# Object Tracking using Correction Filter Method with Adaptive Feature Selection

Xiang Zhang, Yonggang Lu and Jiani Liu
*School of Information Science and Engineering, Lanzhou University, Lanzhou, Gansu, China*

Keywords: Correlation Filters, Lab Color Space, Confidence Score, Feature Selection.

Abstract: Correlation filter based tracking algorithms have shown favourable performance in recent years. Nonetheless, the fixed feature selection and potential model drift limit their effectiveness. In this paper, we propose a novel adaptive feature selection based tracking method which keeps the strong discriminating ability of the correlation filter. The proposed method can automatically select either the HOG feature or color feature for tracking based on the confidence scores of the features in each frame. Firstly, the response map of the color features and the HOG features are extracted respectively using correlation filter. The Lab color space is used to extract the color features which separate the luminance from the color. Secondly, the confidence region and the possible location of the target are estimated using the average peak-to-correlation energy. Thirdly, three criteria are used to select the proper feature for the current frame to perform tracking adaptively. The experimental results demonstrate that the proposed tracker performs superiorly comparing with several state-of-the-art algorithms on the OTB benchmark datasets.



## 1 INTRODUCTION

Visual object tracking is a very active part of the domain in computer vision, which has many applications in areas such as surveillance, automation and robotics (Wu, Lim and Yang, 2013). It is often used to track the object in a series of frames when a target is located in one of them. In many visual tracking tasks, the location of the target is known only in the first frame, and the estimation of the possible locations in other frames need to be made. Many existing algorithms assume that the target location changes little over time, and determine the target within a search window centered at the previous object location, which can be referred to as a motion model (Wang, Shi, Yeung and Jia, 2015). However, these algorithms may not be suitable for handling some complex scenarios, such as illumination variation, scale variation, occlusion and scale variations (Marvasti-Zadeh, Seyed Mojtaba and Cheng, 2019).

Visual object tracking has many algorithms to learn an appearance model of the target using either generative or discriminative methods. The generating model (Kwon and Lee, 2011; Liu, Huang, Yang and Kulikowski, 2011; Sevilla-Lara and Learned-Miller, 2012), such as sparse coding model (Mei, Ling, Wu and Blasch, 2013), transforms the problem of object tracking into the problem of sparse approximation, such as CT tracker (Zhang, Zhang and Yang, 2012). When the noise level is high, the tracking process is prone to the drift of the target frame. Most tracking algorithms use the discriminative methods, and the main idea is to train an online updated classifier, which gives the target position in each frame. In this kind of method, the correlation filter technology performs the complex convolution operation efficiently in the frequency domain, which improves the timeliness of target tracking, so it has attracted much research interest in the research field. In recent years, there are many algorithms showed a fusion of correlation filtering and deep learning. Depth features use different depth convolution features of convolution neural networks, which contain more advanced semantic information. Based on the correlation filtering algorithm of depth features, the target location is determined by calculating the relevant confidence responses of convolution layer features of different depths (Li, Ma and Wu, 2019). Although the depth convolution feature has a strong ability to identify the target, it is hard to understand the feature transformation in the "black box", and it also introduces a high computational complexity such that the resulting algorithm cannot achieve the real-



Zhang, X., Lu, Y. and Liu, J.
Object Tracking using Correction Filter Method with Adaptive Feature Selection.
DOI: 10.5220/0010196604800487
In *Proceedings of the 10th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2021)*, pages 480-487
ISBN: 978-989-758-486-2

time performance (Wang, Zhou, Tian, Hong and Li, 2018). So, it is still necessary to study the correlation filter technology without using the depth convolution features.

