Support Correlation Filters Tracking using Mask Matrix

Zhenyang Su[1,3], Jing Li*[1], Zhiquan Ding[2], Tianqi Qin[2], Yafu Xiao[1]
[1]School of Computer Science, Wuhan University, Wuhan, 430072, China
[2]Sichuan Institute of Aerospace Electronic Equipment,Chengdu, 610100, China
[3]Huanggang Normal University, Huangzhou, 438000, China
zy.su@whu.edu.cn, leejingcn@whu.edu.cn, 252241227@qq.com, tianqiqin2008@163.com, xiaoyafu@whu.edu.cn

***Abstract*—Support correlation filter tracking method uses cyclic sampling to transform the calculation into frequency domain, which solves the problems of sampling and large computation of support vector machine. However, the current method can not exploit the information of backgrounds because all samples are generated by cyclic sampling around the target in the tracking process. To solve this problem, this paper proposes a background awareness support correlation filter tracking method using mask matrix. In the tracking process, the mask matrix is used to extract the patchs densely from background as negative samples, so the background information is used effectively. Experiments on OTB100 database show that compared with Scale Kerneling Supported Correlation Filtering (SKSCF), the proposed algorithm achieves a gain of 4.2% in mean OP and 6.2% AUC score respectively.**

***Index Terms*—SVM; support correlation filter; visual tracking.**

I. Introduction

Target tracking has been widely used in a variety of areas, such as video surveillance, man-machine interaction and robot perception, to name a few [1], [2]. Target tracking is defined as a problem of assessing the state of the target, which includes two parts: to present the original state of the target (such as its location and size) and to estimate the target state in the subsequent frames. Despite its wide applications, target tracking is still a great challenge in practice, which is easily affected by a number of factors, including illumination variation, pose variation, fast motion, occlusion and so on.

In recent years, the target tracking algorithms mainly use machine learning technology, which is widly used in artificial intelligence [3]–[8]. The existing tracking algorithms can be categorized as either generative ones or discriminative ones [9]–[16]. A generative algorithm first learns a model to represent the target object. Then it uses the model for calculating the similarity between the candidate sample and the target object. And finally it takes the most similar sample as the target object. Compared with generative algorithms, discriminative methods have attracted wider attention due to their exploitation of both the target information and its background information. They regard target tracking as a problem of binary classification between target and its background, in which some samples are collected and labeled in each frame for classifier training.

Corresponding author: Jing Li. E-mail:leejingcn@whu.edu.cn.

Recent years have witnessed an explosive popularity of discriminative correlation filters (DCF) for visual tracking [17]–[24] because of their high efficiency and accuracy. Most DCF based trackers use ridge regression or kernel ridge regression as predictors, their failure to exploit the good discrimination of the SVM affects the further improvement of tracking performance. The support correlation filtering tracking methods proposed recently [25], [26] exploit the circulant property of dense samples to accelerate SVM based trackers, making it superior to the traditional correlation filter tracking method. Nevertheless, the support correlation filter tracking methods also suffer boundary effect when they use cyclic shifts of the target as the negative samples.

In view of this, this paper uses mask matrix to crop images, and proposes a background aware support correlation filtering(BASCF) algorithm. The main contributions are as follows:

- The proposed BASCF tracker can collect more samples, and the negative samples came from the backgrounds, so that learned correlation filters can have more discriminative power. The correlation filter can be learned on larger image regions, and the information of backgrounds can be fully utilized.
- In this paper, the Alternative Direction Multiplier Method (ADMM) is used to optimize the solution, which can accelerate the solution to meet the real-time requirements, and can achieve about 40 frames per second on a general PC.

II. Related Work

In the following, we briefly introduce some works most related to this work, and for a detailed survey about visual tracking, please refer to [1], [27].

As the earliest correlation filter tracking method, the Minimum Output Sum of Squared Error (Mosse) algorithm [28] uses the minimum mean square error to constrain the output results, and trains the filter. By utilizing the cyclic properties of the kernel matrix, the Kernel Correlation Filters (KCF) algorithm [17] introduces the kernel technique into the correlation filtering algorithm and has achieved satisfactory results along with the consideration of the multi-channel feature. The Multi-kernel Correlation Filter(MKCF) [29] solves the problem that the manual-setting is not necessarily optimal by introducing

Fig. 1. Different principles for choosing samples. Top: our method; bottom: the SCF based trackers in [25], [26].

multi-kernel on the basis of KCF. The part-based correlation filter [30] is proposed to solve the tracking problem when the object is occluded and the shape changes dramatically to some extent.

