A Robust Parallel Object Tracking Method for Illumination Variations

Shuai Liu, Gaocheng Liu, Huiyu Zhou*

College of Computer Science, Inner Mongolia University, Hohhot, China, 010012

Inner Mongolia Key Laboratory of Social Comuting and Data Processing, Hohhot, China, 010012

***corresponding author. hz143@leicester.ac.uk**

Abstract- Illumination variation often occurs in visual tracking, which has a severe impact on the system performance. Many trackers based on Discriminative correlation filter (DCF) have recently obtained promising performance, showing robustness to illumination variation. However, when the target objects undergo significant appearance variation due to intense illumination variation, the features extracted from the object will not have the ability to be discriminated from the background, which causes the tracking algorithm to lose the target in the scene. In this paper, in order to improve the accuracy and robustness of the Discriminative correlation filter (DCF) trackers under intense illumination variation, we propose a very effective strategy by performing multiple region detection and using alternate templates (MRAT). Based on parallel computation, we are able to perform simultaneous detection of multiple regions, equivalently enlarging the search region. Meanwhile the alternate template is saved by a template update mechanism in order to improve the accuracy of the tracker under strong illumination variation. Experimental results on large-scale public benchmark datasets show the effectiveness of the proposed method compared to state-of-the-art methods.

Keywords—object tracking; illumination variation; parallel computing; correlation filters

1. Introduction

Object tracking is one of the important fields in computer vision, and has a wide range of applications in military security, human-computer interaction, robotics and other fields [1]. The purpose of object tracking is to forecast and track the object's state from video frames when the initial state of the target (such as position and size) is given in the first video frame. In recent years, visual tracking has attracted widespread attention, and a lot of work in this field have been conducted from various perspectives [2, 3]. During the movement of objects, the complex changes of the scene and target appearance, such as illumination change, deformation, occlusion and fast motion, make the object tracking face great challenges.

In order to improve the accuracy of object tracking, researchers have proposed a variety of tracking algorithms [4-6], [50-54]. Discriminative correlation filters have attracted more and more attention recently due to its high speed and robust performance [7-11]. The correlation filter was introduced for the first time by Galton in 1888 [12], after that it has been used to deal with various computer vision difficulties, such as object detection and recognition, attitude detection, and object tracking [49]. The DCF based methods use Fourier Transform to solve the target detection and training process in Fourier domain [13, 14]. These methods significantly reduce necessary calculations and accelerate the tracking process. DCF trackers perform dense sampling by utilizing cyclic shifts operation, which can expand the sample space and improve the tracking performance.

Although many advances have been made in DCF methods, there are still many challenges remaining in real tracking applications. The tracking of the target or background will be affected by dynamic changes, especially intense illumination variation. An example is presented in Figure 1, where the tracked object has been changed very much. In this condition, features of the target cannot be extracted, which cause the failure of tracking. Then the update step is prone to generate a model which has a poor ability to distinguish between the foreground and the background. Illumination changes often occur during tracking, such as the light and dark changes in the indoor lighting and outdoor shaded areas and non - shaded areas appear alternately. Most of the DCF based algorithms use Histogram of Oriented Gradient (HOG) [15] features to improve tracking performance under illumination variation. However, when the object is under intense illumination variation, the performance of DCF based trackers is poor. In order to solve the problem, we propose a novel strategy for visual tracking under intense illumination variation in this paper, which performs multiple regions detection and using alternate template (MRAT) based on parallel computing. The main contributions of our work are summarized below:

First, when undertaking detection, the common idea of the correlation filter-based trackers is to search an object in the area around the target in the previous frame. In order to improve the computing speed, the area will not be too large. With the help of parallel computing, we can perform simultaneous detection of the other 8 sub regions, equivalent to enlarging the search region which helps to improve tracking performance under intense illumination variation. Parallel Computing means that an application which is applied on a parallel machine is decomposed into a plurality of sub-tasks which are assigned to different processors. Each processor can cooperate with each other and execute subtasks in parallel, so as to speed up the process or improve the scalability of the solution [42].

Second, we preserve the template which is obtained from the previous frame as an alternate template. After the next few frames are detected, we use alternate templates for detection to prevent tracking drifts under intense illumination variation.

Third, we evaluate our approach on benchmark datasets with large-scale sequences and extensive experimental results show that the proposed MRAT method performs favorably against the other state-of-the-art tracking algorithms.

The rest of this paper is arranged as following: In section 2, we outlined some of the prior art related to this work. In section, we introduce the framework of the trackers based on correlation filters. The descriptions and details of our proposed methods (MRAT) are introduced in section 4. In section 5, we demonstrate detailed experimental results and analyze for the results. Finally we conclude our work and provide future directions in section 6.

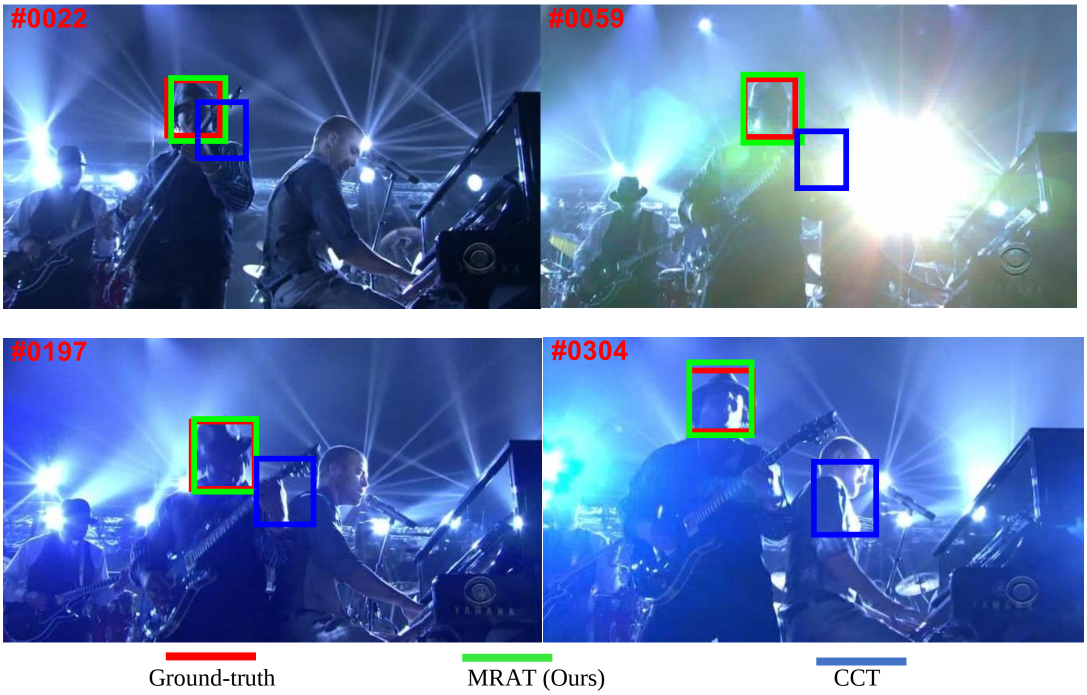


Figure 1. The tracking results of CCT (Zhu et al, BMVC 2015 [34]) and our approach (MRAT) in challenging situation of intense illumination change.

