Infrared object tracking method based on kernel correlation filters

Ying Meng, Chao Ma and Wei An*

1 School of Electronic Science, National University of Defense Technology, Changsha, Hunan, 410003, China
*Corresponding author’s e-mail: machao0408@nudt.edu.cn

Abstract. Benefiting from the superiority visual capacity of infrared imaging system, infrared object tracking is widely applied in military, weapon guidance, security protection and other fields. Recently, correlation filter has attracted considerable attention within the tracking community due to its high computational efficiency. However, traditional correlation filter based trackers use the single feature, which has poor stability in low resolution and complex background. Aiming at the low resolution, poor contrast, low signal-to-noise and insufficient texture information of infrared image, we propose an Infrared object tracking method based on Kernel Correlation Filters, named Inf-KCF. We design a Multi-Feature Fusion structure to improve the expression ability for infrared targets, and separate the target from the background well. Experimental results on public infrared dataset show that, the Inf-KCF achieves promising performance.

1. Introduction

Infrared object tracking is a challenging research topic in computer vision, and is widely applied in various fields, such as military, weapon guidance and security protection. Infrared object tracking algorithms can be divided into two groups: the generative model based trackers and discriminative trackers, according to observation model.

Generative models take the region that best matches the target features as the tracking results, which do not take advantage of the background information, such as Kalman Filter [11], Particle Filter [12] and Mean-Shift [13]. The generative model based trackers could be easily distracted by background regions with similar appearances during tracking. Discriminant methods regard tracking as a binary classification, which perform tracking by separating the target from the background, such as TLD [14] and Struck [15]. The discriminative model based trackers are more stable than the former one.

Recently, correlation filter has attracted considerable attention within the tracking community due to its high computational efficiency. The core idea of these algorithms is to regard object tracking as a process of using the initial frame image to train the correlation filter, and performing correlation filtering on the search area in subsequent frames. The Minimum Output Sum of Squared Error Filter (MOSSE) tracker proposed by Bolme et al. [1] is the seminal work of trackers based on correlation filter, which runs in hundreds of frames per second. But it only uses gray feature to represent the target, which performs poor in the case of background clutters. After that, lots of improved tracking algorithms based on it were proposed. Henriques et al. [2] proposed a Circulant Structure tracking by detection with kernels (CSK), which introduced a kernel function to map linear inseparable input data into a high dimensional linear separable feature space. Cyclic shift is used for dense sampling to make full use of image features. But CSK uses only gray feature, and it cannot deal with the scale variation because of...
the fixed target size. On this foundation, Kernelized Correlation Filters (KCF) \cite{3} tracker efficiently incorporated multi-channel features HOG \cite{8} (Histogram of Oriented Gradient), which can achieve more accurate target description with less computation. Several variations of KCF tracker have been subsequently investigated to improve tracking performance. For example, DSST \cite{7} can adapt to the target scale changes well by building a 3-dimensional filter.

Above trackers are lightweight and less computing burden, and they have advantages both in accuracy and speed. However, they only use a single hand-crafted feature to represent the target appearance, such as HOG \cite{8}, SIFT (Scale Invariant Feature Transform) \cite{9} and CN (Color Naming) \cite{10} features. The representation of hand-crafted features is not effective and robust. For infrared targets with insufficient shape and texture information, it is difficult to get strong discriminative ability by single hand-crafted feature and cannot achieve ideal results.

In this paper, aiming at the problems of low resolution, poor contrast, low signal-to-noise and insufficient texture information of infrared target, we propose an Infrared object tracking method based on Kernel Correlation Filters, named Inf-KCF. To improve the expression ability for infrared targets, we firstly analyse the characteristics of FHOG and gray feature, and then design a Multi-Feature Fusion structure to make full use of them. Finally, the experimental results on infrared object tracking datasets show that Inf-KCF achieves better performance compared with correlation filter based trackers.