The trackers mentioned-above undergo the boundary effects due to periodic repetitions. To address this issue, the Spatial Regularized Correlation Filter (SRDCF) algorithm [31] constrains the filter coefficients to allocates more energy for the central region using a Spatial Gaussian function. The Background Aware Correlation Fiter [32] realized the similar idea with a predefined mask matrix. Very recently, some tracking methods [33]–[35] with deep architectures have achieved good results. Li et al. [33] employ the discriminative power in the gradients to dynamically update the template in the siamese net tracker. Wang et al. [34] proposed a method which perform both object tracking and semi-supervised video object segmentation. However, these methods often lead to poor real time performance because of the deep architectures.

III. SCF Tracking Method

The support correlation filter algorithm [25], [26] introduces the cyclic sampling into support vector machine (SVM) in order to solve the tracking problem. In t frame, the basis sample $x \in R^D$ is selected around the target, and the training samples are represented by the full set of circularly shifted versions x_i. Then the filter f is obtained by solving the following functions

$$\begin{aligned} &\min_{f,b,\xi} \|f\|^2 + C\sum_i \xi_i^2, \\ &s.t.\ y_i(f^T x_i + b) \geq 1 - \xi_i,\ \forall i, \ldots, D. \end{aligned} \tag{1}$$

where $y_i \in R^D$ is the classifier label corresponding to x_i, ξ is the slack variables and b represents a bias. Let $e_i = y_i(f^T x_i + b) + \xi_i - 1$, (1) can be reformulated as

$$\begin{aligned} &\min_{f,b,e} \|f\|^2 + C\sum_i (y_i(f^T x_i + b) - 1 - e_i)^2, \\ &s.t.\ e_i > 0, for\ \forall i = 1, \ldots, D. \end{aligned} \tag{2}$$

Since x_i is a cyclic sample, (2) can be converted to the Fourier domain to accelerate the calculation:

$$\begin{aligned} &\min_{f,b,e} \|f\|^2 + C\left\| y \circ (F^{-1}(\hat{x} \circ \hat{f}) + b\mathbf{1}) - \mathbf{1} - e \right\|_2^2, \\ &s.t.\ e \geq 0 \end{aligned} \tag{3}$$

where $\mathbf{1}$ denotes an all-ones vector, $\circ$ denotes Hadamard product, and $\hat{}$ denotes the Discrete Fourier Transform of a signal. Then the filter f can be computed efficiently in the Fourier domain:

$$\hat{f} = (\hat{x}_c^* \circ \hat{x}_c + 1/C)^{-1} \circ \hat{x}_c \circ \hat{q}_c \tag{4}$$

where $\mathrm{q} = y + y \circ e$, $\bar{\mathrm{q}}$ is the mean of q, and $\mathrm{q_c} = q - \bar{q}$.

IV. Proposed Method

A. Model Establishment

Sample x_i is the cyclic shift form of the basis x. Δ_{i} is the cyclic shift operation, then $x\Delta_{\mathrm{i}}$ is the i_{th} cyclic shift of the basis image x. Consequently, (2) can be expressed as:

$$\begin{aligned} &\min_{f,b,e} \|f\|^2 + C\sum_{i=1}^{D} \left\| y_i(f^T x\Delta_i + b) - 1 - e_i \right\|_2^2, \\ &s.t.\ e_i \geq 0 \end{aligned} \tag{5}$$

Although the trained filters at present are able to discriminate the foreground targets from their shift forms, they are unable to distinguish the real background of the images. Some works [32], [36] extracts patches densely from background using cropping matrix can restrain the boundary effect caused by cyclic sampling. Therefore, this paper introduces mask matrix P to support correlation filtering method to crop the image from a larger image block. P is a binary matrix of $D * T$ dimension, which can crop D dimension pixels ($T \gg D$) from T-dimensional image. Thus the background-aware support correlation filters can be learned by the following objective:

$$\begin{aligned} &\min_{f,b,e} \|f\|^2 + C\sum_{i=1}^{T} \left\| y_i(f^T P x \Delta_i + b) - 1 - e_i \right\|_2^2 \\ &s.t.\ e_i \geq 0 \end{aligned} \tag{6}$$

where $x_i \in R^T$,$y_i \in R^T$, and $f \in R^D$. After the mask matrix P is introduced, the tracking will be sampled in a larger range. As is shown in figure 1, the blue box is T dimension while the red box is D dimension. In each frame, the training samples will be increased from the original D samples to the current T samples, and the sampling samples will also be changed from the original "cyclic shift samples" to samples from the real scene. Both factors greatly enhance the tracking performance.

