exeley INTERNATIONAL JOURNAL ON SMART SENSING AND INTELLIGENT SYSTEMS

Article | DOI: 10.21307/ijssis-2020-001 Issue 1 | Vol. 13 (2020)

Robust single target tracking using determinantal point process observations

S. Hernández[1]* and P. Sallis[2]

[1]Laboratorio de Procesamiento de Información Geoespacial, Universidad Católica del Maule, Av. San Miguel, 3605, Talca, Chile.

[2]Auckland University of Technology, Auckland, New Zealand.

*E-mail: shernandez@ucm.cl

The paper was edited by Anindya Nag.

Received for publication June 5, 2019.

Abstract

The efficiency and robustness of modern visual tracking systems are largely dependent on the object detection system at hand. Bernoulli and Multi-Bernoulli filters have been proposed for visual tracking without explicit detections (image observations). However, these previous approaches do not fully exploit discriminative features for tracking. In this paper, we propose a novel Bernoulli filter with determinantal point processes observations. The proposed observation model can select groups of detections with high detection scores and low correlation among the observed features; thus achieving a robust filter.

Keywords

Visual tracking, Bernoulli filter.

Visual tracking is a challenging computer vision task with applications in human-computer interaction, video surveillance and crowd monitoring among others. Modern visual tracking systems may use complex object detection schemes for estimating the current state of a target in any particular video frame. However, this approach does not fully exploit the temporal structure of the estimation problem. Visual tracking can be also thought of as a dynamic model with observed features and latent states representing the position/velocity of an object (Maggio and Cavallaro, 2011). In this context, the generative model for visual tracking requires not only the correct specification of the model and its parameters but also the ability to capture the variations of the system (Wang et al., 2015).

The Bernoulli filter is a powerful algorithm that allows objects to appear and disappear, using extracted features from the image as observations (Vo et al., 2010). Similar approaches for visual tracking have been proposed (also known in the literature as Track-Before-Detect). Nevertheless, these methods rely on unreliable background subtraction operations or the likelihood function being in a *separable* form (Hoseinnezhad et al., 2012, 2013).

Current state-of-the-art trackers are based on either correlation filters (Bolme et al., 2010), deformable parts models (Hare et al., 2016) or convolutional neural networks (Li et al., 2018). These trackers learn a discriminative model from a single frame and then update the model using new frames. Furthermore, tracking performance can be increased when using more discriminative features such as HOG (Henriques et al., 2015; Solis Montero et al., 2015; Xu et al., 2019). On the other hand, even when Bernoulli filters have demonstrated being useful models for tracking in complex scenarios, it is still hard to rely on such features for increasing their performance.

Related works

The Bernoulli filter is a specialized version of the PHD filter (Mahler, 2003), with the focus on single target tracking. While the original PHD filter is based on a Poisson point process, several extensions have been proposed to cope with non-Poisson distributions. In particular, the Cardinalized PHD filter allows estimating the number of targets using arbitrary distributions and provides improved estimates (Mahler, 2007). The multi-Bernoulli and Poisson multi-Bernoulli mixture filters also allow to approximate the cardinality distribution and become especially well suited when the mean of the multi-target posterior is higher than the variance

(García-Fernández et al., 2018). All of these methods rely on first-order or second-order moments but assume that targets behave independently with each other. Therefore, the authors in Privault and Teoh (2019) propose a second-order filter that accounts for interaction between the targets. The method is based on determinantal point processes (DPP) that take into consideration the correlation among the targets through a kernel function. In Jorquera et al. (2017), the authors propose a determinantal point process for pruning the components of the Gaussian mixture PHD filter. More recently, the authors in Jorquera et al. (2019) compared the PHD filter using determinantal point process observations with other methods for visual multi-target tracking.

The contributions of this paper are twofold. First, the third section provides introductory notions of the Bernoulli filter and then we derive a novel Bernoulli filter using determinantal point process observations (B-DPP filter) for single target tracking in the fourth section. Second, in the fifth section we derive a Sequential Monte Carlo implementation of the B-DPP filter using a truncated likelihood, which can outperform other discriminative trackers in several scenarios.

Point processes for visual object tracking

A *point process* is a random pattern of points in a possibly multi-dimensional space (Kingman, 1993). A simple point process can be defined in one dimension, which is usually times and can be used to describe the random times where the events can occur with no coincident points.

