

Real-Time Tracking in Satellite Videos via Joint Discrimination and Pose Estimation

Xu Chen¹, Haigang Sui^{1,*}

¹ The State Key Laboratory of Information Engineering in Surveying Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China –(chenxu_358@163.com, haigang_sui@263.net)

Commission WG II/10 & WG II/5

KEY WORDS: Real-time Tracking, Satellite Videos, Spatial Weight, Kalman Filter

ABSTRACT:

Object tracking has gained much attention in the field of computer vision and intelligent traffic analysis. Satellite videos are more suitable for long-distance tracking comparing to the road traffic videos. However, most of the state-of-the-art methods produce poor results when applied to satellite videos, due to low resolution of the small target and interference from similar background in satellite videos. In this paper, we present an improved Discriminative Correlation Filter based approach specifically tailored for small objects tracking in satellite videos through applying spatial weight in the filter and estimating the pose by Kalman filter. First, a spatial mask is introduced to encourage the final filter to give different contributions depending on the spatial distance. Furthermore, Kalman filter is incorporated into the approach with the aim of predicting the position when the target run into the large area similar background region. Finally, an efficient strategy for combining the improved DCF tracker and pose estimation is proposed. Experimental results on three satellite videos describing the traffic conditions of three cities demonstrate that the proposed approach can effectively track the targets even though the targets are similar to the background region in a period of time. Compare with other state-of-the-art methods, the final accuracies and speeds of our method achieve the best.

*

1. INTRODUCTION

Object tracking is an attractive research field which is useful for many computer vision and intelligent traffic analysis tasks. Road traffic videos are the main data source for target tracking in traffic surveillance. However, it is difficult to achieve long-distance tracking of a moving target by using road traffic videos as a result of local limited range observed by the traffic camera.

A satellite video is a newly developed earth observation satellite, which can capture continuous image sequences in a long time(Xiao, Wang, Wang, & Ren, 2018). It has received attention in many real-world tasks, such as motion analysis(He, Yi, Cheung, You, & Tang, 2017), moving ship surveillance(Li & Man, 2016), traffic monitoring(Caro-Gutierrez, Bravo-Zanoguera, & González-Navarro, 2017),(Patel & Mishra, 2013). In general, the size of a frame in the satellite video can be up to 3840×2160 pixels. Yet the size of moving objects in satellite video usually can be very small because of the long distance from the camera, consequently the information we can obtain is limited comparing with the background. Meanwhile, the target is very similar to the background, even in some region they are completely indistinguishable.

Prior to 2012 tracking approaches based on generative models are widely applied such as meanshift(Vojir, Nuskova, & Matas, 2013), Kalman filter(Chen, 2012), optical flow(Boroujeni, 2012) and particle filter(Baxter, Leach, & Robertson, 2014). The generative model treats the tracking as a searching problem to find a region most similar to the target. These methods mainly use the information of the target itself such as color, position, feature corners, etc., ignoring the background information. They often produce poor results when lighting changes, deformations or occlusions occur.

Recent years tracking solutions based on correlation filters have achieved great success in object tracking. They treat the tracking problem as a binary classification problem. Many works(Henriques, Caseiro, Martins, & Batista, 2012),(Danelljan, Bhat, Shahbaz Khan, & Felsberg, 2017) based on Discriminative Correlation Filter (DCF) improve continuously accuracy and robustness on tracking benchmarks(Wu, Lim, & Yang, 2013),(Wu, Lim, & Yang, 2015). For instance, Color Attribute(Danelljan, Khan, Felsberg, & Van De Weijer, 2014) is introduced to rich the feature, DSST(Danelljan, Häger, Shahbaz Khan, & Felsberg, 2014) is proposed to solve the problem of scale variations, Spatial Regularization(Danelljan, Hager, Khan, & Felsberg, 2015) is used to reduce the boundary effects. However, most of these methods fail to track the targets effectively when applied to the satellite video, as the background information with target-similar color in the searching box hampers accurate location of the target.

