Target Tracking Algorithm Based on Density Clustering

Jinlong Chen, Qinghao Zeng, and Xingguo Qin

Guangxi Key Laboratory of Trusted Software, Guilin University of Electronic Technology, Guilin 541004, China
641577329@qq.com, 529045053@qq.com

Abstract. The traditional Siamese network-based target tracking algorithm needs to use the convolution feature of the target to scan around the target location when predicting the location of the target in the next frame image, and perform similarity calculation to obtain the similarity score matrix with the highest score. It is the next frame target position. The highest similarity score often does not represent the precise target position of the target, which is often affected by the sliding step size during scanning. Aiming at this problem, this paper proposes a target tracking method based on density clustering. By combining the Siamese network to predict the next frame target position, and adding the target’s motion trajectory information, the direction of the target motion is given more weight, the other directions are given a smaller weight, and finally the target position is predicted by the density clustering method. The results show that the proposed algorithm effectively improves the accuracy of the target location prediction of the Siamese network when tracking targets.

Keywords: Target tracking · Siamese network · Clustering

1 Introduction

In recent years, with the development of science and technology and the increasingly diversified needs of people’s living environment, the rapid development of artificial intelligence technology has been promoted. As an important branch of artificial intelligence, computer vision technology has attracted more and more scholars and experts to study. As an indispensable part of the field of computer vision, target tracking technology has a wide range of applications in video surveillance, autonomous driving, military reconnaissance, medical and life. The target tracking is to predict the position and width and height information of the target in the subsequent frames of the video according to the coordinate information of the target image of the first frame of the video without prior knowledge. The target tracking process can be roughly divided into three steps, namely feature extraction, model building, and adaptive update model. The main problem of target tracking is how to achieve accurate positioning targets in scenes with occlusion, illumination changes, motion blur, complex background and target deformation.

© Springer Nature Switzerland AG 2020
Y. Tan et al. (Eds.): ICSI 2020, LNCS 12145, pp. 556–563, 2020.
https://doi.org/10.1007/978-3-030-53956-6_51
Over the years, target tracking has become a hot issue in the field of computer vision, and more and more related methods, but how to track the position of the target more accurately under various conditions is still a problem. At the current location, the method of target tracking can be roughly divided into two categories. One is the traditional target tracking method based on machine learning. This kind of method usually consists of three parts: appearance model, running model and search strategy. The appearance model is used to estimate the possible state of a specific location. The motion model is used to describe the state information of the target’s possible change with time. The search strategy is used to locate the target position of the next frame. Traditional target tracking methods are characterized by faster speeds, but the accuracy of tracking targets is not sufficient. In the KCF [1] algorithm, an appearance model is established by a discriminative correlation filter, and the target is located according to the correlation of the filter and the target directly. In the Kalman [2] filter algorithm, a sequential random approach is established for linear and Gaussian observation noise, and then statistical methods are used to locate the target, but the nonlinear motion template cannot be tracked. High-order particle filtering is introduced in the HOPF [3] method to track targets, but the classifier training costs are too high and result in data loss. The second type of method is a target tracking method based on deep learning. Deep neural network has made great progress in image classification, target detection and recognition, and target tracking in recent years because of its powerful feature extraction ability. This type of method extracts to the target’s effective feature representation through a large amount of data training, thereby regressing or classifying the target position through these features and neural networks. One of the more representative ones is the SiamFC [4] method, which calculates the similarity matrix of the target and the next frame image through the fully convolved Siamese network [5] to locate the target. Later, the SiamRPN [6] method was added to the RPN structure in the Faster-RCNN [7] to return the target frame width to a more accurate tracking target. Although SiamFC proposed a very constructive idea to introduce the Siamese network to target tracking, it is still not accurate enough for the target positioning, and the target frame of the tracking has always been close to the initial frame, and it is unable to adapt to the change of the target. In order to solve this problem, this paper adds a clustering method, first extracts the target features through the Siamese network, assigns corresponding weights according to the target’s motion trajectory, and then uses the similarity score matrix and density clustering algorithm to locate the target. Tracking the target by clustering instead of the initial frame achieves the goal of improving the accuracy of the target location in the target tracking, and can adapt to the change of the target.

2 Introduction

2.1 Siamese Network

The Siamese network is a depth-based similarity measure that learns similarity measures in data and then uses the learned knowledge to compare and match samples of new unknown categories. The Siamese network is composed of two identical networks. The two networks share weights. Just like a conjoined person, there are two people’s thoughts, but they share a body. In 2016, the Siamese network was applied to the target tracking
technology. In the tracking process, the convolution feature of the target image is used to scan the convolution feature of the next frame image by a certain step size, and then the similarity score of the current target image at each position of the next frame image is calculated. Finally, the position of the target is located according to the similarity score matrix.