2. Related work

2.1 Correlation filters trackers

The correlation filtering algorithms have better tracking outcomes [16]. Bolme et al. [13] proposed a target tracking algorithm by learning a minimum output sum of squared error (MOSSE) correlation filter on a gray scale image. MOSSE applies correlation filtering to object tracking for the first time. This method is not only simple but also fast in tracking speed which is able to reach more than 600 frames per second (fps). Henriques et al. proposed the CSK algorithm [17] which trains a Ridge Regression Classifier [18] by intensive sampling. By combining convolution and cyclic matrix theory, a single-channel kernel-dependent filter is designed. However, the CSK algorithm can only use a single feature for learning and training, and then they proposed a Kernel Correlation Filter algorithm (KCF) [19]. The KCF algorithm extends the single channel kernel correlation filtering to multiple channels which allowed one to use robust features (e.g. HOG). Danelljan et al proposed a color names (CN) algorithm [7] which reduces the dimension of the target color attribute information [20] by doing Principal Component Analysis (PCA) [21] and then inputs it to the CSK tracker after their concatenation with the gray features. In [22], Ma et al. added an online random fern classifier as a long-term tracking re-detection component. Zhou et al. proposed a method which fuses a variety of features and presents an adaptive learning rate and an occlusion detection mechanism to update the target appearance model in presence of occlusion problem [55]. The method also can handle scale changes by using a scale adaptive scheme. In order to solve the scale variation problem, Ding et al. proposed a method called Quadrangle Kernelized Correlation Filters (QKCF) [56], which can simultaneously obtain the location and scale of the target object. Zhang et al. proposed the latent constrained correlation filters (LCCF) and introduced a subspace ADMM algorithm to solve the new learning model [57] which can improve the performance of the tracker very well. A two-camera system based on a RGB-D have been proposed in [58] which enables person detection, tracking and re-entry identification for real-time. The system can be the stepping stone for smart environment applications. Recently, recognizing the deep convolutional neural network (CNN) in the visual recognition tasks on a broad range of success, several tracking methods which combine deep features with correlation filters have been proposed [23,24]. Empirical studies using large target tracking datasets have shown that these CNN-based trackers perform exceptionally well for hand-feature-based methods.

2.2 Research on illumination variation

Illumination changes is a common problem in the tracking process. There are three types of methods commonly used to handle light changes: The first method is to recover as much information as possible from the image to reduce the effects of changes in lighting, such as Histogram equalization [25]. The second method uses light invariant features to describe objects, such as edge features. The last method focuses on describing the changes of images in different lighting environments, and builds models to predict possible images of targets in various lighting environments. To overcome the impact of intense illumination variation and increase the robustness of the visual tracking algorithm, researchers have proposed a number of methods. In [26], Li et al. incorporates the shape, texture, and color features of an image to describe objects. The algorithm is stable for human tracking in the light changing environment, but the real-time performance is poor. As the shape and texture features of object are not very similar before and after illumination change, it has poor tracking results under the illumination change. Chen et al. propose a method that extracts illumination invariant features to deal with tracking under illumination changes [27]. However this method is difficult to use for object tracking in actual scenarios due to the complicated feature extraction. A robust mean-shift tracking algorithm for illumination variation is proposed by Gerald et al. [28]. This method adopts the local background of a target to modify the target model dynamically, and it can reduce the shortcomings that are sensitive to changes in illumination when target modeling based on the color histogram. Zhu et al. establish an efficient target tracking algorithm based on normalized gradient matching which is insensitive to illumination change [29]. However, this method is difficult to be applied to moving target tracking in complex environments. Recently, Histogram of Oriented Gradient (HOG) [15] which constitutes the features by calculating and counting histograms of gradient direction in local region of image is widely used in tracking algorithms.

2.3 Parallel computing

Parallel computing is accompanied by the emergence of parallel machines and becomes an interdisciplinary subject in the last 30 years. It includes parallel machine architecture, compiler system, parallel algorithm, parallel programming, parallel software technology, parallel performance optimization and evaluation, parallel application, etc. [43]. Parallel Computing means that an application which is applied on a parallel machine is decomposed into a plurality of sub-tasks which are assigned to different processors. Each processor can cooperate with each other and execute subtasks in parallel, so as to speed up the solution or improve the scale of the problem [44]. In short, parallel computing uses multiple computing resources to solve a computational problem. A parallel machine consists of three elements: Node, Interconnect network and Memory [45]. Each node is composed of multiple processors, which can be directly input and output and all nodes can connect and communicate with each other through the Interconnect network. Memory consists of multiple memory modules, can be symmetrically distributed with the nodes on both sides of the Interconnect network. In 1966, American scholar Flynn divided computers into four categories according to the instruction flow and data stream of the system.

They are Single Instruction Single Data (SISD), Single Instruction Multiple Data (SIMD), Multiple Instruction Single Data (MISD) and Multiple Instruction Multiple Data (MIMD) [46]. Instruction flow refers to the amount of instruction flow that a computer can process at a single time point. Similarly, data flow refers to a number of data streams that a computer can process at a single time point. Parallel machines belong to SIMD, MISD and MIMD and most of the serial machines belong to SISD. Parallel algorithms are algorithms for solving problems and processing data on a variety of parallel computers. The essence of it is to map multiple tasks into multiprocessors, or to map realistic multidimensional problems into multiprocessors with specific topology [47]. According to different basic computing objects, Parallel algorithm can be divided into many categories such as Numerical parallel algorithm, Non-numerical parallel computation and Synchronized parallel algorithm [48].

3. Correlation Filters

Before introducing our method in detail, we first introduce the traditional the framework of Correlation Filters (CF) tracking. CF trackers first generate a series of samples by doing cyclic shifts based on the target area. This is a dense sampling process which can obtain lots of negative samples. Then, the kernel correlation filter is obtained by training the kernel ridge regression classifier with the circular sample sets. In the tracking process, the target position of the next frame is obtained by the response confidence map of the kernel correlation filter.