B. Optimization algorithm

For efficiency, the filters are generally computed in the Fourier domain. According to the Parseval's theorem, (6) can be represented as

$$\begin{aligned} &\min_{f,\hat{g},b,e} \|f\|^2 + C\left\|\hat{x} \circ \hat{g} + \hat{b}\mathbf{1} - \hat{q}\right\|_2^2 \\ &s.t.\ \hat{g} = \sqrt{T} F P^T f, e \geq 0 \end{aligned} \tag{7}$$

where, g stands for the T-dimensional auxiliary variable matrix and F is a T×T orthogonal complex basis vector matrix, which can transform any T-dimensional signal into Fourier domain (eg.$\hat{a} = \sqrt{T} F a$). The above model in (7) is convex that can be minimized to yield the globally optimal solution via ADMM, and its augmented Lagrangian form is

$$\begin{aligned} &\min_{f,\hat{g},\hat{s},b,e} \|f\|^2 + C\left\|\hat{x} \circ \hat{g} + \hat{b}\mathbf{1} - \hat{q})\right\|_2^2 + \hat{s}^T(\hat{g} - \sqrt{T} F P^T f) \\ &+ \frac{u}{2}\left\|\hat{g} - \sqrt{T} F P^T f\right\|_2^2, \\ &s.t.\ e \geq 0 \end{aligned} \tag{8}$$

where, S is the Lagrange multiplier, u is the penalty coefficient, and (8) can be solved iteratively via the ADMM. The solution to each subproblem is detailed below:

Step 1: update f. Fixing g, s, b, e, we have

$$\begin{aligned} f &= \min_f \|f\|^2 + \hat{s}^T(\hat{g} - \sqrt{T} F P^T f) + \frac{u}{2}\left\|\hat{g} - \sqrt{T} F P^T f\right\|_2^2 \\ &= (u + \frac{1}{CT})^{-1}(ug + s) \end{aligned} \tag{9}$$

where $g = \frac{1}{\sqrt{T}} P F^T \hat{g}$,$s = \frac{1}{\sqrt{T}} P F^T \hat{s}$. Taking the DFT and inverse DFT into account, the complexity of solving f is $\mathcal{O}(N_{Iter} K T \log T)$, where N_{Iter} denotes the numbers of iterations, meanwhile K indicates channel dimension.

Step 2: update $\hat{g}$. Fixing f, s, b, e, we have

$$\begin{aligned} \hat{g} = &\min_{\hat{g}} C\left\|\hat{x} \circ \hat{g} + \hat{b}\mathbf{1} - \hat{q})\right\|_2^2 + \hat{s}^T(\hat{g} - \sqrt{T} F P^T f) \\ &+ \frac{u}{2}\left\|\hat{g} - \sqrt{T} F P^T f\right\|_2^2 \end{aligned} \tag{10}$$

The calculation of (10) is highly time consuming. But, each element of $\hat{y}$ only depends on the element of fiter $\hat{g}$ and sample x , and we define $V_j(g)$ as the value of the j-th element of the fiter g, so (10) can be further divided into D smaller problems, where each of them is defined as:

$$\begin{aligned} \mathrm{V_j}(\hat{g}) = &\min_{V_j(\hat{g})} C\left\|V_j(\hat{x})^T V_j(\hat{g}) + \hat{b} - V_j(\hat{q}))\right\|_2^2 \\ &+ \hat{s}_j^T(V_j(\hat{g}) - V_j(\hat{f})) + \frac{u}{2}\left\|V_j(\hat{g}) - V_j(\hat{f})\right\|_2^2 \end{aligned} \tag{11}$$

where $\hat{f} = \sqrt{D} F P^T f$, set the derivative of (11) be zero, we have

$$\begin{aligned} \mathrm{V_j}(\hat{g}) = &(V_j(\hat{x}) V_j(\hat{x})^T + Tu)^{-1}(V_j(\hat{y}) V_j(\hat{x}) - T V_j(\hat{s}) \\ &+ Tu V_j(\hat{f})) \end{aligned} \tag{12}$$

Since with inverse operation, (12) can be optimized with the Sherman-Morrsion formula as

$$\begin{aligned} \mathrm{V_j}(\hat{g}) = &\frac{1}{u}(V_j(\hat{y}) V_j(\hat{x})/T - V_j(\hat{s}) + u V_j(\hat{f})) \\ &- \frac{V_j(\hat{x})}{um}(V_j(\hat{y}) V_j(\hat{R}_X)/T - V_j(\hat{R}_V) + u V_j(\hat{R}_f)) \end{aligned} \tag{13}$$

where, $V_j(\hat{R}_X) = V_j(\hat{x})^T V_j(\hat{x})$, $V_j(\hat{R}_S) = V_j(\hat{x})^T V_j(\hat{s})$, $V_j(\hat{R}_f) = V_j(\hat{x})^T V_j(\hat{f}) m = V_j(\hat{R}_X) + Tu$. The cost of computing $\hat{g}$ using (13) is $\mathcal{O}(TK)$.

