality and similarity terms, we use a Support Vector Machine classifier with Histograms of Oriented Gradients features [2]. Therefore, our MOT algorithm can be summarized as follows:

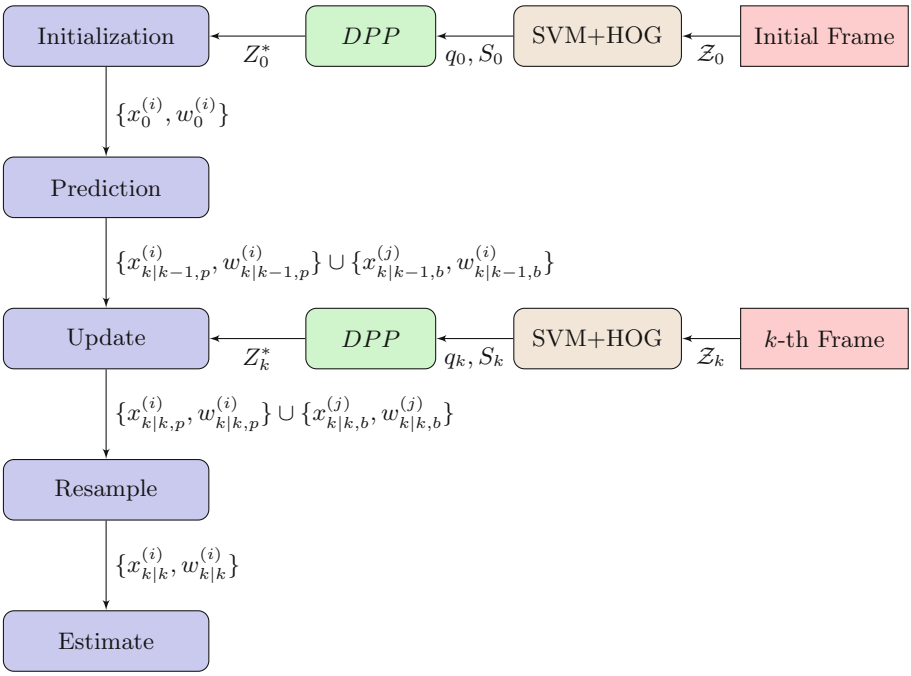


Fig. 1. PHD filter-DPP description.

- **Initialization.** At the initial time $k = 0$, the PHD filter states set $X_0 = \{x_0^{(1)} \ldots, x_0^{(N_0)}\}$ is initialized from the observation set $Z_0^* = \{z_0^{(1)}, \ldots, z_0^{(M_0)}\}$, i.e., DPP results. Also, a weight $w_0^{(i)} = 1/N_0$ is assigned to each one of the states.
- **Prediction.** From each one of the states $x_{k-1}^{(i)}$, we can draw $w_{k|k-1,p}^{(i)} \sim \pi(\cdot|x_{k-1}^{(i)})$ and $w_{k|k-1,p}^{(i)} = p_s w_{k-1}^{(i)}$, where p_s is the probability of survival. We use a measurement driven birth intensity approach. From the observation set Z_k^* (DPP pre-processed detections) a set of new particles is sampled from a birth density $x_{k|k-1,b}^{(j)} \sim b_k(\cdot|z_k^{(j)})$. Then, let v_k be the expected number of newborn objects at time k, weights $w_{k|k-1,b}^{(j)} = v_k/N_k^b$ are assigned to each new born state.

– **Update.** At this step, each state is updated using the DPP observations Z_k^*, according to:

$$w_{k|k,p}^{(i)} = (1 - p_D) w_{k|k-1,p}^{(i)} + \sum_{z \in Z_k^*} \frac{p_D g_k(z|x_{k|k-1,p}^{(i)}) w_{k|k-1,p}^{(i)}}{\mathscr{L}} \tag{4}$$

and

$$\mathscr{L} = \kappa + \sum_{i=1}^{N_k^b} w_{k|k-1,b}^{(i)} + \sum_{i=1}^{N_k} p_D g_k(z|x_{k|k-1,p}^{(i)}) w_{k|k-1,p}^{(i)}, \tag{5}$$

where κ is a constant probability of clutter and p_D is also a constant probability of detection.

– **Resample.** Given a threshold T, the states with weights $w_k^{(i)} < T$ are pruned. Then, the weights for the surviving particles are normalized to get the expected number targets $N_{k|k} = \sum_i w_k^{(i)}$.

