Research on Target Tracking Algorithm Based on Correlation Filtering

Xifeng Guo, Askar Hamdulla and Turdi Tohti

The School of Information Science and Engineering, Xinjiang University, Urumqi 830046, China

email: 1370852737@qq.com  askar@xju.edu.cn  turdy@xju.edu.cn

*Corresponding author’s e-mail: askar@xju.edu.cn

Abstract. Target tracking has always been an important research direction in the field of computer vision. The target tracking method based on correlation filtering has become a research hotspot in the field of target tracking due to its efficiency and robustness. In recent years, a series of new progress has been made in this field, and it has been widely used in automatic driving, video surveillance, human-computer interaction, national defense security and other fields. Mainly summarize the research status in this field. First of all, the main components and research difficulties of current target tracking are introduced. Secondly, the algorithm of target tracking based on correlation filtering is introduced, and the development of the classical algorithm model and its existing problems are analyzed in detail. Finally, some views on the future research direction and development trend of related filtering algorithms are given.

1. Introduction

With the continuous development and progress of modern science and technology, the subject of artificial intelligence has entered a new era, and the use of artificial intelligence technology to replace traditional artificial labor has become a development trend. Computer vision contains many technical researches for different application scenarios. Target tracking algorithm is one of the research directions with practical significance, which is a very challenging problem. With the increase of high-performance computers and the popularization of cameras, the increasing demand for automatic video analysis has aroused people's keen interest in target tracking algorithms.

As an important branch of computer vision, the research and application of target tracking methods are widely used in various fields, such as autonomous driving, intelligent video surveillance and medical fields. The tracking algorithm based on correlation filtering is so widely used that it has become a research hotspot. There are many related papers in this field. It is necessary to summarize and analyze these papers and the development of correlation filtering. This paper mainly combs and discusses the classical algorithms in the field of single target tracking in recent years, and analyzes the advantages and disadvantages of different methods, aiming to provide reference for the next research and development of single target tracking algorithm.

2. The composition and challenges of target tracking

Target tracking usually refers to the initial position of a given target in the first frame, and then estimate the position and shape of the target in the subsequent video sequences, so as to obtain
information such as the target's moving direction and motion trajectory. Its essence is to use feature extraction and feature correlation technology to match the features that are most likely belong to the same target in different frames of images according to the given target image, and then connect the matched targets in the video frame to obtain the target's motion trajectory, and finally achieve goal tracking.

2. Components of target tracking
Naiyan Wang et al [1] divided a tracking system into five parts: motion model, feature extractor, observation model, model update and integration method.

2.1. Motion model
The motion model mainly describes the motion trend and motion state information of the target to be tracked in a continuous image sequence. The model generates some candidate regions or bounding boxes that may contain the target object in the current frame. It is often used in algorithms such as particle filter [2] and Kalman filter [3]. Compared with feature extractor and observation model components, motion model has less impact on the tracking performance, but in the case of rapid motion and large scale changes, it is very important to select the appropriate motion model parameters for the tracking framework, which can not only improve the tracking accuracy, but also reduce the number of target candidate frames and the amount of calculation in model matching, and improve the real-time performance of target tracking.

2.1.2. Feature extractor
The feature extractor is used to extract the features of the target image to be tracked. By calculating the similarity between the image features, the same target in different images is demarcated. Then these matched targets are connected to obtain the target trajectory, and finally achieve the purpose of tracking. Commonly used features are divided into two types: one is the feature manually designed by domain experts, and the other is the deep feature learned from a large number of training samples with supervised information based on a data-driven deep neural network. The commonly used manual features include gray feature (Gray), directional gradient histogram feature (HOG), Haar feature (Haar-like), scale invariant feature (SIFT), etc. Researchers found that the multi-feature fusion method is more robust than a single feature. Therefore, Ning et al. [4] proposed an apparent feature representation method combining color and texture histogram for target tracking. Different from artificially designed features, deep features are learned through a large number of training samples, which are more discriminative than manually designed features. Shallow depth features include appearance features such as textures, while deep features contain highly abstract semantic information such as target categories. Therefore, tracking methods using depth features can usually easily obtain a good effect.

