Small Target Tracking in Satellite Videos Using Background Compensation

Yunming Wang, Taoyang Wang, Guo Zhang, Qian Cheng, and Jia-qi Wu

***Abstract*— Through the use of video technology, satellites can detect dynamic targets and analyze their motion characteristics. Target tracking can extract dynamic information about key ground targets for target monitoring and trajectory prediction by satellite video. Tracking algorithms are affected by target motion characteristics, such as velocity and direction, as well as background characteristics, such as illumination changes, occlusion, and background similarities with the target. However, these problems are seldom studied with satellite video cameras. Current algorithms are unsuitable for satellite video because of the poor texture and color features of the target in satellite video. Therefore, in this article, we enhance target tracking for satellite video technology using two aspects: 1) sample training strategy and 2) sample characterization. We establish a filter training mechanism for the target and background to improve the discrimination ability of the tracking algorithm. We then build a target feature model using a Gabor filter to enhance the contrast between the target and background. Moreover, we propose a tracking state evaluation index to avoid tracking drift. Tracking experiments using nine sets of Jilin-1 satellite videos show that the proposed approach can accurately locate a target under weak feature attributes. Therefore, this article contributes to more robust tracking using satellite video technology.**

***Index Terms*— Correlation filtering, robustness, target tracking satellite video.**

I. Introduction

SATELLITES can shoot high-resolution videos that extract both static and dynamic information from the ground, which is of great significance in real-time monitoring of moving ground targets. The first video satellite was launched by Skybox Imaging in 2013 [1]. In October 2015, China launched the commercial satellite Jinlin-1 [2], equipped with a high-resolution video camera that can shoot full-color video with a resolution of 1 m and an imaging range of up to 4.6 km $\times$ 3.4 km. The satellite image is a 24-bit color image with a spectrum of RGB Bayer in the range 430–720 nm, and the signal to noise ratio is higher than 35 dB. The imaging mode of this satellite is gaze video imaging, and the orbit type is a sun-synchronous orbit with an altitude of 655 km and a 2-day revisit cycle. Moreover, the frame rate is 25 frames/s and the satellite's maximum side swing angle is $\pm 45°$.

Target tracking uses video images to extract the location of a target at different times and guarantees a correct relationship between multiple frames and the same target. It has many civilian applications, including human–computer interaction, medical imaging, video surveillance, self-driving vehicles, and virtual reality [3], as well as several military uses, such as in unmanned reconnaissance aircraft, precision guidance, early air warning, and battlefield warning systems. In recent years, the correlation filtering-based tracking method, which solves the correlation between two signals, has become the mainstream target tracking algorithm due to its high precision and speed. CSK (Exploiting the Circulant Structure of Tracking-by-Detection with Kernels) is a classic target tracking algorithm based on correlation filtering. It uses a rotation matrix for intensive sampling to avoid redundancy or feature insufficiency and employs a fast Fourier transform (FFT) to improve the tracking speed to more than 300 frames/s [4].

Tracking by correlation filtering has been studied under various experimental conditions, but predominantly for camera photographs, which contain rich target features. Nevertheless, the length and width of the ground targets are mostly only a few meters, which are represented as several pixels considering the 1-m resolution of satellite video. Such a small target size results in fewer textural features and also negatively affects color features. Tracking based on correlation filtering uses positive and negative sample features to train the tracking algorithm; therefore, sparse features lead to poor discrimination ability. In addition, satellite observations often involve target occlusion by ground objects, making them prone to tracking drift. Therefore, we suggest that these two aspects should be enhanced in order to achieve accurate target tracking using video satellites through correlation filtering: 1) the discrimination ability of the tracker for weak feature targets and 2) the tracking strategy in order to resist tracking drift.

Manuscript received January 31, 2020; accepted March 2, 2020. Date of publication March 23, 2020; date of current version September 25, 2020. This work was supported in part by the Key Research and Development Program of the Ministry of Science and Technology under Grant 2018YFB0504905, Grant 2018YFC0825803, and Grant 2016YFB0500801, in part by the Major Special Project of High Resolution Earth Observation System under Grant 11-Y20A12-9001-17/18, in part by the National Natural Science Foundation of China under Grant 91538106, Grant 41501503, Grant 41601490, and Grant 41501383, in part by the Development of Space-Based High-Resolution Video Camera and On-Orbit Data Processing and Application Technology under Grant D040107, in part by the Quality Improvement of Chinese Satellite Data and Comprehensive Application Demonstration of Geology and Mineral Resources, and in part by the Open Research Fund of State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing under Grant 15E02. *(Corresponding author: Taoyang Wang.)*

Yunming Wang and Taoyang Wang are with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China (e-mail: wangtaoyang@whu.edu.cn).

Guo Zhang is with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China.

Qian Cheng and Jia-qi Wu are with the School of Geomatics, Liaoning Technical University, Fuxin 123000, China.

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2020.2978512

The existing algorithms mainly aim at the targets with rich features, which have not typically been studied for dim small target in this manner.

In this article, the CSK algorithm is used as the target tracking framework. The performance of this algorithm regarding the aforementioned problems is improved in order to achieve small target tracking using video satellites through background compensation. Tracker training and target feature extraction are the main aspects of correlation filtering; therefore, in order to improve the discrimination ability of the tracker, we propose the following improvements based on these two aspects.

Regarding problem 1), i.e., the discrimination ability of the tracker for small targets because of minimal feature information, CSK cannot be used for satellite videos; therefore, the concept of information compensation is employed to introduce background information into training. In this article, we establish a filter training mechanism based on the target and its background. This mechanism increases the amount of information in the training samples for target tracking and uses background information to enhance the ability of the filter to differentiate the target and background, thereby improving the target tracking robustness. However, gray features of small targets cannot effectively describe small targets. Hamamoto *et al.* [5] proposed a target recognition algorithm based on the Gabor filter and found it to be very suitable for texture representation and separation. Hence, in order to enhance the distinction between positive and negative samples, we introduce the Gabor filter to enhance detailed features of the samples and construct a feature representation model based on the Gabor filter, thereby effectively enhancing the robustness of the tracker.

Regarding problem 2), i.e., the tracking strategy, satellite video targets are easily occluded by high-rise buildings and vegetation. After occlusion, the target tracking will drift, making it difficult to locate when the target will appear again. Therefore, the target tracking status should be evaluated in time to facilitate the redetection and location of the target. As such, we propose a tracking and monitoring index to monitor the target tracking status in realtime. After determining target loss, it utilizes the smoothness of target motion in satellite videos and predicts the target position using the historical trajectory of the target, ensuring the timely location of the target once it reappears.

In summary, this article presents a small target tracking method for satellite video technology using background compensation, named "weak target tracking based on information compensation" (WTIC). A filter training mechanism based on target and background information is established to improve the distinction ability of the filter and a target feature representation model based on the Gabor filter is constructed to enhance the contrast of the prebackground. Finally, a tracking status monitoring index is proposed to evaluate the current tracking state, thus enabling timely prediction of tracking drift and the achievement of real-time robust tracking of moving targets using video satellites.

The remainder of the article is organized as follows: in Section II, we review previous relevant research; in Section III, we introduce the proposed method, including cooperative training of target and background information, the feature representation model based on the Gabor filter, and the monitoring mechanism of target tracking; Section IV contains the analysis of the experimental results and the effect of the tracking algorithm; the conclusions are presented in Section V.

II. Related Literature

In this section, we present a review of the development of target tracking algorithms. Target tracking refers to the process of following a target through successive frames after determining its location in the first frame. The difficulty of target tracking mainly consists of changes in target attributes and changes in the background. Changes of target attributes include appearance distortion, scale changes, rotation, fast motion, and motion blur. Changes to the background include illumination changes, background similarity interference, and motion occlusion. Target tracking algorithms can be divided into generative and discriminative algorithms. Generative algorithms build a model for targets in the current frame and then look for the most similar region to the model in the next frame. Typical examples are the Kalman filter, particle filter, and mean-shift algorithms. Vojir *et al.* [6] proposed the ASMS algorithm, which introduces scale estimation and color histogram features to the mean-shift framework. Moreover, Vojir *et al.* [6] used a target scale that did not change dramatically and the largest possible range of scales as previous information, then added a backward check. Furthermore, incremental learning for robust visual tracking (IVT) involves learning an incremental subspace model, enabling robust tracking when the appearance of the target changes [7].