Step 3: update s. Fixing f, and g, we update s as

$$\hat{s}^{i+1} = \hat{s}^i + \hat{g}^{i+1} - \hat{f}^{i+1} \tag{14}$$

Step 4: update b. Fixing f, and g, s, e, the subproblem b can be represented as

$$\begin{aligned} b &= \min_b C\left\|\hat{x} \circ \hat{g} + \hat{b}\mathbf{1} - \hat{q})\right\|_2^2 \\ &= \bar{q} \end{aligned} \tag{15}$$

Step 5: update e. Fixing f, and g, s, b, the subproblem e can be represented as

$$\begin{aligned} e &= \begin{cases} \min_e C\left\|\hat{x} \circ \hat{g} + \hat{b}\mathbf{1} - \hat{q})\right\|_2^2 \\ \quad s.t.\ e \geq 0 \end{cases} \\ &= \max(y \circ (F^{-1}(\hat{x} \circ \hat{g}) + b\mathbf{1}) - 1, 0) \end{aligned} \tag{16}$$

C. Model updating

Like many other related filtering algorithms, the filters also adopt the online adaptive updating system. It is updated as follows:

$$\hat{x}^t_{\mathrm{mod}\ el} = (1 - \eta)\hat{x}^{t-1}_{\mathrm{mod}\ el} + \eta \hat{x}^* \tag{17}$$

where $\hat{x}$ is the observation model in the t frame and η represents the online learning rate. Replacing $\hat{x}$ by (17), we

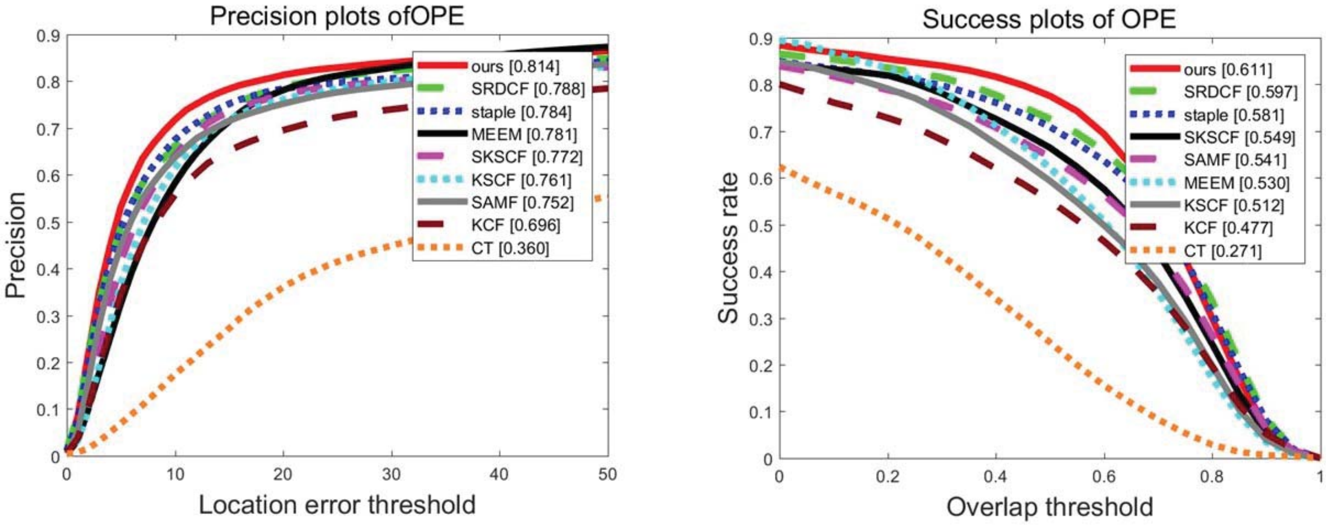

Fig. 2. Overall success and precision plots of OPE of the 9 trackers in OTB100.

can learn the filter $\hat{g}$ through (13). Afterwards, the filter will be used to detect next frame whose size is the same as the filter g.

V. EXPERIMENTS

A. Experimental Setup

We extensively evaluate our tracker on Object Tracking Benchmark (OTB100) [27] , comparing with other eight algorithms, including CT , SAMF [37] , KCF [17] , srdcf [31] , MEEM [38] , staple [18] , skscf [25] , kscf [25] . The benchmark employs one-pass evaluation (OPE) of success plot and precision plot to measure tracking quality. Our method is implemented on a personal computer with Intel i7 CPU (3.60 GHz) and 12 G running memory. It adopts the 31-dimensional HOG characteristics (cell size being 4*4 pixel), and the value of the regularization parameter C is 100. We uses 5 scales whose step length is 1.01, and the learning rate of the model is 0.013. The target location confidence map is used to define y_i as label value. If it is greater than 0.5, it belongs to the section of 1, if less than 0.4, then the section of -1, other cases are 0. The penalty coefficient u of ADMM algorithm is 1, and the iteration step is $\times 10$. We find that the algorithm is able to converge quickly in most videos, so the iterations number is set to 2.

