An improved object tracking algorithm based on adaptive weighted strategy and occlusion detection mechanism

Xiuyan Tian¹,², Haifang Li¹ and Hongxia Deng¹

Abstract

Due to complex background and volatile object shape-appearance in image, the stability and accuracy of tracking algorithm is often disturbed and reduced. So how to accurately and robustly track object in object tracking application is a challenge topic at home and abroad. Built upon the methodologies of compressive tracking and spatio-temporal context, a simple yet robust object tracking method is proposed for solving the drift and occlusion problems in paper. It combines two existing classical ideas into a single framework: adaptive weighted idea and occlusion detection mechanism. In order to weaken interference problems of object background, object area is firstly partitioned into equal-sized sub-patches and the different weight related with location information is assigned for each patch; Then, for improving its robustness, Bhattacharyya distance is adopted to find out these samples with maximum discrimination; In addition, our proposed occlusion detection mechanism is for recapturing the tracked object when occlusion occurs. Many simulation experiments show that our proposed algorithm achieves more favorable performance than these existing state-of-the-art algorithms in handing various challenging infrared videos, especially occlusion and shape deformation.

Keywords

Compressive tracking, adaptive weight, infrared object tracking, maximum discrimination, Bhattacharyya distance, occlusion

Introduction

It is well known that infrared image has anti-interference, good hidden property in all-weather and all-time, and has been widely used in various civilian and military fields, such as aerial defense detecting system, infrared guidance, wide area monitoring. The goal of infrared object tracking is to estimate the location of an object from a sequence of frames given a known initial position.¹ It plays a critical role in many applications such as motion analysis, object recognition, guidance technology, photoelectric measure field and intelligent user interfaces.²,³ Object tracking algorithms have been studied comprehensively in the last decades, but the research for efficient object tracking methods still is a valid challenge due to appearance change caused by occlusion, shape deformation, motion and illumination.³

Essentially there are three types of infrared object tracking method: Optimal-matching, Decision-making and the mixture of them. Model-matching methods focus on finding out the most similar object with maximum response by searching local area. In the field of object tracking, subspace modeling is an effective way to achieve object location, such as Wavelet feature, Transform-domain feature, LBP, LTP and Pixel-level feature.⁴-¹⁰ Early feature-based tracking method relied on fixed basis, while advanced signal processing represent each image patch as a linear combination of a few elements from a basis set called a dictionary. In essence,
they are all data driven optimal-matching methods. Once the object feature is identified, classifier design is the key to determine whether the algorithm can be effective in the tracking process. In fast compression tracking algorithm, the naive Bayesian classifier is used to classify these features of samples extracted from a frame feature space, it is easy to misclassify the edge background interference with high similarity. A weighted multi-instance learning algorithm proposed in Zhang and Song, explored a concept of “Bag of positive-negative” and assigned different weight related to the importance of patch from positive-bag, and hence the robustness of the classifier is improved to a certain extent. In Zhang et al., a dynamic object feature-learning model is proposed to select training feature with strong discrimination ability so as to distinguish the object and its background in the complex background. Real-time object tracking via online discriminative feature selection can automatically select high discrimination features to get a robust classifier. However, there still exist some shortcomings. Due to lack of background information of object, these Optimal-matching algorithms would cause tracking-point drifting, even tracking failure when occlusion occurs.

Decision-making model is based on the correlation property between object and its background to track object. Traditional tracking algorithms make use of the global information of image, such as background subtraction, frame difference algorithm, particle filter and optical flow, where global parameter can be equivalent to object motion parameter. In order to search the background feature related to object region, SIFT or SURF feature is used to extract the feature of background region, increasing greatly the computational complexity. It can be seen that most of decision-making tracking algorithms usually only use some of features from partial region to assist object tracking or estimate the object state. Many tracking algorithms neglecting often the relationship between object and its background have led to lose a lot of important background information. Obviously, the use of background is the key to success. Therefore, spatio-temporal context tracking algorithm proposed by Zhang et al., takes full advantages of all of local-background information of object, and establishes context model with fast Fourier transform, which greatly improves the efficiency of the tracking algorithm.

In this paper, an object tracking method is proposed which is built upon the methodology of “Compression Tracking” and “Spatio-temporal context”. It combines two classical ideas into a single framework: adaptive weighted idea and occlusion detection mechanism. So we propose an adaptive weighted compression tracking algorithm integrated with background information as decision-making. In order to weaken interference problems of object edge, object area is firstly partitioned into equal-sized sub-patches and the different weight related with location information is assigned to each patch; Then, for improving its robustness, Bhattacharyya distance is adopted to find out these samples with maximum discrimination; simultaneously, our proposed occlusion detection mechanism is for continuing to track object when occlusion occurs. Many simulation experiments shows that our proposed algorithm achieves more favorable performance than these existing state-of-the-art algorithms in handing various challenging videos, especially occlusion and shape deformation.

The rest of our paper is organized as follows: we firstly review the related algorithms on fast compression tracking algorithm and spatio-temporal context tracking algorithm in the next section. Our proposed tracking algorithm is detailed in ‘Processed tracking algorithm’ section. In ‘Experimental results’ section, we present the experiment results and performance analysis against existing state-of-the-art algorithm. The last section is the conclusion of the whole dissertation.