Recently deep learning shows its power in many computer vision tasks. The CNN is introduced to extract features for tracking(Nam & Han, 2016), (Bertinetto, Valmadre, Henriques, Vedaldi, & Torr, 2016), (Danelljan, Hager, Khan, & Felsberg, 2016), (Valmadre, Bertinetto, Henriques, Vedaldi, & Torr, 2017). These trackers pre-train the CNN model to get deep features useful for distinguishing the object categories. Nevertheless, there are two obstacles when integrating deep features into tracking models in satellite videos. Firstly, the training data of satellite videos are extremely scarce. It is very difficult to find the labeled satellite videos. The second issue is low resolution of the target in the satellite video. The targets in the satellite videos such as cars and planes occupy only a few pixels and are shape-blur and color-similar. As a result, effective deep features are very hard to obtain through most of the deep models.

* corresponding author : Haigang Sui

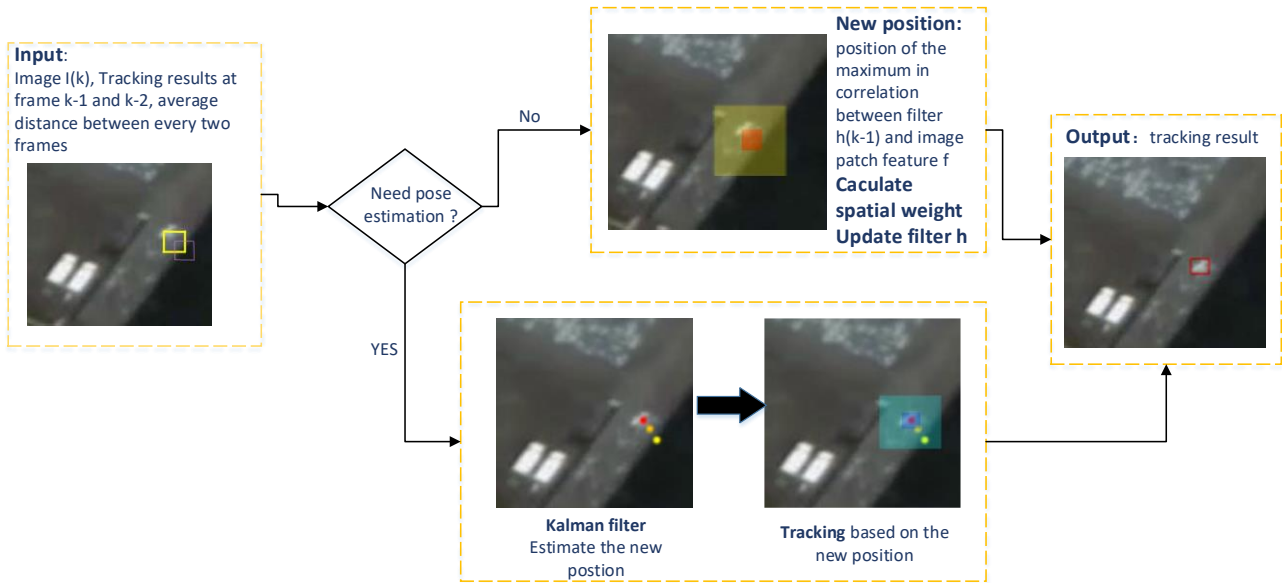


Figure 1 Flowchart of the proposed approach. We introduce the spatial weight to restrict the DCF filter and do the pose estimation through Kalman filter when the tracking result stay at one place.

Even though object tracking has attracted much attention, the tracking of small object is left in the basket. Some moving object detection methods are proposed to fuse with the Kernel Correlation Filter (KCF) to solve the problem (Du, Sun, Cai, Wu, & Du, 2018). High computation of object detection makes the approach operate at low frame rates, thereby infeasible for real-time applications.

In this paper, we present an improved DCF-based method restricted with the spatial weight, at the same time, Kalman filter (Chen, 2012) is combined to estimate the pose. The traditional DCF filter has multi-peak response region when applied to satellite video, primarily due to the similar color and no obvious shape feature. To solve this, we introduce the spatial weight calculated based on the distance from the center to ensure the target region in the previous frame is given higher weights. Most of the state-of-the-art tracking solutions lost the target in the case that the target is obscured for a period of time. The large buildings with similar color of the target often make the target invisible. We estimate the position of the target based on the motion model established by the Kalman filter in the above situation. An effective strategy is employed to decide whether to discriminate or to estimate. Finally, we conduct experiments on three satellite sequences to demonstrate the effectiveness of our approach. An overview flowchart of the proposed approach is shown in Figure 1.