B. Quantitative Analysis

1) Overall Analysis: Figure 2 shows the overall success and precision plots of OPE over all the 100 videos in OTB100, our tracker(BASCF) achieves the best performance with an AUC score of 61.1%, outperforming the support correlation filtering algorithm SKSCF (54.9%) and KSCF(51.2%). The overall precision of our tracking algorithm is 81.4% which is 4.2% and 5.3% respectively higher than the support correlation filtering algorithm SKSCF (77.2%) and KSCF(76.1%). To verify the effectiveness of the BASCF method, we further employ the deep features for BASCF training, yielding an AUC score of 65.3% and a precision score of 86.6%.

2) Attribute Based Analysis: The videos in OTB database are classified into 11 types of attributes according to different environments including illuminate change, scale change, occlusion, shape change, etc. The success and precision plot of each attribute is shown in figure 3 and figure 4 respectively. Our algorithm is superior to the support correlation filtering algorithms in 10 of 11 attributes. Among the precision plot of 11 attributes, our algorithm ranks the first for five times and gets the second place for three times. As for the rest of the attributes, our algorithm presents performance not worse than other excellent algorithms. According to the success plots, our algorithm delivers better performance than any other support correlation filtering algorithms and achieves a satisfactory result in terms of 11 attributes. Specifically, it took the first place for 9 times and the second place twice.

C. Qualitative Analysis

Figure 5 records the tracking results of some video sequences, including couple, lemming and human4 which faces such challenges as deformation, fast motion, scale change, occlusion, out of view and motion blur etc.

1) Couple: The video couple is mainly challenged by factors such as out-of-plane rotation, scale change, shape change, fast motion and disordered background, it proves to be a tracking video with great difficulty. During the process of tracking, algorithm SAMF and CT lost the target at the #17 frame, algorithm KCF at the #37 frame, algorithm KSCF and SKSCF at the #48 frame and Staple lost its target at the #91 frame. In the whole tracking process, only our algorithm and SRDCF algorithm can track the target correctly.

2) Lemming: The video Lemming is mainly challenged by illumination, out-of-plane rotation, scale change, occlusion,