Bernoulli point process

The problem of performing joint detection and estimation of multiple objects has a natural interpretation as a *dynamic point process*, where the stochastic intensity of the model is a space-time function $\lambda(x)$, where $x \in \mathbb{R}^d$ denotes the state space of the target. If we let $\mathbf{B} = B_1 \cup B_2 \cup \ldots \cup B_k$ represent the union of disjoint video frames B_i, the corresponding number of objects on each image can be written as $N(B_1), N(B_2), \ldots, N(B_k)$. The Bernoulli point process for a single object that can randomly appear or disappear takes the form:

$$p\left(N(B_1)=n_1,\ldots,N(B_k)=n_k\right) = \frac{n!}{n_1!\ldots n_k!}\prod_i^k\left(\frac{\lambda(x_i)}{\Lambda(B)}\right)^{n_i} = \frac{n!}{n_1!\ldots n_k!}\prod_i^k p(x_i)^{n_i}$$

where n_i can take either 1 or 0, $n=\sum n_i$ and $\Lambda(\mathbf{B}) = \int_{\mathbf{B}}\lambda(x)dx$. Every subset B_i can take at most one target x with probability q, therefore we can characterize the distribution of the point process $X=\{x\}$ using the following relationship:

$$p(X) = \begin{cases} 1-q & \text{if } X = \emptyset \\ qp(x) & \text{if } X = \{x\} \end{cases} \tag{1}$$

Determinantal point process

In recent years, deep learning approaches have demonstrated outstanding performance in several visual tracking benchmarks (Kristan et al., 2019). These trackers are mostly based on extracted features from a convolutional neural network and an objective loss that minimizes a localization error (Li et al., 2018). However, the detection process is not perfect and false positives and negatives are to be encountered after ranking the top proposals from the convolutional features.

In order to develop an stochastic approach for the single-object observation model, a discrete DPP can be used to capture probabilistic relationships using a kernel matrix $K : \mathcal{Z}\times\mathcal{Z}\mapsto\mathbb{R}$ that measures the similarity among different detections (Lee et al., 2016). Therefore, instead of considering independent detections in a particular frame, the DPP likelihood specifies the joint probability over all 2^n subsets of $\mathcal{Z}$ with distribution:

$$p(Z\subset\mathcal{Z}) = \det(K_Z), \qquad \forall Z\subset\mathcal{Z} \tag{2}$$

where Z is a random subset of $\mathcal{Z}$ and $K_Z \equiv [K_{i,j}]$ for all $i,j\in\mathcal{Z}$. Furthermore, the product density can also be written in terms of a positive definite matrix $L = K(I-K)^{-1}$, such that the probability mass function of Z can be written as:

$$p(Z) = \frac{\det(L_Z)}{\det(I+L)} \tag{3}$$

where I is the identity matrix and L_Z is a sub-matrix of L indexed by the elements of Z.

Bernoulli filter

In this case, a model for detection and estimation of multiple objects can be achieved by the conditional expectation of the posterior point process (random finite set) under transformations (Ristic et al., 2013).

Let $X_k=\{x\}$ be a Bernoulli point process and $Z_k=\{z_1, z_2, \ldots, z_m\}$ a DPP observed from frame K. The result from superposition, translation and thinning

transformations is also a Bernoulli point process $X_k \sim p(X_k|X_{k-1})$ (Kingman, 1993). The predicted point process can be written as the linear superposition of a π_s thinned point process with Markov translation $f(x|x')$ and a π_b Bernoulli birth process. The predicted expected number of targets $N_{k|k-1}$ for a single target with probability of survival $\pi_s(x)$ and spontaneous birth can be written as:

$$N_{k|k-1} = N^s_{k|k-1} + N^b_{k|k-1} \tag{4}$$

where:

$$N^s_{k|k-1} = \pi_s \int f(x \mid x') p_{k-1|k-1}(\{x'\}) dx$$

$$N^b_{k|k-1} = \pi_b \int p_b(x \mid \emptyset) p_{k-1|k-1}(\emptyset) dx.$$

The filtering density of a Bernoulli point process is completely specified by the pair $(p_{k|k-1}, q_{k|k-1})$, which is obtained by:

$$p_{k|k-1}(X') = \begin{cases} 1 - q_{k-1|k-1} & \text{if } X' = \emptyset \\ q_{k-1|k-1} & \text{if } |X'| = 1 \end{cases} \tag{5}$$

Using Equation (5), the probability of existence $q_{k|k-1}$ can be written as:

$$q_{k|k-1} = \pi_b \left(1 - q_{k-1|k-1}\right) + \pi_s q_{k-1|k-1} \tag{6}$$

And the probability of the predicted Bernoulli point process:

$$\begin{aligned} q_{k|k-1} p_{k|k-1}(\{x\}) = {} & \pi_b \left(1 - q_{k-1|k-1}\right) p_b(x) \\ & + \pi_s q_{k-1|k-1} \int f(x \mid x') p_{k-1|k-1}(\{x'\}) dx' \end{aligned} \tag{7}$$