**Related work**

Our work is motivated by recent advances in fast compression tracking algorithm and spatio-temporal context tracking algorithm. Thus, we will only discuss the most related techniques.

In the early tracking works, we always are required to search the most similar patches for a given reference patch from a search window sharing the same center with given reference patch, then group them into an positive sample set. And in practical terms, distance is the sole criterion for judging similarity between the given object area and a sample, namely, the closer the distance of samples is, the more similar their features. Nevertheless, because of appearance changes caused by occlusion, shape deformation and so on, dissimilar samples cannot be used to train a discriminative classifier for distinguishing the object and its surrounding background.

Fast compressive tracking (FCT) algorithm is a fast efficient tracking algorithm based on the Compressed Sensing (CS) theory. It mainly adopted a random matrix $R$ with the size of $n \times h$ to extract the object feature from a frame, where the feature vectors $x \in h$ in the high-dimensional space are projected into a low-dimensional one $V \in n$, namely,

$$V = Rx$$

(1)
where \( n \) is much less than \( h \), and \( R \) satisfies the restricted isometry property (RIP). In other words, if a signal \( x \) is compressible or K-sparse, it is possible to perfectly reconstruct the signal \( x \) from \( V \) by minimizing the sum of the squares of the \( x \) the residuals error, and the compression level depends upon the sparseness. Thus, literature\(^3\) adopts a very sparse random measurement matrix with entries defined as follows:

\[
    r_{ij} = \sqrt{s} \times \begin{cases} 
        1; & p = 1/2s \\
        0; & p = 1 - 1/s \\
        -1; & p = 1/2s 
    \end{cases} \tag{2}
\]

In this entries definition, \( s = n/4 \) and it needs only to compute \( c \) entries for each row in random matrix \( R \), where \( c < 4 \) denotes the number of entries computed for each row. It is observed that good results can be obtained by fixing it in experiments. Therefore, the computational complexity is only \( o(cn) \). After the large set of Haar-like features are compressively sensed with a very sparse measurement matrix, positive or negative samples are adopted to train a naive classifier so as to realize the classification between object and background in the compressed domain efficiently and effectively without the curse of dimensionality, where the object location in next frame is maximum response. Given all low-dimensional features \( f_i \in V \) in space domain \( V \) are independently distributed and can be modeled with a naive Bayes classifier:

\[
    P(f) = \log_2 \left( \frac{\prod_{i=1}^{n} p(f_i | y = 1)p(y = 1)}{\prod_{i=1}^{n} p(f_i | y = 0)p(y = 0)} \right) = \sum_{i=1}^{n} \log_2 \left( \frac{p(f_i | y = 1)}{p(f_i | y = 0)} \right) \tag{3}
\]

where we assume uniform prior, \( p(y = 1) = p(y = 0) \), \( P \) is the probability of distribution, \( f_i \) is the \( i \)th feature in \( V \) and \( y \in \{0, 1\} \) is a binary variable which denotes the positive or negative sample label.

The compressive tracking algorithm takes full advantages of the low-dimensional space feature of the infrared object, and has high robustness when the contour of the object is clear and complete. However, lack of detection mechanism for the tracking objects will lead to the misclassification to the positive and negative samples when the object is under occlusion; it will cause tracking-point drifting, even tracking failure if background information has not yet been fully utilized. The tracker will be distracted to a wrong location. Although the distractor has similar appearance to the object, most of their surrounding background contexts have different appearances which are useful to discriminate object from distractor. So Wen et al.\(^4\) proposed a simple and robust algorithm which exploits the spatio-temporal relationships between the object and its locally dense contexts in a Bayesian framework for visual tracking, which are robust to appearance variations introduced by occlusion, distortion and pose variations. It is observed that the tracking algorithm\(^4\) is one of several algorithms based on decision-making models, making full use of the spatial and background information.

### Processed tracking algorithm

**Adaptive weighted compressive tracking algorithm**

Since the shape of object is usually irregular, it means that the inside of the bounding box contains less background information and more object texture, while the edge of the object brings in more background disturbance directly affecting the classification of positive and negative samples. If classifier model is updated on the basis of misclassification sample, there’s no question that it must affect the tracking accuracy and reliability. Therefore, we propose the strategy of extracted features where object patches are handily assigned weights so as to weaken background interference in paper, as shown in Figure 1. In Zhang and Song,\(^5\) the strong classifier is constructed from all samples where the object location with maximum likelihood probability is selected out and weights based on distance to the center of object are assigned.

Given that the center area of search window with higher confidence has less background information than that of edge area, the searching window is divides \( W \) patches with size of \( N \times N \) from top to down and left to right where the high-dimensional feature vectors of the \( i \)th patch is represent as \( q_{ij}(i,j = 1, \ldots, N) \). In order to extract the low-dimensional feature of object, random measurement matrix generated in equation (2) is adopted to reduce dimensional of feature vector of

![Figure 1. Confidence map: (a) Weight map (b) Sub-patches extraction.](image)
object, and get the low-dimensional compression feature of each patch via equation (1), then these features are classified by naive Bayesian classifier in equation (3) so as to obtain corresponding scores. Meanwhile, according to the above analyses, the closer the patch location is to the center of bounding box, the higher the confidence score is for weakening the background interference, so the confidence weight function of the $k$th patch is defined as

$$C_k = \exp \left( -\frac{1}{2} \sqrt{(k_x - l_x)^2 + (k_y - l_y)^2} \right),$$

where $(l_x, l_y)$ is the center location of bounding area divided with $W$ patches and $(k_x, k_y)$ is 2-D central coordinate of the $k$th patch.