3.1 The ridge regression classifier

Assuming that x_i is the i^{th} sample, which is labeled as y_i. The purpose of training is to find the objective function $f(x) = w^T x$ to minimize the square error between prediction result and real label y_i:

$$\min_w \sum_i (f(x_i) - y_i)^2 + \lambda \|w^2\| \quad (1)$$

Where the first is empirical risk term, the second is the regularization term, and λ is the regularization coefficient of the relationship between the two terms. The solution shown in [30] can obtain the closed form solution of classifier weights based on formula (1):

$$w = (X^T X + \lambda I)^{-1} X^T y \qquad (2)$$

X represents the training sample matrix, y represents the corresponding labels which are generated by a Gaussian function based on the shifted distance, and I is the unit matrix. For ease of handling in frequency domain, X is given Hermitian Transformation, namely $X^H = (X^*)^T$, where X^* is a complex conjugate matrix of X and X^T is the transpose of X. Thus, formula (2) can be transformed into formula (3):

$$w = (X^H X + \lambda I)^{-1} X^H y \qquad (3)$$

The performance of the classifier can be improved by using Kernel function κ, which can map the input x to the high-dimensional feature space $\varphi(x)$. From the Representer Theorem [31], the weight w can be expressed as a linear combination of the sample x_i:

$$w = \sum_i \alpha_i \ \varphi(x_i) \qquad (4)$$

The solution to the kernelized versions of Eq. (4) is given by the following formula, K is the kernel matrix obtained by inputting the samples to the kernel function.

$$\alpha = (K + \lambda I)^{-1} y \qquad (5)$$

3.2 Circulant matrix and confidence map

Assuming there is a vector $x = (x_1 \cdots x_n)$, the circulant matrix has the following form:

$$\mathrm{X} = \mathrm{C}(x) = \begin{bmatrix} x_1, x_2 & \cdots & x_n \\ \vdots & \ddots & \vdots \\ x_2, x_3 & \cdots & x_1 \end{bmatrix} \qquad (6)$$

The circulant matrix has some excellent properties, such as that any circulant matrix can be diagonalized by the Fast Fourier Transform (FFT) of the original vector. We can obtain the solution of Eq. (5) in the Fourier domain as follows.

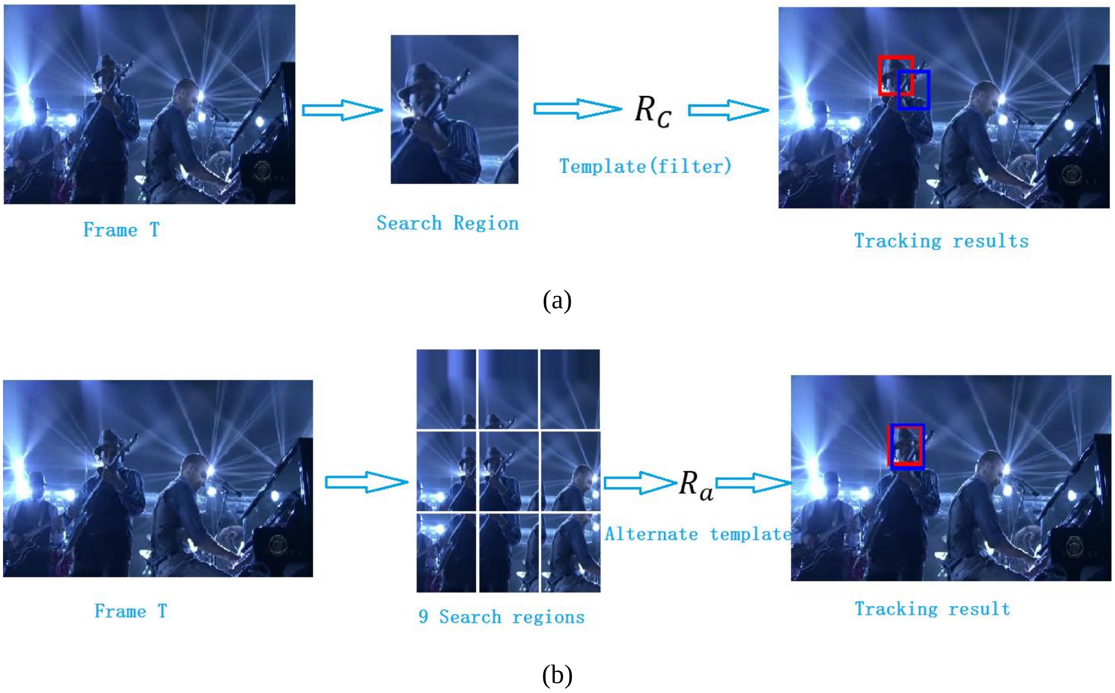
$$\hat{\alpha} = \frac{\hat{y}}{\hat{\kappa}^{xx} + \lambda} \qquad (7)$$

Where κ^{xx} is the first row of the kernel function matrix $\mathrm{K} = \mathrm{C}(\kappa^{xx})$, $\hat{\alpha}$ and $\hat{y}$ are Discrete Fourier transform (DFT) of α and y. Then we circulate the search area z of the current frame to get the circulant test samples and it can evaluate every cyclic patch of z by:

$$f(z) = \mathcal{F}^{-1}\left(\hat{k}^{xz} \odot \hat{\alpha}\right) \ (8)$$

Where $f(z)$ are the Confidence map for all the cyclic shifts of z and $\mathcal{F}^{-1}$ represents the Inverse Fourier transform. According to the correlation filtering principle, the maximum confidence value is selected as the target in the current frame.

4. Proposed approach



(a)

(b)

Figure 2. The flowchart of our strategy when frame at T moment under the intense illumination variation. The red box is the ground truth and the blue box is the tracking result of the trackers. Figure 2(a) shows the tracking result of the baseline algorithm. R_c is the template (context regression model) and it will be updated every frame. Figure 2(b) shows the tracking result of our approach. The R_a is the alternate

template which is saved before intense illumination occurs.

```
Algorithm 1: The strategy of improved tracking algorithm
 Input : Initial target bounding box $x_1$; the threshold $\mathcal{T}_a$, the threshold $\mathcal{T}_b$, the parameter $\mathrm{K}$ and the parameter $\rho$.
 Output : Estimated object position $\mathrm{X}_t = (x_t, y_t)$, context regression model $R_c$.
 repeat
   Detect: According to the position $(x_{t-1}, y_{t-1})$ to crop out the searching window at frame t and extract the features; Compute the correlation response $y_t$ using $R_c$ and (8) to estimate the new position$(x_t, y_t)$.
   If $\max y_t < \mathcal{T}_a$ then
     Based on the position $(x_{t-1}, y_{t-1})$ to calculate other 8 search regions according to Eqs.10a-10h
     Do Detect simultaneously for the 8 search regions based on parallel computing and obtain their correlation responses $y_t^i$ $(i = 1 \cdots 8)$
     $y_m$ = Max $(y_t^i)$
       If $y_m > \mathcal{T}_b$
         $y_t = y_m$
       Else
         $y_t = y_t$
       end
   end
    Save $R_c$ as the alternate template $R_a$
    Update $R_c$ according to the Eqs.11a-11b
   If $\mathrm{t} < \mathrm{frame} < \mathrm{t} + \mathrm{K}$ then
       Do Detect, using $R_c$ to compute the correlation map $y_t$
       Update $R_c$ according to the Eqs.11a-11b
    Else if $\mathrm{frame} = \mathrm{t} + \mathrm{K}$ then
       When do Detect, using $R_a$ to compute the correlation map $y_t$
         If $\max y_t < \mathcal{T}_a$ then
           Based on the position $(x_{t-1}, y_{t-1})$ to calculate other 8 search regions according to Eqs.10a-10h
           Do Detect simultaneously based on parallel computing for the 8 search regions and obtain their correlation responses $y_t^i$ $(i = 1 \cdots 8)$
           $y_m$ = Max $(y_t^i)$
           If $y_m > \mathcal{T}_b$
             $y_t = y_m$
           Else
             $y_t = y_t$
           end
       end
       Let $R_c = R_a$
       Update the $R_c$ according to the Eqs.11a-11b
   End
 until End of video sequences;
```

In this section, we describe the proposed approach which performs multi-regions detection and uses alternate template (MRAT) in detail. The main purpose of our method is to improve the robustness and accuracy of the correlation filter-based trackers for handling intense illumination change. First, we analyze the confidence of the objects obtained under intense illumination changes. Then, we combine parallel computation to detect multiple regions. Finally, we introduce our model update mechanism which saves an alternate template when illumination changes occur. We outlined our strategy in algorithm 1 and display the flowchart of our approach in Figure 2.