The discriminative model mainly distinguishes the target from the background by training the classifier. The convolution theorem shows that the time-consuming convolution operation can be converted into an efficient element dot product operation in the Fourier domain. Based on the observation, the correlation filtering technology is introduced into object tracking. Bolme et al. (2010) applied the correlation filtering algorithm to the target tracking task for the first time, and proposed the minimum mean square error (MOSSE) tracking algorithm which can perform tracking with high speed. However, the tracking accuracy of the algorithm is not ideal because of the grayscale characteristics of the single channel. Henriques et al. (2012, 2015) added kernel function to the framework of correlation filtering algorithm, replaced the grayscale feature of single channel with multi-channel HOG feature, and proposed kernel correlation filter (KCF) tracking algorithm. Through the cyclic matrix, the sampling problem in the training stage of the correlation filter template is equivalent to the cyclic shift operation of the Eigen matrix to complete the intensive sampling of the training samples. The cyclic sampling effectively increases the number of samples, and the robustness of the tracker is further improved. As for the scale adaptation, Danelljan et al. (2014) put forward the DSST method on the basis of MOSEE, which uses two single filters to estimate the position and scale. While the single feature or multi-feature fusion is usually used to feature extraction in the most trackers, Danelljan et al. (2014a, 2014b) used characteristic Color Names to extend CSK algorithm, which have obtained good results in the tracking of color video sequence. In the aspect of multi-feature fusion, HOG features and Color features are superimposed directly in a work (Li and Zhu, 2014). The Staple algorithm (Bertinetto, Valmadre and Golodetz, 2016) fuses HOG features and color histogram features (Possegger, Mauthner and Bischof, 2015) with weighting proportions of 0.7, which effectively promotes the robustness of tracking, but lacks pertinence to specific video scenes.

In the process of tracking, if the discrimination power of one feature is very different from another feature, using the two features together may result in low tracking accuracy. To deal with this problem, a correlation filter tracker based on adaptive feature selection using the confidence score of response map is proposed. In this paper, we calculate the confidence score of the response map using correlation filter for each feature (HOG feature and color feature in Lab color space), which can be used to select the features adaptively. For example, when the object is deformed, the HOG feature will be affected a lot. The color feature that is less affected than the HOG features will be selected for tracking. In the other case when the color feature of the object is disturbed a lot, the HOG feature will be selected for tracking. The difference between this method and other correlation filtering methods is that the more suitable features are selected according to the actual situation of each frame, which increases the flexibility in the feature selection in the complex scene, and can describe the target more accurately. The algorithm is tested on the evaluation criteria of OTB benchmark dataset and 100 video sequences, and its performance is compared with many mainstream algorithms.

The architecture of this paper is arranged as follows: Some related works are briefly reviewed in Section 2. Then, we introduce the proposed method based on the confidence score in HOG feature and color feature in Lab color space in Section 3. The experiment results are shown in Section 4. Finally, conclusions are given in Section 5.

# 2 RELATED WORK

## 2.1 Correlation Filters

Correlation filtering was first used in signal processing to describe the correlation between two signals. Initially, correlation filtering was applied to visual object tracking in grayscale images by Bolme et al. (2010). Later, the extension to multiple feature channels and HOG feature achieved the state of the art performance (Henriques, Caseiro, Martins, 2012). In 2014, DSST (Danelljan, Häger and Khan, 2014), a multi-scale template for Discriminative Scale-Space Tracking on the basis of MOSSE is used to deal with the scale change, with two filters to track the position change and the scale change respectively. The position filter is used to determine the new target position, and the scale filter is used for scale estimation. One deficiency of Correlation Filters is that they are constrained to learn from all circular shifts. Several recent works have sought to resolve this issue, and one of work, called Spatially Regularized Formulation (SRDCF) (Danelljan, Hager and Shahbaz, 2015), has demonstrated excellent tracking results. However, this improvement is achieved at the cost of real-time operation. The Staple algorithm (Bertinetto, Valmadre and Golodetz, 2016) propose a simple combination of template and histogram scores that are learnt independently to preserve real-time operation. The resulting tracker outperforms significantly more complex state-of-the-arts trackers in several benchmarks, but there is still the problem of sacrificing some of the characteristic advantages.

The tracking-by-detection paradigm is used in the Staple algorithm to calculate the response graph matrix in the correlation filters as:

$$\mathrm{F}_{w,h}\left(t\right)=F_{w,h}\left(T\left(x_{t},p\right);\theta_{\mathrm{t}\text{-}1}\right) \tag{1}$$

where $x_t$ represents the *t-th* frame image, w and h represent the position in the frame, T represents an image transformation function, p represents a rectangular window which gives a target location in $x_t$, F represents a candidate target, and $\theta$ is the model parameter. Let $S_t$ represents the set of candidate targets of the *t-th* frame, the score F($w$, $h$) of all candidate target F $\in$ $S_t$ is the response map matrix of the frame. The parameter $\theta$ is calculated by:

$$\theta_{\mathrm{t}\text{-}1}=\arg\min_{\theta\in Q}\{\mathrm{L}\left(\theta;x_{t-1}\right)+\lambda\mathrm{R}\left(\theta\right)\} \tag{2}$$