2. Methods

2.1. Kernel Correlation Filter Method

KCF uses the circulant matrix around the target to collect positive and negative samples, and uses ridge regression to train the target detector. And successfully uses the diagonalization property of circulant matrix in Fourier space to transform the matrix operation into Hadamard product of vector, which greatly reduces the amount of calculation and makes the tracker meet the real-time requirements. The linear ridge regression model constructed by KCF using sample set \{\text{x}_1, \text{x}_2, \text{x}_3, ..., \text{x}_n\} is as follows.

\[
\text{loss} = \min_w \sum_i (f(x_i) - y_i) + \lambda ||w||^2
\]

Where \(y_i\) denotes the label of \(x_i\) sample, \(f\) denotes the filter, \(w\) denotes the parameters of filter, \(\lambda\) denotes the regularization coefficient. And the results can be expressed as,

\[
W = (X^TX + \lambda I)^{-1}X^Ty
\]

Where \(X\) denotes the cyclic matrix of samples, \(I\) denotes identity matrix. In order to improve the discriminative ability of tracker, the sample set is mapped to high-dimensional space by nonlinear function, and a kernel function is introduced to obtain the final ridge regression results as follows.

\[
\alpha = (K + \lambda I)^{-1}y
\]

Where \(K\) denotes the kernel correlation matrix. With the help of the properties of circulant matrix, the solution in frequency domain is expressed as follows, where \(\Lambda\) denotes Fourier-transform, and \(\hat{k}_{xx}\) denotes the autocorrelation of sample \(x\) in Fourier domain.

\[
\hat{\alpha} = \frac{\hat{y}}{\hat{k}_{xx} + \lambda}
\]

2.2. FHOG feature

Gray feature is the normalization of pixels, we won’t introduce it in detail. FHOG is an improved feature descriptor based on HOG. The main process of extracting FHOG feature is shown in Fig.2. Firstly, we divide the image into small connected regions named cell unit, then collect the direction histogram of the gradient of each pixel in the cell unit. Several cell units are combined into blocks, and we normalize
the histogram of the gradient in blocks. Finally, these histograms are combined to form the feature descriptor.

FHOG is a local feature descriptor, which is used to describe the shape edge contour of target. It can deal with small morphological changes, translation and rotation of target well. However, in complex environment, the gradient feature of target region is not obvious, and the description ability of FHOG is weak, which easily cause failure in tracking.

2.3. Multi-Feature fusion

Feature extraction is a vital factor for tracking performance. Aiming at the problem that single feature extracted by KCF tracker cannot express the target well, we propose a method to fuse multi-feature from the feature level and decision level respectively to improve the tracking performance.

In this paper, we fuse FHOG features and gray features at the feature level. For the image of size $M \times N$, the fusion method is as follows: Firstly, we add Hamming window to the image, and then extract the HOG feature of dimension $M/4 \times N/4 \times 31$ and gray feature of dimension $M \times N$ respectively. In order to maintain the consistency of dimensions, we transformer the gray feature to $M/4 \times N/4$. Finally, we concatenate HOG feature and gray feature to get $M/4 \times N/4 \times 32$ fusion feature.

3. Results & Discussion

3.1. Dataset

We evaluate our tracker on the dataset for infrared detection and tracking of dim-small aircraft targets underground / air background proposed by Hui et al [16], which consists of 22 sequences, 30 trajectory, 16177 frames and 16944 targets, and cover sky, ground and other scenes. The image resolution is $256 \times 256$, as shown in Figure 1.

![Figure 1. dataset samples.](image)

3.2. Evaluation criteria

In this paper, we adopt two evaluation criteria: precision plot and success plot. The precision plot shows the ratio of frames whose average center location error (CLE) is less than a given threshold when the threshold varies from 0 to 50 pixels. CLE is the Euclidean distance between the center of output bounding box and the center of ground truth, which is expressed by the following formula.

$$S_{err} = \sqrt{(x_d - x_t)^2 + (y_d - y_t)^2}$$  \hspace{1cm} (5)

where $x_d$ and $y_d$ denote the abscissa and ordinate of the output bounding box center predicted by tracker, $x_t$ and $y_t$ denote the abscissa and ordinate of ground truth center.

Success plot is the percentage of successful frames, where success means IoU (intersection over union) between the tracking results $R^{gt}$ and ground truth $R^{tr}$ is greater than given threshold.
\[ \text{IoU} = \frac{|R^t \cap R^r|}{|R^t \cup R^r|} \]  

(6)

3.3. Comparison with the state-of-the-art trackers

We compare our tracker with several trackers, including MOSSE [1], KCF [2], Staple [3], STRCF [4], BACF [5], DAT [6]. These trackers are ranked by the area under the success rate curve (AUC) and precision. Fig.2 shows that our tracker achieves 53.7% AUC and 56.8% precision, which is 6.8% AUC and 10.2% precision higher than KCF tracker.