– **Target State Estimation.** Cluster particles $x_{k|k}^{(i)}$ using the E-M algorithm to get $N_{k|k}$ states. Connect each one of the states at time k to only one of the tracks collected until time $k-1$.

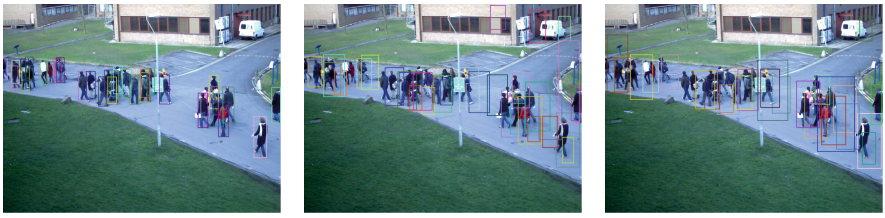
4 Experiments

Firstly, we discuss the experimental settings for evaluating the proposed and existing methods, and then present the empirical results.

4.1 Experimental Settings

In order to demonstrate the advantages of the proposed method, we use the OSPA distance proposed by Schuhmacher et al. [14], which can be interpreted as a combination of two components referred to "localization" and "cardinality" errors.

Also, the proposed and existing methods were evaluated on sequence S1L1 (see Fig. 2) from PETS2009 [4] dataset, which contains 190 frames and exhibit regular crowd movement in a relatively dense queue in two directions.



(a) (b) (c)

Fig. 2. Sequence S1L1 (PETS2009). (a) Ground-truth, (b) DPP, (c) NMS

4.2 Evaluation Results

In Table 1 the DPP parameters are shown, where α and β are the weight constants needed to balance detection scores of different detectors, ϵ is the acceptance parameter for Algorithm 1, λ is the penalty constant for overlap between detection and μ determine the importance of each detection feature descriptor. Also, Table 2 shows parameters of the SVM+HOG pedestrian detector, where G is the coefficient to regulate the NMS threshold and H is the threshold distance between detection features and SVM classifying plane. Furthermore, in Table 3 PHD Filter parameters are shown, where N_0 is the initial number of states, v_k is the expected number of newborn particles, κ is the probability of clutter, p_S is the survival probability and T is the resample threshold.

Table 1. DPP

Parameter	Values
Alpha α	0.9
Beta β	1.1
Epsilon ϵ	0.1
Lambda λ	-0.1
Mu μ	0.2, 0.4, 0.6, 0.8, 1.0

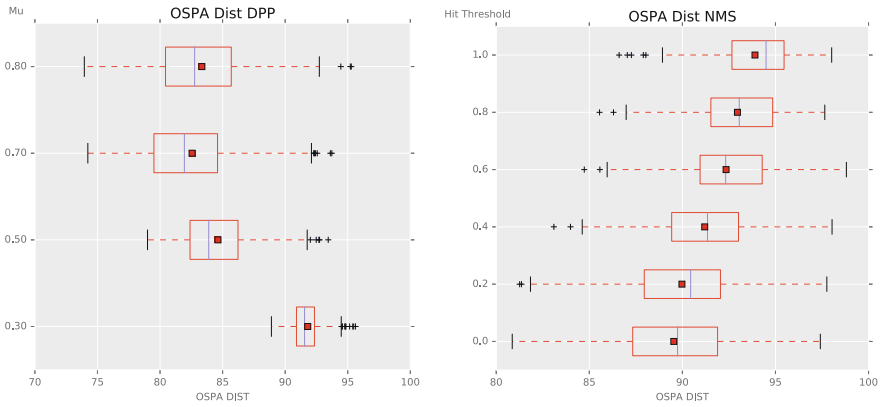
Table 2. SVM-HOG

Parameter	Values
Group Threshold G	1
Hit Threshold H	0.0, 0.2, 0.4, 0.6, 0.8, 1.0

Table 3. PHD Filter

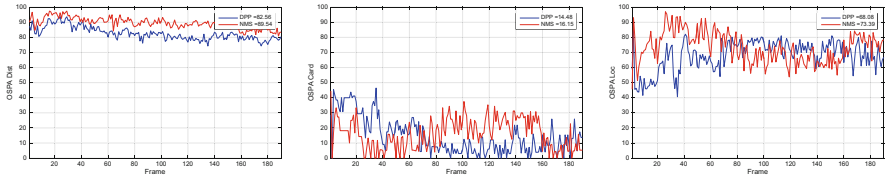
Parameter	Values
Initial particles number N_0	100
Newborn objects number v_k	100
Probability of survival p_S	0.9
Probability of clutter κ	10^{-3}
Probability of detection p_D	0.7
Resample threshold T	1000