where the L($\theta$; $x_{t-1}$) represents the loss function, $X_t=X\{x_i,p_i\}_{i=1}^{t}$ represents the sequence of historical frame tracking results, R($\theta$) represents a regular term to prevent overfitting and the $\lambda$ is regular term parameter. The loss function is a linear combination of all sample losses, as shown in (3) and (4):

$$L\left(\theta,x_{t}\right)=\sum\nolimits_{i=1}^{t}\omega_{i}\ell\left(x_{i},p_{i},\theta\right) \tag{3}$$

$$\ell\left(x,p,\theta\right)=d\left(p,argmax_{q\in S_t}f\left(T\left(x,q\right);\theta\right)\right) \tag{4}$$

where d($p$, $q$) defines the cost of choosing rectangle $q$ when the correct rectangle is $p$. In our paper, we use the $h$ and $\beta$ represent the specific model parameters of HOG feature and color feature respectively which is solved by two independent ridge regression problems such as (5) and (6):

$$\mathrm{h}_{\tau}=\arg\min_{h}\left\{L_{temp1}\left(h;X_{t}\right)+\frac{1}{2}\lambda_{temp1}\left\|h\right\|^{2}\right\} \tag{5}$$

$$\beta_{\mathrm{t}}=\arg\min_{\beta}\left\{L_{temp2}\left(\beta;X_{t}\right)+\frac{1}{2}\lambda_{temp2}\left\|\beta\right\|^{2}\right\} \tag{6}$$

## 2.2 Color Features in Lab Color Space

At present, the input and output of most image capture devices are based on RGB color space, the traditional method is to covert the color images into RGB color space and extract the color features. But the three channels in RGB color space contain luminance information, and there is a strong correlation among them. Therefore, it may not be possible to obtain the desired effect by using these directly. Different from the most common RGB color space, Lab color space does not rely on light or pigments. It is a color space determined by the International Lighting Committee (CIE), which can theoretically describe all colors in nature. In the 3-channel component of Lab color space, $L$ represents luminance whose range is [0, 100] and the luminance increases with the numerical value. The ranges of $a$ and $b$ channel are both [-128, 127]. In our experiments, the color features in the Lab color space is used for the object tracking.

# 3 THE PROPOSED APPROACH

In this section, we first present the problem formulation of the adaptive feature selection method based on the confidence score. Then, we design an algorithm that builds up a bridge between the known correlation filter and our problem formulation.

How to judge whether the tracking result is accurate is a very important problem, because this determines the update strategy of the model. Many algorithms, such as KCF, DSST, SRDCF, Staple, do not judge the reliability of tracking results, and the results of each frame are updated either immediately or every N frames. This is unreliable, especially when the target is occluded, or the tracking has not followed well, and then update the model, which will only make the tracker more and more unable to recognize the target, which is the problem of model drift. In the correlation filter tracking, ideally, the response map of each frame is a single peak, and the corresponding position of the maximum value of the response map is the tracking result position of the current frame. The sharper the single spike in the response map means that the tracking result is more credible; conversely, the flatter the single spike means the less credible the tracking result is. However, because the actual scene is disturbed by complex factors, the actual response map may be multi-peak, and the correct target position can be at the highest peak of the response map. It can also be at the secondary peak or other peak.

In order to evaluate the effectiveness accurately, the average peak-to-correlation energy and the maximum response score $F_{max}$ of the response map are used together. The $F_{max}$ is defined as:

$$F_{max}=max\left(F_{w,h}\left(T\left(x_t,p\right);\theta_{t-1}\right)\right) \tag{7}$$

Generally, the value of the highest point $F_{max}$ in response map represents the result of center position, but, if only this condition is used for feature selection, the phenomenon of model drift will occur when there are multiple peaks in the response map. The $F_{max}$ does not reflect the degree of oscillation of the response map, a new criterion called average peak-to-correlation energy (APCE) is used, which is:

$$APCE=\frac{\left|F_{\max}-F_{\min}\right|^2}{mean(\sum_{w,h}(F_{w,h}-F_{\min})^2)} \tag{8}$$

where $F_{max}$, $F_{min}$, $F_{w,h}$ represents respectively the response at the highest, lowest and the *w-th* row and *h-th* column scores of F(*w, h*). This criterion can reflect the degree of oscillation of the response map. When the APCE suddenly changes, the target is occluded, or the target is lost in this feature which is unreliable in the frame.