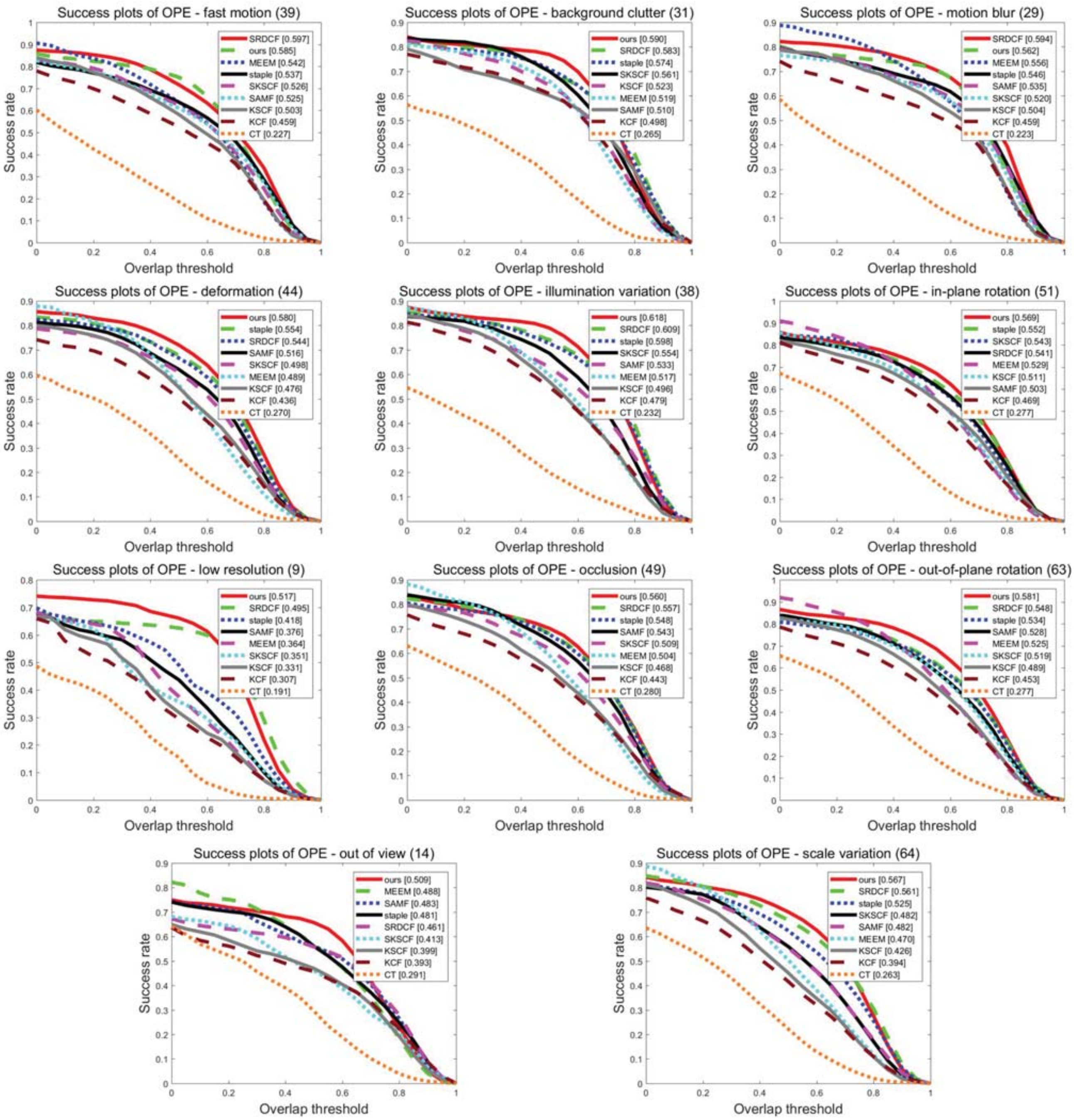

Fig. 3. Success plots of videos with different attributes.