It is obvious from equation (4) about weight that there are larger weight in the center location of bounding area, so their confidence scores have been bigger than that of edge area. The true core idea of feature-matching based object tracking algorithm is the discrimination between the object feature and its background. Due to occlusion, background interference, motion blur and illumination changes, it does not ensure that the extracted low-dimensional feature, especially from edge area, has a good separability when the random measurement matrix is used to extract the low-dimension feature in compression tracking. Thus, Babenko et al.\textsuperscript{13} creatively proposed a concept of “Bag of Instance”, and log-likelihood function of Bags are computed for selecting out weak classifier, then the best ones from all classifiers in pool are greedily picked to form a strong classifiers. On the basis of Babenko et al.,\textsuperscript{13} Wen et al.\textsuperscript{4} further improved the classifier and increased its robustness, which assigned different weight to these selected weak classifier. A dynamic object feature-learning model proposed in Tian et al.\textsuperscript{15} used high divergence between two classes to train classifier. Diaconius and Freedman in Tian et al.\textsuperscript{14} have proved that the random projection of high-dimensional random vector are approximately Gaussian distribution, and it can be described by using 4 parameters as follows

$$p(f_i | y = 1) \sim N(\mu_i^1, \sigma_i^1)$$

$$p(f_i | y = 0) \sim N(\mu_i^0, \sigma_i^0)$$

where $\mu_i^1 (\mu_i^0)$ and $\sigma_i^1 (\sigma_i^0)$ are mean and standard deviation of the $i$th feature with positive (negative) sample.

It is well known that these features not only include high discrimination feature but also have relative weak or redundancy feature. In order to magnify discrimination between all these feature and improve the accuracy of object tracking, we introduce Bhattacharyya distance to measure the separability between two probability distributions for low-dimensional features of positive (negative) samples, so that feature with larger separability can be adaptively selected to train a classifier. As shown in Figure 2, all samples are manifested as Gaussian distribution, where we also note that there are big differences between features of positive-sample 2, 3 and that of negative-sample and it was more difficult to distinguish the positive-sample 2 and negative-sample. Let us suppose the probability density distribution (PDF) of the $k$th positive (negative) sample is $p_1(p_2)$, so Bhattacharyya distance can be written as

$$\begin{align*}
B_D(p_1, p_2) &= -\ln(B_C(p_1, p_2)) \\
B_C(p_1, p_2) &= \int \sqrt{p_1(x)p_2(x)}dx
\end{align*}$$

In either case, $0 \leq B_C \leq 1$, $0 \leq B_D \leq \infty$, where $B_C$ is expressed as Bhattacharyya coefficient and $B_D$ is the Bhattacharyya distance between distributions. All Bhattacharyya distances $B_D$ are listed in descending order, then the top $m$-dimension features are chosen to train classifier so as to improve robustness.

In addition, a naive Bayesian classifier based on adaptive confidence-weighted is proposed, and shown as follows.

$$H(f) = \sum_{k=1}^{W} C_k \log_2 \left( \prod_{i=1}^{m} \frac{p(f_{ki} | y = 1)p(y = 1)}{\prod_{i=1}^{m} p(f_{ki} = 0)p(y = 0)} \right)$$

$$= \sum_{k=1}^{W} C_k \sum_{i=1}^{m} \log_2 \left( \frac{p(f_{ki} | y = 1)}{p(f_{ki} | y = 0)} \right)$$

Figure 2 Schematic diagram for sample feature probability.
where we assume uniform prior, \( p(y = 1) = p(y = 0) \), and \( f_{kj} \) is the \( i \)th low-dimension features of the \( k \)th in bounding box, about \( m \) features in all. The classification scores of all patches in tracking bounding-box are multiplied by the corresponding weights, and then added up to find maximal response as current object location. Finally, 4 parameters in equation (5) are added up to find maximal response as current object multiplied by the corresponding weights, and then classification scores of all patches in tracking bounding-box are added up. The 4 parameters in equation (5) are defined as follows:

\[
\begin{align*}
\mu_k^l & = \frac{1}{m} \sum_{i=1}^{m-1} f_{ki}(q) \\
\sigma_k^l & = \left\{ \frac{1}{m} \sum_{i=1}^{m-1} (f_{ki}(q) - \mu_k^l)^2 \right\}^{1/2}
\end{align*}
\] (8)

where, learning factor is \( 0 < \lambda < 1 \), and its value is expressed as the updating rate of parameter, \( \mu_k^l \) and \( \sigma_k^l \) are defined respectively as

\[
\mu_k^l = \frac{1}{m} \sum_{i=1}^{m-1} f_{ki}(q) \\
\sigma_k^l = \sqrt{\frac{1}{m} \sum_{i=1}^{m-1} (f_{ki}(q) - \mu_k^l)^2}
\] (9)

where \( m \) denotes the number of features, \( \mu_k^l \) and \( \sigma_k^l \) is the mean and standard deviation for the \( k \)th patch respectively. Thus, due to our proposed adaptive weighted compression tracking algorithm, our classifier performs robustly and weakens interference problems of object edge.