2. METHODOLOGY

2.1 The Single Channel Discriminative Correlation Filters

We base our method on the DCF tracker, which is the basis of many trackers providing good performance in a recent evaluation. Single channel is used in the correlation filters, seeing that targets in satellite video have a low resolution and their colors are similar to the backgrounds. Here we provide a brief introduction of the single channel discriminative correlation filters.

The DCF tracker determines the target position by find a function that minimizes the squared error over the correlation response and their desired output as solving a ridge regression problem,

$$\varepsilon(h) = \frac{1}{2} \| f * h - y \|^2 + \frac{\lambda}{2} \| h \|^2 \quad (1)$$

where f is the grayscale feature of samples and h is the corresponding target templates, y is a 2D Gaussian function centered at the target location, λ is a regularization term.

The lost function (1) can be expressed as a Hadamard product in the Fourier domain to improve tracking efficiency,

$$\varepsilon(h) = \frac{1}{2} \| \text{diag}(\hat{f})\hat{h} - \hat{y} \|^2 + \frac{\lambda}{2} \| \hat{h} \|^2 \quad (2)$$

where \hat{f} , \hat{h} , \hat{y} are the Fourier transforms of f , h , y and \hat{h} is the complex-conjugate operator of \hat{h} . The solution to the lost function (2) is

$$\hat{h} = (\text{diag}(\hat{f})\hat{y}) \odot^{-1} (\text{diag}(\hat{f})\hat{f} + \lambda) \quad (3)$$

where \odot^{-1} is element-wise division. The new position of the target is thought to be maximum in the final response.

2.2 Spatial Weight Mask

The color of targets are similar to the background in satellite video, consequently we get a multi-peak response in the filter. To alleviate the problem induced by the similar background, we introduce a spatial weight mask m . The element in the mask is between one and zero depending on the location in the sample. Then we get a constraint function $h = m \odot h$, where \odot represents element-wise product. After adding the constraint, the solution to the lost function is not closed-form. An iterative method is proposed to solve the constraint problem in Correlation Filter with limited boundaries (Italiano, Sim, & Lucey, 2015).

We introduce a variable $h_s = m \odot h$ and handle the closed-form problem through an Augmented Lagrangian Method (Boyd, Parikh, Chu, Peleato, & Eckstein, 2011). The augmented Lagrangian of the constraint loss function can be formed as

$$\zeta(\widehat{h}_s, h, \widehat{I} | m) = \frac{1}{2} \| \text{diag}(\widehat{f}) \widehat{h}_s - \widehat{y} \|^2 + \frac{\lambda}{2} \| h_s \|^2 + \left[\widehat{I}^H (\widehat{h}_s - \widehat{h}_m) + \overline{\widehat{I}^H (\widehat{h}_s - \widehat{h}_m)} \right] + \mu \| \widehat{h}_s - \widehat{h}_m \|^2 \quad (4)$$

where \widehat{I} is a complex Lagrange multiplier, \widehat{I}^H is the hermite operator of \widehat{I} and $\mu > 0$, $\lambda = \frac{\mu}{100.0}$. A Gauss-Seidel optimization strategy can be applied to the augmented Lagrangian objective. At each iteration:

$$\widehat{h}_s^{i+1} = \frac{\arg \min}{h_s} \zeta(\widehat{h}_s, h^i, \widehat{I}^i | m) \quad (5)$$

$$h^{i+1} = \frac{\arg \min}{h} \zeta(\widehat{h}_s^{i+1}, h, \widehat{I}^i | m) \quad (6)$$

$$\widehat{I}^{i+1} = \widehat{I}^i + \mu (\widehat{h}_s^{i+1} - \widehat{h}_m) \quad (7)$$

The minimizations to the solution can be computed,

$$\widehat{h}_s^{i+1} = (\widehat{f} \odot \widehat{y} + (\mu \widehat{h}_s^i - \widehat{I}^i)) \odot^{-1} (\widehat{f} \odot \widehat{f} + \mu^i) \quad (8)$$

$$h^{i+1} = m \odot \mathcal{F}^{-1} \left[\widehat{I}^i + \mu^i \widehat{h}_s^{i+1} \right] / \left(\frac{\lambda}{2D} + \mu^i \right) \quad (9)$$

And the μ value is updated through $\mu^{i+1} = \beta \mu^i$.