When we begin a new tracking using the discriminative methods, the goal is to learn a suitable classifier which can extract the target from the background in real time. In the first frame, the location of the target is known. It can directly get the HOG feature and color feature using the correlation filter. In this paper, we use the confidence score to evaluate the reliability of each feature and select the suitable feature adaptively during the target tracking.

Assume there are t frames $X = [X_1, X_2, \dots, X_t]$. Give the location of the target in $X_1$, we learn a discriminative correlation filters in HOG feature and get the score $F_{HOG}$ at each pixel in the response map with (1). Then the color of RGB image is converted to the Lab color space. We also learn a correlation filters in color feature to computed the score $F_{color}$ with (1). Because the location of the target in the first frame is known, the (9) is used to get a standard confidence score $M_{HOG}$ and $M_{color}$ in the first frame to measure the reliability of the following frame. The threshold is set to evaluate the reliability of the next frame, which is:

$$\begin{aligned} threshold_{HOG} &= 30\% * M_{HOG} \\ threshold_{color} &= 30\% * M_{color} \end{aligned} \tag{9}$$

If the score of the next frame is greater than the threshold, the frame is considered to be a valid frame and added to the benchmark queue, otherwise the frame is considered invalid, which will cause the model drift and is discarded. The benchmark queue stores the scores of three valid frames and takes the average score as the benchmark:

$$S_{base}=\sum\nolimits_{t=1}^{n}\frac{APCE^t}{n} \tag{10}$$

In this paper, we use three indicators to select the optimal features adaptively. First, the score of the response map APEC represents the confidence degree of the current frame. We put the two kinds of scores under the same standard and compare them. Because the $M_{HOG}$ and $M_{color}$ is the standard score, we set a ratio to transform the $APCE_{color}$ to $APCE_{HOG}$, which is:

$$ratio^t=\frac{APCE^t_{color}*M_{HOG}-M_{color}*APCE^t_{HOG}}{M_{color}} \tag{11}$$

If the ratio>0, the color feature is better than the HOG feature. Second, a rate of changing the score is set to measure the confidence degree, which is:

$$change\_rate^t=\frac{APCE^t-S_{base}}{S_{base}} \tag{12}$$

When the rate of the HOG feature is greater than the color feature, we choose the stable feature of color feature in this case. Otherwise, the other one will be selected. Finally, we calculate the position offset of the maximum value of the response map $F_{max}$ in the current frame from the previous frame, which is:

$$\Delta^t=\sqrt{\left(x_t-x_{t-1}\right)^2+\left(y_t-y_{t-1}\right)^2} \tag{13}$$

Because a normal tracking changes smoothly between neighboring frames, if there is a sudden change in the offset of the maximum value, the model drift is considered to have occurred. We choose the smaller distance of each feature by comparing it separately. A confidence score is designed to measure the three criteria to select the suitable feature, which is:

$$\begin{cases} C\_Score^t = Sign(\alpha)+Sign(\beta)+Sign(\gamma), \\ \alpha=ratio^t, \\ \beta=change\_rate^t_{color}-change\_rate^t_{HOG}, \\ \gamma=\Delta^t_{color}-\Delta^t_{HOG} \end{cases} \tag{14}$$

where the Sign(x) is the symbolic function. In the definition of the *C_Score*, if the score is positive, the color feature is better than the HOG feature in at least two criteria, we select the color feature in the current frame for tracking and model update. Otherwise, the HOG feature is selected. An overview of the proposed method is summarized in Algorithm 1.

Algorithm 1: The proposed tracking algorithm.

Input: Frames {It}, initial target location $p_1$
Output: Target locations of each frame $\{p_t, t \neq 1\}$
1: REPEAT
2: Crop an image region from the last location $p_{t-1}$ and extract its response map in HOG feature.
3: Convert the RGB image to the Lab color space and extract the response map.
4: Calculate the $F_{max}$ and the APCE from the response map via (7) and (8).
5: Calculate three adaptively selected conditions via (11), (12) and (13).
6: IF the C_Score via (14) is positive THEN
7: Select the response map of color feature to update the model
8: ELSE
9: Select the response map of HOG feature to update the model
10: END IF
11: UNTIL end of video sequence.