As shown in Fig. 1, Network1 and Network2 are a group of neural networks with the same structure and shared weights. In this paper, a convolutional neural network (CNN) is used. The input of the network is two different images, and the corresponding convolution feature representation is obtained through the neural network output. The vector, by calculating the distance between the two vectors, obtains the similarity between the images. The more similar the two images are, the smaller the distance will be, and the bigger the difference. The distance is calculated as shown in Eq. (1):

$$D_w(x_1, x_2) = \|G_w(x_1) - G_w(x_2)\|$$  \hspace{1cm} (1)

Fig. 1. Siamese network.

Where x_1 and x_2 are two pictures of the same size input by the network, w is the weight of the network training, G_w (x) is the vector obtained by the picture through the network, and D_w (x_1, x_2) is the vector of the two network outputs. The Euclidean distance between the two is used to judge the similarity between the two images. In the target tracking, you need to enter the current target and the next frame image. In the vector outputted by the network, the vector obtained by the network convolution of the target will be smaller than the vector obtained by convolving the next frame image. The vector requiring the target is scanned on the vector of the next frame according to a certain step size to obtain a similarity score matrix. And in order to cope with the change of the target’s movement, it will be calculated on multiple scales when calculating the similarity score.
2.2 Density-Based Clustering Method

Clustering is the process of dividing a collection of data objects into one or more similar clusters, so that each object in the same cluster has a high similarity, and objects in different clusters have higher dissimilarity. According to the scale of clustering, clustering methods can be roughly divided into three categories, namely distance-based, interconnect-based, and density-based. The distance-based clustering method usually evaluates the similarity between the targets by the distance between the targets. The density-based clustering method mainly uses the density function to determine the similarity between targets. Interconnectivity-based clustering methods typically combine highly connected objects into one class based on a graph or hypergraph model. In 2014, the Alex Rodriguez [8] team published a text on Science about density-based clustering, which argues that cluster centers are surrounded by lower-density neighbors and farther away from other denser points. The method proposes two keywords, namely local density and distance. The local density is the number of other points within a certain distance from the point. The distance is the smallest cluster of points to any other denser point.

\[
\rho_i = \sum_j F(d_{ij} - d_c) \quad (2)
\]

\[
F(x) = \begin{cases} 
1, & x < 0; \\
0, & x \geq 0; 
\end{cases} \quad (3)
\]

Where \( \rho_i \) is the density of point \( i \), \( d_{ij} \) is the distance from point \( i \) to point \( j \), and \( d_c \) is a truncation distance. If the difference between \( d_{ij} \) and \( d_c \) is less than 0, \( F(x) \) is equal to 1, otherwise \( F(x) \) is equal to 0. However, in target tracking, only the target position of the next frame needs to be located, so only one cluster needs to be clustered (Fig. 2).

In the method of this paper, the motion trajectory information before the target is added in the clustering process, so the clustering part of this paper will be clustered twice. The first clustering is to cluster the target’s scoring matrix with the target’s trajectory information to get the exact location of the target. The second clustering is a clustering of the target’s width and height, and obtaining the target’s accurate width and height information.

\[
\rho_i = \sum_j F(\varphi_{ij}(s_{ij}, w_{ij}) - \varphi_c(s, w)) \quad (4)
\]
Where $s_{ij}$ is the difference between the similarities of the i point and the j point, and $w_{ij}$ is the difference between the weights of the i point and the j point. In this paper, the motion trajectory information of the target is added to the prediction target process, and the density calculation changes the distance into the difference between the density and the weight product. For weight setting, since the target tracking is a continuous process, the weight of the direction before the target should be larger, and the weight of the opposite direction should be smaller. At the same time, we also cluster the width and height, combined with the width and height information of the prediction frame, combined with the weight of the prediction frame, to calculate the final target frame, so that the target tracking is more accurate.

The overall structure of the proposed method is shown in Fig. 3. The current target image and the target image of the next frame are input to the network, and the similarity score matrix of the target on the next frame image is calculated by the Siamese network. Then, according to the motion trajectory before the target, the scoring matrix is given corresponding weights. Finally, the density clustering algorithm is used to cluster the target position and width and height, and the exact position of the target is obtained, and the moving direction information of the target is updated.