Discriminative algorithms train the classifier through machine learning. They take the target as the positive sample and the background as the negative sample in the current frame. The optimal region is then searched by the trained classifier in the current frame. Hare *et al.* [8] presented a framework for adaptive visual object tracking. By explicitly allowing the output space to express the needs of the tracker, the method can avoid the need for an intermediate classification step. At the same time, the algorithm introduces a budgeting mechanism to allow for real-time applications [8]. Deng *et al.* [9] presented an efficient and robust tracking algorithm by exploiting the fast learning and classification capabilities of an extreme learning machine. Specifically, a fast learning-based image feature extractor is developed using ELM-AE. While an accurate appearance model is efficiently established by exploiting ELM/OS-ELM for feature classification and updating.

After Bolme *et al.* [10] first applied correlation filtering to target tracking, its use has become widespread. The CSK algorithm proposes dense sampling and core techniques to optimize MOSSE, the KCF algorithm uses multichannel HOG features based on CSK [11], and the CN algorithm uses multichannel color features to optimize CSK [12]. Danelljan *et al.* [13] introduced a scale change into the CSK algorithm and proposed the DSST algorithm. More recently, Lukežic *et al.* [14] proposed a method of correlation filtering combined with color probability, which achieved spatial and channel reliability. Danelljan *et al*. [15] also proposed

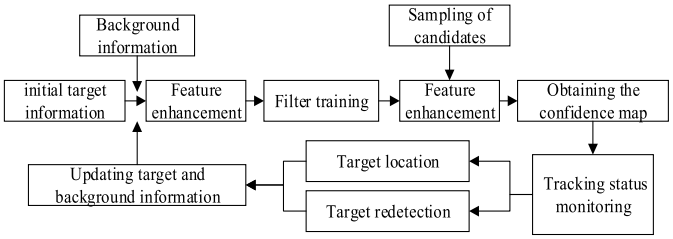


Fig. 1. Flowchart of the proposed target tracking process.

the C-COT algorithm, which combines space regularization and adaptive sample weights to extract the depth features of multilayer convolution and then transforms them into a continuous space domain. As a result, it easily integrates multiresolution feature maps and maintains high positioning accuracy.

Considering the efficiency and accuracy of the satellite video processing, the algorithm with a large amount of calculation is not suitable. So, we improved the algorithm on the basis of CSK. CSK avoids redundancy or insufficiency of features and ensures the rational application of computational resources. Additionally, it uses the kernel trick to improve the ability to discriminate the samples [4]. However, in satellite videos, the target is small, and the information is insufficient to support tracking using CSK. Therefore, our proposed method overcomes these difficulties through the following three aspects:

1) establishing a training mechanism combining foreground and background information;
2) using a feature representation model with the Gabor filter to enhance the contrast between the foreground and background;
3) avoiding target drift by a target tracking state monitoring index, which rapidly reflects the tracking state of the target.

Fig. 1 shows a flowchart of the tracking process proposed in this article. After getting the initial information of the target, first, add the background information around the target into the training samples, then use the Gabor filter to enhance the characteristics of all samples, and finally complete the training of the filter. After training, the filter begins to track the target and evaluate each result. If the target is lost, the historical data are used to complete the redetection. The result of tracking will be used as a new sample to update the filter.

III. Proposed Method

In this section, we introduce the principle of the CSK algorithm and our proposed algorithm (WTIC), which includes three key aspects: a filter training mechanism for the target and background information, a target feature representation model using the Gabor filter, and tracking status monitoring indicators (TCMIs).

A. Theoretical Formulation of the CSK Algorithm

CSK is divided into two parts: classifier training and target detection. In the training part, the CSK tracker would learn a filter. The confidence is the highest in the center of the target and zero in the infinite distance, showing the reverse relationship with the distance. The purpose of training is to form a classifier ω on the set IP so as to detect the subsequent images.

The error between the response value $f(u)$ of the filter ω to the training sample u and the predefined Gaussian regression label r is minimized. The calculation method of $f(u)$ is shown in the following equation:

$$f(u_i) = \langle \omega, u_i \rangle + b. \tag{1}$$

It is worth mentioning that CSK uses the characteristics of the cyclic matrix to obtain a training sample set IP through simple base u. CSK uses the RLS method to train classifiers

$$\min_\omega \sum_i^n L(r_i, f(u_i)) + \tau \|\omega\|^2 \tag{2}$$

$$L(r_i, f(u_i)) = (r_i - f(u_i))^2 \tag{3}$$

where $L(*)$ is the loss function and the classifier parameter, ω, minimizes the loss function. $\tau\|\omega\|^2$ is a regular term to prevent the classifier from over-fitting. τ takes a value of 0.01.

In low-dimensional space, nonlinear features are usually difficult to classify. Therefore, CSK introduces the kernel trick to project samples into feature space. In this way, the nonlinear features can be transformed into linear ones, which can be further classified. According to the kernel theorem of the support vector machine, ω can be expressed as follows:

$$\omega = \sum_i^n \alpha_i \varphi(u_i). \tag{4}$$

In (4), u_i is the sample and $\varphi(u_i)$ projects u_i into the feature space of high dimension. We only need to solve the classifier α. CSK combines (4) with (2) and uses the Lagrange maximum method to solve the minimum. The classifier α is expressed as follows [4]:

$$\alpha = (K + \tau I)^{-1} r. \tag{5}$$

Here, K is the kernel function generated by training samples u. I is the unit array. As K is a cyclic matrix, it can be expressed by the first line k of the matrix. The classifier parameter α can be solved by the attributes of the cyclic matrix as follows [4]:

$$\alpha = F^{-1}\left(\frac{F(r)}{F(k) + \tau}\right) \tag{6}$$

where F and F^{-1} represent the Fourier transform and inverse Fourier transform, respectively. Because of the reduction of matrix dimensions and the application of FFT, the calculation time is greatly shortened.

In the detection phase, y is sampled from the current image. The confidence map is solved using the classifier α according to (1) and (4) as per the following formula:

$$r' = \sum_i^n \alpha_i k(u_i, y). \tag{7}$$

Again, using the loop structure, Fourier transform can be used to complete the detection as follows:

$$\hat{r} = F^{-1}\left(F\left(\bar{k}\right) F(\alpha)\right). \tag{8}$$

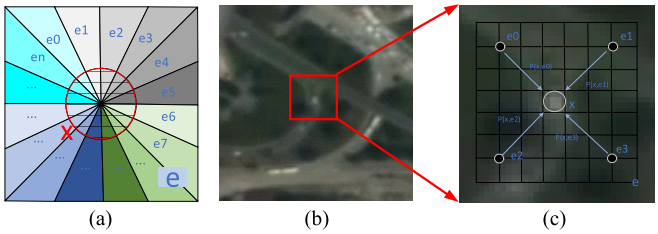


Fig. 2. (a) Venn diagram of background influence. (b) Position of the target in the image. (c) Background constraint on the target $P(x, e_i)$.

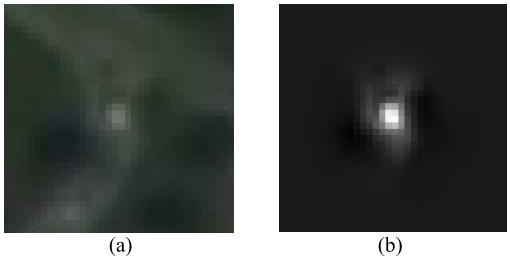


Fig. 3. (a) Image of the target. (b) $P(e_i)$ indicates the influence of the background e_i on the target. The farther the position is, the lower the weight.

In (8), $\bar{k}$ is the kernel function of the current block z and the base block u, as shown in the following equation:

$$\bar{k}_i = k(z, P^i u) \tag{9}$$

where $P^i u$ is the representation of the cyclic matrix, because this article mainly uses the tracking framework of CSK, so the loop structure is used to refer to CSK [4].

After the response value $\hat{r}$ is normalized, the value in $\hat{r}$ indicates the probability that the target is located at the pixel. The maximum value of $\hat{r}$ will be the target position of the current frame. Finally, the classifier α is updated in (6) and linear interpolation.

B. Target and Background Cooperative Training

CSK cannot be used in satellite video directly because the moving target accounts only for a few pixels in the satellite video, and very little target information cannot achieve target tracking. We applied the concept of information compensation to satellite videos that contain less target information [16]–[18].