**Background decision-making mechanism**

The compressive tracking algorithm takes full advantages of the low-dimensional space feature of the object, and has high robustness when the contour of the object is clear and complete. However, lack of detection mechanism for the tracking objects will lead to the misclassification to the positive or negative samples when the object is under occlusion; That background information has not yet been fully utilized will cause tracking-point drift, even tracking failure. To solve this problem, the object detection mechanism is proposed in paper. The classification score of edge patch in the tracking bounding vertices is usually represented the detection threshold. The positive and negative samples will stop updating if the classification score of the edge patch is lower than the threshold.

When the occlusion has been detected by detection module, the relationship between the object and local background are modeled firstly by the conditional probability function of the spatial context model, where the PDF is defined as follows:

\[
P(x|c(z), o) = h^{\omega}(x - z)
\] (10)

Notice that \( c(z) = (I(z), z) \) is the context feature of gray features \( I(z) \) at location \( z \), and \( h^{\omega}(x - z) \) is denoted as the relative distance and direction function between object location \( x \) and its local context location \( z \), which shows the spatial relationship between an object and its spatial context. In additional, the tracking model often is trained by online learning, and incrementally updated with samples from observations in recent frames. The focus of attention(FoA) theory in the biological vision system indicates the closer the distance to the object is and the greater the attention is. Using this theory, the object location of next frame is predicted, whose prior probability function model is defined as follows:

\[
\begin{align*}
P(c(z)|o) &= I(z)\omega_o(z - x') \\
\omega_o(z) &= \frac{a e^{-|z|^2}}{\pi}
\end{align*}
\] (11)

where gray features \( I(z) \) described the appearance of the local context \( z \), \( \omega_o \) and \( x' \) are denoted as weighted function and object location, respectively. The nearer an local context \( z \) is to \( x \), the bigger its weight is; \( \sigma \) depends on the size of background area.

Simplifying the equation (11), confidence model of the object prior probability is written as

\[
c(x) = P(x|o) = be^{-|x|^\beta}
\] (12)

where \( s \) is scale parameter and \( \beta \) means sharp degree parameter with maximum value in confidence map.

According to the spatio-temporal relationship between object and background in context prior model(11) and object confidence model(12), our proposed tracking model can be expressed as:

\[
c(x) = be^{-|x|^\beta} = \sum_{z \in \Omega(z)} h^{\omega}(x - z)I(z)\omega_o(z - x') = h^{\omega}(x) \otimes (I(x)\omega_o(x - x'))
\] (13)

Since a time-domain convolution is equivalent to its frequency-domain multiplication, so we can get

\[
F(be^{-|x|^\beta}) = F(h^{\omega}(x)) \circ F(I(x)\omega_o(x - x'))
\] (14)
where, $\cdot$ denotes dot product. In conclusion, after above analysis, spatio-temporal context model is transformed into

$$h^{sc}(x) = F^{-1}\left(\frac{F(be^{-\frac{-(x-x^*)^2}{2\sigma^2}})}{F(I(x)\omega_\sigma(x-x^*))}\right)$$  (15)

whose updated equation is

$$H_{t+1} = (1 - \rho)H_t + \rho h^{sc}_t$$  (16)

After the spatio-temporal context model is transformed by Fourier inversion, we can obtain the confidence value of each pixel-point, and maximum value is formed by Fourier inversion, we can obtain the confidence value of each pixel-point, and maximum value is formed by Fourier inversion. The object template will stop mechanism is built on 4 patches centered around track-point drifting, even tracking failure, a object detection of positive and negative samples could cause tracking failure. After the spatio-temporal context model is transformed, the object area is firstly partitioned into equal-sized sub-patches, then the sparse matrix in formula (2) is adopted to extract the top $m$ low-dimension feature vectors with the larger Bhattacharyya distances; next these positive and negative samples are classified by classifier in equation (7) so as to obtain classification scores; finally the maximum score is defined as the location $l_i$ of coarse positioning.

1. The tracking bounding-box is searched and generated from a $r$ radius around the search center $l_{t-1}$ with step-length $\lambda_c$, and object area is firstly partitioned into equal-sized sub-patches, then the sparse matrix in formula (2) is adopted to extract the top $m$ low-dimension feature vectors with the larger Bhattacharyya distances; next these positive and negative samples are classified by classifier in equation (7) so as to obtain classification scores; finally the maximum score is defined as the location $l_i$ of coarse positioning.

2. Within the scope of $D^{ij} = \{z||l(z) - \ell_i^t|| < \gamma_j\}$ around the location $l_i^t$, the bounding-box is generated with step-length $\lambda_{ij} (\lambda_{ij} << \lambda_c)$ for an accurate search; the top $m$ low-dimension features with the larger Bhattacharyya distances are projected by classifier $H(f)$, so the object location $l_i$ with maximum response can be obtained.