4.1 The confidence of target under illumination variation

In the detection part of the tracking process, most of the CF trackers will take dense sampling through cyclic shift operations around the target region in the last frame. Then it will generate many candidate windows and each candidate window will be described by the appearance model. The appearance model is an important part of tracking algorithms, which mainly consists of two parts: visual representation and statistical modeling [32]. Visual representation uses different types of visual features to describe objects. Statistical modeling exploits statistical learning techniques to build the object models. Finally, we use Eq. (8) to compute the confidence of each window which represents the possibility of being the target and the one with the highest score is chosen to be the target. When the time interval is very small, the changes between the two consecutive frames are not very large, and the background around the target remains possibly unchanged. For this reason, the background information surrounding the target is helpful to find the target and the appearance model in the CF trackers usually larger than the target area as figure 3 shows. The visual features easily change dramatically when objects are in the strong light. In this case, the confidence score of the object will be very low. We first compute the similarity values S_v of two image patches which have the same foreground target, but the background is completely different. When the tracking failure occurs under intense illumination, the confidence score of the object is lower than a threshold T_a which can be obtained by Eq.9.

$$T_a = \frac{S_v}{\eta} \quad (9)$$

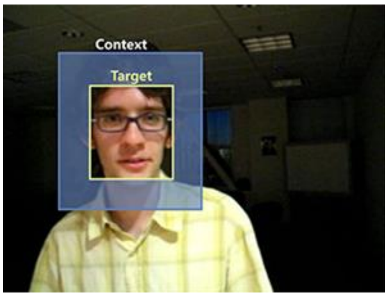


Figure 3. The image patch containing the target, which is in the yellow rectangular and the background context (the Blue areas) surrounding it.

4.2 Multi regions detection based on parallel computing

In order to make an accurate tracking of the object under intense illumination, we perform detection of the other 8 sub regions with the help of parallel computing simultaneously. Parallel computing can solve a computational problem by using multiple computing resources simultaneously. It is assumed that the target detection part in the tracking process is considered as a computational problem. The traditional method is to search objects in one region. Now, we perform detection in another 8 regions and they have the same size as the original search region. In this way, we divide the computational problem of the detection part into 9 small sub problems that can be executed concurrently. Therefore, the detection of sub regions can be executed on different processors simultaneously to maintain high computational efficiency. If the other 8 search regions are adjacent to the original search region, this means that the 9 search regions are adjacent to each other and have no intersection, which can lead to the following problems. When the part of the tracking target is beyond the boundary of the search region, it causes the target to be divided into two search regions and both of the two search regions fail to detect the target which ultimately lead to tracking failure. In order to get over this situation, we consider when the target is just in the boundary of one search area. The search area adjacent to it intersects with this region, and the intersected part just contains the target. Then the center distance of the two search regions is calculated in this case to obtain the location of the other 8 search regions. As mentioned above, the search region is ρ times larger than the target. The position of the target in the previous frame is $P^{t-1}(x,y)$, the x and y are the central coordinate of the target, then the location of other 8 search regions can be calculated, see Eqs.10a-10h.

$$P_1^t = (x, y+(\rho-1)n) \quad (10a) \quad P_2^t = (x, y-(\rho-1)n) \quad (10b)$$
$$P_3^t = (x+(\rho-1)m, y) \quad (10c) \quad P_4^t = (x-(\rho-1)m, y) \quad (10d)$$
$$P_5^t = (x+(\rho-1)m, y+(\rho-1)n) \quad (10e) \quad P_6^t = (x-(\rho-1)m, y-(\rho-1)n) \quad (10f)$$
$$P_7^t = (x+(\rho-1)m, y-(\rho-1)n) \quad (10g) \quad P_8^t = (x-(\rho-1)m, y+(\rho-1)n) \quad (10h)$$

Where m and n are the size of the target box, and $P_n^t\ (n = 1 \cdots 8)$ are the locations of the search regions in the current frame. The results of the 9 search regions are shown in Figure 4.

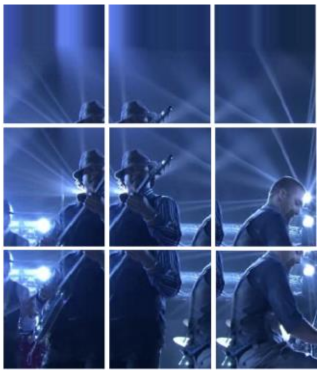


Figure 4. The results of 9 search regions. The region in the center is the area which methods based on Discriminative Correlation Filter (DCF) use to detect objects. Other 8 search regions are generated according to the middle one by Eqs.10a-10h.

After detecting in 9 regions simultaneously, each regions will give a candidate box to be the target with the confidence which represents the possibility of becoming the target. Then we choose a box with the highest confidence of these nine candidate boxes to be the final tracking result of this frame. We've added multiple search regions, and it is equivalent to expanding the search area which will lead to more background clutters. Background clutters means that the background has the similar color or texture as the target are included. Therefore, in order to prevent tracking error targets under illumination variation, the confidence of the 8 extra search regions should be higher than a threshold $\mathcal{T}_b$. In this way, they can be chosen as the final tracking result.