## 4 EXPERIMENTAL RESULTS AND ANALYSIS

In this section, the proposed method is verified on two benchmark datasets, OTB13 (Wu, Lim and Yang, 2013) and OTB15 (Wu, Lim and Yang, 2015), and compared with several recent algorithms such as Staple, LMCF (Wang, Liu and Huang, 2017), SRDCF, DSST (Danelljan, Hager and Khan, 2017) and KCF. It is implemented in MATLAB with an Intel Core i7 3.60GHz CPU and 8GB of RAM.

We follow the evaluation standard provided by the benchmark OTB15 which includes 100 video sequences with various targets and backgrounds. In OTB15, four indices are used to evaluate all the compared algorithms with one-pass evaluation (OPE) such as bounding box overlap, center location error, distance precision and overlap precision. We evaluate the trackers according to the result with an error threshold of 20 pixels for the precision plots. For the success plots, the trackers are evaluated by the AUC scores.

Figure 1 shows the performance of our method with the other correlation filters on OTB-15. The proposed performs significantly better than the other methods. In the precision plot, our tracker performs 6% better than the Staple algorithm. The tracker also shows 7% better than the Staple in the success plot. For a more specific analysis, the performance of our tracker approach can be affected by several challenges as shown in Figure 2. It shows the performance of tracking method for various challenging attributes provided in the benchmark OTB-15 such as Illumination Variation (IV), Scale Variation (SV), Occlusion (OCC), Deformation (DEF), Motion Blur (MB), Fast Motion (FM), In-Plane-Rotation (IPR), Out-of-Plane-Rotation (OPR), Out-of-View (OV), Background Clutters (BC), Low Resolution (LR). Our method is effective in BC, DEF, IV, OPR and OCC compared to the existing approaches and more robust than the compared trackers for deformable object. Because the color feature of our algorithm is in Lab space, it can better

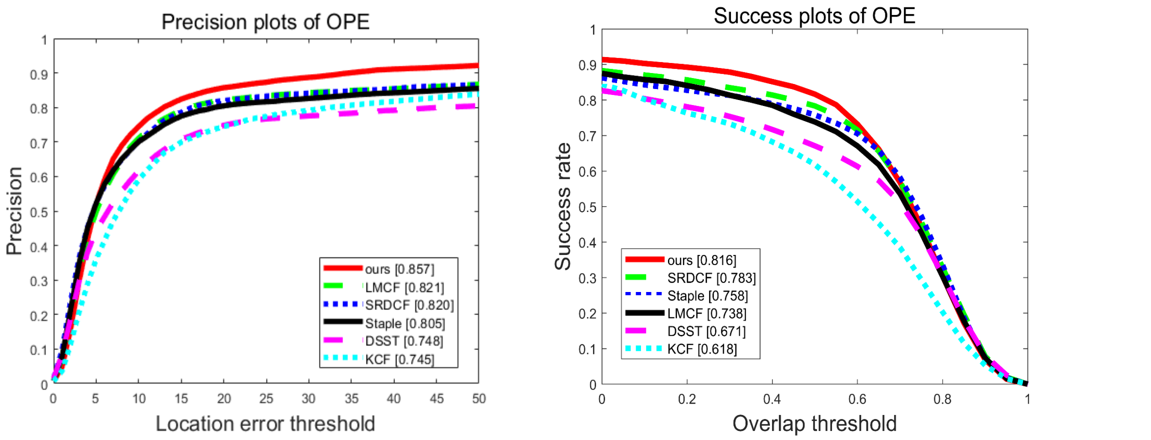


Figure 1: The precision plots(left) and success plots(right) of OPE on OTB-15. The numbers in the legend shows the precision scores and AUC scores for each tracker.