We use the background to enhance the tracker discernment. There is a spatial relationship $P(x|e_i)$ between the background e_i of each position and the target position x, which shows that the influence $P(x, e_i)$ of the background e_i on the target position x. We use a spatial relationship to constrain the probability $P(x)$ of the target appearing at position x, which would satisfy the following formula:

$$P(x) = \sum_{e_i \in e} P(x, e_i). \tag{10}$$

We use the Venn diagram [Fig. 2(a)] to illustrate the role of background e. In Fig. 2(a), the shadow area in the circle is the probability of the target appearing in x. The background e_i of each part will affect the probability $P(x, e_i)$, that is, the area of $x \cap e_i$. We use Fig. 2(c) to simply describe the influence of each point background on the target.

In addition, according to the conditional probability, the spatial relationship $P(x|e_i)$ satisfies the relation of the following equation:

$$\mathrm{P}(x|e_i) = \frac{P(x, e_i)}{P(e_i)} \tag{11}$$

where $P(e_i)$ is expressed as the weight of background e_i. The farther the background is, the smaller the constraint on the target. We use Fig. 3(b) to show the influence on the background, and the far background weight is 0.

According to the above, the confidence con (x) of the target appearing in position x is the sum of the probabilities under the background constraint. It can be expressed as follows:

$$\mathrm{con}(x) = \sum_{e_i \in e} P(x \mid e_i)\, P(e_i). \tag{12}$$

The spatial relationship is the mapping of position x and background e_i relative position, which can be expressed as follows:

$$P(x \mid e_i) = d(x - e_i) \tag{13}$$

where $d(x - e_i)$ represents a mapping of the spatial relationship between x and e_i.

$P(e_i)$ represents the weight of the background as described above. It is determined by distance factor $W(e_i)$ and gray factor, gray (e_i), as shown in the following equation:

$$P(e_i) = \mathrm{gray}(e_i) W(e_i). \tag{14}$$

The distance factor $W(e_i)$ is determined by the distance between the background e_i and the center $\bar{x}$, which is defined as follows:

$$W(e_i) = a e^{-\frac{|e_i - \bar{x}|^2}{\sigma^2}}. \tag{15}$$

Based on the above, target tracking can also be represented using the following formula:

$$\begin{aligned} \mathrm{con}(x) &= \sum_{e_i \in e} d(x - e_i)\, \mathrm{gray}(e_i)\, W(e_i) \\ &= \sum_{e_i \in e} d(x - e_i) \mathrm{gray}(e_i)\, a e^{-\frac{|e_i - \bar{x}|^2}{\sigma^2}}. \end{aligned} \tag{16}$$

Equation (16) can be further represented by a convolution operation

$$\begin{aligned} \mathrm{con}(x) &= (d(x)) \otimes (\mathrm{gray}(x) W(x)) \\ &= (d(x)) \otimes \left(\mathrm{gray}(x) a e^{-\frac{|x - \bar{x}|^2}{\sigma^2}} \right). \end{aligned} \tag{17}$$

After the confidence map is calculated, the maximum value is the position of the target. The above formula is still applicable to the CSK algorithm framework. After conversion to the frequency domain, the confidence can be expressed as follows:

$$F(\mathrm{con}(x)) = F(d(x)) \odot F(\mathrm{gray}(x) W(x)) \tag{18}$$

$$\mathrm{con}(x) = F^{-1}(F(d(x)) \odot F(\mathrm{gray}(x) W(x)). \tag{19}$$

In the tracking framework, the target tracker needs to initialize the classifier in the first frame. After several frames of processing are completed, historical data are used to update the classifier. This article uses the same update method as in CSK to update the spatial relationship $d(x)$ between the target and the background.

In training, the image slice centered on the target is used as the training sample. The confidence map is set to a similar normal distribution. The training expression of target tracking is as follows:

$$d(x)^{\text{update}} = F^{-1}\left(\frac{F\left(be^{-\frac{(x-\bar{x})^2}{2\sigma^2}}\right)}{F\left(\text{gray}\left(x\right)W\left(x\right)\right)}\right) \tag{20}$$

$$d^{\text{new}} = (1-\rho)d^{\text{old}} + \rho d^{\text{update}}. \tag{21}$$

Combining the old spatial relationship d^{old} with the new spatial relationship d^{update} can ensure the correctness of the tracker. In (21), d^{new} is the updated spatial relationship, and ρ denotes the learning rate that balances the tradeoff for historical target information and the current target state.

C. Gabor-Based Feature Representation Model

To make the CSK algorithm applicable to satellite videos, in this article, we introduced the Gabor filter to the CSK algorithm to enhance the features of targets in satellite videos and the reliability of the CSK tracking method.

In image processing, Gabor function is a linear filter for edge extraction, which is suitable for texture representation and separation. In practical applications, more attention is paid to the characteristics of the local range of signals.

The Gabor function is also similar to the sensory field of mammalian retinal neurons and is thus mostly used in the fields of image processing, understanding, and recognition [19], [20]. In research into multiresolution representation based on Gabor filters, Manjunath and Ma [21] noted that the latter can be considered as orientation and scale tunable edge and line (bar) detectors, and the statistics of these microfeatures in a given region are often used to characterize the underlying textural information. In addition, Zhang *et al.* [22] proposed the robust visual tracking via a basis matching method and used a set of Gabor basis functions to train a target model based on target images using a set of Gabor basis functions.

In the spatial domain, a 2-D Gabor filter is the product of a sinusoidal plane wave and a Gauss kernel function. The former is a tuning function, and the latter is a window function, as shown in the following equation:

$$g\left(x, y; \lambda, \theta, \psi, \sigma, \gamma\right) = e^{-\frac{x'^2+\gamma^2y^2}{2\sigma^2}}e^{i\left(2\pi\frac{x'}{\lambda}+\psi\right)}. \tag{22}$$

It can be divided into real and imaginary parts as below

$$\begin{cases} g_{\text{real}}\left(x, y; \lambda, \theta, \psi, \sigma, \gamma\right) = e^{-\frac{x'^2+\gamma^2y^2}{2\sigma^2}}\cos\left(2\pi\dfrac{x'}{\lambda}+\psi\right) \\ g_{\text{imag}}\left(x, y; \lambda, \theta, \psi, \sigma, \gamma\right) = e^{-\frac{x'^2+\gamma^2y^2}{2\sigma^2}}\sin\left(2\pi\dfrac{x'}{\lambda}+\psi\right). \end{cases} \tag{23}$$

Among which

$$\begin{cases} x' = x\cos\theta + y\sin\theta \\ y' = -x\sin\theta + y\cos\theta. \end{cases} \tag{24}$$

We set the details of the above parameters. we set λ to 2 considering that our target size is below 8×8 pixels. In this article, the θ value is related to the way in which Gabor filtering is used. Furthermore, because the target is at the center of the photograph, we set the phase offset ψ to 0. Finally, we set the aspect ratio γ to 1 and the bandwidth σ to 2π because the shape of the moving target is nearly square.

After extracting target samples or candidate regions, we employ the Gabor filter to process image blocks. The vehicle is weak and the surrounding background is disturbed in satellite videos; therefore, we advocate a filtering strategy to enhance the distinguishability between the target and the background. Specifically, we use the direction of the target movement to complete the Gabor filtering. The features in the direction in which the target advances are preserved while those in other directions are suppressed.

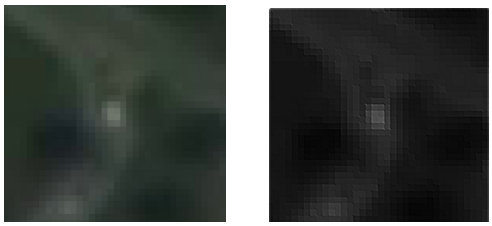
We illustrate the filtered image as below. The small point-shaped target is more prominent than the background and thus improves the distinguishability between the target and the background.

D. Target Tracking State Monitoring

Determining the accuracy of the tracking results is a key issue. This step determines the updating strategy of the model. In the case of a tracking failure or poor accuracy, error information is introduced into the training and will inevitably affect the accuracy of the model. Grossberg [23] suggested that a critical problem in model updates is related to the stability–plasticity dilemma; thus, model updates should be stable to avoid drifting problems in which small errors accumulate and the model adapts to other objects. The model also requires plasticity to effectively assimilate new information derived during tracking [23].

Many algorithms such as the kernelized correlation filter (KCF) [11] and Accurate Scale Estimation for Robust Visual Tracking (DSST) [13] do not determine the reliability of the tracking results. Training typically employs the results of each frame or the results of an interval of N frames, such as the multidomain network (MDNet) [24] and tubelets with convolutional neural networks (TCNNs) [25]. However, these methods are risky. When the target is occluded or the tracker has failed, the training model will affect the ability of the tracker to make discriminations [26], [27].