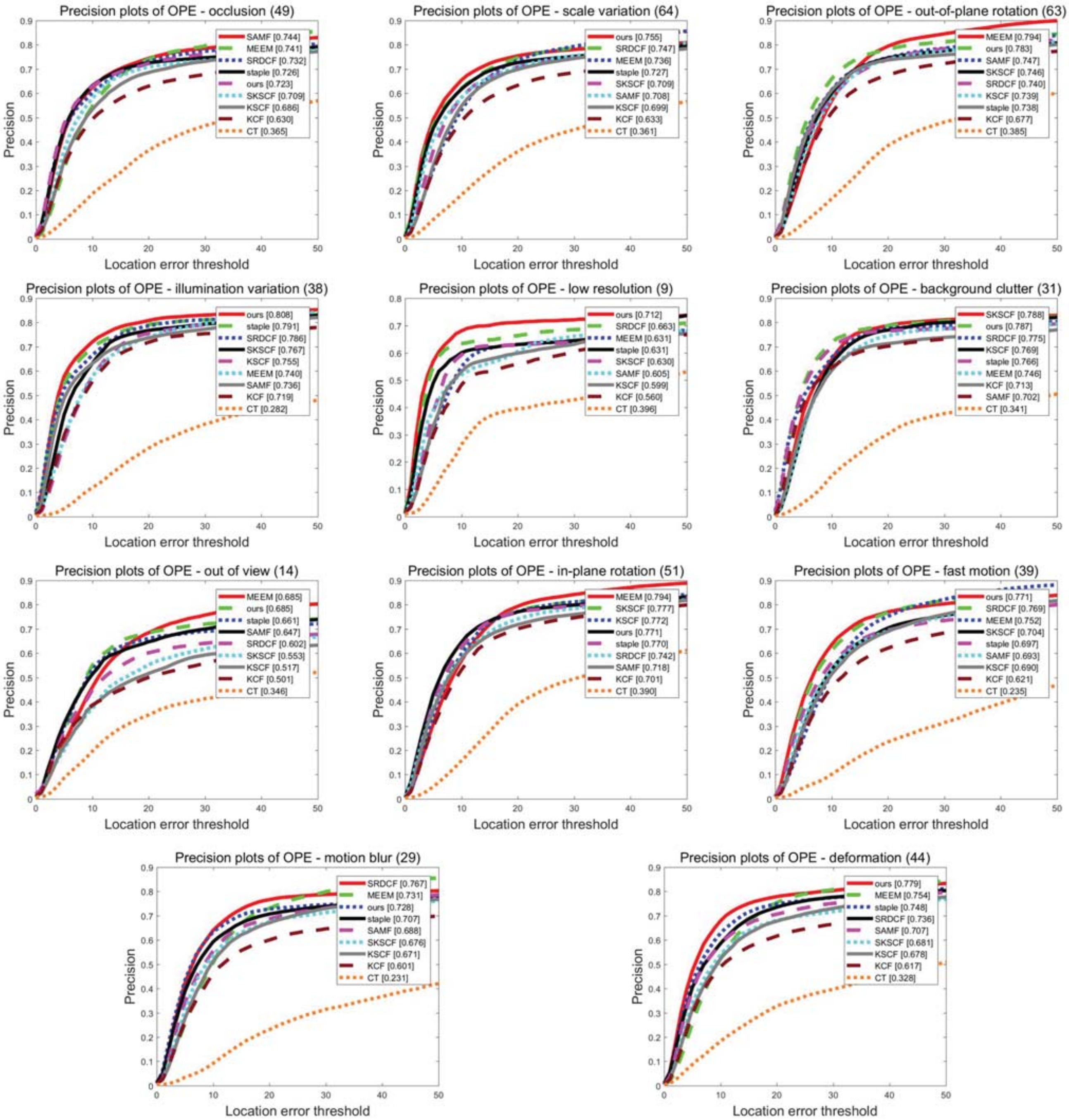

Fig. 4. Precision plots of videos with different attributes.

Fig. 5. Qualitative comparative results of the 9 trackers over 3 sequences *couple,lemming, human4.*

motion blur, out of view and so on. Due to the changes like out-of-plane rotation and motion blur in the early stage, the algorithm CT began to drift from #230 frame. In the video, severe occlusion occurred during #333 and #365, and when the target completely comes out from the occlusion at #375, algorithms like KSCF, SKSCF, KCF, SRDCF and Staple have completely drifted to the occlusion. Affected by the factors like scale change, motion blur and out-of-plane rotation, the target boxes in algorithms MEEM and SAMF appear to be continuously small. As a result, only a part of targets can be tracked.

3) human4: The video human4 is mainly challenged by such factors as illumination change, occlusion, shape change and scale change and so on. Since the target is blocked by the signs, the algorithm CT started to drift from #129. Subsequently, due to the occlusion from the traffic lights (#342), trees (#361) and street light poles (#465), algorithm MEEM, algorithm KCF and algorithm KSCF also lost their targets successively. By the time of frame #645, our algorithm and algorithm SRDCF are the only two algorithms that can track the target correctly.

VI. Conclusion

In this work, we presents a support correlation filtering algorithm with background aware ability. Compared with traditional support correlation filtering algorithm, our algorithm is able to collect more training samples, mostly from the backgrounds of real scenes, and train filters with stronger discriminant power. As is shown by the experiments on the OTB100 database,the proposed method achieves a favorable performance against various state-of-the-art trackers. In the future we will extend the proposed tracker to a variety of video analysis tasks, such as person re-ID, activity recognition.

References

[1] X. Li, W. Hu, C. Shen, Z. Zhang, A. Dick, and A. V. D. Hengel, "A survey of appearance models in visual object tracking," *ACM transactions on Intelligent Systems and Technology*, vol. 4, no. 4, p. 58, 2013.

[2] Y. Song, C. Ma, X. Wu, L. Gong, L. Bao, W. Zuo, C. Shen, R. Lau, and M.-H. Yang, "VITAL: VIsual Tracking via Adversarial Learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8990–8999.

[3] J. Wu, S. Pan, X. Zhu, C. Zhang, and X. Wu, "Multi-instance learning with discriminative bag mapping," *IEEE Transactions on Knowledge & Data Engineering*, pp. 1065–1080, 2018.

[4] C. Jie, B. Wang, and D. Brown, "Similarity based leaf image retrieval using multiscale r-angle description," *Information Sciences*, vol. 374, pp. 51–64, 2016.

[5] W.-X. Lu, C. Zhou, and J. Wu, "Big social network influence maximization via recursively estimating influence spread," *Knowledge-Based Systems*, vol. 113, pp. 143 – 154, 2016.

[6] J. Cao, Z. Wu, J. Wu, and W. Liu, "Towards information-theoretic k-means clustering for image indexing," *Signal Processing*, vol. 93, no. 7, pp. 2026–2037, 2013.

[7] L. Gao, C. Zhou, J. Wu, and Y. Hu, "Collaborative dynamic sparse topic regression with user profile evolution for item recommendation," in *The Thirty-First Conference on Artificial Intelligence (AAAI-17)*, 2017.

[8] Y. Zhang, J. Wu, C. Zhou, and Z. Cai, "Instance cloned extreme learning machine," *Pattern Recognition*, vol. 68, pp. 52–65, 2017.

[9] K. Zhang, L. Zhang, and M.-H. Yang, "Real-time compressive tracking," in *European Conference on Computer Vision*. Springer, 2012, pp. 864–877.

[10] X. Jia, H. Lu, and M.-H. Yang, "Visual tracking via adaptive structural local sparse appearance model," in *Proceedings of the IEEE Conference on Computer vision and pattern recognition*. IEEE, 2012, pp. 1822–1829.

