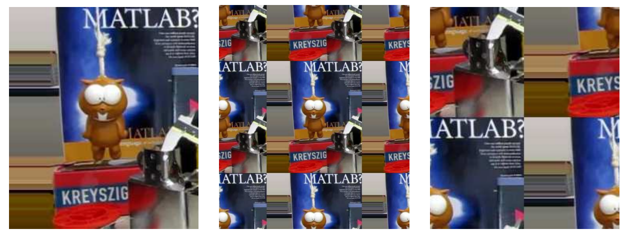
Extended Kernelized Correlation Tracking with Target Enhancement and Sample Selection

Peng Liu, Chang Liu, Wei Zhao*, Xianglong Tang
Department of Computer Science
Harbin Institute of Technology, Harbin, China
*Corresponding author
Email: {pengliu, 15b903045, zhaowei, tangxl}@hit.edu.cn

***Abstract*—In this paper, we address the problem of fast motion and bound effect about the popular high-speed correlation filters-based trackers. Such trackers are facing with the contradiction between extended detection region and reduced precision. To improve the robustness against fast motion, we firstly propose a tracker with extended region. In addition, in order for adapting different region sizes and target sizes, we introduce the target enhancement strategy to increase the effect of the target in learning a discriminative regression. Furthermore, a novel sample selection mechanism is established to drop the error samples generated by the circular structure of correlation filters. Our approach enlarges the detection region, improves the tracking accuracy and preserves the significant kernel structure of the correlation filters. Moreover, extensive experimental results in a recent benchmark datasets show that our proposed method have a promising performance compared to the state-of-art methods.**



(a) Base sample (b) Circular samples (c) An error sample

Fig. 1. In KCF, all samples (b) are circular shifts of the base sample (a). Inevitably, error samples are produced such as (c)

I. Introduction

Visual tracking has been playing an important role in computer vision because it can be widely applied in military, surveillance and human computer interface. A recent review [1] has pointed out the main challenges in visual tracking such as fast motion, deformation, occlusion and background clutter. And [2] established a popular benchmark and evaluated a host of state-of-the-art trackers. Most existed trackers are generative [3], [4] or discriminative [5]–[7].

Recently, great progress has been made in object tracking since the kernelized correlation filters (KCF) [8] was introduced. The KCF utilized the circular samples to train a ridge regression model to detect the target, and gained outstanding performance with high speed. Correlation filter based trackers [9]–[18] achieved comparable performances to deep learning trackers [19], [20] and real-time speeds.

However, the KCF is limited by the detection region because it must be the same size with the training region. The KCF fails when the target is located in the boundary or out of the detection region, because it is sensitive to fast motion. That issue can be solved by enlarging the detection region. However, the KCF considers the background as a part of the target. The background information increases as the region is extended, whereas the target information remains unchanged. Consequently, enlarging the detection region will bring down the tracking precision. Besides, bound effect is another main drawback of KCF as referred in [9], [10]. A great deal of error samples are involved in the KCF. The examples of circular samples and error samples are demonstrated in Fig.1. [9] and [10] solve the bound effect in different ways. However, they both destroyed the kernel structure of the KCF.

To overcome the region limitation and bound effect of the KCF, we propose the EKCF, an extended kernelized correlation tracker with a larger search region and constrain the background with target enhancement and sample selection. Firstly, the detection region is extended to adapt the objects with fast motion. Secondly, we enhance the effect of target to adapt the extended region. Thirdly, we propose to drop the error samples by constraining the kernel coefficients. In nonlinear regression, kernel trick is introduced and the classification plane is the linear combination of all training samples. We introduce the kernel constrained correlation filters that constrain the coefficients of error samples to be zeros, so that error samples will not be included in the classification plane. The EKCF can adapt fast motion, reduce bound effect and preserve the kernel trick of the KCF. Besides, quantitative and qualitative experimental results in the most popular benchmark [2] exhibit the state-of-art performance of the EKCF. The main contributions of this paper can be summarized as follows:

- We introduce a target enhancement method for kernel function that can degrade the background and increase the effect of the target in tracking process to adapt different sizes of regions.
- We propose a kernel constrained correlation filters learning method. Samples are selected by constraining their kernel weights to reduce the effect of error samples in the KCF.
- The detection region is extended to improve the robustness against fast motion. And combining the target enhancement

and the sample selection can reduce the bound effect and preserve the kernel structure of the KCF.

The rest of this paper is organized as follows. Related work is firstly introduced in Sect. 2. Then the proposed method are presented in Sect. 3. The environmental results are described and analysed in Sect. 4. Finally, Sect. 5 contains our conclusions.