3. Stop updating positive and negative samples when occlusion detected by occlusion detection mechanism occurs, then the spatio-temporal context relationship between object and its background is modeled and obtained the classifier score $H(f)$; finally, the equation (17) is applied to obtain the maximum confidence value, namely object location $l_i$.

4. If $H(f) > H_T(f)$, then the tracked object has magically reappeared; Stop the spatio-temporal context model and recover the updating of positive and negative samples.

5. The sample is updated, where the range of positive samples is $D_{pos} = \{z||l(z) - l_i|| < l_{pos}\}$ close to object location, and the range of negative samples is $D_{neg} = \{z||l(z) - l_i|| > l_{neg}\}$ far away from object; then sparse matrix is adopted for reducing the dimensionality of samples, where their compressed features are denoted as $f(D_{pos})$ and $f(D_{neg})$, respectively.

The main framework of proposed tracking scheme

In order to reduce the computational complexity, a coarse-to-fine sliding window search strategy is adopted, significantly reducing computational cost. To conclude this section, we will summarize the main step of the proposed object tracking method. The concrete process is shown as follows:

**Input:** Given a current image frame $t$ with a object location $l_{t-1}$ in the last frame, the search range for candidate region of object is set to 0-$r$, the range of positive samples is 0-$l_{pos}$, and the range of negative samples is $\min_{neg} - \max_{neg}$.

**Output:** Tracking location $l_i$.
Experimental results

In this section, the performance of our proposed tracking scheme is tested and analyzed. Due to the lack of object reference value in infrared image, it is very difficult to evaluate quantitatively our proposed algorithm. Therefore, the infrared image sequence obtained by the uncooled infrared detector is adopted in our experiment to measure manually the object position and scale so as to facilitate the quantitative and quantitative analysis of our tracking performance, where 56 researchers in related fields have calibrated and measured the position and size of the object frame by frame, and then averaged as the baseline data of the infrared target. In order to more fully verify the tracking ability of the proposed method, 12 nature image sequences with reference value is adopted to analyze qualitatively and quantitatively. The main innovation in the framework is the combination of adaptive weighted idea and occlusion detection mechanism.

Experiments configuration

All of the experiments are run under MATLAB v7.8 (R2009a) on PCs with an Intel quad-core i5 CPU at 3.3 GHz and 4 GB memory. Center location ratio (CLR) and Overlap ratio (OR) based on the ground truth are used to evaluate quantitatively the performance of tracking method. Our proposed algorithm is evaluated with other 4 state-of-the-art algorithms on 12 challenging sequences with publicly available. It is noted that all quantitative evaluation of proposed method are averaged over six independent trials. In addition, the infrared image sequences captured by infrared detector are used to analyzed qualitatively.

The 4 evaluated trackers are the Fast Compressive Tracker (FCT),\(^3\) the spatio-temporal context tracker (STC),\(^4\) MIL tracker,\(^5\) and TLD tracker,\(^8\) respectively. It is worth noticing that the most challenging sequences from the existing works are used for evaluation. All parameters in comparative algorithms are fixed for all the experiments to demonstrate the robustness and stability of our proposed tracking method.

Parameters setup

The range of positive sample is set to \(l_{\text{pos}} = 4\), which means we select positive samples within a range of 4 pixels around the modified candidate object and generate 45 positive samples; The range of negative sample is set to \(\min l_{\text{neg}} = 8\) and \(\max l_{\text{neg}} = 30\) where 50 negative samples can be selected randomly; In the process of coarse tracking, the range of search and the search step are \(r_c = 25\) and \(\lambda_c = 4\), respectively; In the process of fine tracking, the range of search is \(r_c = 10\), and the search step is \(\lambda_c = 1\); 100 samples in FCT with low dimensional features are selected for training, while our proposed algorithm use equation (6) to select 50 low dimensional features for training the classifier. The threshold of classification score for 4 sub-patches centered around tracking bounding vertices is set to zero, which means the presence of occlusion on the edge of the object. In addition, as for \(N\), the greater it is, the smaller is the interference of background within the tracking bounding-box, but it will increase the complexity of the algorithm and is more sensitive to the background interference. Therefore, \(N\) cannot exceed the limited value. In our algorithm, \(N = 5\), which indicates that the tracking bounding-box is divided into 25 small sub-patches. All the results are based on the source codes or executables released by the original authors. The default parameters are employed in the comparison algorithms.

Comparison of quantitative evaluation for nature sequences

In order to evaluate the effectiveness of our proposed algorithm, the success rate and the center location error are adopted to measure the accuracy of tracker. The former metric shows the coverage rate between the tracked bounding-box \(R_T\) and the ground-truth bounding-box \(R_G\), defined as follows:

\[
\eta = \frac{\text{area}(R_T \cap R_G)}{\text{area}(R_T \cup R_G)}
\]

where the size of \(\eta\) is denoted as the tracking accuracy, so the result of one frame is considered as a success if \(\eta > 0.5\). The success rate on twelve test sequences are reported in Table 1 for different tracker. The other metric CLE is defined as the euclidean distance between the central locations of the tracked objects and the ground truth, as also shown in Table 1. The best performance and the second best are highlighted by red and blue color in each cell, respectively. From Table 1, we can conclude that the proposed adaptive weighted idea and occlusion detection mechanism method significantly outperforms other compared algorithms.