The spatial weight mask is target location centered. If the pixel is within the original width and height, the value of the pixel is set to one, otherwise it is set to zero. We also try to decrease the weight value progressively based on the distance from the center. Results indicate that the spatial weight mask having weights within the original box of the target set to one performs best.

2.3 Pose Prediction

Aiming at the problem that targets are hard to distinguish from the color-similar background region, Kalman filtering feedback mechanism is adopted to estimate the target position. Kalman filter can predict the state sequence of a dynamic system having the minimum mean square error. The state equation and observation equation is the base of Kalman filter, the estimated value and the observation value are used to update the state variable so as to predict the movement state.

The state equation and the measurement equation are

$$X(k) = AX(k-1) + BU(k) + W(k) \quad (10)$$

$$Z(k) = HX(k) + V(k) \quad (11)$$

Where $X(k)$, $X(k-1)$ are the system states at times k and $k-1$, $U(k)$ is the control amount at time k . $Z(k)$ is the observation value at time k . $W(k)$ and $V(k)$ are the noises. A, B and H are the parameter of the system.

The whole process of Kalman filter is a recursive calculation process, it continuously predict and update. The predict function are

$$X(k|k-1) = AX(k-1|k-1) + BU(k), \quad (12)$$

$$P(k|k-1) = AP(k-1|k-1)A' + Q. \quad (13)$$

The update function are

$$Kg(k) = P(k|k-1)H' / (HP(k|k-1)H' + R), \quad (14)$$

$$X(k|k) = X(k|k-1) + Kg(k)(Z(k) - HX(k|k-1)), \quad (15)$$

$$P(k|k) = (I - Kg(k)H)P(k|k-1). \quad (16)$$

We use the tracking results as the measurement to regress the Kalman filter, the predicted value is used as target location when targets and the surrounding region are exactly the same.

2.4 Tracking with Spatial Weight and Pose Prediction

We introduce a strategy for combining the improved DCF tracker and pose estimation in this section. When targets reach the region having highly similarity, the tracking results of DCF with spatial weight stay at the old positions. So the average distance between two frames are selected as a standard to decide the ratio of the position estimated by Kalman filter.

The accuracy of the Kalman filter is great improved after regression of some frames. Therefore we do the pose estimation in every Nth frame after Ns frame. The parameter Ni determines how often pose estimation is done. The parameter R_k is adopted to calculate the new possible position,

$$P_t' = R_k \bullet P_k + (1.0 - R_k) \bullet P_{t-1} \quad (17)$$

where P_k is the position estimated by Kalman filter, P_{t-1} is the position tracked at frame t-1. If distance between position at frame t-1 and position at frame t-2 less than the average distance between two frames, R_k is set to a high value, otherwise R_k is set to a small value. R_k must be at least 0 but not more than 1.

Instead of using P_t' as the position at frame t, we track again centered on this position. The tracking result is the new position of the target. The new position is used as the measure value to update the Kalman filter.

Algorithm 1 The proposed tracking algorithm

Input: Image It, object position on previous frame P_{t-1} and P_{t-2} , average distance between two frames, filter h_{t-1}

Method:

- 1: Set the parameter R_k according to the distance between the position of frame P_{t-1} and the position P_{t-2} . Update the average distance.
 - 2: Calculate the new possible position P_t' using P_{t-1} and the position P_k estimated by Kalman filter.
 - 3: Extract image patch feature f centered on position P_{t-1} .
 - 4: New target position P_t : position of the maximum in correlation between h_{t-1} and image patch feature f .
 - 5: Calculate the spatial weight mask m , estimate a new filter h_t using m .
 - 6: Update filter $h_t = (1 - \eta)h_{t-1} + \eta h_t'$.
 - 7: Update Kalman filter using P_t as the measurement value.
-

Output: tracking location P_t at frame t

3. EXPERIMENT

We validate our proposed approach by performing experiments on three sequences. These sequences describing the traffic conditions of three cities are provided by Chang Guang Satellite Technology Co. Ltd. The image sizes are 3840×2160 pixels. Challenging problems such as low resolution, shape blur, background clutter and occlusion are posed in these data sets. Details of the data sets are shown in

Figure 2. We select planes and cars as our targets.