II. Related Work

Visual tracking approaches include generative or discriminative methods. Generative methods [3], [4] search the region that are most similar to the target appearance in the current frame. Discriminative methods [5]–[7] train a classifier to discriminate the target from the background. Since MOOSE [21] was proposed by Bolme, correlation filter has promoted the development of visual tracking. Features and scales are improved by Danelljan [13], [15]. In CSK [14] and the later KCF [8], Henriques utilized circular samples to interpret the correlation filter and introduced the kernel trick for nonlinear regression. The KCF is also improved with part-based model [16], [17], long-term model [18], and CNN features [22].

The KCF has both advantages in precision and speed. However, the bound effect, which is discussed in [9], is the main drawback of correlation filter based trackers. [9] learned a spatially regularized correlation filters and [10] limited the boundary of the samples. However, they both solve the bound effect in linear regression, the kernel trick can not work.

In [23], Liu introduced a sample selection strategy into visual tracking. The LASSO regularization [24] was used to automatically select samples with sparsity for better update. Yao [25] applied the LASSO in the KCF to use the latent sparsity of circular samples.

III. The Proposed Method

In this section, we proposed an extended kernelized correlation tracker with target enhancement and sample selection. Firstly, we introduce the KCF tracker, and then highlight the target in the classifier. Next, the kernel constrained correlation filter is presented to select effective samples. At last, the implement details and complexity is explained. The framework of our proposed method is shown in Fig.2.

A. Kernelized Correlation Filters

The KCF tracker [8] demonstrated outstanding performance. The KCF follows the discriminative framework, but obtains massive samples by circulating a base sample and achieves high speed by solving the regression in the Fourier domain.

The KCF trains the samples to find a function $f(\mathbf{z}) = \mathbf{w}^T\mathbf{z}$ that minimizes the ridge regression loss,

$$\min_{\mathbf{w}} ||\mathbf{y} - \mathbf{X}\mathbf{w}||_2^2 + \lambda||\mathbf{w}||_2^2 \tag{1}$$

where $\mathbf{X}$ denotes all training samples and $\mathbf{X} = C(\mathbf{x})$ is the circulant matrix of a base sample $\mathbf{x}$, $\mathbf{y}$ is the labels vector of the samples, and λ is a regularization parameter.

The KCF maps the inputs to a non-linear feature-space $\varphi(\mathbf{x})$ by introducing the kernel trick: $\mathbf{w} = \sum_i \alpha_i \varphi(\mathbf{x}_i)$. Then the solution of the regression is:

$$\hat{\boldsymbol{\alpha}} = \frac{\hat{\mathbf{y}}}{\hat{\mathbf{k}}^{\mathbf{xx}} + \lambda} \tag{2}$$

where $\hat{\boldsymbol{\alpha}}$ is the DFT of $\boldsymbol{\alpha}$, $\hat{\mathbf{k}}^{\mathbf{xx}}$ is the kernel correlation of $\mathbf{x}$, and the kernel matrix of all samples subjects to $K^{\mathbf{x}} = C(\mathbf{k}^{\mathbf{xx}})$. The fraction denotes element-wise division. And in detection, the new base sample $\mathbf{z}$ is detected by:

$$\hat{\mathbf{y}} = \hat{\mathbf{k}}^{\mathbf{xz}} \odot \hat{\boldsymbol{\alpha}} \tag{3}$$

where $\hat{\mathbf{k}}^{\mathbf{xz}}$ is the kernel correlation of $\mathbf{x}$ and $\mathbf{z}$, the kernel matrix of all samples subjects to $K^{\mathbf{z}} = C(\mathbf{k}^{\mathbf{xz}})$, and $\odot$ denotes dot product. The target is located where the value in $\mathbf{y}$ is maximum.

B. Target Enhancement

The KCF performs well in a limited region, but it cannot find and locate the target out of the region. It is necessary to extend the detection region. However, the training region must be the same size as the detection region. We observe that the features of the background surrounding the target have the same effect with the ones of the target in training. The effect of the background increases as the sampling region is extended. This phenomenon will bring down the accuracy of the regression. In the end, the tracker will drift to background. For example, we tested the KCF with different region sizes in OTB100 [2]. The precision score were 69.2%, 67.4%, 62.5%, and 56.5% with regions of sizes 2.5, three, 3.5, and four times the target, respectively. In the failure cases, the KCF obtained the maximum responses in background instead of the target when the target moves. Hence, we propose a target enhancement strategy for correlation filters to highlight the effect of the target. Our target enhancement strategy can adapt different region sizes and target sizes.