To avoid the template contamination caused by inaccurate tracking results, we added the judgment of the tracking results. As shown in Fig. 5, the peak of the response map is more prominent compared to the background, assuming that the response map follows a Gaussian distribution with spikes and slight tails throughout the search window. However, under the influence of some unusual properties and the sample noise, the tracker can easily lose the target and the response map will contain multiple peaks or unusual shapes. At this point, the peak becomes less obvious.



(a) (b)

Fig. 4. (a) Original image. (b) Image processed using Gabor filtering. The background noise of the image is suppressed, and the target is highlighted.

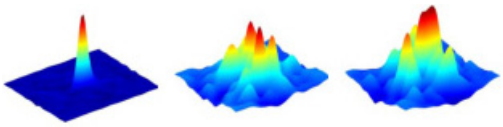


Fig. 5. Confidence maps of different state trackers.

Based on the relationship between confidence map and tracking results, we propose the TCMI to judge the tracking condition for the current frame. Specifically, the value of TCMI is determined by the peak intensity in the confidence map. The detailed formula is shown as follows:

$$\mathrm{TCMI} = \frac{|\mathrm{con}_{\max} - \mathrm{con}_{\min}|}{3 \times \mathrm{sqrt}\left(\frac{\sum_{w,h}(\mathrm{con}_{w,h} - \mathrm{con}_{\min})^2}{w \times h}\right)}. \tag{25}$$

Here, $\mathrm{con}_{\max}$ and $\mathrm{con}_{\min}$ denote the maximum and minimum values, respectively, in the response map. Thus, we record each value in the response map and calculate the standard deviation for the confidence map. For the correlation-based tracker, the response map would have a sharp peak and slight tail once the tracking condition is ideal; therefore, finding the maximum and minimum deviations should be sufficient. In this way, we can monitor the tracking condition by calculating the proposed TCMI. Based on the results of multiple experiments, we take 0.6 as the threshold to determine whether or not tracking failure occurs. The full algorithm is detailed below.

Algorithm

Input: Current image **I**, Initial target position **pos_t-1**, spatial relationship after training **d**.
Output: The target location **pos_t**, Updated spatial relationship **d_update**, History track **pos_history[]**.
Repeat:
Extract the image block **ui** from the image **I**, and convert **ui** into a grayscale image.
if not the first frame
  Calculate the current direction $\boldsymbol{\theta}$ using the history track **pos_history[]**,then use the filter $\mathbf{g}(\boldsymbol{\theta})$ to enhance the **ui** (Equation 22).
else
  Use median filtering to enhance ui
end if
Use **W(ui)** to weight **ui** (Equation 14)
Use **d** to solve the confidence map **con** (Equation 19).
Solve the maximum value **con_max** of the confidence map, solve the target tentative position **pos_tmp**
Calculate the standard deviation **con_standard** of the confidence map. Calculate the **TCMI**. (Equation 25)
if TCMI suggests that the target tracking is successful, then
  Update target location **pos_t**;
  Add **pos_t** to the historical track **pos_history[]**
  Extract the slice **ui** where the target is located;
  Calculate the current direction θ using the history trajectory **pos_history[]**, then use the Gabor filter $\mathbf{g}(\boldsymbol{\theta})$ to enhance **ui** (Equation 23)
  Use the **ui** processed by **W(ui)** to solve the spatial relationship **d_update** in the current **ui** (Equation 21) and combine it with **d** to update spatial relationship **d** (Equation 21).
else
  Use the result of the last n frames **pos_history[n]** to predict the location of the target **pos_t**. n = 10 in the experiment, all use less than 10.
end if
until the end of video sequence.

IV. Experimental Results and Analysis

In this article, nine Jilin-1 satellite video data sets were used to verify the proposed algorithm (WTIC). The resolution of the Jilin-1 satellite video data sets is 1 m, and the tracking target is a vehicle. The experiment included satellite videos of Libya, Long Beach, Valencia, Ankara, Istanbul, Atlanta, and Mexico. The videos could be broadly divided into three groups with different challenging attributes, namely background clutter, turning targets, and short-term occlusion. All videos were shot from March to May 2017. The target size is approximately 8×8 pixels given the resolution of the satellite data.

We used a MATLAB 2015b platform to verify the algorithm. The computer was configured as an Intel i5-CPU (GHz) with 16-GB memory and running Windows 10. We conducted the quantitative comparison experiments with 12 state-of-the-art trackers, namely, the L1 tracker using accelerated proximal gradient approach (L1APG) [28], sum of template and pixel-wise learners (Staple) [29], multitask tracking (MTT [32], a convolutional network-based tracker (CNT) [31], CSK [4], KCF [11], sparsity-based collaborative model (SCM) [30], discriminative correlation filter tracker with channel and spatial reliability (CSR-DCF) [14], continuous convolution operator tracker (CCOT) [15], online robust image alignment (ORIA) [33], incremental learning for robust visual tracking (IVT) [7], and distribution fields for tracking (DFT) [34]. The quantitative experiments compared the tracking accuracy, tracking success rate, and overall performance. However, as the last two algorithms failed to track further, we flagged the first 12. These trackers were selected because the abovementioned tracking algorithms have achieved success in related conferences or competitions in the tracking community. Thus, via comparisons with classical methods, we can highlight how this method is much better for target tracking in satellite video.

A. Qualitative Experiment

To better evaluate our approach, we conducted target tracking experiments in representative experimental scenes based on the characteristics of the satellite videos. The experimental scenes included similar backgrounds, rotation, and occlusion. We selected only the top seven tracking algorithms from the qualitative experiments due to the low resolution of the satellite data. Furthermore, we have provided the necessary description for each data set, including the time of acquisition, side swing angle, and the direction in which the target was moving. The number of frames in different videos is also different, and in most cases, the tracking effect is similar. Therefore, Fig. 6 shows several typical frames of the video for analysis. For better visualization, we have labeled the frame number in Fig. 6(a)–(c).

1) Similar Targets and Backgrounds: Three data sets were arranged according to the similarities between the target and the background (Fig. 6).

Poor discrimination between the target and the background greatly affected tracker discrimination and resulted in part of the tracking algorithm losing the target. In experiments with greater similarity, more trackers lost the targets. Most of these methods extract features from different objects and backgrounds but were ineffective for weak features. SCM is the most typical example; it failed to track objects or vehicles when they exhibited high similarities to the background. However, the target feature representation model constructed using the Gabor filter effectively strengthened the description of the target and improved the contrast between the target and the background (Fig. 6). Our approach ensured that the target was tracked accurately in the experimental video (Fig. 6).

2) Target Rotation: Fig. 7 shows the experimental results for target rotation. While the target was changing direction, the tracker still located the target accurately because the texture features of the satellite video target are simple; hence, the impact of the direction change is smaller. The target was tracked accurately using the feature adapted for rotation. However, the background changes were greater than the linear motion, and some algorithms exhibited tracking drift or loss of the target. When the background changed, the tracker required better discrimination capability. When the background changed substantially, the majority of trackers (such as staple and MTT) exhibited peak confidence offset or target scale change. In this article, the feature representation model used the Gabor filter, which is adaptable for rotation, and the target and background information were used to train the tracker to minimize the impact of background interference. As a result, the tracker effectively followed the target, indicating enhanced robustness of moving target tracking.

3) Target Occlusion: Fig. 8 shows the experimental results of target tracking under target occlusion along with a description for each sequence. It should also be mentioned that since the occlusion is between a few frames, we show only the keyframes where the target enters the occlusion.