2.1.3. Observation model
The role of the observation model is to predict the state of the target based on candidate state provided by the apparent model and the motion model. Most tracking methods mainly focus on this piece of design. Generally, target tracking can be divided into generative method and discriminant method according to the model category. The idea of the generative tracking method is to first extract the target features to learn the appearance model that represents the target, then use it to search the image
area for pattern matching, and finally find area in the image that best matches the model as the target. Among them, the representative methods include subspace, sparse coding and dictionary learning. It emphatically describes the distribution of target appearance features, and has strong feature characterization capabilities, but it ignores the image background information, which makes the model drift easily when encountering interference such as occlusion, resulting in tracking failure. The discriminant method is to solve the target tracking as a classification or regression problem, with target information as positive samples and background information as negative samples. It trains a discriminant function to separate the target and background, so as to achieve the purpose of tracking. Therefore, it is also called target tracking method based on detection [5]. The method makes full use of target information and background information, and improves the defect of generative method, which is more robust than generative method. It has gradually become the mainstream in the field of target tracking. However, it is easily affected by sample data, such as uneven positive and negative samples, and duplication of positive samples. The commonly used theoretical methods of this model include logistic regression, ridge regression, support vector machines, correlation filtering and so on. In the discriminant algorithm, the discriminant tracking method based on correlation filtering gradually stands out with advantages of fast speed and good effect, which has attracted the attention of the majority of researchers.

2.1.4. Model update
In order to adapt to the changes of object appearance and prevent the tracking process from drifting, model updating is mainly aimed at the appearance model and motion model updating. But so far, there is no unified adaptive model updating standard. In the process of model updating, we usually face the problems of model updating selection strategy and update interval selection. If the model is updated too frequently, it may lead to the problem of large amount of calculation and poor real-time performance; Conversely, if the update is too slow, it may make the change too fast, which would result in tracking frame drift and excessive noise. Therefore, many models use the combination of long-term and short-term updates to solve this problem. Compared with single update strategy, the performance of this strategy is improved. How to choose a better model updating method to deal with various changing situations so as to achieve long-time real-time tracking has always been the focus and difficulty in the field of target tracking.

2.1.5. Integration method
The integration method can improve the prediction accuracy of the model, and it is often regarded as an effective means to improve the tracking accuracy. It can be generally divided into two categories: selecting the best one among multiple prediction results, or using the weighted average of all prediction results.

2.2. Challenges in target tracking
After years of research by researchers at home and abroad, a large number of classic and excellent algorithms have emerged. Target tracking technology has also made great progress, and there are many practical applications in engineering projects. However, due to the variety of object motion, background changes and many other interference factors, the current tracking algorithm is far from being able to meet the needs of practical application in accuracy, robustness and real-time, still facing a series of challenges. In the OTB100 dataset, Wu et al. summarized these difficulties caused by deformation, Illumination variation, fast movement, background clutters, rotation, scale variation, and occlusion[6]. For the existing target tracking algorithms, there are few that can handle both background changes and target changes. Therefore, how to establish a stable target appearance model to deal with these difficulties is still the focus and difficulty of future research.

3. Introduction to related filtering tracking algorithm
In recent years, target tracking algorithms based on correlation filtering have attracted much attention
due to their excellent speed and accuracy. This type of algorithm not only shows quite good performance in complex scenes such as occlusion and deformation, but also is much faster than other types of tracking algorithms in speed. Next, we will introduce the principle and development of the algorithm.

3.1. Principle of Correlation Filter Tracking Algorithm
The basic idea of correlation filter tracking is to design a filter template, which is used to perform correlation operation with the target candidate region, and the position of the maximum output response is the target position of the current frame. According to the research of some tracking algorithms in recent years, the general framework of target tracking based on correlation filtering is [7]: firstly, the image block is extracted from the target position determined in the initial frame, and the correlation filter is trained from the given image block. Then, in the subsequent image sequence, in order to effectively represent the appearance features, some features of the target image block can be extracted though manual features, depth features or fusion features. In order to smooth the boundary and prevent the image edge problem caused by the target feature in the process of Fourier transform, cosine window operation is usually performed on the image to make the pixel value closed to the edge close to zero. This operation can also increase the weight of the central target. Use Discrete Fourier Transform (DFT) to perform effective correlation operations. The response graph is obtained by inverse Fourier transform (IFFT), and the position of the maximum response score is the new position of the target in the current frame. Finally, the target region at the latest position is extracted to update the correlation filter.