According to the experimental results in Table 1, our proposed algorithm enjoys the highest success rate in most sequences. Our proposed tracking algorithm is the most efficient algorithm among all evaluated methods except for the STC method which is down only 0.1 in some sequences. In addition, our algorithm achieve much better results than the FCT algorithm in terms of center location error, shown the effectiveness of using adaptive weighted idea.
Table 1. Quantitative comparison for nature sequences.

| Sequences index | Success rate | Center location error |
|-----------------|--------------|-----------------------|
| | TLD | KCF | MIL | FCT | Our | TLD | KCF | MIL | FCT | Our |
| Person | 0.515 | 0.607 | 0.612 | **0.552*** | **0.753*** | 8.8 | 10.1 | 18.8 | 8.6 | **6.5*** |
| Car4 | 0.805 | **0.922*** | 0.539 | **0.882*** | 0.881 | 5.7 | 5.2 | 3.7 | **3.0*** | 4.4 |
| Carl1 | 0.725 | 0.808 | 0.391 | **0.782*** | **0.843*** | 1.7 | 1.9 | 1.8 | 2.2 | **1.6*** |
| Surfer3 | 0.457 | **0.543*** | 0.504 | 0.516 | **0.737*** | 25.3 | 11.3 | 22.7 | 21.4 | **4.3*** |
| DavidInDoor | 0.757 | **0.800*** | 0.398 | 0.782 | **0.860*** | 5.1 | 3.6 | 3.7 | 3.7 | 4.0 |
| Faceocc2 | 0.822 | **0.836*** | 0.718 | 0.835 | **0.836*** | 4.5 | 6.9 | 4.2 | 4.0 | **3.8*** |
| Girl | **0.716*** | 0.593 | 0.623 | 0.485 | **0.710*** | 12.7 | 19.0 | 41.8 | **12.4*** | 12.5 |
| Jumping | 0.682 | 0.687 | 0.109 | **0.712*** | **0.794*** | 5.0 | 8.2 | 4.7 | 5.0 | **4.6*** |
| Occlusion1 | 0.879 | **0.931*** | 0.290 | 0.934 | **0.892*** | 7.0 | 9.1 | **3.4*** | 4.7 | 6.1 |
| Lake | 0.797 | **0.822*** | 0.387 | **0.878*** | 0.752 | 5.3 | 4.9 | **3.3*** | 4.7 | 7.1 |
| Bird1 | 0.506 | **0.663*** | 0.352 | **0.626*** | 0.548 | 3.6 | 2.9 | 2.4 | **1.7*** | 3.2 |
| Woman | **0.751*** | 0.701 | 0.619 | 0.657 | **0.820*** | 12.3 | 20.2 | 66.9 | 10.9 | **2.2*** |

Even on a certain challenging video sequences containing occlusion, shape deformation, blur motion and illumination, such as Singer1, Faceocc2 and Occlusion1, our proposed scheme outperforms others existing algorithms. That demonstrates adaptive weighted idea and occlusion detection mechanism are very suitable for tracking and recapturing the object with occlusion. For Surfer3, when the man is easily lost within the sea, performance of proposed method outperforms other trackers. The reason is that adaptive weighted idea can make sure the classifier without disruption, and can recapture the object when object first appears. No matter how strong the occlusion is, it is easy to capture the corresponding local pattern of object. Therefore, our scheme always outperforms the FCT and STC.

It is not surprising that our method shows outstanding performances for the test sequences containing background occlusion and shape deformation. For Surfer3 and Bird1, our proposed method outperforms the TLD tracker and MIL tracker, and it is superior to FCT by 0.05—0.1 and to CST by 0.04. For Jumping, our proposed method outperforms the FCT, STC, the TLD tracker and MIL tracker, and it is superior to FCT by 0.08, to STC by 0.1 and to TLD by 0.1.

Compared with the success rate(SR), the center location error(CLE) is a popular evaluation for measuring the tracking accuracy of object. To understand how the combination of adaptive weighted idea and occlusion detection mechanism tracks a object in stable, we also compare the metric of the proposed algorithm and other leading tacker techniques in the literature at different videos. The CLE results are also shown in Table 1. The best performance and the second best are highlighted by red and green font in each cell, respectively. We conclude the proposed method has achieved competitive CLE performance to the FCT and STC for the most test sequences. The success owe to the fact the proposed method aims to construct optimum classifier and weaken interference from background information. In addition, our tracker can accurately recapture object when it appears.

Note that the parameters in our tracker have not been optimized for speed and for quality in this experiment. They are set just on basis of experience. Optimized parameters may further improve the performances. This is another direction of future research. It is natural conclusion that the proposed method is faster than FCT and STC. The reason is that our algorithm reduces data dimension and remove the multi-scale filtering module.