When an object is occluded, the target disappears for a short time; thus, the target is lost and error information becomes incorporated into the training of the target tracker. This leads to increased errors and reduced tracking robustness. After a short period of complete occlusion of the target, most algorithms lost the target or exhibited decreased accuracy (Fig. 8). When the target was occluded several times, some trackers, such as L1APG, could not resist occlusion interference. However, the proposed condition monitoring index effectively evaluated the tracking state of the target and rapidly eliminated the erroneous information. Subsequently, the tracker was able to estimate the position of the target using the smoothness characteristics of the satellite video target motion and simultaneously

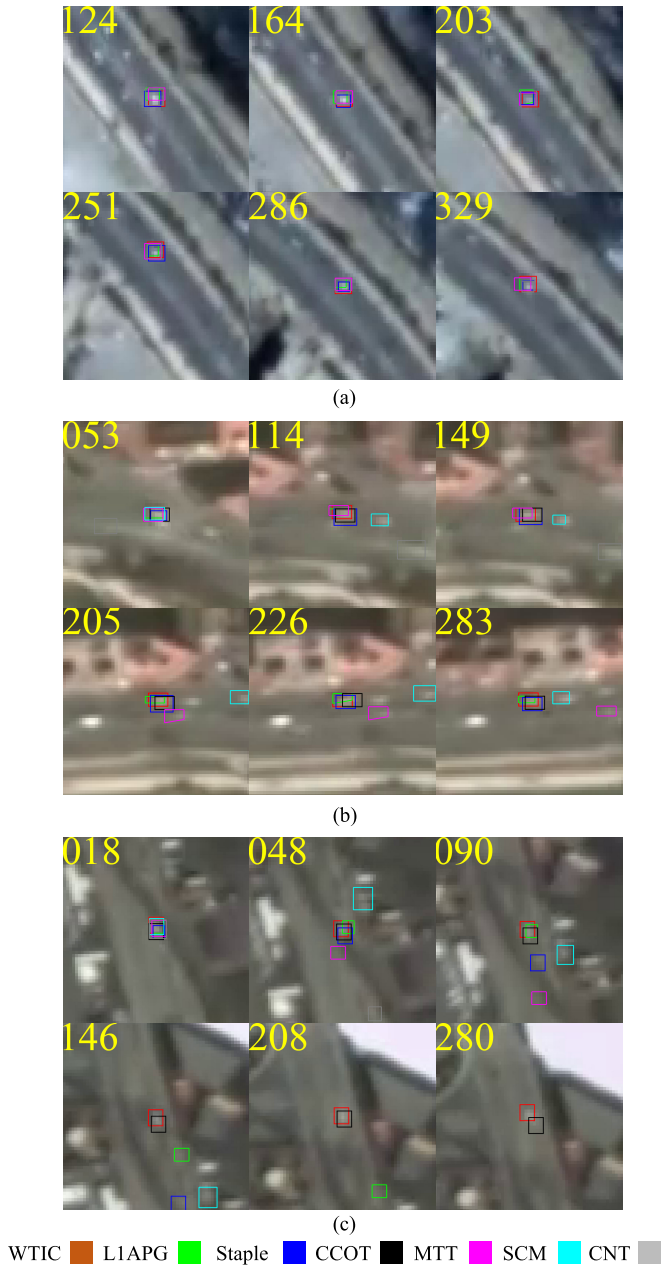


Fig. 6. Experimental results for a similar target and background. (a) Libya: video shot on May 20, 2017, with a side swing angle 2.1256°. The target is moving from the bottom to the upper left of the images. (b) Valencia: video shot on March 7, 2017, with a side swing angle of 1.6501°. The target was moving from right to left. (c) Long Beach: video shot on April 3, 2017, with a side swing angle of 3.0424°. The target is moving from the top to bottom of the image. Each algorithm uses the corresponding color to mark the target.

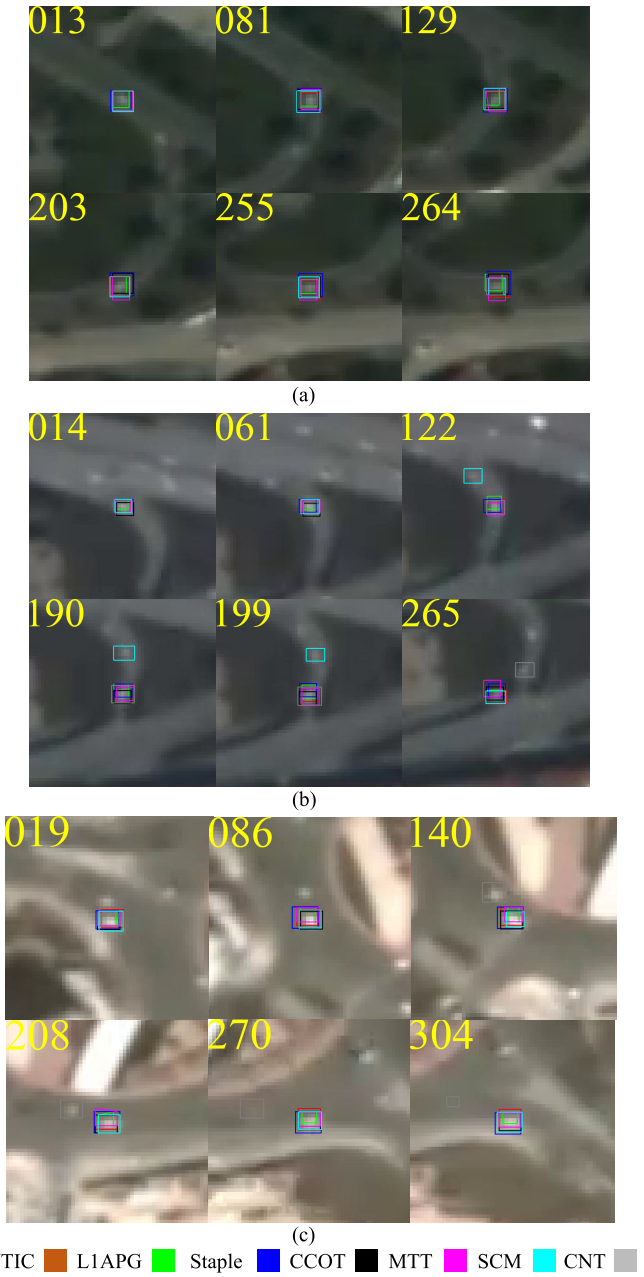


Fig. 7. Experimental results for target rotation. (a) Ankara: shot on April 29, 2017, with side swing angle 14.702°. The target turns from the top to left. (b) Istanbul: shot on April 14, 2017, with a side swing angle of 21.016°. The target turns from the top to the left. (c) Valencia: shot on March 7, 2017 with side swing angle of 1.6501°. The target performs an S-shaped movement from the top left to the bottom right. Each algorithm uses the corresponding color to mark the target.

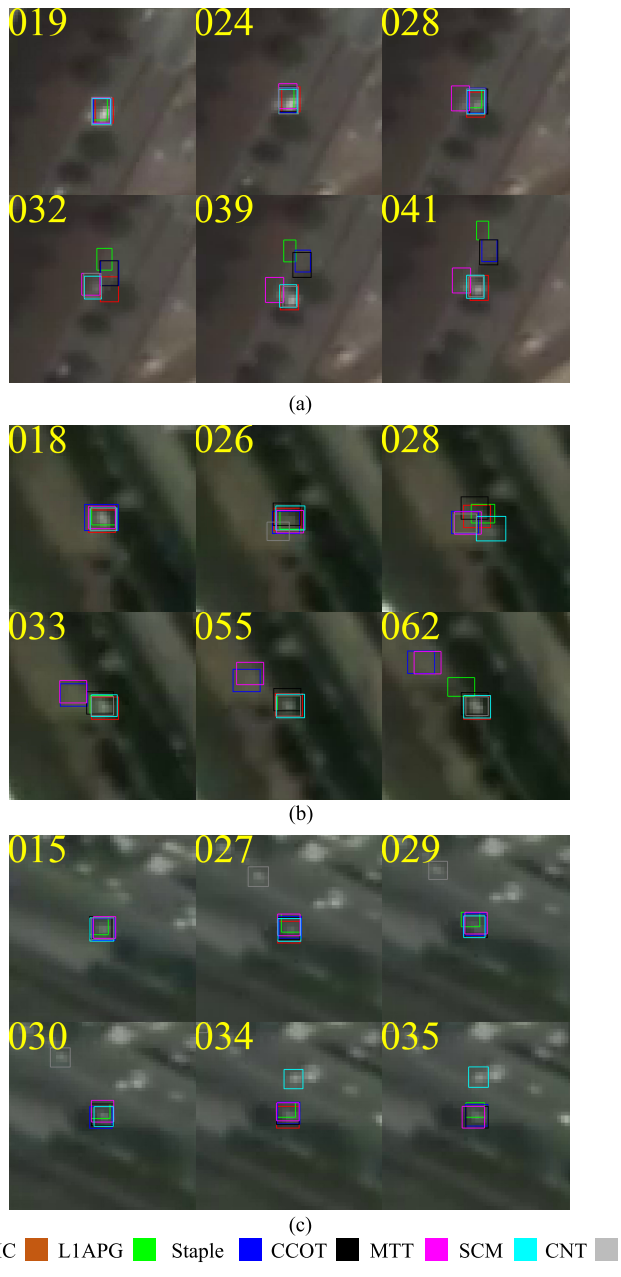


Fig. 8. Experimental results for target occlusion. (a) Atlanta: shot on May 3, 2017, with side swing angle is -2.9679°. The target moves for bottom to top right. The occlusion begins on frame 24 and ends on frame 39. Specifically, the target is fully-occluded on frame 32. (b) Guadalajara: shot on April 2, 2017, with a side swing angle of -33.1742°. The target moves from top left to bottom right. The target moves from top left to bottom right. The occlusion starts on frame 26 and ends on frame 33. Specifically, the vehicle is fully occluded by the vegetation from frame 28. Similarly, the target is under partial occlusion from frame 55 and reappears at frame 62. (c) Ankara: shot on April 29, 2017, with a side swing angle of 14.702°. The target moves from the top left to the right. The target begins to enter the occlusion at frame 15, is occluded at frame 27, 29, and reappears at frame 34. Each algorithm uses the corresponding color to mark the target.

redetect the target. This ensured timely redetection after occlusion as well as tracking accuracy (Fig. 8).