Figure 2 Three data sets in the experiment. Traffic scenes in the USA, Mexico and Spain are shown respectively. The planes are selected as the targets in A and B, while the car is chosen as the target in C.

3.1 Implementation

Raw Grayscale feature is used in the correlation filter, since moving objects in satellite videos have no obvious shape features and their colors are similar to the background. The image region area of the samples was set to 42 times the target area. We multiply samples by a Hann window (Bolme, Beveridge, Draper, & Lui, 2010). We set the augmented Lagrangian optimization parameters μ to 5 and β to 3. We do 4 times Gauss-Seidel iterations and the filter fusing rate is set to $\eta = 0.02$. We do the pose estimation in every three frames after 25 frame. R_k is set to 0.95 when the distance between the last two frames less than the average distance between every two frames, else R_k is set to 0.2. All parameters do not need to be fine-tuned and were consistent across all experiments. The proposed approach is implemented in a C++ OpenCV library. Our implementation runs at 1500 frame per second on an Intel Core i7-6700HQ 2.60GHz CPU with 8GB RAM.

3.2 State-of-the-art Comparison

We compare our method with CSRT (Lukežič, Vojtř, Zajc, Matas, & Kristan, 2017), ECO-HC (Danelljan et al., 2017), STAPLE (Bertinetto, Valmadre, Golodetz, Miksik, & Torr, 2015), LCT (Ma, Yang, Zhang, & Yang, 2015) and KCF (João F. Henriques, Caseiro Rui, Martins Pedro, & Batista Horge, 2015), these approaches provide excellent results in literature. ECO-HC, STAPLE and LCT are provided with the benchmark evaluation. CSRT and KCF are provided with OpenCV library.

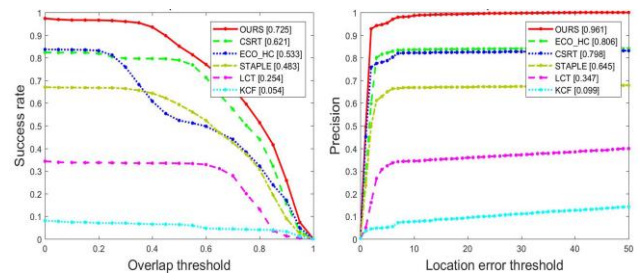


Figure 3 Success rate and Precision Plots. The legend shows the AUC score of each tracker.

Figure 3 shows a comparison with the mentioned methods on the three sequences using overlap rate and distance precision using the method provided with the benchmark. Among the compared trackers our method achieved the highest success rate with the AUC score of 0.725, while CSRT, ECO-HC, STAPLE, LCT, KCF get the AUC scores of 0.621, 0.533, 0.483, 0.254, 0.054. In the Precision plot, the best approach is OURS proposed in this letter. One of the most important reason leading to the good performance is that we do the pose estimation when targets disappear in the background, especially targets have the same color with the large buildings in the background. The sequence of the USA has a short video clips that targets have the same color with the surrounding background region, the other methods except ours fail to track the target effectively in such a situation.