With kernel trick, the features $\mathbf{x}$ are mapped to a non-linear feature space, $\varphi(\mathbf{x})$ and $\varphi^T(\mathbf{x})\varphi(\mathbf{x}') = k(\mathbf{x}, \mathbf{x}')$, which is calculated with a kernel function k. We discover that the features with small absolute values have less effect in training the regression and detecting the target. Namely, the features with greater values will have more effect in the regression. As a result, we map the features of the samples to a target enhanced space, $\varphi_p(\mathbf{x}_i) = \varphi(\mathbf{p}_i \odot \mathbf{x}_i)$. So $\varphi_p(\mathbf{X}) = \varphi(\mathbf{P} \odot \mathbf{X})$. The values in $\mathbf{P}$ are defined close to ones when the corresponding values in $\mathbf{X}$ belong to the target. Otherwise, the values in $\mathbf{P}$ are defined close to zeros when those in $\mathbf{X}$ belong to the background. Specifically, we define $\mathbf{p} = mask(\mathbf{x}, \mathbf{bx})$ and for all circular samples, $\mathbf{P} = C(\mathbf{p})$, where $\mathbf{bx}$ is the bounding box of the target in $\mathbf{x}$, and $mask$ is a function to determine whether the features belong to the target. Here, we simply use a Gaussian window to determine the $mask$. For example, when the size of a 2D $\mathbf{x}$ is with height H and width W,

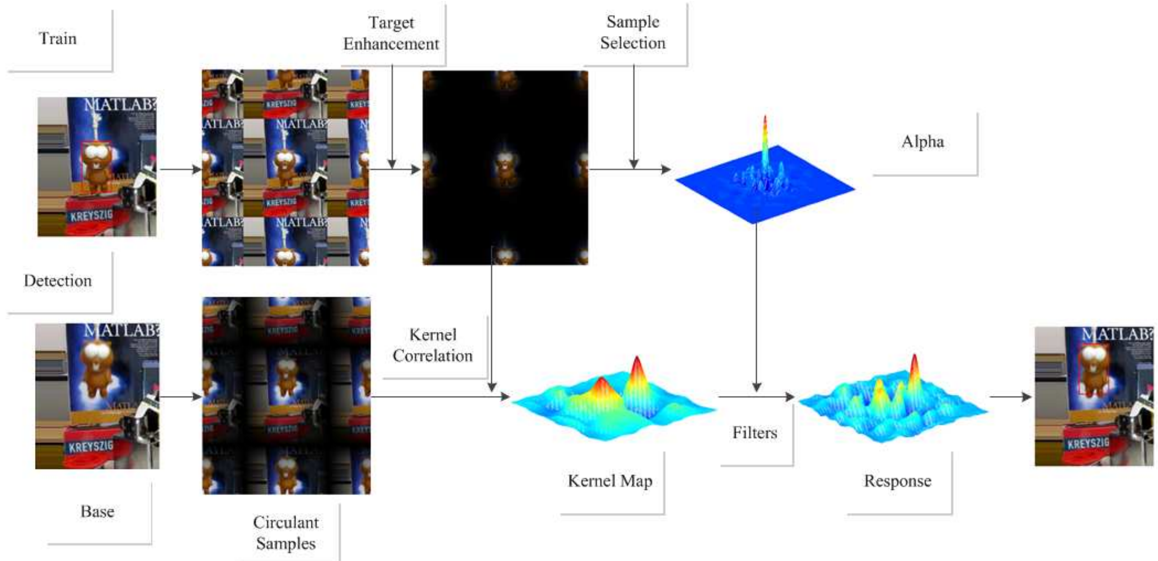


Fig. 2. The framework of our proposed tracking method

and the size of the target in $\mathbf{x}$ is h and w, the $mask$ can be calculated by,

$$mask(\mathbf{x}) = gausswin(H, \frac{w}{W}\theta)' gausswin(W, \frac{h}{H}\theta) \quad (4)$$

where $gausswin(N, \theta)$ means a Gaussian window of length N with parameter θ.

Accordingly, we can obtain that $\varphi_p(\mathbf{X}) = C(\varphi(\mathbf{p} \odot \mathbf{x}))$ is a circulant matrix. In training, the target enhanced kernel matrix is $K_p^{\mathbf{x}} = C(\mathbf{k}_p^{\mathbf{xx}})$, where the $K_p^{\mathbf{x}}(i,j) = \varphi_p^T(\mathbf{x}_i)\varphi_p(\mathbf{x}_j) = k(\mathbf{p}_i \odot \mathbf{x}_i, \mathbf{p}_i \odot \mathbf{x}_j)$. So the solution of the target enhanced KCF is,

$$\hat{\boldsymbol{\alpha}} = \frac{\hat{\mathbf{y}}}{\hat{\mathbf{k}}_p^{\mathbf{xx}} + \lambda} \quad (5)$$

Furthermore, in detection, $K_p^{\mathbf{z}}(i,j) = \varphi_p^T(\mathbf{x}_i)\varphi(\mathbf{z}_j) = k(\mathbf{p}_i \odot \mathbf{x}_i, \mathbf{z}_j)$ will degrade the background and highlight the effect of the target area.