B. Quantitative Experiment

In this section, we used the one-pass-evaluation, which is a common evaluation method used in OTB proposed in [35], to evaluate the performance of the proposed tracker (WTIC) and other related trackers. The quantitative evaluation uses two criteria: success rate and precision, which are defined below.

Precision: The percentage of frames whose center position error is less than the predefined threshold. The center position error indicates the distance between the center of the tracking result and the center of the boundary box. In this article, we set 3-pixel as the determination threshold.

Success Rate: The percentage of frames the overlap rate of the tracking area and boundary box of which is greater than the threshold value. The area under the curve (AUC) indicates that the tracking performance of all thresholds varies

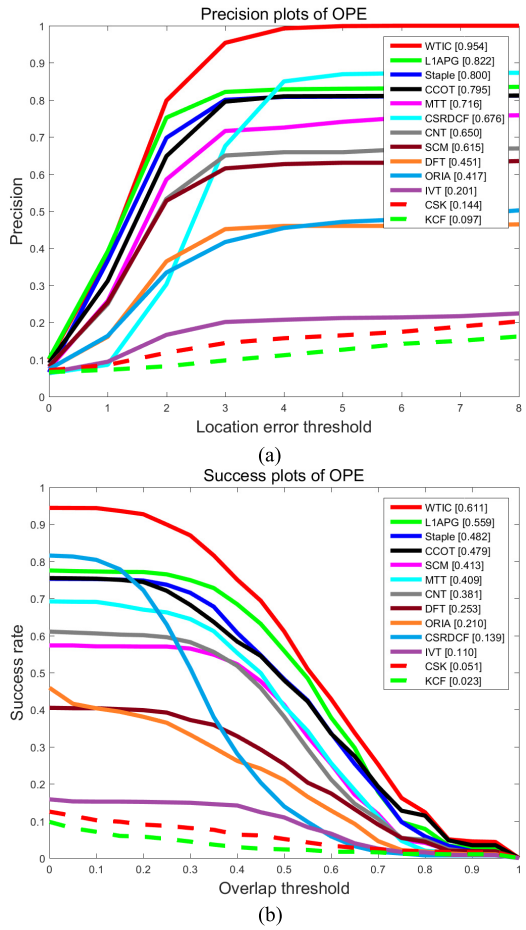


Fig. 9. (a) Precision and (b) success rate of the top ten algorithms compared in this article.

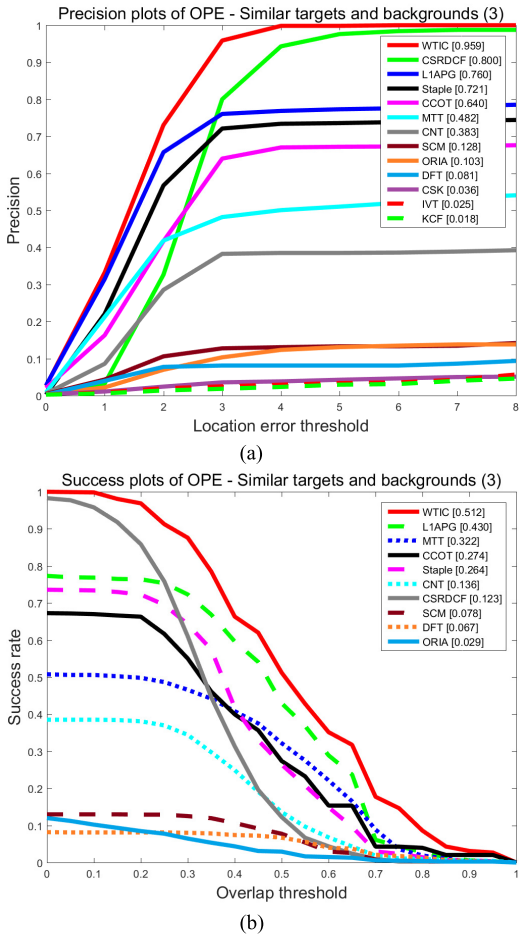


Fig. 10. (a) Precision and (b) success rate of the top ten algorithms for similar targets and backgrounds.

from 0 to 1. Considering that targets in satellite videos are small (the largest targets only occupy 8 × 8 pixels), there may be some labeling inaccuracy since all the ground truths are labeled manually. After conducting extensive experiments, we found that the success rate is easily affected by the labeling error when the overlap rate is larger than 0.5. Therefore, we set the threshold for the success rate as 0.5.

Figs. 9–12 show the effect of a quantitative experiment. We use different colors to sort different performance methods. Red is the best algorithm.

1) Overall Performance: Fig. 9 shows the ten groups of algorithms with the best performance in many trackers. Notably, the approach proposed in this article (WTIC) shows the top performance according to both success rate and precision. The success rate of WTIC is 61.1%. However, the CSK algorithm, which is the basis of WTIC, does not appear in the top ten algorithms. This indicates that our improvements to the algorithm provide a stronger ability to distinguish targets. In addition, the accuracy of our tracker is better than that of CCOT, CSR-DCF, and Staple, which exhibit excellent performance in visual object tracking (VOT). The VOT challenge is a tracking competition with a precisely defined and repeatable method of comparing short-term trackers as well as a common platform to discuss the evaluation and advancements made in the field of visual tracking. Satellite video targets have weak features; therefore, in order to robustly track the target, it is necessary to strengthen the description of the target features. In CCOT and other algorithms, multidimensional features are used to describe the target, which cannot describe the small targets of satellite videos. In this article, Gabor filtering is used to enhance the detailed features of the target and make full use of the target information, which is shown to have a significant effect on accurate tracking.

2) Similar Targets and Backgrounds: In order to further evaluate WTIC, it was compared with other algorithms under different experimental conditions. Fig. 10 shows the results of the experiment with similar targets and backgrounds, which are most common in satellite videos. Among all tracking algorithms, WTIC ranks first for precision and success rate. The precision and success rate of WTIC are 0.954 and 0.611, both of which are higher than those of the second best algorithm. CSR-DCF uses spatial confidence to ensure tracking precision but tracking drift, caused by weak features, reduces the success rate. Other trackers (L1APG, Staple, CCOT, and MTT) gradually lose their targets due to background interference. When the target is similar to the background, the target exhibits weak features; therefore, the Gabor filter enhances the feature representation of the target and improves the contrast between the target and the background.

3) Target Rotation: Fig. 11 shows the results for a change in the direction of target motion. Among all the experimental

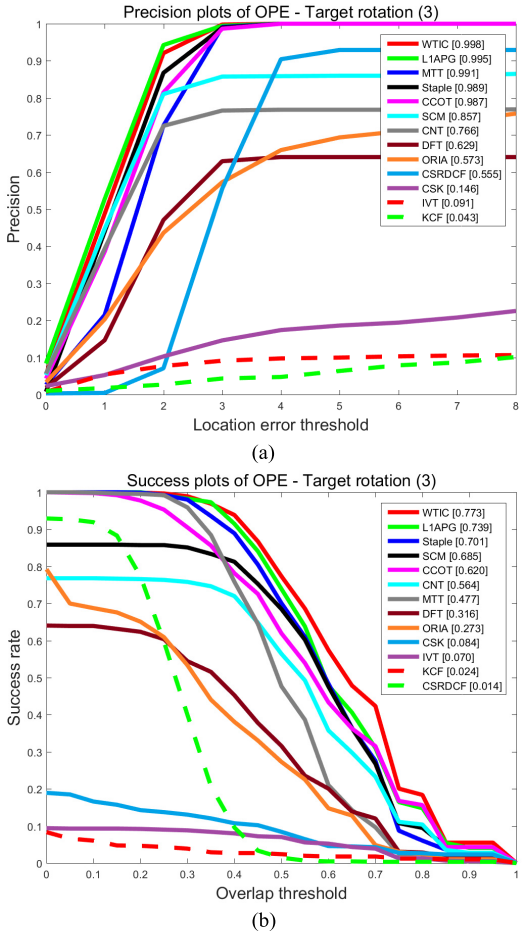


Fig. 11. (a) Precision and (b) success rate of the top ten algorithms for target rotation.