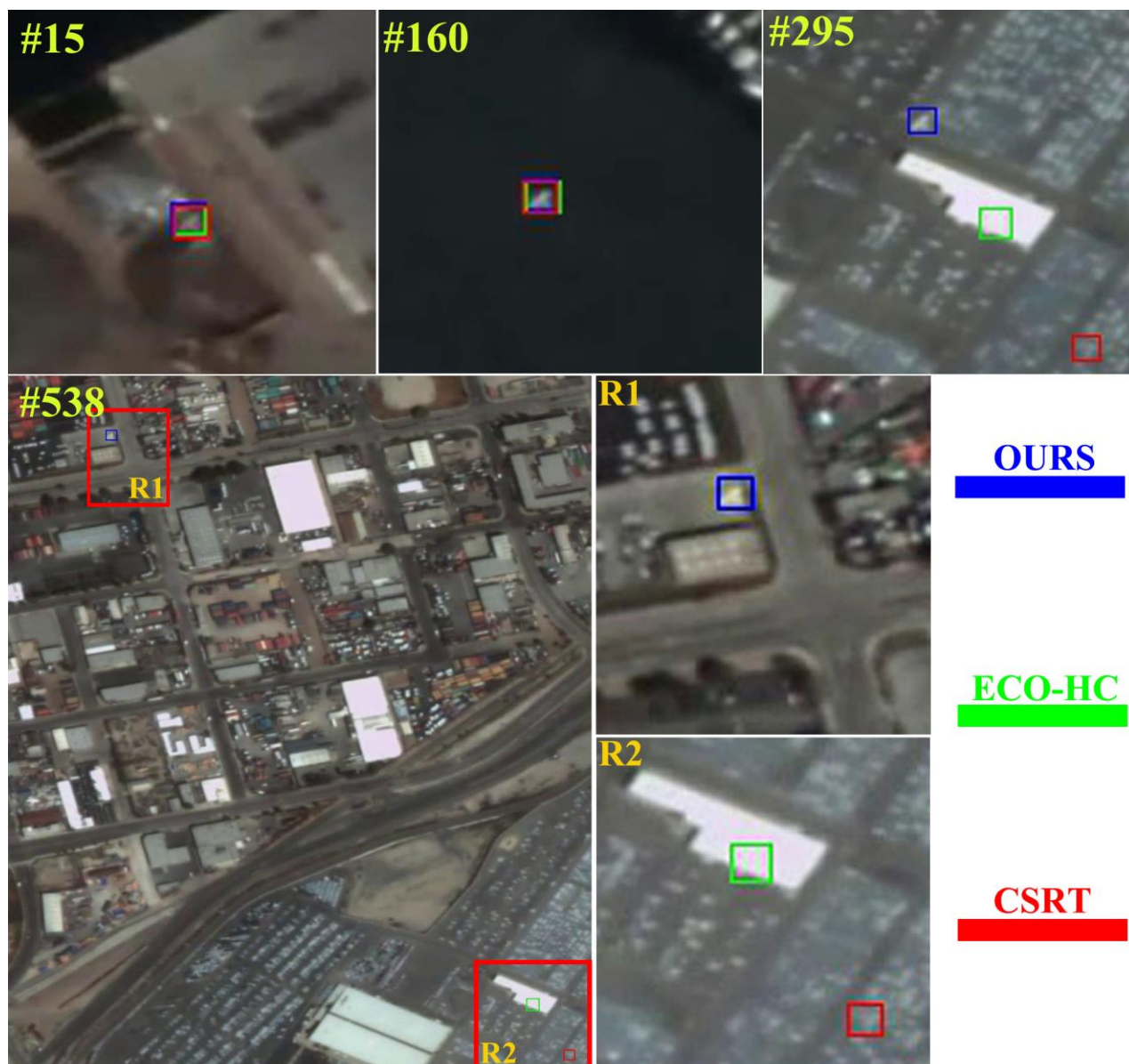


Figure 4 Example tracking results of different methods on the USA sequence. The number in the top left corner shows the frame number in the video. Our tracker is the only tracker tracking the target correctly from the beginning to the end.

Figure 4. ECO-HC and CSRT are selected besides ours method as they get better performance comparing to STAPLE, LCT and KCF. In the USA sequence, ours can track the target accurately until the end of the video. We can see the ECO-HC tracking box staying in the similar background region obviously at the 295 frame, in addition CSRT has lost the target before then.

Another important criterion for tracking targets in satellite video is tracking speed. Table I shows the speed comparison among the six approaches. OURS has a remarkable speed advantage. The reasons why the fps of the proposed method is so high are as follows. The targets in satellite video are almost scale invariant as a result of the long distance from the camera, so we do not estimate scale variations. Moreover, only gray feature is employed, considering that targets occupy only a small number of pixels without evident shape features and the color is similar to the background. Contrary to the usual experience, small

We draw the tracking results of some frames belonging to the experiment data sets in objects tracking in satellite videos do not tend to benefit from HOG and CN feature.