![Success plots and Precision plots—comparison with state-of-the-art trackers.](image)

3.4. Contrast analysis of features

In order to verify the effectiveness of the proposed Inf-KCF, we discuss the influence of different features in this section. Table 1 shows the performance comparison of different features in public infrared object tracking dataset. It demonstrates that the proposed structure improves the tracking performance in terms of precision and AUC effectively.

| Feature | HOG  | Gray | HOG+Gray |
|---------|------|------|----------|
| Precision | 0.466 | 0.551 | 0.568 |
| AUC      | 0.469 | 0.527 | 0.537 |

4. Conclusions

In this paper, we propose an Infrared object tracking method based on Kernel Correlation Filters, named Inf-KCF. By analyzing the characteristics of HOG and gray feature, we design a Multi-Feature Fusion structure to improve the expression ability of the model for infrared targets. Experimental results on public infrared dataset show that, the Inf-KCF achieves promising performance and separate the target from the background well than other correlation filter based trackers. In the follow-up work, the stability and accuracy of the algorithm will be further improved. Performance comparison of different features.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 62002372.
References

[1] Bolme D S, Beveridge J R, Draper B A, et al.: Visual Object Tracking using Adaptive Correlation Filters[M], 2010 IEEE Conference on Computer Vision and Pattern Recognition, 2010: 2544-2550.

[2] Henriques J F, Caseiro R, Martins P, et al. High-Speed Tracking with Kernelized Correlation Filters[J]. 2014.

[3] Bertinetto L, Valmadre J, Golodetz S, et al. Staple: Complementary Learners for Real-Time Tracking[C]// Computer Vision & Pattern Recognition. IEEE, 2016.

[4] Li F, Tian C, Zuo W, et al. Learning Spatial-Temporal Regularized Correlation Filters for Visual Tracking[J]. IEEE, 2018.

[5] Galoogahi H K, Fagg A, Lucey S. Learning Background-Aware Correlation Filters for Visual Tracking[J]. 2017 IEEE International Conference on Computer Vision (ICCV), 2017.

[6] Shi P. Deep Attentive Tracking via Reciprocative Learning[J]. 2018.

[7] Danelljan M, Hager G, Khan F S, et al. Discriminative Scale Space Tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, PP(99):1-1.

[8] N Dalal, B Triggs. Histograms of Oriented Gradients for Human Detection [J]. IEEE Computer Society Conference on Computer Vision & Pattern Recognition, 2005, 1(12): 886-893.

[9] Lowe D G. Distinctive Image Features from Scale-Invariant Keypoints [J]. International Journal of Computer Vision, 2004, 60(2):91-110.

[10] Danelljan M, Hager G, Khan F S, et al. Discriminative Scale Space Tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, PP(99):1-1.

[11] Basar T. A New Approach to Linear Filtering and Prediction Problems[M]. Wiley-IEEE Press, 2009.

[12] Okuma K, Taleghani A, Freitas D N, et al. A Boosted Particle Filter: Multitarget Detection and Tracking[J]. 8th European Conference on Computer Vision (ECCV) ; Prague, Czech, May, 2004.

[13] D Comaniciu, Meer P. Mean shift: a robust approach toward feature space analysis[J]. IEEE Trans Pattern Analysis & Machine Intelligence, 2002, 24(5):603-619.

[14] Kalal Z, Mikolajczyk K, Matas J. Tracking-learning-detection[J]. IEEE Transactions on Software Engineering, 2011, 37(7):1409-1422.

[15] Hare S, Saffari A, Torr P. Struck: Structured output tracking with kernels[C]// IEEE International Conference on Computer Vision, ICCV 2011, Barcelona, Spain, November 6-13, 2011. IEEE, 2011.

[16] Bingwei Hui, Zhiyong Song, Hongqi Fan, et al. Target detection and tracking dataset of infrared image in ground / air background [J]. Chinese scientific data: Chinese English network edition, 5 (3): 12. (Chinese)