C. Sample Selection in KCF

We have pointed out the error samples in Fig.1. Since $\mathbf{w} = \sum_i \alpha_i \varphi(\mathbf{x}_i)$, we drop the error samples by constraining them to obtain zero coefficients, namely, $\alpha_i = 0$ if $\mathbf{x}_i$ is not selected.

All samples are $\mathbf{X} = C(\mathbf{x})$, and each sample $\mathbf{x}_i$ is a cyclic shift of the base $\mathbf{x}$. The error in $\mathbf{x}_i$ increases as the shift increases. Besides, samples far away from the target has little effect on the regression. Hence, the samples with small shifts are selected to train the regression. We use a matrix $\mathbf{m}$, whose values are zeros if the corresponding samples are selected and ones otherwise.

The regression problem in (1) is described in its dual space as:

$$\min_{\boldsymbol{\alpha}} ||\boldsymbol{y} - \boldsymbol{K}\boldsymbol{\alpha}||_2^2 + \lambda \boldsymbol{\alpha}^T \mathbf{K} \boldsymbol{\alpha} \quad (6)$$

where $\mathbf{K}$ is the kernel matrix, with element $k_{ij} = \varphi(\mathbf{x}_i)^T \varphi(\mathbf{x}_j)$.

Different from [25], the constrain matrix $\mathbf{m}$ is involved in the regularization. Namely, We conduct the subjective sample selection in a latent way. To constrain the $\boldsymbol{\alpha}$ with the matrix $\mathbf{m}$, the regression problem is:

$$\min_{\boldsymbol{\alpha}} ||\boldsymbol{y} - \boldsymbol{K}\boldsymbol{\alpha}||_2^2 + \lambda \boldsymbol{\alpha}^T \mathbf{K} \boldsymbol{\alpha} + \tau ||\mathbf{m} \odot \boldsymbol{\alpha}||_1 \quad (7)$$

where τ is the constrain parameter.

For convenient, we relax the ι_1-norm $\boldsymbol{\alpha}$ by introducing another variable $\boldsymbol{\beta}$:

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta}} ||\boldsymbol{y} - \boldsymbol{K}\boldsymbol{\alpha}||_2^2 + \lambda \boldsymbol{\alpha}^T \mathbf{K} \boldsymbol{\alpha} + \tau ||\mathbf{m} \odot \boldsymbol{\beta}||_1 + \mu ||\boldsymbol{\alpha} - \boldsymbol{\beta}||_2^2 \quad (8)$$

The optimization problem can be divided into two subproblems with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$:

$$\min_{\boldsymbol{\alpha}} ||\boldsymbol{y} - \boldsymbol{K}\boldsymbol{\alpha}||_2^2 + \lambda \boldsymbol{\alpha}^T \mathbf{K} \boldsymbol{\alpha} + \mu ||\boldsymbol{\alpha} - \boldsymbol{\beta}||_2^2 \quad (9)$$

$$\min_{\boldsymbol{\beta}} \tau ||\mathbf{m} \odot \boldsymbol{\beta}||_1 + \mu ||\boldsymbol{\alpha} - \boldsymbol{\beta}||_2^2 \quad (10)$$

Eq. (9) can be solved with least squares, the solution has a closed-form:

$$\boldsymbol{\alpha} = (\mathbf{K}^H \mathbf{K} + \lambda \mathbf{K} + \mu \mathbf{I})^{-1} (\mathbf{K}^H \mathbf{y} + \mu \boldsymbol{\beta}) \quad (11)$$

With the circulant nature of $\mathbf{K}$ in [8], Eq. (11) can be solved in Fourier domain by:

$$\hat{\boldsymbol{\alpha}} = \frac{\hat{\mathbf{k}}^{\mathbf{xx}} \odot \hat{\mathbf{y}} + \mu \hat{\beta}}{\hat{\mathbf{k}}^{\mathbf{xx}} \odot conj(\hat{\mathbf{k}}^{\mathbf{xx}}) + \lambda \hat{\mathbf{k}}^{\mathbf{xx}} + \mu} \quad (12)$$

As for Eq. (10), for each element in $\boldsymbol{\beta}$, it is a standard ι_1-regularized least squares problem, the solution is as follows:

$$\boldsymbol{\beta} = \delta(\frac{\tau}{2\mu}\mathbf{m}, \boldsymbol{\alpha}) \quad (13)$$

where δ denotes the shrinkage operator, $\delta(\epsilon, x) = sign(x) max(0, |x| - \epsilon)$.