4. CONCLUSIONS

In this paper, we propose a novel approach with spatial weights and pose estimation for tracking small objects in satellite videos. It is hard to extract strong features from the small targets in the satellite videos, on the other hand, large area color-similar background is a major obstacle. By exploiting the spatial weight mask, we can get a single-peak response filter. Kalman filter for pose estimation is introduced to avoid interference from the large area color-similar region. We further propose a strategy to combine the discrimination and pose estimation based on the average distance between frames. Experiments on three satellite videos show the tracker can produce superior tracking performance in terms of both accuracy and speed.

Table I AVERAGE SPEED(IN FPS UNITS)

	OURS	CSRT	ECO-HC	STAPLE	LCT	KCF
USA	1300.61	12.0144	49.0603	81.3512	56.38	49.9394
Mexico	614.287	10.7943	45.4434	72.7624	13.5	50.6377
Spain	2586.69	12.9614	50.7821	77.613	61.41	49.7636
Average	1500.529	11.9234	48.4286	77.2422	43.7633	50.1136

ACKNOWLEDGEMENTS (OPTIONAL)

This work was supported by the National Key Research and Development Program of China (No. 2016YFB0502600), and the National Natural Science Foundation of China under Grant Nos. 41771457.

REFERENCES

- Baxter, R., Leach, M., & Robertson, N. ,2014. Tracking with Intent. *2014 Sensor Signal Processing for Defence, SSPD 2014*.
- Bertinetto, L., Valmadre, J., Golodetz, S., Miksik, O., & Torr, P. ,2015. Staple: Complementary Learners for Real-Time Tracking. *ArXiv:1512.01355 [Cs.CV]*. <https://doi.org/10.1109/CVPR.2016.156>
- Bertinetto, L., Valmadre, J., Henriques, J. F., Vedaldi, A., & Torr, P. H. S. ,2016. Fully-convolutional siamese networks for object tracking. *Computer Vision – ECCV 2016 Workshops, 9914 LNCS*, 850–865. https://doi.org/10.1007/978-3-319-48881-3_56
- Bolme, D. S., Beveridge, J. R., Draper, B. a., & Lui, Y. M. L. Y. M. ,2010. Visual object tracking using adaptive correlation filters. *Computer Vision and Pattern Recognition CVPR 2010 IEEE Conference On*, 2544–2550. <https://doi.org/10.1109/CVPR.2010.5539960>
- Boroujeni, N. S. ,2012. Fast obstacle detection using targeted optical flow. ... (*ICIP*), *2012 19th IEEE ...*, 2(3), 65–68. <https://doi.org/10.1109/ICIP.2012.6466796>
- Boyd, S., Parikh, N., Chu, E., Peleato, B., & Eckstein, J. ,2011. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Found. Trends Mach. Learn.*, 3(1), 1–122. <https://doi.org/10.1561/22000000016>
- Caro-Gutierrez, J., Bravo-Zanoguera, M. E., & González-Navarro, F. F. ,2017. Methodology for automatic collection of vehicle traffic data by object tracking. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10061 LNAI, 482–493. https://doi.org/10.1007/978-3-319-62434-1_39
- Chen, S. Y. ,2012. Kalman filter for robot vision: A survey. *IEEE Transactions on Industrial Electronics*, 59(11), 4409–4420. <https://doi.org/10.1109/TIE.2011.2162714>
- Danelljan, M., Bhat, G., Shahbaz Khan, F., & Felsberg, M. ,2017. ECO: Efficient convolution operators for tracking. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017-Janua*, 6931–6939. <https://doi.org/10.1109/CVPR.2017.733>
- Danelljan, M., Hager, G., Khan, F. S., & Felsberg, M. ,2015. Learning spatially regularized correlation filters for visual tracking. *Proceedings of the IEEE International Conference on Computer Vision, 2015 Inter*, 4310–4318. <https://doi.org/10.1109/ICCV.2015.490>
- Danelljan, M., Hager, G., Khan, F. S., & Felsberg, M. ,2016. Convolutional Features for Correlation Filter Based Visual Tracking. *Proceedings of the IEEE International Conference on Computer Vision, 2016-Febru*, 621–629. <https://doi.org/10.1109/ICCVW.2015.84>
- Danelljan, M., Häger, G., Shahbaz Khan, F., & Felsberg, M. ,2014. Accurate Scale Estimation for Robust Visual Tracking. *Proceedings of the British Machine Vision Conference 2014*, (January), 65.1-65.11. <https://doi.org/10.5244/C.28.65>
- Danelljan, M., Khan, F. S., Felsberg, M., & Van De Weijer, J. ,2014. Adaptive color attributes for real-time visual tracking. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1090–1097. <https://doi.org/10.1109/CVPR.2014.143>
- Du, B., Sun, Y., Cai, S., Wu, C., & Du, Q. ,2018. Object Tracking in Satellite Videos by Fusing the Kernel Correlation Filter and the Three-Frame-Difference Algorithm. *IEEE Geoscience and Remote Sensing Letters*, 15(2), 165–172. <https://doi.org/10.1109/LGRS.2017.2776899>
- He, Z., Yi, S., Cheung, Y. M., You, X., & Tang, Y. Y. ,2017. Robust Object Tracking via Key Patch Sparse Representation. *IEEE Transactions on Cybernetics*, 47(2), 354–364. <https://doi.org/10.1109/TCYB.2016.2514714>
- Henriques, J. F., Caseiro, R., Martins, P., & Batista, J. ,2012. Exploiting the Circulant Structure of Tracking-by-Detection with Kernels. In A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, & C. Schmid (Eds.), *Computer Vision -- ECCV 2012* (pp. 702–715). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Italiano, I., Sim, T., & Lucey, S. ,2015. Correlation Filters with Limited Boundaries. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4630–4638.
- João F. Henriques, Caseiro Rui, Martins Pedro, & Batista Horge. ,2015. Kernelized Correlation Filters. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, 37(3), 583–596. <https://doi.org/10.1109/TPAMI.2014.2345390>
- Li, H., & Man, Y. ,2016. Moving Ship Detection Based on Visual Saliency for Video Satellite. *Geoscience & Remote Sensing Symposium*, 1248–1250. <https://doi.org/10.1109/IGARSS.2016.7729316>
- Lukežič, A., Vojjiř, T., Zajc, L. Č., Matas, J., & Kristan, M. ,2017. Discriminative correlation filter with channel and spatial reliability. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017-Janua*, 4847–4856. <https://doi.org/10.1109/CVPR.2017.515>
- Ma, C., Yang, X., Zhang, C., & Yang, M. H. ,2015. Long-term correlation tracking. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 07-12-June*, 5388–5396. <https://doi.org/10.1109/CVPR.2015.7299177>
- Nam, H., & Han, B. ,2016. Learning Multi-Domain Convolutional Neural Networks for Visual Tracking. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4293–4302. <https://doi.org/10.1109/CVPR.2016.465>
- Patel, S., & Mishra, A. ,2013. Moving Object Tracking Techniques: A Critical Review. *Indian Journal of Computer Science & Engineering*, 4(2), 95–102. Retrieved from <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Moving Object Tracking Techniques : A Critical Review#0>