4.3 Alternate template

Because the target changes constantly in the process of tracking, we need to update the model in time so that it can adapt to the change of the target. In DCF trackers, each frame can produce a correlation filter (template) according to training schemes described above. Most of the correlation filter-based trackers adopt the real-time template (filter) updates, they update the correlation filter in every frame by combining the filter obtained from the previous frame with existing trained filter. The template is updated by Eqs.11a-11b, which linearly interpolates the obtained values of α and x with the ones from the previous frame, to provide the tracker with some memory of the target information in the previous frame to improve the robustness of tracking. Where t is current frame and θ is interpolation weights. The $\hat{\alpha}$ and $\hat{x}$ are Discrete Fourier transform (DFT) of α and x. The $\hat{x}^t$ is the object appearance model trained by the current frame. The $\hat{\alpha}^t$ is the dual variable in the current frame. One problem with this approach is that when the template does not describe the target well under intense illumination conditions, it will cause the tracking to fail owing to the feature of the target has changed a lot and it can't match the template very well. In order to improve the performance of the tracker under illumination change, we have made improvements to the template update process by using alternate template. In the current frame, when illumination change occurs, the confidence values of the selected target will be lower than the threshold T_a, we save the correlation filter which has been trained from the previous frame as an alternate template R_a. An intense lighting event in the field of view can last for a period of time, after K frames covering the duration of this event, the intense light condition would have passed and the target will recover from it. At this point, we use the alternate template to detect the object and update the template. During the K frames, the existing template (context regression model) R_c is used to detect the target and update the template. Meanwhile the template R_a is in reserve and not updated.

$$\hat{X}^t = (1-\theta)\hat{X}^{t-1} + \theta\hat{x}^t \qquad (11a)$$

$$\hat{A}^t = (1-\theta)\hat{A}^{t-1} + \theta\hat{\alpha}^t \qquad (11b)$$

5. Experiment and Analysis

We use a well-known benchmark dataset – Online object tracking: A benchmark (OTB2013) [33] to evaluate our proposed algorithm. The benchmark dataset contains 50 challenging video sequences with 11 attributes, these are: 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) and low resolution (LR) respectively. Each attribute represents a specific challenging factor in object tracking. One sequence may be annotated with many attributes. OTB2013 uses one-pass evaluation (OPE) to evaluate tracers, which is referred to an initial test sequence, from the initial position of first frame to the last frame to evaluate trackers. When doing quantitative analysis, two metrics: precision and success plots are used. The precision metric computes the distance precision (DP) rate of frames whose center location is within some certain distance with the ground truth location. The success metric computes the overlap success (OS) rate between the tracked and ground truth bounding boxes. Usually, it report the area under curve (AUC) of the success plot and precision score at 20 pixels threshold corresponding to the one-pass evaluation for each tracking method.

Setup parameters.

The search region where most tracking algorithms use is 2.5 times larger the target, therefore the ρ which we use here is 2.5. We believed that when the target's confidence is less than 1.5 times the value S_v, the tracking results under this setting would not be considered reliable and require re-detection, hence we set $\eta = 1.5$. The value S_v which can be obtained by the method mentioned in section 3.2 is set to 0.4. Based on Eq. 1, the threshold T_a was set to 0.25. The threshold $\mathcal{T}_b$ is set to 0.27, and if the confidence of other 8 regions is higher than it, we can adopt their results. For the template update, we set $\mathrm{K} = 4$. This meant that after 4 frames, we would use the alternate template and no longer used the current template for testing and updating templates.

5.1 Overall performance

We conduct experiments on the OTB-2013 with comparisons to the baseline tracker (CCT) [34] which is one of the top performing trackers and other 7 state-of-the-art trackers, including ACFN-selNet [35], Staple [36], BIT[37], fDSST [38], SiamFC_3s [39], RPT [40] and CFNet-conv3 [41]. The performance results are illustrated in Figure 5 and Table 1. For the overall performance, the MRAT (Ours) obtains the best performance with an average DP of 83.4% in precision plots of OPE. In the success plots, our method MRAT also achieves excellent performance with the OS of 61.7% among these existing trackers. Compared with baseline the algorithm, our approach performs favorably against the CCT method with DP of 79.5% and OS of 59.4%. The results proves that our approach achieves good performance. The CFNet-conv3 uses the deep learning methods to represent object appearance and its performance with an OS of 61.0% is slightly weaker than our method. Overall, our method, MRAT, is shown to perform well both from the precision and success plots. This is unlike several other algorithms which only achieve good performance in either the precision or success plots such as ACFN-selNet tracker, which obtains the second-best score of 82.3% for DP but performs very poorly with the OS of 58.9% in the success plots. In addition, due to the adoption of a parallel framework, our method is able to achieve real-time tracking with speed of 28 frames per second (fps).

Table 1.The results of our method and other 8 state-of-the-art trackers. The two best trackers are shown coloured in red and blue.

	CCT [34]	ACFN-selNet [35]	Staple [36]	BIT [37]	fDSST [38]	SiamFC_3s [39]	RPT [40]	CFNet-conv3 [41]	MRAT (ours)
DP (%)	79.5	82.3	79.3	81.6	80.2	80.9	81.0	82.2	83.4
OS (%)	59.4	58.9	60.0	59.3	59.5	60.8	57.8	61.0	61.7

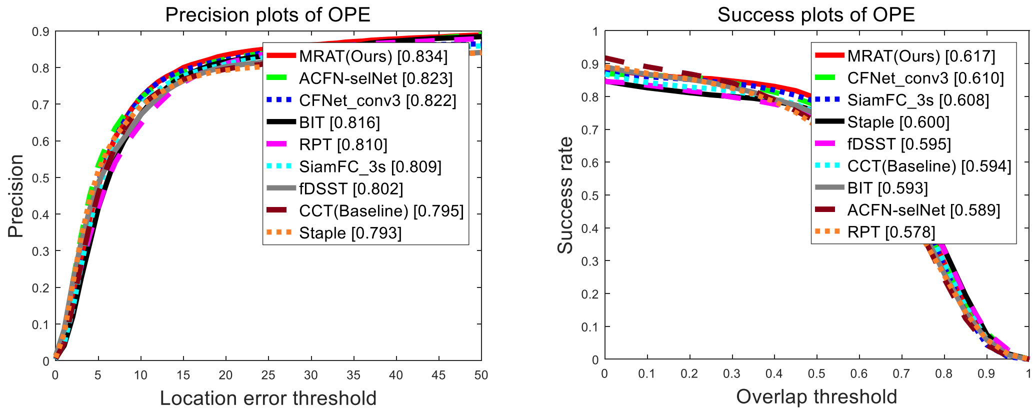
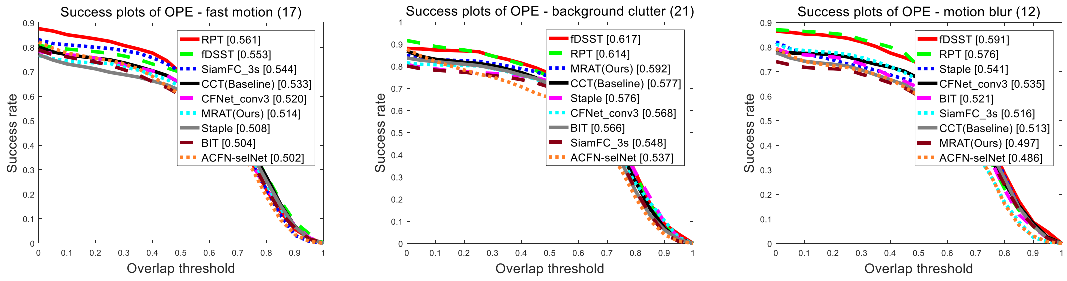


Figure 5. Precision and success plots using OPE on OTB2013. The numbers in the legend indicate the representative precisions at 20 pixels for precision plots, and the AUC scores for success plots. Our approach MRAT performs favorably against the other state-of-the-art trackers.