Accordingly, the kernel constrained regression can be solved iteratively by optimizing $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ alternately. We define that the iteration is converged when the average difference of the absolute value of $\boldsymbol{\alpha}$ elements is very small.

The KCF uses a Hann window $coswindow = hann(H)'hann(W)$ to decrease the effect of error samples. But the Hann window will result in the tracker preferring detecting a closer object as the target. Owing to the sample selection strategy, we use $coswindow^{0.5}$ that make the tracker can detect target far from the center and do not worry about the effect of error samples.

D. Implement Details and Complexity

In this section, we display the implement details such as the motion model, the update strategy, the scales strategy and the location method. The complexity of our methods is also explained.

We extend the sample region of the KCF to solve fast motion and occlusion. However, the more the sample region is extended, the more distractors or similar objects are included in the detection region. When the target occurs appearance variations, the regression may detect the distractors as the target. Although we reduce the error in training process, it is inevitable that the single regression cannot discriminate the distractors in the detection. To achieve a compromise for the region, we propose a novel region setting method. Our method is different with existing CF-based trackers [8]–[13], whose region sizes are several times the size of the target. When the aspect ratio of the target is great, such as a person, the detection region is unreasonable as shown in Fig. 3(a). It is more reasonable to think the target has the same velocity in different directions. Namely the distance between the target's boundary and the region's boundary is a fixed value d, as shown in Fig. 3(b). For all videos, we obtain the d by keeping a stationary area $(1 + paddings)$ times of the target area. So, our motion model is more robust to fast motion and distractors when the target has a big aspect ratio. Our region can not be applied to the existing CF-based trackers, because the proportion of the correct and error samples will decrease in the direction of greater target size, such as the vertical direction in Fig. 3(b). Our tracker can adapt different region sizes, because the error samples are dropped by the sample selection strategy.

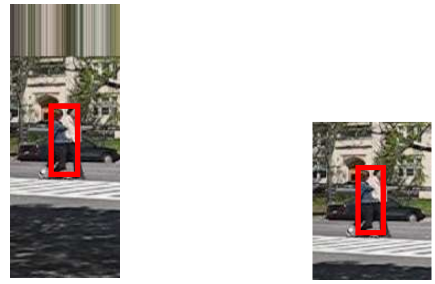
Because the target is tracked in successive frames, the model parameters are updated frame by frame to adapt the appearance variations of the target,

$$\hat{\mathbf{x}}_t = (1-\eta)\hat{\mathbf{x}}_{t-1} + \eta\hat{\mathbf{x}} \tag{14}$$

$$\hat{\boldsymbol{\alpha}}_t = (1-\eta)\hat{\boldsymbol{\alpha}}_{t-1} + \eta\hat{\boldsymbol{\alpha}} \tag{15}$$

To estimate the target's scale, we apply the scale adaptive strategy similar to [12]. The samples $\mathbf{z}_s$ with different scales are extracted and resized to the same size with training samples. Then the location and the scale of the target are obtained simultaneously:

$$x, s = \underset{x\in\mathbf{z}_s, s\in Scales}{argmax}\ \mathbf{y}(\mathbf{z}_s) \tag{16}$$



(a) The region in other CF-based trackers (b) The region in our method

Fig. 3. The examples of the regions in our and other CF-based trackers

where $Scales$ is the scale set.

The KCF has a computational complexity of $O(Nlog(N))$, where N is the quantity of each floor of the base feature vector $\mathbf{x}$. For target enhancement, the computational complexity is $O(N)$. For sample selection and kernel constrained correlation filter learning, the computational complexity is $O(DNlog(N))$, where D is the number of iterations. In detection, the computational complexity is $O(SNlog(N))$,where S is the number of scales for detection. So the computational complexity of our method is $O((D + S)Nlog(N))$.

IV. Experimental Results and Analysis

A. Experimental Setting

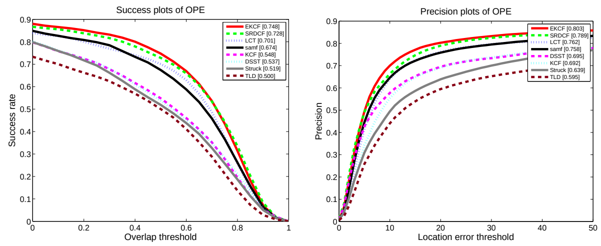
The proposed tracker EKCF is evaluated in all 100 video sequences in OTB100 [2], and compared with 7 state-of-the-art trackers: SRDCF [9], LCT [18], SAMF [12], KCF [8], DSST [13], TLD [6], Struck [7]. We use the HOG [26] and Colornames [15] features. Our tracker ran at a speed of 11.9fps using MATLAB R2012 on a 3.3 GHZ Intel Core i3 PC with 4 GB RAM.