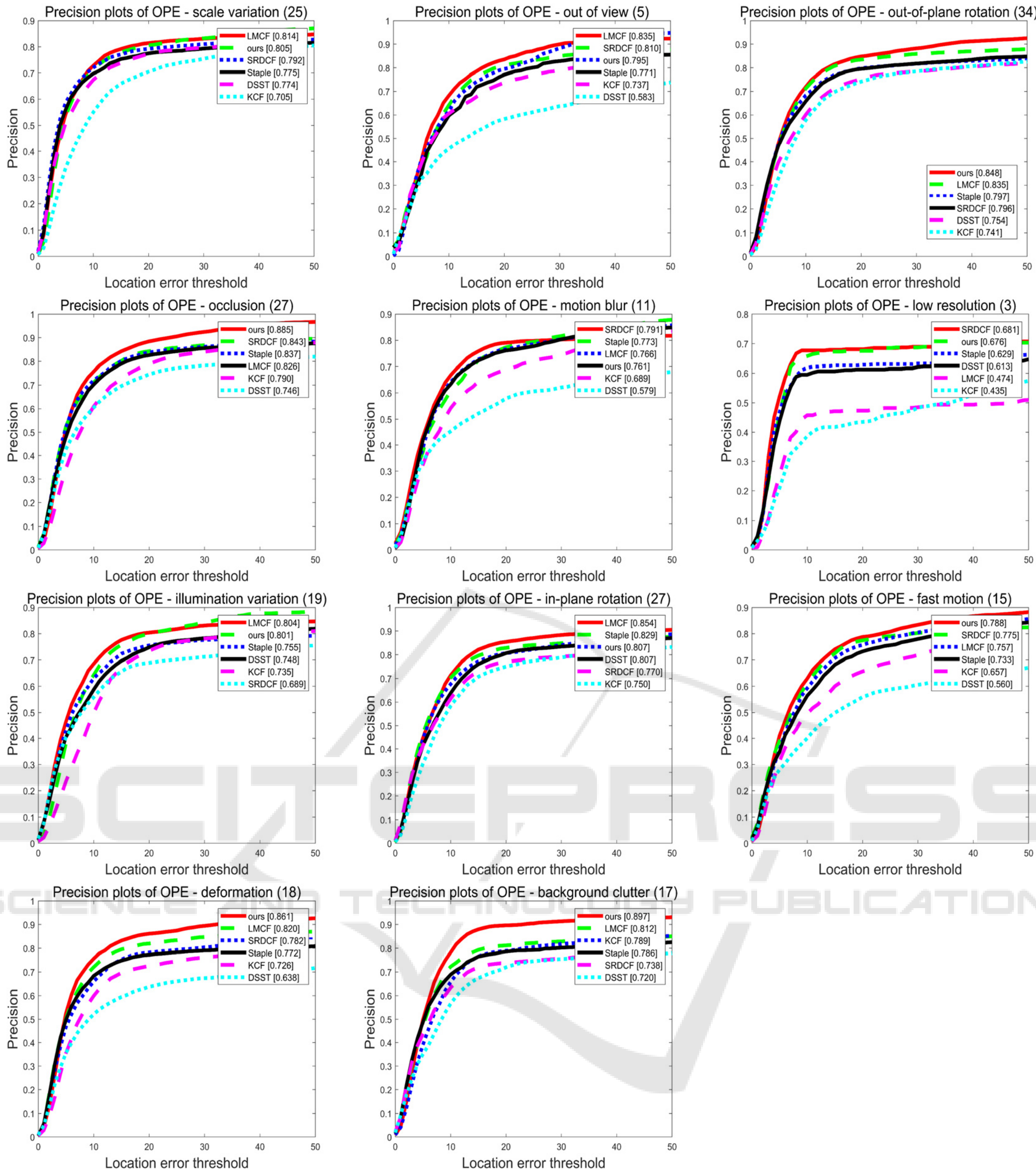




Figure 2: The success plots for 11 challenging attributes including background clutter, illumination variation, occlusion, deformation, out-of-plane rotation, out-of-view, scale variation, in-plane rotation, motion blur, fast motion, low resolution. The proposed tracker performs best or second best in almost all the attributes.

recognize the change of the different frame and enhance the tracking effect than other correlation filtering algorithms in RGB color space in IV. In addition, if the color feature does not change much, the HOG feature can enhance the tracking effect in the DEF. It can also be seen from Figure 2 that our tracker achieves better performance than the other trackers when the object suffers from fast motion and background clutters.

Figure 3 shows the results of different tracking algorithms. It can be seen that the proposed approach can handle different situations well. There are four types of video tracking results shown in Figure 3. The first row is the skiing sequence which represents a deformation scenario. In the video sequence the athlete is in a state of high-speed and deformation. The color feature of the player changes obviously in each frame and the HOG feature is not useful. The proposed method automatically selects the color features for tracking. Compared with other trackers,