3.2. Correlation filtering tracking algorithm and its development
In order to show the development of the target tracking algorithm of correlation filtering clearly, it is expounded from the following several angles.

3.2.1. Feature representation
In 2010, Bolme et al.[8] proposed the Minimum Output Sum of Squared Error (MOSSE) model in their CVPR work, and used the correlation filtering algorithm for target tracking for the first time. In order to obtain a faster algorithm execution speed, the work of correlation calculation is converted to the Fourier frequency domain space, which greatly improves the speed and efficiency of the algorithm. The tracking speed reaches 669 frames per second, which significantly exceeds the advanced algorithms at the time, and has caused a huge response in the target tracking field. Since then, a large number of tracking algorithms based on correlation filtering have appeared. Henriques of Oxford University proposed a cyclic structure kernel (Circulant Structure Kernel, CSK) tracking method [9]. The principle is to find the frequency domain resonance position of two consecutive frames, and use the kernel function for Kernel mapping. The cyclic matrix and kernel function are introduced to improve the operation speed. It is worth mentioning that both MOOSE and CSK tracker use single channel grayscale features, and the representation ability of gray features is not enough to deal with the situation that the background is complex or the color of the target is similar to the background.

Later, Henriques et al. proposed a new kernelized correlation filter KCF (Kernelized Correlation Filter) [10] based on CSK, which extended the single channel feature to multi-channel directional gradient histogram feature, and used the gradient information of the image to improve the performance of the algorithm. After that, Martin Danelljan et al. [11] extended CSK with CN color features as a multi-channel color tracker, combined with an adaptive dimensionality reduction strategy, which reduced the computational cost and improved the performance of the algorithm.

Considering the limitations of using single feature, more and more researchers have focused on how to integrate different features to improve the robustness of the algorithm while ensuring real-time tracking. Yang Li et al. proposed the SAMF tracker from the perspective of feature optimization [12], and meanwhile modeled Gray, CN and HOG features to improve the robustness of the tracker in complicated environments. Taking advantage of the complementary advantages of HOG and CN, the
Staple algorithm proposed in [13] comprehensively utilizes color features that are robust to shape changes and HOG features that are robust to occlusion conditions, which improves the performance of the model when dealing with different complex tracking scenes. In the framework of ridge regression, these two clues with complementary characteristics are combined. Not only the accuracy is improved, but also the calculation efficiency is significantly improved, and the frame rate reaches 80FPS. Because manual features are difficult to adapt to various changes of the target, while deep learning technology can directly learn features from the original data without manual intervention, and it has powerful feature processing and representation learning performance, there are many researches on correlation filtering methods combined deep features.

3.2.2. Scale evaluation

In view of the traditional correlation filter trackers usually adopt fixed size windows, it is difficult to adapt to the scale change of the target in complex scenes, a scale adaptive (SAMF) algorithm based on multi-feature fusion is proposed in literature [12]. Based on KCF tracking framework, the filter is detected on the multi-scale image block to obtain the response map. The position with the largest response is taken as the target position and the best scale. SAMF can be regarded as a global optimization process, which only needs one filter without additional training and storage. Martin Danelljan et al. proposed a new robust scale estimation method DSST (Discriminative Scale Space Tracker) [14]. The method of feature scale pyramid is used in the model, which improves the target tracking effect in the scene of scale change. Different from SAMF, DSST algorithm can be regarded as a process of local optimization. It uses a distributed exhaustive strategy, that is, the translation tracking is performed first, and then the scale estimation is carried out. Since the DSST algorithm uses a position filter and a scale filter for target location and scale estimation respectively, and the two filters are relatively independent, so the scale estimation can be easily transplanted to other tracking algorithms. Nevertheless, the DSST scale pool contains 33 scales, which reduces the computational efficiency. In order to solve the complexity problem of the tracker, dimension reduction and QR decomposition are used in reference [15] to reduce the computational complexity. In addition to the exhaustive scale pooling method, literature [16] proposed a real-time block adaptive kernel correlation filter RPAC, which decomposes target tracking into five components, each part has an independent filter, and estimates the change of target scale by calculating the change of the maximum response score in each response graph.