The parameters we used are $paddings = 10$, $\theta = 15$, $\eta = 0.01$, $Scales = \{0.985, 0.99, 0.995, 1, 1.005, 1.01, 1.015\}$, $\lambda = 10^{-4}$, $\tau = 1$, and $\mu = 5 \times 10^{-5}$.

B. Quantitative Evaluation

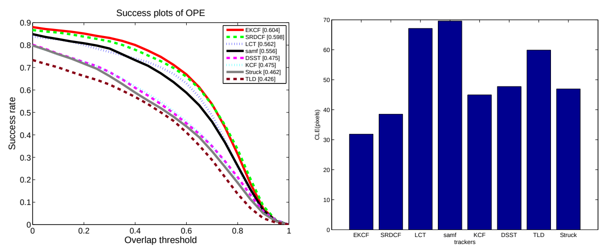
To assess the performance, we used four criteria: the overlap success (OS) rate at an overlap threshold of 0.5, the distance precision (DP) at a threshold of 20 pixels, the area under curve (AUC) of OS and the center location error (CLE). We report the overall performance for one-pass evaluation (OPE).

The curves of overlap success rate and distance precision are shown in Fig. 5. Our method yields the best performance with the success rate 74.8% at a overlap threshold of 0.5 and the distance precision 80.3% at a location error threshold of 20 pixels. The AUC of OS and the average center location error are exhibited in Fig. 5. The proposed EKCF has a maximum overlap success AUC score of 60.4% and a minimum average center location error. The Struck and TLD are two typical



(a) The overlap success rate curve (b) The distance precision curve

Fig. 4. The overall performance of OPE in OTB100. The legends present the values of OS at a threshold of 0.5 and the values of DP at a threshold of 20



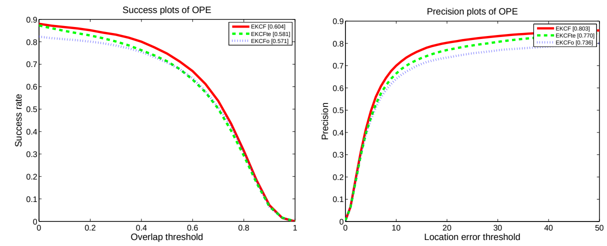
(a) The overlap success rate curve (b) The average center location error histogram

Fig. 5. The AUC of OS and the average CLE of the eight trackers

tracking-by-detection methods and illustrated first-class performance in several years ago. The DSST, KCF, and SAMF are all correlation filters based methods and exhibited promising performance in the recent two years. The SRDCF enlarges the region of KCF, whereas the LCT utilizes a detector to detect the target in the whole image when the target is lost. Their methods have been proved to be outstanding [27]. Our proposed methods outperforms the SRDCF and the LCT in four criteria. We also enlarge the region of the KCF, but we overcome the bound effect in nonlinear space by constraining the kernel and mapping the features to a target enhanced space. Hence, the EKCF gets a state-of-art experimental results, and keeps the nature of the original KCF. Furthermore, our method restrains the bound effect successfully as the SRDCF does. But the kernel nature is preserved in our method, whereas the SRDCF can only solve the bound effect in linear space. Also the speed of our method (11.9fps) is more than twice that of the SRDCF (5.3fps).

The SAMF improves the original KCF with features integration and scales adaption. We also utilize the two improvement in our tracker. The performance of the SAMF declines as the region was extended, but our EKCF raises the performance as the AUC score was from 0.556 to 0.604. We improve the SAMF with an extended region by using the target enhancement and sample selection strategy.

The effect of the target enhancement and the sample selection is displayed in Fig. 6 by testing our method with different components. The EKCFo is our proposed EKCF without target enhancement and sample selection. The EKCFte is our proposed EKCF without sample selection. From the curves, the target enhancement strategy increased 1% AUC score and



(a) The overlap success rate curve (b) The distance precision curve

Fig. 6. The OSAUC and DP of the EKCF, EKCFte and EKCFo

3.4% DP score. The sample selection strategy improved the EKCFte by 2.3% AUC score and 3.3% DP score. Both the target enhancement and the sample selection play significant roles in the EKCF. The target enhancement can decrease the effect of the great deal of background when the region is extended, and the sample selection can reduce the negative effect caused by error samples.