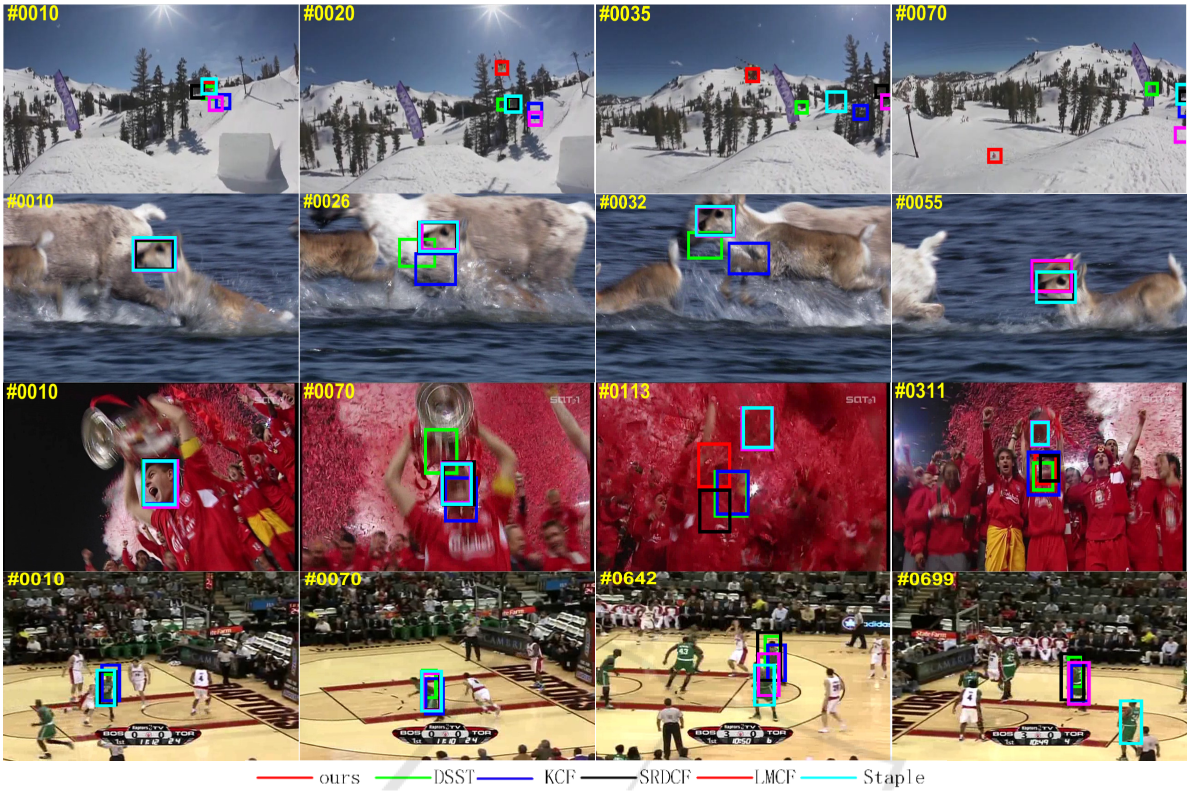


Figure 3: Several representative frames of different tracking algorithms. Our tracker exhibits robustness in challenging scenarios like deformation (row 1), fast motion (row 2), background clutter (row 3), occlusion (row 4). These sequences come from the OTB15 benchmark (Skiing, Deer, Soccer and Basketball). The red rectangle indicates the bounding box obtained by the proposed tracker.

it does not fuse the color feature with the HOG feature and improves the tracking effect. The second row is the deer sequence which represents the fast motion scenario. In the video sequence a deer runs in a high speed. The color feature is also automatically selected by the proposed method. The third row shows the results of the soccer sequence which represents a background clutters scenario. The target is the player dressed in red in the video sequence. From 70*th* frame to 113*th* frame, the background color also becomes red, so the APEC score of the color feature is lower than that of the HOG feature. The proposed tracker only uses the HOG feature for tracking and produces more accurate tracking results, while the other trackers use both the color feature and HOG feature and the tracking performance is deteriorated, resulting in the loss of tracking targets. The fourth row is the basketball sequence which represents an occlusion scenario. The target is partially or fully occluded in the video sequence. When the player is occluded, the color feature is affected and the HOG feature is automatically selected by the propose method to deal with this situation.

These results demonstrate that the proposed tracker is able to handle various categories of objects by selecting the feature adaptively.

## 5 CONCLUSIONS

In this paper, to improve the precision and success in target tracking, a method is proposed to select a feature adaptively based on a confidence score. The confidence score can be used to select the most suitable feature for tracking at each frame adaptively. The adaptive feature selection at the frame level is shown to be effective for improving the robustness in tracking. One disadvantage of the method is the introduction of the additional computation of the confidence score, although the real time tracking can still be performed. We will continue to work on improving the speed of the proposed method in the future work.