3.2.3. Border effect

The false samples generated by boundary effect will result in the poor discrimination ability of the classifier, which has a serious impact on the performance of the tracker. In order to solve the problem of background response, and expand the image search area as much as possible to improve the effect, the Danelljan team designed a spatial regularization discriminant correlation filter SRDCF[17] for tracking. The model can only not suppress the background response and expand the search, but also deal with the problem of image boundary effect. On the basis of SRDCF, literature[18] introduced time regularization into the framework of SRDCF, and proposed a spatio-temporal regularization correlation filter STRCF. STRCF based on online passive attack learning(PA) can not only approach the SRDCF form on multiple training images reasonably, but also has stronger robustness than SRDCF under the condition of large appearance change.

Also inspired by SRDCF is the ASPCF [19] algorithm published in 19 years. Similar to SRDCF, ASRCF adds adaptive spatial regularization to solve boundary effects and noise. At the same time, ASRCF uses the idea of combining HOG features and convolution features to improve speed. The CFLB algorithm based on the gray-scale feature MOOSE proposed in [20] uses large detection and update image blocks. It trains the correlation filter with a relatively small scope, and directly fills the edges of the filter with zeros. In addition, ADMM is used to ensure the correct filter size. BACF [21] also deals with the boundary effect and proposes the idea of using real negative samples to expand the
searched area, and achieved excellent results, which largely indicates that sample quality has a great influence on the performance of the tracker.

3.2.4. Long-term tracking
In order to solve the problem of visual tracking that the object's appearance changes significantly due to deformation, sudden movement, serious occlusion and the object beyond the line of sight, from the perspective of long-term tracking, reference [22] combines tracking and detection algorithm, and adds an improved on-line learning mechanism to make the overall tracking more stable and effective. However, it can't make full use of temporal motion cues like LCT [23] method, so it can't well track targets with obvious deformation and fast motion. In addition, the TLD method updates its detector frame by frame, which leads to drift and wrong target re-detection. In reference [23], a correlation filter responsible for target confidence is introduced based on DSST. LCT realizes translation estimation by modeling temporal context related information, using appearance information to construct scale pyramid to realize scale estimation, and trains online random ferns detector to realize re-detection in case of target loss. It greatly enhances the robustness of the target in the case of large area occlusion and moving out of view. On the basis of LCT, the author proposes ILCT method [24], which uses support vector machine instead of online random ferns detection as re-detection module.

4. Problems and future development trends of correlation filtering tracking algorithms
Although the target tracking algorithm based on correlation filtering has made great progress, there are still some problems to be solved.

1) Model update strategy and update time interval selection problem: If the update is too frequent, it will lead to a large amount of model calculations, thus affecting the real-time problem, and may cause the loss of some feature information; if the update is too slow, the tracking frame may drift due to the feature changing too fast.

2) At present, most single target tracking algorithms are focused on short-term target tracking task. Although a few researchers have studied long-term tracking, its essence is still to introduce re-detection modules or other correction modules into the short-term tracking framework, and the lack of targeted long-term tracking framework is worthy of further study.

3) How to choose the size of the search box: if the search box is too small, it is difficult to detect the fast moving targets; If it's too large, it will introduce a lot of useless background information, even some backgrounds similar to the target, which will cause interference to the target tracking process, resulting in the degradation of the model effect and tracking drift.

Future research on related filtering and tracking algorithms can be carried out from the following directions: Since features have an important influence on target tracking algorithms, how to develop good and effective tracking features is still an open issue; In addition, how to update the target model reasonably so that it can not only adapt to the appearance changes of the target, but also avoid the tracking failure caused by model drift as far as possible still needs further research.

5. Conclusions
Although visual target tracking technology has made great progress in recent years, there is still a huge gap between computer target tracking system and human visual system in some complex realistic scenes. Therefore, the real fast, accurate and universal target tracking research is still far from complete. It is believed that with the joint efforts of many researchers, more robust and long-term stable target tracking algorithms will continue to emerge, and greater achievements will be made in the field of target tracking in the future.