If we let Z_k be the observations that contain both false detections and target originated measurements, the update equation considers the probability of observing the target with probability of detection π_d under clutter (e.g. false positives). From (Mahler, 2003, 2007), the multi-target likelihood function for the *standard measurement model* (Poisson distributed clutter with density $\kappa_p(Z_k) = e^{-\lambda} \prod_i \lambda f_c(z_i)$ and Bernoulli probability of detection π_d) can be written as:

$$p(Z_k \mid X_k) = \kappa_p(Z_k) (1 - \pi_d)^{|X_k|} \sum_{\sigma} \prod_i \frac{\pi_d \, p\left(z_{\sigma_i} \mid x_i\right)}{(1 - \pi_d) \lambda f_c\left(z_{\sigma_i}\right)} \tag{8}$$

The likelihood term in Equation (8) considers all possible locations and location-to-track associations σ, so most of the terms will be canceled. The likelihood term becomes:

$$p(Z_k \mid \{x\}) = \kappa_p(Z_k)(1 - \pi_d) + \pi_d \sum_{z \in Z_k} \prod_i \frac{p(z_i \mid x)}{\lambda f_c(z_i)} \tag{9}$$

The Bayes update equation takes the form:

$$p(X_k \mid Z_k) = \frac{p(Z_k \mid X_k) p(X_k \mid Z_{1:k-1})}{p(Z_k \mid Z_{1:k-1})} \tag{10}$$

The denominator of Equation (10) can be written as:

$$\begin{aligned} p(Z_k \mid Z_{1:k-1}) = f_c(Z_k) \Bigg\{ & 1 - q_{k|k-1} + q_{k|k-1} (1 - \pi_d)^{M_k} \\ & + \sum_{Z \in Z_k} \psi_k \frac{\int \prod_i p(z_i \mid x) p_{k|k-1}(x) dx}{\prod_j f_c(z_j)} \Bigg\} \end{aligned} \tag{11}$$

where:

$$\psi_k = \frac{M_k!}{(M_k - |Z|)!} \frac{\pi_d^{|Z|}}{(1 - \pi_d)^{|Z| - M_k}}.$$

The updated binomial point process can be derived as follows:

$$q_{k|k} = \frac{1 - \Delta_k}{1 - q_{k|k-1} \Delta_k} q_{k|k-1} \tag{12}$$

where:

$$\Delta_k = 1 - (1 - \pi_d)^{M_k} - \sum_{z \in Z_k} \psi_k \frac{\int \prod_i p(z_i \mid x) p_{k|k-1}(x) dx}{\prod_j \lambda f_c(z_j)} \tag{13}$$

and

$$p_{k|k}(x) = \frac{(1 - \pi_d)^{M_k} + \sum_{z \in Z_k} \psi_k \frac{\int \prod_i p(z_i \mid x) p_{k|k-1}(x) dx}{\prod_j f_c(z_j)}}{1 - \Delta_k} \tag{14}$$

Determinantal filter

Let $X_k = \{x\}$ be a Bernoulli point process and $Z_k = \{z_1, z_2, \ldots, z_m\}$ a DPP observed at frame K. The result from superposition, translation and thinning transformations is also a Bernoulli point process

$X_k \sim p(X_k|X_{k-1})$ (Ristic et al., 2013). The predicted point process can be written as the linear superposition of a π_s thinned point process with Markov translation $f(x|x')$ and a π_b Bernoulli birth process. In order to measure the quality of the observations, we must introduce a random variable L such that $p(L|Z) \propto det(L(Z))$, where $L(Z)$ is a positive definite kernel matrix that depends on the observed features Z. The $L(Z)$ kernel can be written as a Gram matrix:

$$L_{ij}(Z) = g_x(z_i)\phi(z_i)^T \phi(z_j) g_x(z_j) \tag{15}$$

$$= g_x(z_i) S_{ij}(Z) g_x(z_j) \tag{16}$$

The function $g_x(z_i) = \sum_c p(z_i|c)p(c|x)$ is used to model the quality of the item z_i and $S(Z)$ the diversity of the set Z. If we let W be a subset of detections arising from the target (Reuter et al., 2013):

$$\eta(W \mid \{x\}) \approx \begin{cases} (1-\pi_d) & \text{if } W = \emptyset \\ \pi_d g_x(w_1) \det(S_{w_1}) & \text{if } |W| = 1 \\ |W|! \pi_d^m \prod_i g_x^2(w_i) \det(S_W) & \text{if } |W| = m \end{cases} \tag{17}$$

The DPP Z can be treated as the union of two independent sets $Z = C \cup W$, where $C = \{c_1, \ldots, c_m\}$ represents clutter. The clutter density becomes:

$$\kappa_d(C) = |C|! \prod_i f_c^2(c_i) \det(S_C) \tag{18}$$

The likelihood function for the standard measurement model using determinantal observations becomes:

$$P(Z \mid \{x\}) = \sum_{W \subseteq Z} \eta(W \mid \{x\}) \kappa_d(Z \setminus W) \tag{19}$$

$$p(Z_k \mid \{x\}) = \kappa_d(Z_k) \Bigg[(1-\pi_d)^{M_k} + \sum_{Z \in Z_k} \frac{|Z|!(M_k - |Z|)!}{M_k!} \pi_d^{|Z|} (1-\pi_d)^{M_k - |Z|} \prod_i \left[\frac{g_x(z_i)}{f_c(z_i)} \right]^2 \frac{\det(S_Z)\det(S_{Z_k \setminus Z})}{\det(S_{Z_k})} \Bigg] \tag{20}$$