The objects in the Girl, Singer1, Occlusion1 and Carl sequences are partially occluded at times. The object in the women sequence also undergoes rotation which makes the tracking tasks difficult. Only our proposed algorithm can successfully track object in most frames.

The Occlusion1 sequence has deformation and heavy occlusion, shown in Figure 3. All the other trackers fail to successfully track the object except for our proposed algorithms. Full sequences are adopted to better evaluate the performance of all algorithms. Only our proposed algorithm is able to achieve favorable tracking results in terms of both accuracy and success rate. This can be attributed to the use of adaptive weighted idea and occlusion detection mechanism.

According to the results of comparison experiments above, it is not hard to see that our proposed algorithm has the best stability among various existing state-of-the-art object tracking algorithms in the process of object tracking. In the Boat and the Car, the objects are occluded and interfered by background. In these cases, our proposed algorithm can still realize accurate tracking, which shows that our adaptive weighted method, can effectively diminish the interference from
background information. In the Woman with self-similarity for background interference and the person with illumination change, only our algorithm can realize the whole tracking process, which indicates that our strategy in this paper can effectively enlarge the difference of features between the object and background. In the Skating and the Suffer with serious occlusion, our proposed algorithm has the highest success rate of tracking, which shows that our proposed occlusion detection mechanism can accurately detect the occlusions, and recapture object when it appears. In Boat, All algorithms drift between frame 100–127 due to a rapid movement followed by an occlusion, however our proposed algorithm is able to recapture back on at frame 132 and continue tracking the object until the end of the sequence.

Comparison of quantitative evaluation for infrared sequences
Due to the lack of object reference value in infrared image, so the infrared image sequence obtained is adopted in our experiment to measure manually the object position and scale so as to evaluate the accuracy, stability and coverage of our proposed tracking algorithm, as shown in Figure 4 and Table 2. Sequence 1 (the first row of Figure 4) is white-hot image from the infrared detector, where the object is a van and the tracking process occlusion, background interference, rotation and similarity background. Sequence 2 (the second row of Figure 4) is a black-hot image from the infrared detector, where the object is a truck and the tracking process appears pose changes, occlusion, blur and other interference. The TLD and MIL tracking algorithms will tracking-point drifting due to the interference of the poles and weeds. With the accumulation of the error, the tracking bounding-box will gradually deviate from the object, which is mainly because the generalization ability of the tracking model is not enough and cannot be fully adapted complex background and significant appearance changes. Our proposed algorithm, FCT and STC algorithm can track the object, but our bounding-box always surrounds the object and STC has a certain drift. Qualitative analysis shows that our proposed
method has better tracking stability than other existing state-of-the-art algorithms in handling various challenging infrared videos, especially occlusion and shape deformation.

Table 2 shows the quantitative comparison of the different tracking algorithms for infrared images. It can be seen from the data in Table 2 that the algorithm proposed in this paper has achieved good tracking effect for 6 different infrared scenes. For example, in the tracking process of IR sequence 1, the tracking accuracy of the other three methods is better, except for TLD and MIL, mainly because the TLD and MIL methods begin to drift in the 139th frame. The SFC algorithm has the drift and the STC algorithm will occasionally jump due to similar background disturbances, which results in a greater influence on the OR and the CLE. For Sequence 2, our proposed method outperforms the FCT, STC, the TLD tracker and MIL tracker, and it is superior to FCT by 0.03, to STC by 0.08 and to TLD by 0.17. Our proposed algorithm in this paper can track the object better when it is occluded, and obtain good tracking precision, which is mainly due to the generalization ability of the multi-cue model.

Conclusion

In combination with infrared object tracking based on compression algorithm, our proposed algorithm in this paper mainly utilize sub-patch feature extraction to reduce the size of the features matrix, and decrease the storage space of random matrix. The weight of candidate samples depend on the location of feature patches, which weakens the interference from background. We use Bhattacharyya distance as the measure to adaptively extract the features of object so as to enlarge the difference between the object and its background and improve the robustness of classifiers. Meanwhile, we set a detection mechanism on the edge of the object; when the object is occluded, the object and local background can be modeled through spatio-temporal context algorithm so as to achieve an fine tracking. Compared with various existing state-of-the-art object tracking algorithms, many experimental results show that our proposed infrared object tracking algorithm has a good stability and high robustness in the process of object tracking.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Natural Science Foundation of Shanxi Province (No.2014021022-5), the Technological Project of State Grid Corporation of China (No.5205301500).

ORCID iD

Xiuyan Tian https://orcid.org/0000-0001-6124-8008

References

1. Tian X, Li H and Deng H. An improved two-steps saliency detection algorithm based on binarized normed gradients and nuclear norm model in video sequences. J Inform Hiding Multimedia Signal Process 2018; 9: 841–852.
2. Zheng C, Chen J, Yang X, et al. Adaptive fusion tracking based on optimized co-training framework. J Infrared Millim Waves 2016; 35: 496–504.
3. Zhang K, Zhang L and Yang M-H. Fast compressive tracking. IEEE Trans Pattern Anal Mach Intell 2014; 36: 2002–2015.
4. Wen L, Cai Z, Lei Z, et al. Online spatio-temporal structure context learning for visual tracking. In: European Conference on Computer Vision (ECCV), 2012, pp. 716–729.
5. Zhang KH and Song HH. Real-time visual tracking via online weighted multiple instance learning. Pattern Recogn 2013; 46: 397–411.
6. Zhang K, Zhang L, Yang M-H, et al. Robust object tracking via active feature selection. IEEE Trans Circuits Syst Video Technol 2013; 23: 1957–1967.