We note the correlation between the performance of the PHD filter-DPP method and the μ parameter (see Fig. 3a). Best performance of the proposed approach is achieved when $\mu = 0.7$, then we can conclude that even implementing strong features (HOG), solely using appearance individualness is not robust



(a) (b)

Fig. 3. OSPA distance comparison



(a) OSPA Dist (b) OSPA Card (c) OSPA Loc

Fig. 4. Performance comparison between DPP v/s NMS

enough. Also, from Fig. 3b we can see a high correlation between H parameter and NMS performance. Best performance is achieved when $H = 0.0$.

Therefore, we set parameters specified by Tables 1, 2 and 3 and optimal values for μ and H ($\mu = 0.7$, $H = 0.0$, see Fig. 3). Thus, we note that proposed approach outperforms to the existing method (see Fig. 4).

Both implementations were developed by using C++ with OpenCV 3.1 and can be found in the author repository[1].

5 Conclusion

In this paper, we present a novel multi-target tracking method based on tracking-by-detection approach using DPP to introduce individualness and similarity between detections. The results show that the suppression of redundant detections using the proposed method outperforms NMS.

Acknowledgments. This work was supported by CONICYT/FONDECYT grant, project **Robust Multi-Target Tracking using Discrete Visual Features**, code 11140598.

References

1. Choi, W.: Near-online multi-target tracking with aggregated local flow descriptor. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3029–3037 (2015)
2. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, vol. 1, pp. 886–893. IEEE (2005)
3. Fagot-Bouquet, L., Audigier, R., Dhome, Y., Lerasle, F.: Improving multi-frame data association with sparse representations for robust near-online multi-object tracking. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9912, pp. 774–790. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46484-8_47

[1] http://github.com/fjorqueraurible/multitarget-tracking.

4. Ferryman, J., Shahrokni, A.: Pets 2009: dataset and challenge. In: 2009 Twelfth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS-Winter), pp. 1–6. IEEE (2009)
5. Kieritz, H., Becker, S., Hübner, W., Arens, M.: Online multi-person tracking using integral channel features. In: 2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 122–130. IEEE (2016)
6. Kim, C., Li, F., Ciptadi, A., Rehg, J.M.: Multiple hypothesis tracking revisited. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 4696–4704 (2015)
7. Leal-Taixé, L., Milan, A., Schindler, K., Cremers, D., Reid, I., Roth, S.: Tracking the trackers: an analysis of the state of the art in multiple object tracking. arXiv preprint arXiv:1704.02781 (2017)
8. Lee, D., Cha, G., Yang, M.-H., Oh, S.: Individualness and determinantal point processes for pedestrian detection. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9910, pp. 330–346. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46466-4_20
9. Mahler, R.P.: Multitarget Bayes filtering via first-order multitarget moments. IEEE Trans. Aerosp. Electron. Syst. **39**(4), 1152–1178 (2003)
10. Pirsiavash, H., Ramanan, D., Fowlkes, C.C.: Globally-optimal greedy algorithms for tracking a variable number of objects. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1201–1208. IEEE (2011)
11. Ristic, B., Clark, D., Vo, B.N., Vo, B.T.: Adaptive target birth intensity for PHD and CPHD filters. IEEE Trans. Aerosp. Electron. Syst. **48**(2), 1656–1668 (2012)
12. Ristic, B.: Particle Filters for Random Set Models. Springer, Heidelberg (2013). https://doi.org/10.1007/978-1-4614-6316-0
13. Sadeghian, A., Alahi, A., Savarese, S.: Tracking the untrackable: learning to track multiple cues with long-term dependencies. arXiv preprint arXiv:1701.01909 (2017)
14. Schuhmacher, D., Vo, B.T., Vo, B.N.: A consistent metric for performance evaluation of multi-object filters. IEEE Trans. Signal Process. **56**(8), 3447–3457 (2008)
15. Tang, S., Andres, B., Andriluka, M., Schiele, B.: Multi-person tracking by multicut and deep matching. In: Hua, G., Jégou, H. (eds.) ECCV 2016. LNCS, vol. 9914, pp. 100–111. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-48881-3_8
16. Vo, B.N., Singh, S., Doucet, A.: Sequential Monte Carlo methods for multitarget filtering with random finite sets. IEEE Trans. Aerosp. Electron. Syst. **41**(4), 1224–1245 (2005)