Now, we want to derive the posterior distribution for Bernoulli point process given DPP observations:

$$p(Z_k) = \kappa_d(Z_k) \Bigg[(1-q_{k|k-1}) + q_{k|k-1}(1-\pi_d)^{M_k} + \sum_{Z \in Z_k} \Xi_k \frac{\int \prod_i g_x^2(z_i) p_{k|k-1}(x) dx}{\prod_j f_c^2(z_j)} \Bigg] \tag{21}$$

where:

$$\Xi_k = \frac{|Z|!(M_k - |Z|)!}{M_k!} \frac{\pi_d^{|Z|}}{(1-\pi_d)^{|Z| - M_k}} \frac{\det(S_Z)\det(S_{Z_k \setminus Z})}{\det(S_{Z_k})}.$$

The updated binomial point process can now be derived as follows:

$$q_{k|k} = \frac{1 - \breve{\Delta}_k}{1 - q_{k|k-1}\breve{\Delta}_k} q_{k|k-1} \tag{22}$$

where:

$$\breve{\Delta}_k = 1 - (1-\pi_d)^{M_k} - \sum_{z \in Z_k} \Xi_k \frac{\int \prod_i p(z_i \mid x) p_{k|k-1}(x) dx}{\prod_j f_c(z_j)} \tag{23}$$

and:

$$p_{k|k}(x) = \frac{(1-\pi_d)^{M_k} + \sum_{z \in Z_k} \Xi_k \frac{\int \prod_i g_x(z_i) p_{k|k-1}(x) dx}{\prod_j f_c(z_j)}}{1 - \breve{\Delta}_k} \tag{24}$$

Approximated Bernoulli determinantal filter

In practice, it is difficult to store and compute the power set with all possible configurations of Z_k in the likelihood term (see Equation (20)). An approximation can be constructed by truncating the likelihood and focusing only on the more likely elements. Let $Z_k^* = \arg\max_{Z \subset Z_k} \eta(Z|\{x\})$ be a subset of Z_k whose elements are detections arising from the target. The likelihood becomes:

$$p(Z_k^* \mid \{x\}) = |Z_k^*|! \prod_i g_x^2(z_i) \det(Z_k^*) \tag{25}$$

DPPs have been proposed in the literature as an alternative to other object refinement techniques such as non-maximum suppression (Lee et al., 2016). These methods operate over object proposals and eliminate redundant detections. For DPPs, mode

finding can be tackled using the following greedy algorithm (Kulesza and Taskar, 2011):

Algorithm 1 Greedy Mode Finding

Require: $q_x, S, \mathcal{Z}, \epsilon$

$Z_k^* = \emptyset$

while $Z_k^* \neq \emptyset$ **do**

$z^* = \arg\max_{z \subset Z_k} \left(\prod_{i \in Z_k^* \cup \{z\}} g_x^2(z_i) \right) \det(Z_k^* \cup \{z\})$

$Z = Z_k^* \cup \{z^*\}$

if $p(Z)/(Z_k^*) > \epsilon$ **then**

$Z_k^* = Z$

else

Break

end if

end while

Conversely, by using the truncated likelihood from Equation (25), the Sequential Monte Carlo algorithm for the Bernoulli filter can be used to estimate the single-target posterior (Ristic, 2013).

Experimental results

In order to demonstrate the advantages of the proposed model updating approach over other discriminative approaches, we evaluate the tracking results on six challenging video sequences from the Visual Object Challenge 2014 (VOT) data set[1]. The proposed SMC implementation uses local binary patterns (LBP) as observed features and a simple observation model $p(z_i|c) \propto exp(-\frac{D_k^2}{2\sigma_o^2})$, with $D_k = dist[z_c, z_k]$ and z_c being a reference LBP histogram (Czyz et al., 2007). The state x_k is configured as a 4-dimensional rectangle including the left-most position, width and height of the target. The dynamic model uses a random walk and the parameters of the model are held fixed for all sequences. The B-DPP filter is implemented in the C++ language using the OpenCV library. The parameters for the B-DPP filter are determined empirically and shown in Table 1.

The parameter setting the Greedy Mode Finding algorithm is described in Table 2.

The sequence *jogging* is a challenging example containing full occlusions, rotations and background clutter. Figure 1 shows one frame of the sequence and the estimates using the proposed approach and other state-of-the-art methods.