| Sequences Index | Success rate | Center location error |
|-----------------|--------------|-----------------------|
|                 | TLD | STC | MIL | FCT | Our | TLD | STC | MIL | FCT | Our |
| Sequence 1      | 0.281 | 0.70 | 0.112 | 0.719 | **0.865** | 47.3 | 14.4 | 64.9 | 9.1 | **6.1** |
| Sequence 2      | 0.308 | 0.67 | 0.317 | 0.802 | **0.810** | 31.7 | 9.2  | 37.4 | 7.4 | **2.8** |
| Sequence 3      | 0.590 | **0.817** | 0.619 | 0.771 | 0.798 | 12.5 | **6.3** | 12.6 | 9.3 | **8.2** |
| Sequence 4      | 0.218 | 0.71 | 0.452 | 0.700 | **0.712** | 74.6 | 11.4 | 27.1 | **8.2** | 8.6 |
| Sequence 5      | 0.659 | 0.45 | 0.677 | 0.581 | **0.687** | 8.7  | 4.6  | 9.4  | 7.1 | **4.8** |
| Sequence 6      | 0.612 | 0.90 | 0.801 | 0.872 | **0.910** | 12.22 | 3.9  | 4.7  | 4.1 | **2.8** |
7. Zhang KH, Zhang L and Yang MH. Real-time object tracking via online discriminative feature selection. *IEEE Trans Image Process* 2013; 22: 4664–4677.

8. Kalal Z, Matas J and Mikolajczyk K. Pn learning: bootstrapping binary classifiers by structural constraints. In: *IEEE conference on computer vision and pattern recognition (CVPR)*, 2010, pp. 49–56.

9. Han Z, Jiao J, Zhang B, et al. Visual object tracking via sample-based adaptive sparse representation (AdaSR). *Pattern Recogn* 2011; 44: 2170–2183.

10. Pan G, Chen G, Kang W, et al. Correlation filter tracker with Siamese: a robust and real-time object tracking framework. *Neurocomputing* 2019; 358: 33–43. SEP.17

11. Ke B, Zheng H, Chen L, et al. Multi-object tracking by joint detection and identification learning. *Neural Process Lett* 2019; 50: 283–296.

12. Henriques JF, Caseiro R, Martins P, et al. High-speed tracking with kernelized correlation filters. *IEEE Trans Pattern Anal Mach Intell* 2015; 37: 583–596.

13. Babenko B, Yang M and Belongie S. Visual tracking with online multiple instance learning. In: *IEEE conference on computer vision and pattern recognition*, 2013, pp.983–900.

14. Tian X, Li H and Deng H. Object tracking algorithm based on improved context model in combination with detection mechanism for suspected objects. *Multimedia Tools Appl* 2019; 78: 16907–16922.

15. Tian X, Li H and Deng H. Object tracking algorithm based on improved Siamese convolutional networks combined with deep contour extraction and object detection under airborne platform. *J Imaging Ence Technol* 2020; 12: 146–156.

16. ZhuangLu B, Xiao HZ, et al. Low-rank sparse learning for robust visual tracking. *IEEE Trans Image Process* 2014; 23: 1872–1881.

17. Zeng X, Xu L, Cen Y, et al. Dual-scale weighted structural local sparse appearance model for object tracking. *IET Computer Vision* 2019; 13: 146–156.

18. Fantacci C, Vo B-N, Vo B-T, et al. Robust fusion for multisensor multiobject tracking. *IEEE Signal Processing Letters* 2018; 25: 640–641.

19. Guo Y, Zhang S and Bao F. Structured robust correlation filter with $L_{(2,1)}$ norm for object tracking. *J Electron Imaging* 2019; 28: 1.

20. Jia X, Lu H and Yang MH. Visual tracking via adaptive structural local sparse appearance model. In: *Proceedings of IEEE computer society conference on computer vision and pattern recognition (CVPR)*, 2012, pp.1822–1829.

21. Zhong W, Lu H and Yang MH. Robust object tracking via sparsity-based collaborative model. In: *Proceedings of IEEE computer society conference on computer vision and pattern recognition (CVPR)*, 2012, pp.1838–1845.

22. Yoon J. and Hong Structural constraint data association for online multi-object tracking. *Int J Comput Vision* 2018; 12: 685–691.

23. Emami P, et al. Machine learning methods for data association in multi-object tracking. *ACM Comput Surv* 2020; 24: 63–85.

24. Qiao L and Zhang W. Double-Loop integral terminal sliding mode tracking control for UUVs with adaptive dynamic compensation of uncertainties and disturbances. *IEEE J Oceanic Eng* 2019; 44: 29–25.

25. Wen C, Meng H and Qisheng W. A novel hard decision based simultaneous target tracking and classification approach. *Sensors* 2018; 18: 622.