5.2 The performance on the subsets with different challenging attributes

In this section, we further analyze the performance of MRAT based on the attributes of image sequences in terms of 11 challenging factors. These attributes are useful for analyzing the performance of trackers in different aspects. For each challenging attributes, we do the performance evaluation on subsets with the challenging attributes in the benchmark. For example, when we evaluate the performance of the algorithms on the challenging attribute of scale variation. In the benchmark, there are 29 video sequences that are tagged with this challenging attribute, so we will perform experiment to evaluate the performance of the algorithms on these 29 video sequences. Here, we provide the overlap success plots in Figure 6, and the details of the attribute-based performance is shown in Table 2. The main purpose of our paper is to deal with tracking problem under intense illumination variation conditions, and the proposed MRTA method achieves remarkable improvements over the baseline method (CCT) that it has about 3.87% improvement in OS. In the sequences which contain illumination variation, our method MRAT is the second-best algorithm, with the OS of 58.3% among all the algorithms. The fDSST which proposes a novel scale adaptive tracking method has a slightly improved performance with an OS of 59.1%. From the results, it clearly shows that it is able to deal with the tracking well under the conditions of illumination variation and our method achieves superior performance and robustness in these scenarios. We note that the proposed tracking method also performs well in dealing with challenging factors including background clutters, deformation, out-of-plane rotation, in-plane rotation, out-of-view, occlusion and scale variation.



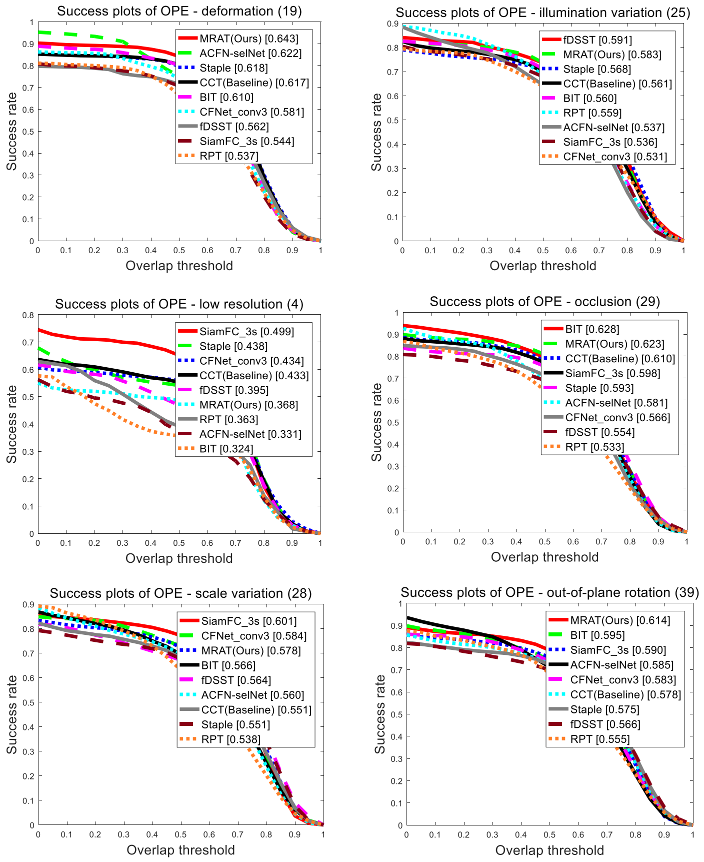


Figure 6. The success plots over 11 tracking challenging attributes of FM, BC, MB, OPR, DEF, SV, IV, OCC, OV, LR and IPR. The AUC values of each tracker are included in the legend. Our improved algorithm gets outstanding performance compared with other 8 algorithms especially in scenarios with illumination variation (IV).

Table 2: OS scores of our approaches compare with other 8 outstanding trackers on 11 Challenging attributes in OTB-2013. First, second, and third trackers are marked in red, blue, and green fonts, respectively.

	FM	BC	MB	DEF	IV	OCC	OPR	OV	SV	LR	IPR
CCT [34]	53.3	57.7	51.3	61.7	56.1	61.0	57.8	63.0	55.1	43.3	54.7
ACFN-selNet[35]	50.2	53.7	48.6	62.2	53.7	58.1	58.5	60.9	56.0	33.1	56.0
Staple [36]	50.8	57.6	54.1	61.8	56.8	59.3	57.5	54.7	55.1	43.8	58.0
BIT[37]	50.4	56.6	52.1	61.0	56.0	62.8	59.5	55.2	56.6	32.4	55.7
fDSST [38]	55.3	61.7	59.1	56.2	59.1	55.4	56.6	55.6	56.4	39.5	57.7
SiamFC_3s [39]	54.4	54.8	51.6	54.4	53.6	59.8	59.0	63.5	60.1	49.9	57.0
RPT [40]	56.1	61.4	57.6	53.7	55.9	53.3	55.5	57.6	53.8	36.3	56.9
CFNet-conv3 [41]	52.0	56.8	53.5	58.1	53.1	56.6	58.3	42.3	58.4	43.4	56.5
MRAT (proposed)	51.4	59.2	49.7	64.3	58.3	62.3	61.4	61.2	57.8	36.8	58.9

5.3 The tracking visualization results on challenging sequences

We compare our algorithm to the basic trackers on some challenging sequences, which are mainly under intense illumination changes as shown in Figure 7. As we can see, when intense illumination changes occur, the baseline tracker (CCT) obviously can't keep up with the target, but our MRAT can deal with this kind of scene very well. The results in the figure demonstrate that our method has an improved tracking effect in the case of intense illumination changes.