C. Qualitative Evaluation

Moreover, the videos in OTB100 are labeled with different attributes. The performance of the eight trackers in the videos under six main challenges are given in Fig. 7. Our EKCF achieves the best success rate in low resolution, scale variation, and occlusion attributes with the AUC score 0.538, 0.577, and 0.571, respectively. In illumination variation and fast motion attributes, the EKCF yields promising performance that is slightly lower than the SRDCF because the SRDCF utilizes a larger region than our EKCF. However, the EKCF outperforms the SRDCF in videos with low resolution, scale variation, occlusion and out of view because we reduce the errors caused by the extended region. In the videos featuring out of view, the SAMF gains the best, and the proposed EKCF ranks the second. It is because the EKCF will track a similar object when the target is lost because of the large region. But, the EKCF improves the SAMF in other situations.

The tracking results of the eight trackers in nine challenging video sequences are shown in Fig. 8. Fast motion, deformation, occlusion, scales variation and background clutters are included in the nine video sequences. In the $Couple$ sequence, fast motion occurs, our EKCF and the SRDCF can track successfully owing to the extended region. LCT and TLD can also follow the target because they use an extra detector. In the sequence $Human3$, the bounding box contains too many background, most trackers can not discriminate the target from the background. Owing to the target enhancement strategy, our EKCF is the only tracker that can track the target successfully. In $Lemming$, $Football$, and $Kitesurf$ sequences, occlusion and distractors in background are severe. Most trackers drift to other objects, whereas the EKCF has a high accuracy. In $Carscale$, most trackers can follow the target, but can not adapt the variation of target aspect ratio. In $Human4$, the target moves smoothly and background is complex. Extending the region of detection will increase the possibility of failure. But our EKCF does not fail until most parts of the target is

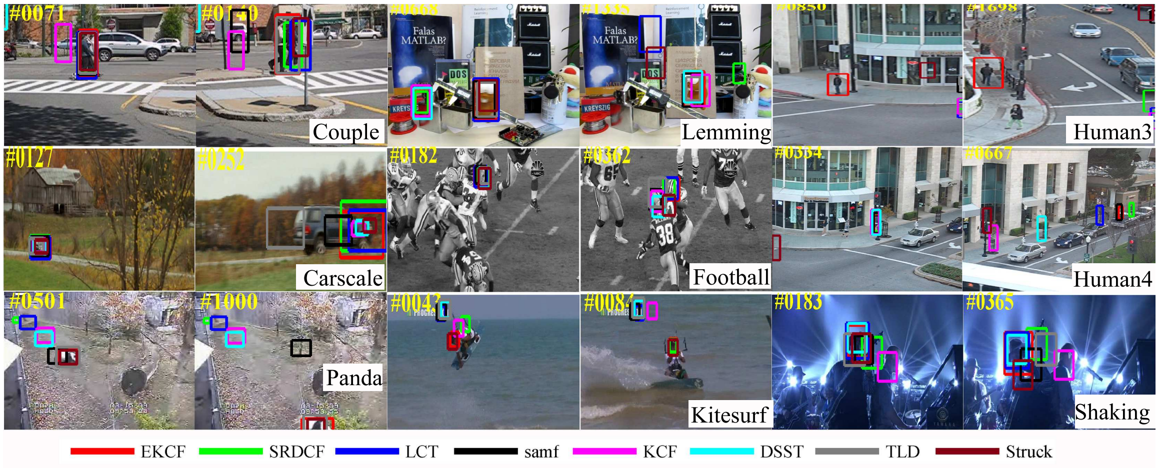


Fig. 8. Tracking results in 9 challenging video sequences

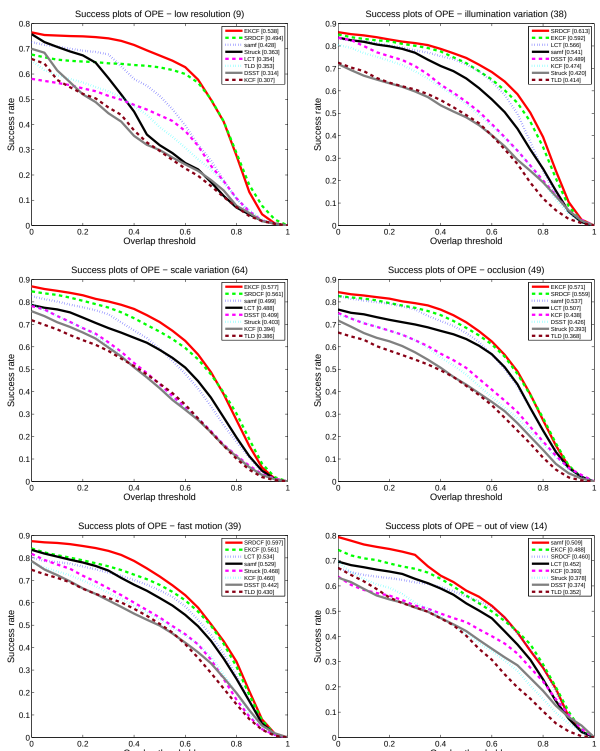


Fig. 7. The OSAUC in six main challenges.

occluded. The performance of the EKCF is not brought down by the extended region because of the target enhancement and sample selection strategies. In *Panda*, target deformation is severe, our EKCF and the Struck show high precision in tracking. Besides, the EKCF adapts the scales variation perfectly. In *Shaking*, the EKCF shows the robustness against background clutters and illumination variation.