Acknowledgments
This work is supported by Natural Science Foundation of China (No.61563049).
References

[1] Wang Naiyan, Shi Jianping, Yeung Dit - Yan, et al. Understanding and diagnosing visual tracking systems [C]. In IEEE ICCV, 2015.

[2] Salmond D J, Birch H. A particle filter for track-before-detect [C]// Proceedings of the 2001 American Control Conference, 2001, DOI: 10.1109 /ACC.2001.946220.

[3] Aidala V. Kalman filter behavior in bearings-only tracking applications [J]. IEEE Transactions on Aerospace and Electronic Systems, 2007,15(1) : 29-39.

[4] Ning J F, Zhang L, Zhang D, et al. Robust object tracking using joint color-texture histogram [J]. International Journal of Pattern Recognition and Artificial Intelligence, 2009, 23(7) : 1245-1263.

[5] Collins R T, Liu Y, Leordeanu M. Online selection of discriminative tracking features [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(10) : 1631-1643.

[6] Wu Yi, Lim J, Yang M. Object tracking benchmark [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2015, 37(9) : 1834-1848.

[7] CHEN Z, HONG Z, TAO D. An experimental survey on correlation filter-based tracking [J]. Computer Science, 2015, 53(6025) : 68-83.

[8] Bolme D S, Beveridge J R, Draper B A, et al. Visual object tracking using adaptive correlation filters [C]. In IEEE CVPR, 2010.

[9] HENRIQUES J F, CASEIRO R, MARTINS P, et al. Exploiting the circulant structure of tracking - by - detection with kernels [C]. In IEEE ECCV, 2012.

[10] HENRIQUES J F, CASEIRO R, MARTINS P, et al. High speed tracking with kernelized correlation filters [C]. IEEE TPAMI, 2015.

[11] DANELLJAN M, SHAHBAZ KHAN F, FELSBERG M, et al. Adaptive color attributes for real-time visual tracking [C]. In IEEE CVPR 2014.

[12] LI Y, ZHU J. A scale adaptive kernel correlation filter tracker with feature integration [C] // European conference on computer vision. Springer, Cham, 2014:254-265.

[13] BERTINETTO L, VALMADRE J, GOLODETZ S, et al. Staple: Complementary learners for real-time tracking [C] // Proceedings of the IEEE conference on computer vision and pattern recognition, 2016: 1401-1409.

[14] DANELLJAN M, HAGER G, KHAN F S, et al. Accurate scale estimation for robust visual tracking [C]. In BMVC (British Machine Vision Conference ), 2014.

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

[16] LIU T, WANG G, YANG Q. Real-time part-based visual tracking via adaptive correlation filters[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 4902-4912.

[17] Danelljan M, Hager G, Khan F S, et al. Learning spatially regularized correlation filters for visual tracking [C]//IEEE International Conference on Computer Vision (IC-CV) , 2015: 4310-4318.

[18] LI F, TIAN C, ZUO W, et al. Learning spatial-temporal regularized correlation filters for visual tracking [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 4904-4913.

[19] Dai Kenan, Wang Dong, Lu Huchuan, et al. Visual tracking via adaptive spatially-regularized correlation filters [C] // IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 4665-4674.

[20] GALOOGAHI H K, SIM T, LUCEY S. Correlation filters with limited boundaries [C]. In CVPR, 2015.

[21] GALOOGAHI H K, FAGG A, LUCEY S. Learning background - aware correlation filters for visual tracking [C]. In arXiv : 1703. 04590v2 , 2017.
[22] Kalal Z, Mikolajczyk K, and Matas J. Tracking-learningdetection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(7):1409–1422, 2012.

[23] MA C, YANG X, ZHANG C, et al. Long-term correlation tracking [C] // Computer Vision & Pattern Recognition. Boston, MA, 2015: 5388-5396.

[24] MA C, HUANG J B, YANG X, et al. Adaptive correlation filters with long-term and short-term memory for object tracking [J]. International Journal of Computer Vision, 2018, 126(8): 771-796.