## ACKNOWLEDGMENT

This work is supported by the National Key R&D Program of China (Grants No. 2017YFE0111900, 2018YFB1003205).

## REFERENCES

Wu, Y. , Lim, J. , & Yang, M. H. . (2013). Online Object Tracking: A Benchmark. Computer Vision & Pattern Recognition. IEEE.

Wang, N. , Shi, J. , Yeung, D. Y. , & Jia, J. . (2015). Understanding and diagnosing visual tracking systems.

Marvasti-Zadeh, Seyed Mojtaba & Cheng, Li & Ghanei-Yakhdan, Hossein & Kasaei, Shohreh. (2019). Deep Learning for Visual Tracking: A Comprehensive Survey.

Sevilla-Lara, L. , & Learned-Miller, E. . (2012). Distribution fields for tracking. Computer Vision & Pattern Recognition. IEEE.

Liu, B. , Huang, J. , Yang, L. , & Kulikowski, C. A. . (2011). Robust tracking using local sparse appearance model and K-selection. IEEE Conference on Computer Vision & Pattern Recognition. IEEE.

Kwon, J. , & Lee, K. M. . (2011). Tracking by Sampling Trackers. International Conference on Computer Vision. IEEE.

Mei, X. , Ling, H. , Wu, Y. , & Blasch, E. P. . (2013). Efficient minimum error bounded particle resampling l1 tracker with occlusion detection. IEEE Transactions on Image Processing, 22(7), 2661-2675.

Zhang, K. , Zhang, L. , & Yang, M. H. . (2012). Real-Time Compressive Tracking. European Conference on Computer Vision. Springer, Berlin, Heidelberg.

Li, X. , Ma, C. , Wu, B. , He, Z. , & Yang, M. H. . (2019). Target-Aware Deep Tracking. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE.

Wang, N. , Zhou, W. , Tian, Q. , Hong, R. , & Li, H. . (2018). Multi-Cue Correlation Filters for Robust Visual Tracking. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE.

Bolme, D. S. , Beveridge, J. R. , Draper, B. A. , & Lui, Y. M. . (2010). Visual object tracking using adaptive correlation filters. The Twenty-Third IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2010, San Francisco, CA, USA, 13-18 June 2010. IEEE.

Henriques, J. F. , Caseiro, R. , Martins, P. , & Batista, J. . (2012). Exploiting the circulant structure of tracking-by-detection with kernels.

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

Danelljan, M. , Gustav Häger, Khan, F. S. , & Felsberg, M. . (2014). Accurate Scale Estimation for Robust Visual Tracking. British Machine Vision Conference.

Danelljan, M. , Khan, F. S. , Felsberg, M. , & Weijer, J. V. D. . (2014). Adaptive Color Attributes for Real-Time Visual Tracking. IEEE Conference on Computer Vision & Pattern Recognition. IEEE.

Cai, Z. , Wen, L. , Lei, Z. , Vasconcelos, N. , & Li, S. Z. . (2014). Robust deformable and occluded object tracking with dynamic graph. IEEE Transactions on Image Processing, 23(12).

Li, Y. , & Zhu, J. . (2014). A scale adaptive kernel correlation filter tracker with feature integration.

Bertinetto, L. , Valmadre, J. , Golodetz, S. , Miksik, O. , & Torr, P. H. S. . (2016). Staple: Complementary Learners for Real-Time Tracking. Computer Vision & Pattern Recognition. IEEE.

Possegger, H. , Mauthner, T. , & Bischof, H. . (2015). In defense of color-based model-free tracking. Computer Vision & Pattern Recognition. IEEE.

Danelljan, M. , Hger, G. , Khan, F. S. , & Felsberg, M. . (2016). Learning spatially regularized correlation filters for visual tracking.

Wang, M. , Liu, Y. , & Huang, Z. . (2017). Large margin object tracking with circulant feature maps.

Wu, Y. , Lim, J. , & Yang, M. H. . (2015). Object tracking benchmark. IEEE Transactions on Pattern Analysis & Machine Intelligence, 37(9), 1834-1848.

Babenko, B. , Yang, M. H. , & Belongie, S. . (2011). Robust object tracking with online multiple instance learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(8), 1619-1632.

Danelljan, M. , Hager, G. , Khan, F. S. , & Felsberg, M. . (2016). Discriminative scale space tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, PP(99), 1-1.