Valmadre, J., Bertinetto, L., Henriques, J. F., Vedaldi, A., & Torr, P. H. S. .2017. End-to-end representation learning for Correlation Filter based tracking. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, Hawaii, USA*, 5000–5008. <https://doi.org/10.1109/CVPR.2017.531>

Vojir, T., Noskova, J., & Matas, J. .2013. Robust Scale-Adaptive Mean-Shift for Tracking. In J.-K. Kämäräinen & M. Koskela (Eds.), *Image Analysis* (pp. 652–663). Berlin, Heidelberg: Springer Berlin Heidelberg.

Wu, Y., Lim, J., & Yang, M.-H. .2013. 1- Object Tracking -Online Object Tracking: A Benchmark. *2013 IEEE Conference on Computer Vision and Pattern Recognition*, 2411–2418. <https://doi.org/10.1109/CVPR.2013.312>

Wu, Y., Lim, J., & Yang, M. H. .2015. Object tracking benchmark. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(9), 1834–1848. <https://doi.org/10.1109/TPAMI.2014.2388226>

Xiao, A., Wang, Z., Wang, L., & Ren, Y. .2018. Super-resolution for "Jilin-1" satellite video imagery via a convolutional network. *Sensors (Switzerland)*, 18(4), 1–10. <https://doi.org/10.3390/s18041194>