The Bernoulli DPP filter maintains a balance between the observed features and the quality of the observations (see Figure 1). The observation model uses a simple histogram comparison and no template update is performed, so the model is not robust to object deformation or rotation. Even that, as seen in Figure 1 the Bernoulli-DPP tracker achieves good performance in cases such as full occlusion where the other discriminative tracking methods fail. Performance is measured using widely used precision and success metrics[2].

The precision metric describes the percentage of frames whose center location error is below a given threshold. Table 3 shows the overall precision metric averaged over all sequences on five different runs for each one of the algorithms.

The success measure accounts for bounding box overlap. Table 4 shows the number of success frames whose overlap is above some threshold, averaged over the sequences on five different runs. Quantitative analysis shows improved performance for the proposed approach when compared to the discriminative trackers in six different video sequences.

Figure 2 shows the precision metric against the location error threshold for all of the six tested sequences. The red line indicates the best performing method among the four different algorithms. Since the *bolt* and *jogging* sequences have background clutters (the background near the target has similar appearance as the target), the proposed Bernoulli DPP tracker reduces redundant observations and improves precision.

Figure 3 shows the ratio of the frames whose tracked box has more overlap with the ground-truth box than a threshold. The success metric can be associated with the tracker algorithm ability to

Table 1. Particle Bernoulli-DPP filter.

Particle Bernoulli-DPP filter	
Number of particles N	100
Uniform birth probability (π_b)	0.1
Uniform survival probability (π_s)	0.99
Newborn particles (N_b)	0
Standard deviation for observation model (σ_o)	20.4
Covariance matrix for dynamic model ($\sigma_x \times 1$)	3.0×1

Table 2. Greedy mode finding.

Greedy mode finding	
Acceptance ratio ε	0.7

[1] www.votchallenge.net/vot2014/

[2] http://cvlab.hanyang.ac.kr/tracker_benchmark/benchmark_v10.html

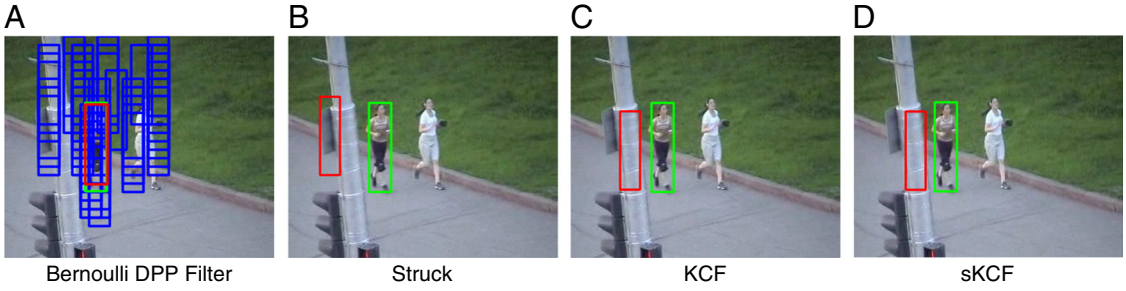

Figure 1: Frame 85 of the *jogging* sequence. At each frame, a greedy mode finding step is performed using Algorithm 1. Rectangles represent ground-truth, state estimates and DPP observations.

Table 3. Average precision (th = 20).

Sequence	DPP	KCF	sKCF	Struck
Ball	0.309	0.289	0.246	*0.372*
Bolt	*0.083*	0.017	0.017	0.026
Diving	0.073	0.082	0.087	*0.091*
Gymnastics	*0.710*	0.425	0.425	0.435
Jogging	*0.707*	0.231	0.231	0.228
Polarbear	*0.946*	0.857	0.916	0.844

th = threshold

Table 4. Average success (th = 0.5).

Sequence	DPP	KCF	sKCF	Struck
Ball	0.206	*0.211*	0.201	0.128
Bolt	*0.031*	0.011	0.011	0.017
Diving	*0.183*	0.110	0.114	0.151
Gymnastics	*0.560*	0.415	0.420	0.425
Jogging	0.205	*0.225*	*0.225*	*0.225*
Polarbear	0.749	0.747	*0.760*	0.712