[11] K. Zhang, L. Zhang, and M.-H. Yang, "Fast compressive tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 10, pp. 2002–2015, 2014.

[12] W. Zhong, H. Lu, and M.-H. Yang, "Robust object tracking via sparsity-based collaborative model," in *Proceedings of the IEEE Conference on Computer vision and pattern recognition*. IEEE, 2012, pp. 1838–1845.

[13] W. Chen, K. Zhang, and Q. Liu, "Robust visual tracking via patch based kernel correlation filters with adaptive multiple feature ensemble," *Neurocomputing*, vol. 214, pp. 607–617, 2016.

[14] J. Yang, K. Zhang, and Q. Liu, "Robust object tracking by online fisher discrimination boosting feature selection," *Computer Vision and Image Understanding*, vol. 153, pp. 100–108, 2016.

[15] K. Zhang, L. Zhang, and M.-H. Yang, "Real-time object tracking via online discriminative feature selection," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 4664–4677, 2013.

[16] K. Zhang, Q. Liu, Y. Wu, and M.-H. Yang, "Robust visual tracking via convolutional networks without training," *IEEE Transactions on Image Processing*, vol. 25, no. 4, pp. 1779–1792, 2016.

[17] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 583–596, 2015.

[18] L. Bertinetto, J. Valmadre, S. Golodetz, O. Miksik, and P. H. Torr, "Staple: Complementary learners for real-time tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1401–1409.

[19] Y. Qi, L. Qin, J. Zhang, S. Zhang, Q. Huang, and M.-H. Yang, "Structure-aware local sparse coding for visual tracking," *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 3857–3869, 2018.

[20] Y. Qi, S. Zhang, L. Qin, Q. Huang, H. Yao, J. Lim, and M.-H. Yang, "Hedging deep features for visual tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018.

[21] C. Sun, D. Wang, H. Lu, and M.-H. Yang, "Correlation Tracking via Joint Discrimination and Reliability Learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 489–497.

[22] F. Li, C. Tian, W. Zuo, L. Zhang, and M.-H. Yang, "Learning spatial-temporal regularized correlation filters for visual tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 479–487.

[23] T. Zhang, S. Liu, C. Xu, B. Liu, and M.-H. Yang, "Correlation particle filter for visual tracking," *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 2676–2687, 2018.

[24] T. Zhang, C. Xu, and M.-H. Yang, "Learning Multi-task Correlation Particle Filters for Visual Tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018.

[25] W. Zuo, X. Wu, L. Lin, L. Zhang, and M. H. Yang, "Learning support correlation filters for visual tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1158–1172, 2018.

[26] Z. SU, J. LI, J. CHANG, B. DU, and Y. XIAO, "Real-time visual tracking using complementary kernel support correlation filters," *Frontiers of Computer Science*, vol. 14, no. 2, pp. 417–429, 2020.

[27] Y. Wu, J. Lim, and M.-H. Yang, "Object tracking benchmark," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1834–1848, 2015.

[28] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, "Visual object tracking using adaptive correlation filters," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2010, pp. 2544–2550.

[29] M. Tang, B. Yu, F. Zhang, and J. Wang, "High-speed tracking with multi-kernel correlation filters," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4874–4883.

[30] S. Liu, T. Zhang, X. Cao, and C. Xu, "Structural correlation filter for robust visual tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4312–4320.

[31] M. Danelljan, G. Hager, F. Shahbaz Khan, and M. Felsberg, "Learning spatially regularized correlation filters for visual tracking," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 4310–4318.

[32] H. Kiani Galoogahi, A. Fagg, and S. Lucey, "Learning background-aware correlation filters for visual tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1135–1143.

[33] P. Li, B. Chen, W. Ouyang, D. Wang, X. Yang, and H. Lu, "Gradnet: Gradient-guided network for visual object tracking," in *ICCV*, October 2019.

[34] Q. Wang, L. Zhang, L. Bertinetto, W. Hu, and P. H. Torr, "Fast online object tracking and segmentation: A unifying approach," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2019.

[35] X. Li, C. Ma, B. Wu, Z. He, and M. Yang, "Target-aware deep tracking," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1369–1378.

[36] H. K. Galoogahi, T. Sim, and S. Lucey, "Correlation filters with limited boundaries," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[37] Y. Li and J. Zhu, "A scale adaptive kernel correlation filter tracker with feature integration," in *European Conference on Computer Vision*. Springer, 2014, pp. 254–265.

[38] J. Zhang, S. Ma, and S. Sclaroff, "Meem: robust tracking via multiple experts using entropy minimization," in *European Conference on Computer Vision*. Springer, 2014, pp. 188–203.