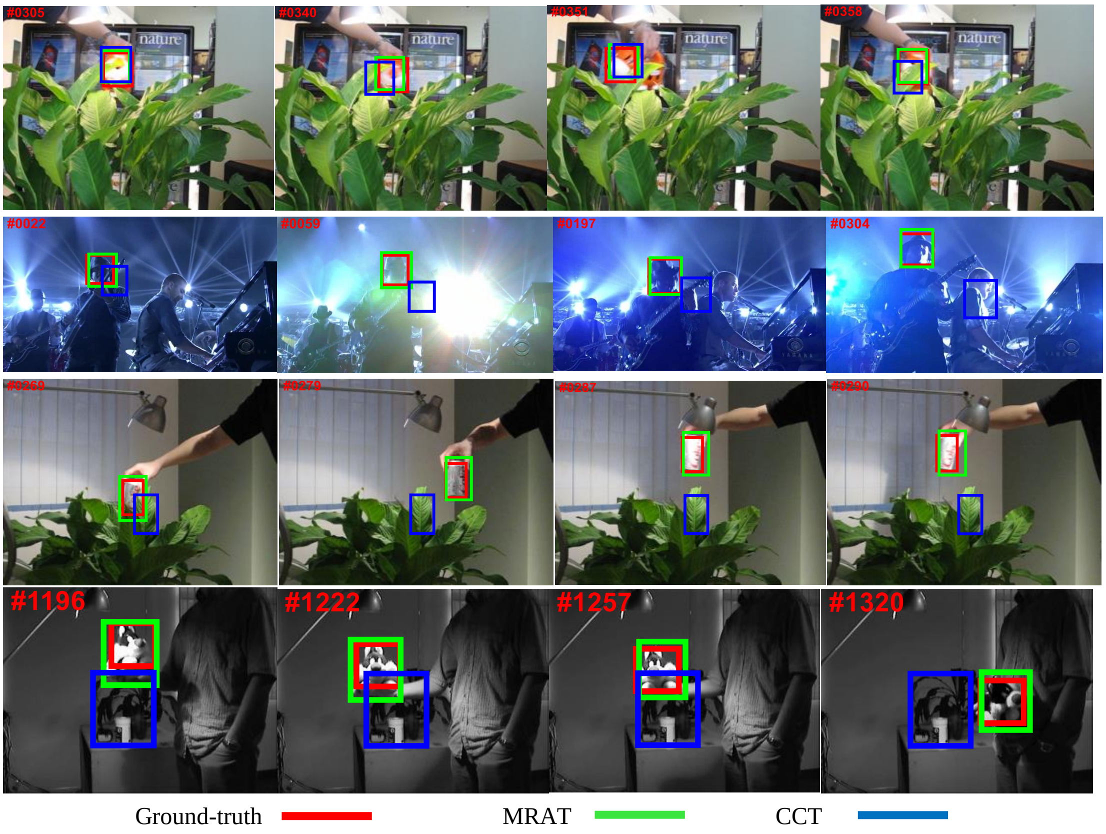


Figure 7. Tracking results of ours (MRAT) and CCT (Baseline) on four challenging sequences. The names of them from top to down are Shaking, Tiger1, Sylvester and Tiger2.

6. Conclusion

Object tracking is the key technology of intelligent video surveillance, and it is the basis of computer intelligent recognition and analysis. Illumination change is a common challenge in intelligent surveillance systems. In order to deal with target tracking tasks in varying lighting conditions, we put forward a novel strategy for object tracking in this paper, which performs multi regions detection and uses an alternate template (MRAT) based method. Further, In order to improve the speed of calculation and meet the requirements of real-time tracking, we combine our method with parallel computing. We conducted extensive experiments comparing our approach with other state of the art tracking algorithms. The results show that, the proposed approach has improvement of adaptability and robustness for intense illumination variation. Moreover, our strategy (MRAT) can be combined with other strategies such as taking advantage of divide-and-conquer strategy in order to track multiple objects based on the single object tracking algorithms. This is able to provide a unifying platform to deliver this new breed of emerging services and enable the development of new urban surveillance and monitoring applications. Our future work will aim to address other challenges such as fast motion in tracking to achieve more accurate tracking.

Acknowledgement

This work is supported by National Natural Science Foundation of China (No:61502254), Natural Science Foundation of Inner Mongolia [No. 2014BS0602].

Reference

[1] Yilmaz A, Javed O, Shah M. Object tracking: A survey[J]. Acm computing surveys (CSUR), 2006, 38(4): 13.

[2] Wang N, Shi J, Yeung D Y, et al. Understanding and diagnosing visual tracking systems[C]//Proceedings of the IEEE International Conference on Computer Vision. 2015: 3101-3109.

[3] Yang H, Shao L, Zheng F, et al. Recent advances and trends in visual tracking: A review[J]. Neurocomputing, 2011, 74(18): 3823-3831.

[4] Kalal Z, Mikolajczyk K, Matas J. Tracking-Learning-Detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2012, 34(7):1409-1422.

[5] Hare S, Saffari A, Torr P H S. Struck: Structured output tracking with kernels[C]// IEEE International Conference on Computer Vision. IEEE, 2012:263-270.

[6] Zhang J, Ma S, Sclaroff S. MEEM: robust tracking via multiple experts using entropy minimization[C]//European Conference on Computer Vision. Springer, Cham, 2014: 188-203.

[7] Danelljan M, Shahbaz Khan F, Felsberg M, et al. Adaptive color attributes for real-time visual tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 1090-1097.

[8] Montero A S, Lang J, Laganiere R. Scalable kernel correlation filter with sparse feature integration[C]//Computer Vision Workshop (ICCVW), 2015 IEEE International Conference on. IEEE, 2015: 587-594.

[9] Zhang K, Zhang L, Liu Q, et al. Fast visual tracking via dense spatio-temporal context learning[C]//European Conference on Computer Vision. Springer, Cham, 2014: 127-141.

[10] Bibi A, Mueller M, Ghanem B. Target response adaptation for correlation filter tracking[C]//European Conference on Computer Vision. Springer International Publishing, 2016: 419-433.

[11] Danelljan M, Hager G, Shahbaz Khan F, et al. Adaptive decontamination of the training set: A unified formulation for discriminative visual tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 1430-1438.

[12] Kumar B V K V, Mahalanobis A, Juday R D. Correlation pattern recognition[M]. Cambridge University Press, 2005.

[13] Bolme D S, Beveridge J R, Draper B A, et al. Visual object tracking using adaptive correlation filters[C]// Computer Vision and Pattern Recognition. IEEE, 2010:2544-2550.

[14] Danelljan M, Robinson A, Khan F S, et al. Beyond Correlation Filters: Learning Continuous Convolution Operators for Visual Tracking[C]// European Conference on Computer Vision. Springer, Cham, 2016:472-488.

[15] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. IEEE, 2005, 1: 886-893.

[16] Chen Z, Hong Z, Tao D. An experimental survey on correlation filter-based tracking[J]. arXiv preprint arXiv:1509.05520, 2015.

[17] Henriques J F, Caseiro R, Martins P, et al. Exploiting the circulant structure of tracking-by-detection with kernels[C]//European conference on computer vision. Springer, Berlin, Heidelberg, 2012: 702-715.

[18] Rifkin R, Yeo G, Poggio T. Regularized least-squares classification[J]. Nato Science Series Sub Series III Computer and Systems Sciences, 2003, 190: 131-154.

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

[20] Van De Weijer J, Schmid C, Verbeek J, et al. Learning color names for real-world applications[J]. IEEE Transactions on Image Processing, 2009, 18(7): 1512-1523.

[21] Hotelling H. Analysis of a complex of statistical variables into principal components[J]. Journal of educational psychology, 1933, 24(6): 417.

[22] Ma C, Yang X, Zhang C, et al. Long-term correlation tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 5388-5396.

[23] Ma C, Huang J B, Yang X, et al. Hierarchical convolutional features for visual tracking[C]//Proceedings of the IEEE International Conference on Computer Vision. 2015: 3074-3082.

[24] Qi Y, Zhang S, Qin L, et al. Hedged deep tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 4303-4311.