th = threshold

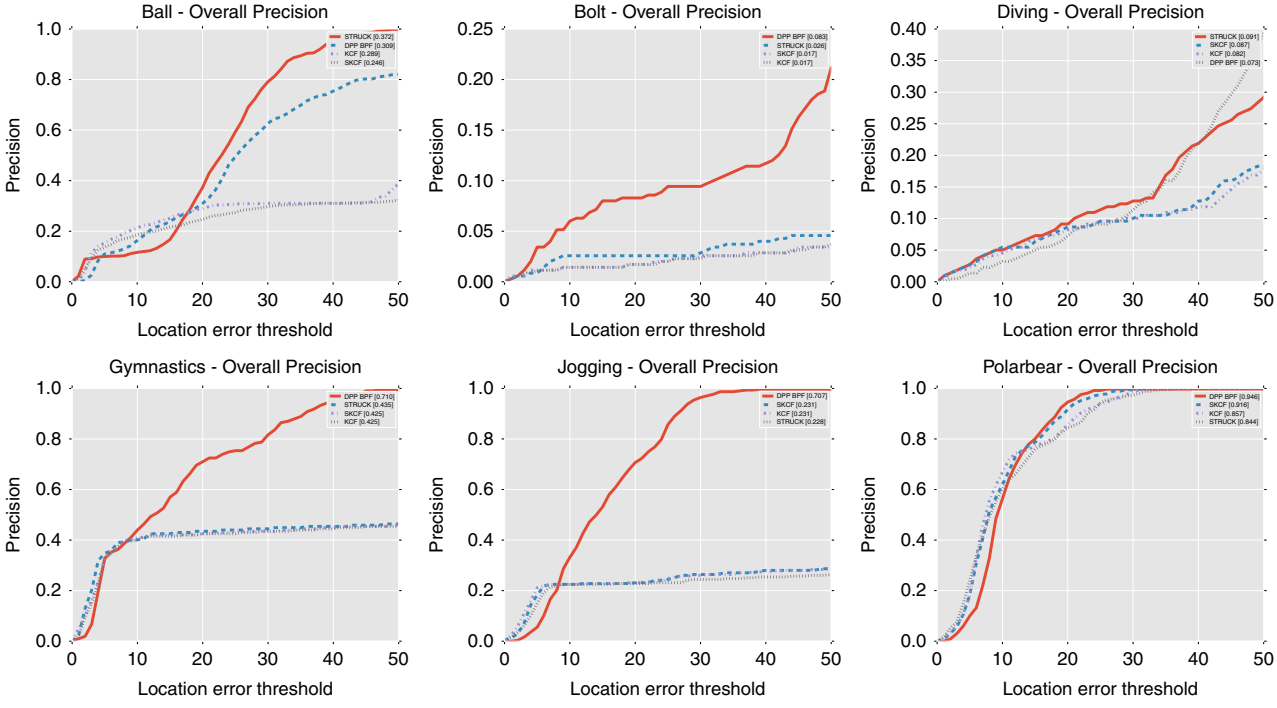

Figure 2: Overall precision plots for the visual tracking sequences.

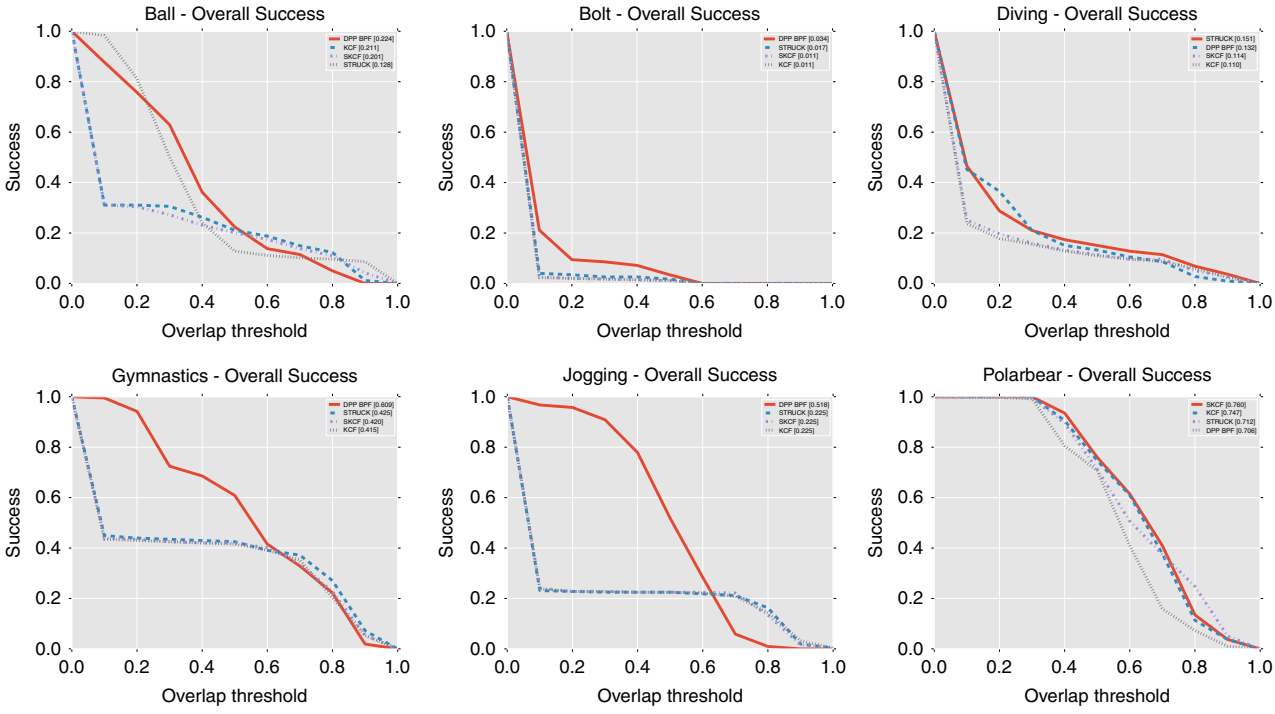

Figure 3: Overall success plots for the visual tracking sequences.