V. Conclusion

In this paper, we propose an extended kernelized correlation tracker with target enhancement and sample selection. The proposed method can track target with fast motion by implementing an extended detection region. We analyse the reason about the precision decrease caused by extended region. We propose a target enhancement strategy for the circular samples to adapt different region sizes, so that the precision will not decrease as the region size changes. Our method can solve the bound effect in the non-linear space by selecting useful samples. Experimental results on challenging video sequences demonstrate the effectiveness and favorable performance of our tracker.

Acknowledgment

This work is partly supported by the National Natural Science Foundation of China (Grant No. 61671175 ,61370162 and 61672190).

References

[1] Y. Wu, J. Lim, and M.-H. Yang, "Online object tracking: A benchmark," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 2411–2418.

[2] ——, "Object tracking benchmark," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1834–1848, 2015.

[3] S. Oron, A. Bar-Hillel, D. Levi, and S. Avidan, "Locally orderless tracking," *International Journal of Computer Vision*, vol. 111, no. 2, pp. 213–228, 2015.

[4] D. A. Ross, J. Lim, R.-S. Lin, and M.-H. Yang, "Incremental learning for robust visual tracking," *International journal of computer vision*, vol. 77, no. 1, pp. 125–141, 2008.

[5] B. Babenko, M.-H. Yang, and S. Belongie, "Visual tracking with online multiple instance learning," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 983–990.

[6] Z. Kalal, K. Mikolajczyk, and J. Matas, "Tracking-learning-detection," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 7, pp. 1409–1422, 2012.

[7] S. Hare, S. Golodetz, A. Saffari, V. Vineet, M.-M. Cheng, S. L. Hicks, and P. H. Torr, "Struck: Structured output tracking with kernels," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 10, pp. 2096–2109, 2016.

[8] 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.

[9] 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.

[10] H. Kiani Galoogahi, T. Sim, and S. Lucey, "Correlation filters with limited boundaries," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4630–4638.

[11] 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.

[12] 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.

[13] M. Danelljan, G. Hager, F. S. Khan, and M. Felsberg, "Discriminative scale space tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016.

[14] J. Henriques, R. Caseiro, P. Martins, and J. Batista, "Exploiting the circulant structure of tracking-by-detection with kernels," *Computer Vision–ECCV 2012*, pp. 702–715, 2012.

[15] M. Danelljan, F. Shahbaz Khan, M. Felsberg, and J. Van de Weijer, "Adaptive color attributes for real-time visual tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1090–1097.

[16] T. Liu, G. Wang, and Q. Yang, "Real-time part-based visual tracking via adaptive correlation filters," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4902–4912.

[17] 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.

[18] C. Ma, X. Yang, C. Zhang, and M.-H. Yang, "Long-term correlation tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5388–5396.

[19] H. Li, Y. Li, and F. Porikli, "Deeptrack: Learning discriminative feature representations online for robust visual tracking," *IEEE Transactions on Image Processing*, vol. 25, no. 4, pp. 1834–1848, 2016.

[20] N. Wang and D.-Y. Yeung, "Learning a deep compact image representation for visual tracking," in *Advances in neural information processing systems*, 2013, pp. 809–817.

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

[22] C. Ma, J.-B. Huang, X. Yang, and M.-H. Yang, "Hierarchical convolutional features for visual tracking," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 3074–3082.

[23] Q. Liu, X. Ma, W. Ou, and Q. Zhou, "Visual object tracking with online sample selection via lasso regularization," *Signal, Image and Video Processing*, pp. 1–8, 2017.

[24] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[25] Y. Sui, G. Wang, and L. Zhang, "Correlation filter learning toward peak strength for visual tracking," *IEEE Transactions on Cybernetics*, 2017.

[26] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1. IEEE, 2005, pp. 886–893.

[27] M. Kristan, J. Matas, and A. Leonardis, "The visual object tracking vot2015 challenge results," in *Proceedings of the IEEE international conference on computer vision workshops*, 2015, pp. 1–23.