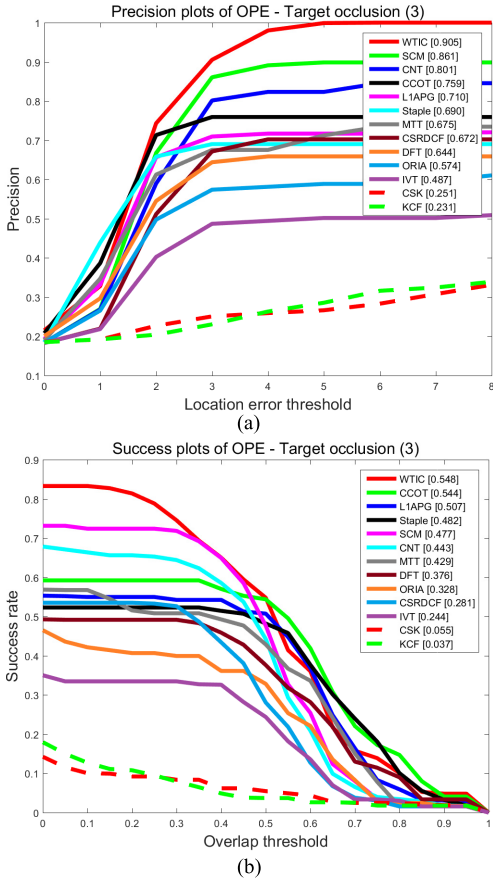


Fig. 12. (a) Precision and (b) success rate of the top ten algorithms for target occlusion.

tracking algorithms, our algorithm ranks first in precision and success rate (0.99 and 0.77, respectively). When small targets rotate, the target does not change dramatically; therefore, we can conclude that rotation interference does not have a substantial influence. In fact, most algorithms do not lose the target during the experiment. However, the background changes when the target changes direction, which reduces the accuracy of most trackers. For example, CSR-DCF does not lose track of the target when the threshold of precision is large, but the precision decreases rapidly when this threshold is reduced. CSR-DCF relies on spatial confidence to ensure the robustness of the tracker, but cannot adapt rapidly when the background changes. On the contrary, the Gabor filter enhances the description of the target allowing the tracker to adapt to the rotation of the target. Moreover, introducing the background in training to account for and address the interference caused by background changes is shown to be a reasonable and effective approach.

4) Target Occlusion: Fig. 12 shows the results for target occlusion. Again, WTIC ranks first in both precision and success rate (0.905 and 0.548, respectively). During target occlusion, the target disappears for a short time, and the tracking algorithm introduces error samples into the training. For example, the tracking accuracy of CCOT and Staple gradually decreases to approximately 0.73 due to this updating mechanism. However, the proposed evaluation index of the tracking status ensures timely redetection after the target is lost. Then, using a smooth target trajectory in the satellite video, WTIC uses target motion information to predict the next position of the target after occlusion. The experimental results show that the algorithm can effectively avoid target occlusion.

TABLE I

TARGET AND BACKGROUND COOPERATIVE TRAINING-SUCCESS RATE

Attributes\Methods	WTIC	WTIC without Cooperative-Training
Background clutter	0.512	0.004
Target rotation	0.773	0.027
Target occlusion	0.548	0.119
Overall	0.611	0.050

C. Ablation Study

In this section, we will validate the key components of the proposed tracker to see how they contribute to improving the overall performance. Generally, our WTIC tracker has three constitute components, including collaborative training, Gabor filtering, and TCMIs.

1) Target and Background Cooperative Training: First, to verify the capability of the cooperative training scheme, we conducted a quantitative experiment with and without this scheme based on our tracker. The overall performance for different environments is shown in Tables I and II.

TABLE II
TARGET AND BACKGROUND COOPERATIVE TRAINING-PRECISION RATE

Attributes\Methods	WTIC	WTIC without Cooperative-Training
Background clutter	0.959	0.008
Target rotation	0.998	0.052
Target occlusion	0.905	0.375
Overall	0.954	0.145

TABLE III
GABOR-BASED FEATURE REPRESENTATION MODEL-SUCCESS RATE

Attributes\Methods	WTIC	WTIC without Gabor feature
Background clutter	0.512	0.141
Target rotation	0.773	0.385
Target occlusion	0.548	0.430
Overall	0.611	0.319

TABLE IV
GABOR-BASED FEATURE REPRESENTATION MODEL-PRECISION RATE

Attributes\Methods	WTIC	WTIC without Gabor feature
Background clutter	0.959	0.344
Target rotation	0.998	0.640
Target occlusion	0.905	0.612
Overall	0.954	0.532

TABLE V
TARGET TRACKING STATE MONITORING-SUCCESS RATE

Attributes\Methods	WTIC	WTIC without TCMI indicator
Background clutter	0.512	0.512
Target rotation	0.773	0.773
Target occlusion	0.548	0.517
Overall	0.611	0.601

TABLE VI
TARGET TRACKING STATE MONITORING-PRECISION RATE

Attributes\Methods	WTIC	WTIC without TCMI indicator
Background clutter	0.959	0.959
Target rotation	0.998	0.998
Target occlusion	0.905	0.793
Overall	0.954	0.916

Both the precision and success rate have greatly improved for three scenarios with the help of the cooperative training of target and background, which could further highlight its significance for this scheme.

2) Gabor-Based Feature Representation Model: Second, we have disabled the Gabor filter to verify the effectiveness of this key component. As shown in Tables III and IV, the precision rate decreases from 0.954 to 0.532, and the success rate reduces from 0.611 to 0.319. Furthermore, without the employment of the Gabor feature, the proposed tracker performs poorly under challenging conditions, especially background clutter. We attribute this phenomenon to the better appearance modeling ability for Gabor feature.

3) Target Tracking State Monitoring: Finally, we analyze the effectiveness of the proposed TCMI. Similarly, we have conducted comparison experiments with and without TCMIs. As shown in Tables V and VI, the tracking results vary only in the case of occlusion, since the tracking condition monitoring index is designed to continue to maintain the correct tracking state after the target is occluded by the ground object. As is evident, the tracking accuracy and precision would drop significantly under occlusion scenarios with the TCMIs disabled.

V. CONCLUSION

In this article, a robust tracking algorithm (WTIC) is proposed for a small target with Jilin-1 satellite video data. The proposed approach effectively and robustly tracks vehicles under three challenging environments: background clutter, target rotation, and target occlusion. The following conclusions can be drawn.

First, considering the small size of satellite video targets, a tracking strategy of increasing sample information should be adopted. Herein, we trained the tracker using both the target and background, with excellent results. Second, the weak features of satellite video targets should be strengthened. In this article, we constructed a feature representation model using the Gabor filter to successfully improve the representation of target features. Third, as satellite video targets can be occluded by ground objects, we proposed an evaluation index to rapidly determine the state of the tracker and ensure redetection of the target after occlusion. The WTIC algorithm exhibited better performance for small target tracking of satellite video than CCOT, CSR-DCF, and Staple, which are all excellent algorithms used in VOT 2016 and VOT 2017, and achieved robust tracking.

Finally, our tracker can provide technical support for the implementation of urban road monitoring and tracking for key targets. This article also provides an important foundation for target motion state estimation through sequential space-based remote sensing.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their constructive comments and suggestions. They would also like to thank Yuqi Han from the Beijing Institute of Technology, Beijing, China, for his encouraging and insightful advice.