maintain long-term tracks. Since the Bernoulli DPP filter accounts for missed detections, the proposed approach improves the area under the curve of the success metric in 67% of the tested sequences.

Conclusions

In this paper, a novel algorithm for joint detection and tracking a single object in video has been presented. The proposed approach takes into account the detection score and the similarity of the observed features. Then, a Bayesian filter using a Bernoulli point process estimates the state of the target from a diverse subset of object proposals. Experimental evaluations show that the results are comparable to other state-of-the-art techniques for visual tracking in only 6 of the 25 sequences of the data set. In this paper, we only considered a simple observation model (distance to a reference LBP histogram), which might hinder the performance of this approach in the overall data set. This observation model is not robust to scale and rotation changes and no model updating strategies are considered in this paper. Nevertheless, our model is expected to increase its performance when using a more complex observation model (such as deep learning features), model updating and ensemble post-processing techniques for combining the output from different tracking schemes.

Acknowledgments

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

Literature Cited

Bolme, D. S., Beveridge, J. R., Draper, B. A. and Lui, Y. M. 2010. Visual object tracking using adaptive correlation filters. 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, pp. 2544–2550.

Czyz, J., Ristic, B. and Macq, B. 2007. A particle filter for joint detection and tracking of color objects. *Image and Vision Computing* 25(8): 1271–1281.

García-Fernández, A. F., Williams, J. L., Granström, K. and Svensson, L. 2018. Poisson multi-bernoulli mixture filter: direct derivation and implementation. *IEEE Transactions on Aerospace and Electronic Systems* 54 August, pp. 1883–1901.

Hare, S., Golodetz, S., Saffari, A., Vineet, V., Cheng, M.-M., Hicks, S. L. and Torr, P. H. 2016. Struck: structured output tracking with kernels. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38(10): 2096–2109.

Henriques, J. F., Caseiro, R., Martins, P. and Batista, J. 2015. High-speed tracking with kernelized correlation filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37(3): 583–596.

Hoseinnezhad, R., Vo, B. N. and Vo, B. T. 2013. Visual tracking in background subtracted image sequences via multi-bernoulli filtering. *IEEE Transactions on Signal Processing* 61 January, pp. 392–397.

Hoseinnezhad, R., Vo, B.-N., Vo, B.-T. and Suter, D. 2012. Visual tracking of numerous targets via multibernoulli filtering of image data. *Pattern Recognition* 45(10): 3625–3635.

Jorquera, F., Hernández, S. and Vergara, D. 2017. Multi target tracking using determinantal point processes. in Mendoza, M. and Velastin, S. (Eds), *Iberoamerican Congress on Pattern Recognition* Springer International Publishing, Cham, pp. 323–330.

Jorquera, F., Hernández, S. and Vergara, D. 2019. Probability hypothesis density filter using determinantal point processes for multi object tracking. *Computer Vision and Image Understanding* 183: 33–41.

Kingman, J. F. C. 1993. *Poisson Processes*, Clarendon Press, Oxford.