[25] Sun X, Rosin P L, Martin R R, et al. Bas-relief generation using adaptive histogram equalization[J]. IEEE transactions on visualization and computer graphics, 2009, 15(4): 642-653.

[26] Li L, Huang W, Gu I Y H, et al. Statistical modeling of complex backgrounds for foreground object detection[J]. IEEE Transactions on Image Processing, 2004, 13(11): 1459-1472.

[27] Chen L H, Yang Y H, Chen C S, et al. Illumination invariant feature extraction based on natural images statistics—Taking face images as an example[C]//Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011: 681-688.

[28]Silveira G, Malis E. Real-time visual tracking under arbitrary illumination changes[C]//Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on. IEEE, 2007: 1-6.

[29] Zhu G, Zhang S, Chen X, et al. Efficient illumination insensitive object tracking by normalized gradient matching[J]. IEEE Signal Processing Letters, 2007, 14(12): 944-947.

[30] Rifkin R, Yeo G, Poggio T. Regularized least-squares classification[J]. Nato Science Series Sub Series III Computer and Systems Sciences, 2003, 190: 131-154.

[31] Schölkopf B, Smola A J. Learning with kernels: support vector machines, regularization, optimization, and beyond[M]. MIT press, 2002.

[32] Li X, Hu W, Shen C, et al. A survey of appearance models in visual object tracking[J]. Acm Transactions on Intelligent Systems & Technology, 2013, 4(4):58.

[33] 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.

[34] Zhu G, Wang J, Wu Y, et al. Collaborative Correlation Tracking[C]//BMVC. 2015: 184.1-184.12.

[35] Choi J, Chang H J, Yun S, et al. Attentional Correlation Filter Network for Adaptive Visual Tracking[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2017.

[36] Bertinetto L, Valmadre J, Golodetz S, et al. Staple: Complementary learners for real-time tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 1401-1409.

[37] Cai B, Xu X, Xing X, et al. BIT: Biologically inspired tracker[J]. IEEE Transactions on Image Processing, 2016, 25(3): 1327-1339.

[38] Danelljan M, Häger G, Khan F S, et al. Discriminative scale space tracking[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(8): 1561-1575.

[39] Bertinetto L, Valmadre J, Henriques J F, et al. Fully-convolutional siamese networks for object tracking[C]//European Conference on Computer Vision. Springer International Publishing, 2016: 850-865.

[40] Li Y, Zhu J, Hoi S C H. Reliable patch trackers: Robust visual tracking by exploiting reliable patches[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 353-361.

[41] Valmadre J, Bertinetto L, Henriques J F, et al. End-to-end representation learning for Correlation Filter based tracking[J]. arXiv preprint arXiv:1704.06036, 2017.

[42] Kumar V. Introduction to parallel computing: design and analysis of algorithms[J]. 2003, 22(2):N12.

[43] Dongarra J, Foster I, Fox G, et al. Sourcebook of parallel computing[M]. Morgan Kaufmann Publishers Inc. 2003.

[44] Yu Y, Gunda P K, Isard M. Distributed aggregation for data-parallel computing: interfaces and implementations[C]// ACM Sigops, Symposium on Operating Systems Principles. ACM, 2009:247-260.

[45] Culler D E, Singh J P, Gupta A. Parallel computer architecture: a hardware/software approach[M]. Gulf Professional Publishing, 1999.

[46] Flynn M J. Some computer organizations and their effectiveness[J]. IEEE transactions on computers, 1972, 100(9): 948-960.

[47] Grama A. Introduction to parallel computing[M]. Pearson Education, 2003.

[48] Kumar V P, Gupta A. Analyzing scalability of parallel algorithms and architectures[J]. Journal of parallel and distributed computing,

1994, 22(3): 379-391.

[49] Kwon J, Lee K M. Visual tracking decomposition[C]// Computer Vision and Pattern Recognition. IEEE, 2010:1269-1276.

[50] Lan X, Ma A J, Yuen P C. Multi-cue Visual Tracking Using Robust Feature-Level Fusion Based on Joint Sparse Representation[C]// Computer Vision and Pattern Recognition. IEEE, 2014:1194-1201.

[51] Lan X, Ma A J, Yuen P C, et al. Joint Sparse Representation and Robust Feature-Level Fusion for Multi-Cue Visual Tracking[J]. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2015, 24(12):5826.

[52] Lan X, Zhang S, Yuen P C. Robust joint discriminative feature learning for visual tracking[C]// International Joint Conference on Artificial Intelligence. AAAI Press, 2016:3403-3410.

[53] Lan X, Yuen P C, Chellappa R. Robust MIL-Based Feature Template Learning for Object Tracking[C]//AAAI. 2017: 4118-4125.

[54] Lan X, Zhang S, Yuen P C, et al. Learning Common and Feature-Specific Patterns: A Novel Multiple-Sparse-Representation-based Tracker[J]. IEEE Transactions on Image Processing, 2017.

[55] Zhou T, Zhu M, Zeng D, et al. Scale Adaptive Kernelized Correlation Filter Tracker with Feature Fusion[J]. Mathematical Problems in Engineering, 2017, 2017:1-8.

[56] Ding G, Chen W, Zhao S, et al. Real-Time Scalable Visual Tracking via Quadrangle Kernelized Correlation Filters[J]. IEEE Transactions on Intelligent Transportation Systems, 2017, 19(1):140-150.

[57] Zhang B, Luan S, Chen C, et al. Latent Constrained Correlation Filter[J]. IEEE Transactions on Image Processing, 2017, PP(99):1-1.

[58] Han J, Pauwels E J, Zeeuw P M D, et al. Employing a RGB-D sensor for real-time tracking of humans across multiple re-entries in a smart environment[J]. IEEE Transactions on Consumer Electronics, 2012, 58(2):255-263.

Shuai Liu received his bachelor's (2002), master's (2004) and doctoral (2011) degrees from the College of Computer Science and Technology, Jilin University. He now works as an Associate Professor in the College of Computer Science, Inner Mongolia University. He has published more than 20 papers in Elsevier, Springer, Oxford and other publishers. He is now interested in image processing, computer vision and fractals. He now acts as an (associate/guest) editor in IEEE Access, Mobile Networks and Application and so on. He also acts as a reviewer of IEEE Transactions on Fuzzy Systems, IEEE Transactions on Image Processing, Information Sciences, Applied Mathematics and Computation, IEEE Acces, and so.

Gaocheng Liu is a graduate student studying at College of Computer Science, Inner Mongolia University, Hohhot, China. His mainly study domain is image processing and computer vision.

Huiyu Zhou obtained a BEng degree in Radio Technology from Huazhong University of Science and Technology of China, and a MSc degree in Biomedical Engineering from University of Dundee of United Kingdom, respectively. He was then awarded a Doctor of Philosophy degree in Computer Vision from Heriot-Watt University, Edinburgh, United Kingdom. Dr. Zhou is currently leading the Biomedical Image Processing Lab at University of Leicester, United Kingdom. He has published widely in the field.