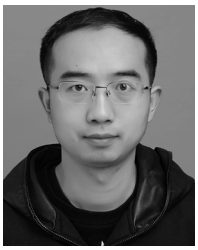
REFERENCES

[1] P. d'Angelo, G. Máttyus, and P. Reinartz, "Skybox image and video product evaluation," *Int. J. Image Data Fusion*, vol. 7, no. 1, pp. 3–18, Nov. 2015.
[2] X. Zhang, J. Xiang, and Y. Zhang, "Tracking imaging attitude control of video satellite for cooperative space object," in *Proc. IMCEC*, Xi'an, China, Oct. 2016, pp. 429–434.
[3] Y. Han, C. Deng, Z. Zhang, J. Li, and B. Zhao, "Adaptive feature representation for visual tracking," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Beijing, China, Sep. 2017, pp. 1867–1870.
[4] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "Exploiting the circulant structure of tracking-by-detection with kernels," in *Proc. ECCV*, Florence, Italy, Oct. 2012, pp. 702–715.
[5] Y. Hamamoto, S. Uchimura, M. Watanabe, T. Yasuda, Y. Mitani, and S. Tomita, "A Gabor filter-based method for recognizing handwritten numerals," *Pattern Recognit.*, vol. 31, no. 4, pp. 395–400, Apr. 1998.
[6] T. Vojir, J. Noskova, and J. Matas, "Robust scale-adaptive mean-shift for tracking," *Pattern Recognit. Lett.*, vol. 49, pp. 250–258, Nov. 2014.
[7] D. A. Ross, J. Lim, R.-S. Lin, and M.-H. Yang, "Incremental learning for robust visual tracking," *Int. J. Comput. Vis.*, vol. 77, nos. 1–3, pp. 125–141, May 2008.
[8] S. Hare *et al.*, "Struck: Structured output tracking with kernels," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 10, pp. 2096–2109, Oct. 2016.

[9] C. Deng, Y. Han, and B. Zhao, "High-performance visual tracking with extreme learning machine framework," *IEEE Trans. Cybern.*, to be published, doi: 10.1109/TCYB.2018.2886580.
[10] D. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, "Visual object tracking using adaptive correlation filters," in *Proc. CVPR*, San Francisco, CA, USA, Jun. 2010, pp. 2544–2550.
[11] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 583–596, Mar. 2015.
[12] M. Danelljan, F. S. Khan, M. Felsberg, and J. V. D. Weijer, "Adaptive color attributes for real-time visual tracking," in *Proc. CVPR*, Columbus, OH, USA, Jun. 2014, pp. 1090–1097.
[13] M. Danelljan, G. Häger, F. Shahbaz Khan, and M. Felsberg, "Accurate scale estimation for robust visual tracking," in *Proc. BMVC*, Nottingham, U.K., Sep. 2014, pp. 1–11.
[14] A. Lukežič, T. Vojír, L. Č. Zajc, J. Matas, and M. Kristan, "Discriminative correlation filter tracker with channel and spatial reliability," *Int. J. Comput. Vis.*, vol. 126, no. 7, pp. 671–688, Jul. 2017.
[15] M. Danelljan *et al.*, "Beyond correlation filters: Learning continuous convolution operators for visual tracking," in *Proc. ECCV*, Oct. 2016, pp. 472–488.
[16] Y. Han, C. Deng, B. Zhao, and B. Zhao, "Spatial-temporal context-aware tracking," *IEEE Signal Process. Lett.*, vol. 26, no. 3, pp. 500–504, Mar. 2019.
[17] S. Haner and I. Y. Gu, "Combining foreground/background feature points and anisotropic mean shift for enhanced visual object tracking," in *Proc. 20th Int. Conf. Pattern Recognit.*, Istanbul, Turkey, 2010, pp. 3488–3491.
[18] S. Sun and W. Kang, "Self-adaptive visual tracker based on background information," in *Proc. 6th Int. Conf. Instrum. Meas., Comput., Commun. Control (IMCCC)*, Harbin, China, Jul. 2016, pp. 1003–1008.
[19] L.-K. Zhang and H. Zhao, "Real time mean shift tracking using the Gabor wavelet," in *Proc. Int. Conf. Mechatronics Autom.*, Harbin, China, Aug. 2007, pp. 1617–1621.
[20] R. Mehrotra, K. R. Namuduri, and N. Ranganathan, "Gabor filter-based edge detection," *Pattern Recognit.*, vol. 25, no. 12, pp. 1479–1494, Dec. 1992.
[21] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 8, pp. 837–842, Aug. 1996.
[22] S. Zhang, X. Lan, Y. Qi, and P. C. Yuen, "Robust visual tracking via basis matching," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 3, pp. 421–430, Mar. 2017.
[23] S. Grossberg, "Competitive learning: From interactive activation to adaptive resonance," *Cognit. Sci.*, vol. 11, no. 1, pp. 23–63, Jan. 1987.
[24] H. Nam and B. Han, "Learning multi-domain convolutional neural networks for visual tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 4293–4302.
[25] K. Kang *et al.*, "T-CNN: Tubelets with convolutional neural networks for object detection from videos," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 10, pp. 2896–2907, Oct. 2018.
[26] M. Lourenco and J. P. Barreto, "Tracking feature points in uncalibrated images with radial distortion," in *Proc. ECCV*, in Lecture Notes in Computer Science, vol. 7575, 2012, pp. 1–14.
[27] Y. Han, C. Deng, B. Zhao, and D. Tao, "State-aware anti-drift object tracking," *IEEE Trans. Image Process.*, vol. 28, no. 8, pp. 4075–4086, Aug. 2019.
[28] C. Bao, Y. Wu, H. Ling, and H. Ji, "Real time robust ℓ_1 tracker using accelerated proximal gradient approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Providence, RI, USA, Jun. 2012, pp. 1830–1837.
[29] L. Bertinetto, J. Valmadre, S. Golodetz, O. Miksik, and P. H. S. Torr, "Staple: Complementary learners for real-time tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 1401–1409.
[30] W. Zhong, H. Lu, and M.-H. Yang, "Robust object tracking via sparsity-based collaborative model," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 1838–1845.
[31] K. Zhang, Q. Liu, Y. Wu, and M.-H. Yang, "Robust visual tracking via convolutional networks without training," *IEEE Trans. Image Process.*, vol. 25, no. 4, pp. 1779–1792, Apr. 2016.
[32] T. Zhang, B. Ghanem, S. Liu, and N. Ahuja, "Robust visual tracking via multi-task sparse learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Providence, RI, USA, Jun. 2012, pp. 2042–2049.
[33] Y. Wu, B. Shen, and H. Ling, "Online robust image alignment via iterative convex optimization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Providence, RI, USA, Jun. 2012, pp. 1808–1814.
[34] L. Sevilla-Lara and E. Learned-Miller, "Distribution fields for tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Providence, RI, USA, Jun. 2012, pp. 1910–1917.
[35] Y. Wu, J. Lim, and M.-H. Yang, "Online object tracking: A benchmark," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Portland, OR, USA, Jun. 2013, pp. 2411–2418.

Yunming Wang received the B.E. degree and bachelor's degree in management in 2017. He is pursuing the master's degree with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China.

Under the guidance of Professor Taoyang Wang at Wuhan University, he participated in several research projects. His research interests mainly focus on visual tracking, target detection, and machine learning.



Taoyang Wang received the B.E. and Ph.D.E. degrees in photogrammetry and remote sensing from the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China, in 2007 and 2012, respectively.

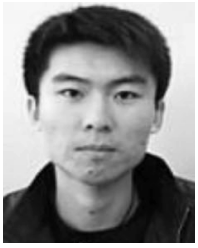
He has been with the School of Remote Sensing and Information Engineering, Wuhan University, since 2014 and where he joined as an Associate Research Fellow in 2015. His doctoral dissertation concerned the block adjustment of high resolution satellite remote sensing imagery. His research interests include space photogrammetry, geometry processing of spaceborne optical/SAR/InSAR imagery, and target detection and recognition based on satellite video.

Guo Zhang received the B.E. and Ph.D.E. degrees in photogrammetry and remote sensing from the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China, in 2000 and 2005, respectively.

He has been with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, since 2005, where he joined as a Professor in 2011. His doctoral dissertation concerned the rectification for high resolution remote sensing image under lack of ground control points. His research interests include space photogrammetry, geometry processing of spaceborne optical/SAR/InSAR imagery, altimetry, and high accuracy image matching.

Qian Cheng received the B.E. degree from the School of Geomatics, Liaoning Technical University, Fuxin, China, in 2017, where he is pursuing the master's degree in photogrammetry and remote sensing with the School of Geomatics.

His major interests include geometry processing of spaceborne imagery and high accuracy image matching.



Jia-qi Wu was born in 1985. He is pursuing the Ph.D. degree with the School of Geomatics, Liaoning Technical University, Fuxin, China.

His research interests include satellite video data processing.