Kristan, M., Leonardis, A., Matas, J., Felsberg, M., Pflugfelder, R., Zajc, L. Č., Vojír, T., Bhat, G., Lukežič, A., Eldesokey, A., Fernández, G., García-Martín, Á., Iglesias-Arias, Á., Alatan, A. A., González-García, A., Petrosino, A., Memarmoghadam, A., Vedaldi, A., Muhič, A., He, A., Smeulders, A., Perera, A. G., Li, B., Chen, B., Kim, C., Xu, C., Xiong, C., Tian, C., Luo, C., Sun, C., Hao, C., Kim, D., Mishra, D., Chen, D., Wang, D., Wee, D., Gavves, E., Gundogdu, E., Velasco-Salido, E., Khan, F. S., Yang, F., Zhao, F., Li, F., Battistone, F., De Ath, G., Subrahmanyam, G. R. K. S., Bastos, G., Ling, H., Galoogahi, H. K., Lee, H., Li, H., Zhao, H., Fan, H., Zhang, H., Possegger, H., Li, H., Lu, H., Zhi, H., Li, H., Lee, H., Chang, H. J., Drummond, I., Valmadre, J., Martin, J. S., Chahl, J., Choi, J. Y., Li, J., Wang, J., Qi, J., Sung, J., Johnander, J., Henriques, J., Choi, J., van de Weijer, J., Herranz, J. R., Martínez, J. M., Kittler, J., Zhuang, J., Gao, J., Grm, K., Zhang, L., Wang, L., Yang, L., Rout, L., Si, L., Bertinetto, L., Chu, L., Che, M., Maresca, M. E., Danelljan, M., Yang, M.-H., Abdelpakey, M., Shehata, M., Kang, M., Lee, N., Wang, N., Miksik, O., Moallem, P., Vicente-Moñivar, P., Senna, P., Li, P., Torr, P., Raju, P. M., Ruihe, Q., Wang, Q., Zhou, Q., Guo, Q., Martín-Nieto, R., Gorthi, R. K., Tao, R., Bowden, R., Everson, R., Wang, R., Yun, S., Choi, S., Vivas, S., Bai, S., Huang, S., Wu, S., Hadfield, S., Wang, S., Golodetz, S., Ming, T., Xu, T., Zhang, T., Fischer, T., Santopietro, V., Štruc, V., Wei, W., Zuo, W., Feng, W., Wu, W., Zou, W., Hu, W., Zhou, W., Zeng, W., Zhang, X., Wu, X., Wu, X.-J., Tian, X., Li, Y., Lu, Y., Law, Y. W., Wu, Y., Demiris, Y., Yang, Y., Jiao, Y., Li, Y., Zhang, Y., Sun, Y., Zhang, Z., Zhu, Z., Feng, Z.-H., Wang, Z. and He, Z. 2019. The sixth visual object tracking vot2018 challenge results. in Leal-Taixé, L. and Roth, S. (Eds), *Computer Vision – ECCV 2018 Workshops* Springer International Publishing, Cham, pp. 3–53.

Kulesza, A. and Taskar, B. 2011. Learning determinantal point processes. Proceedings of the Twenty-Seventh Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-11), AUAI Press, Corvallis, OR, pp. 419–427.

Lee, D., Cha, G., Yang, M.-H. and Oh, S. 2016. Individualness and determinantal point processes for pedestrian detection. in Leibe, B., Matas, J., Sebe, N. and Welling, M. (Eds), *European Conference on Computer Vision* Springer International Publishing, Cham, pp. 330–346.

Li, B., Yan, J., Wu, W., Zhu, Z. and Hu, X. 2018. High performance visual tracking with siamese region proposal network. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8971–8980.

Li, P., Wang, D., Wang, L. and Lu, H. 2018. Deep visual tracking: review and experimental comparison. *Pattern Recognition* 76: 323–338.

Maggio, E. and Cavallaro, A. 2011. *Video Tracking: Theory and Practice*, John Wiley & Sons, Bridgewater, NJ.

Mahler, R. 2007. Phd filters of higher order in target number. *IEEE Transactions on Aerospace and Electronic Systems* 43(4): pp. 1523–1543.

Mahler, R. P. S. 2003. Multitarget bayes filtering via first-order multitarget moments. *IEEE Transactions on Aerospace and Electronic Systems* 39(4): 1152–1178.

Mahler, R. P. S. 2007. *Statistical Multisource-Multitarget Information Fusion* Artech House, Inc.

Privault, N. and Teoh, T. 2019. Second order multi-object filtering with target interaction using determinantal point processes. Tech. Rep. arXiv:1906.06522 [math. PR], ArXiV, June.

Reuter, S., Wilking, B., Wiest, J., Munz, M. and Dietmayer, K. 2013. Real-time multi-object tracking using random finite sets. *IEEE Transactions on Aerospace and Electronic Systems* 49(4): 2666–2678.

Ristic, B. 2013. *Multi-Object Particle Filters* Springer New York, New York, NY, pp. 53–84.

Ristic, B., Vo, B. T., Vo, B. N. and Farina, A. 2013. A tutorial on bernoulli filters: Theory, implementation and applications. *IEEE Transactions on Signal Processing* 61 July, pp. 3406–3430.

Solis Montero, A., Lang, J. and Laganiere, R. 2015. Scalable kernel correlation filter with sparse feature integration. 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, December, pp. 587–594.

Vo, B.-N., Vo, B.-T., Pham, N.-T. and Suter, D. 2010. Joint detection and estimation of multiple objects from image observations. *IEEE Transactions on Signal Processing* 58(10): 5129–5141.

Wang, N., Shi, J., Yeung, D.-Y. and Jia, J. 2015. Understanding and diagnosing visual tracking systems. The IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, December, pp. 3101–3109, doi: 10.1109/ICCV.2015.355.

Xu, T., Feng, Z.-H., Wu, X.-J. and Kittler, J. 2019. Learning adaptive discriminative correlation filters via temporal consistency preserving spatial feature selection for robust visual object tracking. *IEEE Transactions on Image Processing* 28(11): 5596–5609.