![Overall structure of the algorithm.](image)

### 3 Experiment

The training and detection data of the method in this paper are from the public VOT dataset, and we randomly select part of the video data for training. The algorithm in this paper mainly improves on the traditional target tracking algorithm based on Siamese network, hoping to track the target more accurately (Fig. 4).

We randomly selected part of the video data from the VOT dataset to test on our method, SiamFC method, and particle filtering-based target tracking method. The image marked with a red box in the picture is the real position of the target during video tracking. The image with a yellow box is the tracking effect of our algorithm. The image with a blue box is the effect of the SiamFC method. The effect of particle filter on target tracking
method. It can be seen from the figure that although the SiamFC method and particle filtering-based target tracking method can track the target during the tracking process, the tracking accuracy is low, and the tracking frame is always close to the frame marked by the initial frame size, little change. In the method of this paper, both the center point and the width and height of the target are clustered. As can be seen from the figure, the box obtained by the clustering can well adapt to the change of the target (Figs. 5 and 6).

We compared Particle filter, SiamFC and our method. Statistics on accuracy and IOU. It can be seen that our method is 5%–10% higher in accuracy than Particle filter, slightly better than SiamFC. However, our method on IOU is about 10% higher than Particle filter and SiamFC. It can be seen that this paper proposes to add clustering and target motion trajectory information, which can make tracking more accurate in target tracking.
4 Conclusion

For the traditional Siamese network-based target tracking method, when tracking the target, the target position prediction is affected by the sliding step and the accuracy is low. This paper proposes the motion rule information before the predicted target position is added to the target, and better predicts the target position. And the density-based clustering method is added to locate the target frame, and the target center position and the target width and height are respectively clustered to better predict the position of the target in the next frame image. The experimental results show that the proposed method can effectively improve the accuracy of the prediction of the target position of the next frame without losing the success rate of tracking. In the next phase of the study, we can introduce the idea of online learning in the tracking process, while tracking the target, while updating the network weights, and constantly learning the new characteristics of the target in the movement process.
Acknowledgements. The National Natural Science Foundation of China (61866007), the Natural Science Foundation of Guangxi District (2018GXNSFDA138006). Research and application demonstration of intelligent recommendation system for tourist routes based on user preferences Guangxi Scientific Research and Technology Development Plan Project (Guike AB17195027). Research and application demonstration of key technologies for indoor positioning and location service of intelligent guides in karst cave scenic areas Guangxi Scientific Research and Technology Development Plan Project (Guike AB18221011). Funded by the research project of Guangxi Trusted Software Key Laboratory (No. kx201620). (Guilin Key Laboratory of Trustworthy Software, Guilin University of Electronic Technology, Guilin, Guangxi 541004). Guilin Scientific Research and Technology Development Plan Project 20180107-4.

References

1. Henriques, J.F., Caseiro, R., Martins, P.: High-speed tracking with kernelized correlation filters. IEEE Trans. Pattern Anal. Mach. Intell. 37, 583–596 (2015)
2. Bi, H.Y., Ma, J.W., Wang, F.J.: An improved particle filter algorithm based on ensemble Kalman filter and Markov chain Monte Carlo method. J. Sel. Top. Appl. Earth Obs. Remote Sens. 8(2), 447–459 (2017)
3. Xiao, Q.L., Fang, Y.L., Liu, Q.: Online machine health prognostics based on modified duration-dependent hidden semi-Markov model and high-order particle filtering. Int. J. Adv. Manuf. Technol. 94, 1–15 (2017)
4. Bertinetto, L., Valmadre, J., Henriques, J.F., Vedaldi, A., Torr, P.H.S.: Fully-convolutional siamese networks for object tracking. In: Hua, G., Jégou, H. (eds.) ECCV 2016. LNCS, vol. 9914, pp. 850–865. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-48881-3_56
5. Chopra, S. Hadsell, R. LeCun, Y.: Learning a similarity metric discriminatively, with application to face verification. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 539–546 (2005)
6. Li, B., Yan, J.J., Wu, W.: High performance visual tracking with siamese region proposal network. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 8971–8980 (2018)
7. Ren, S.Q., He, K., Girshick, R.: Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. 1137–1149 (2017)
8. Alex, R., Alessandro, L.: Clustering by fast search and find of density peaks. Science 344, 1492–1496 (2014)
9. Bo, L. Wei, W. Qiang, W.: SiamRPN ++: evolution of siamese visual tracking with very deep networks. In: Proceeding of 2019 IEEE Conference on Computer Vision and Pattern Recognition (2019)
10. Wang, N., Yeung, D.Y.: Learning a deep compact image representation for visual tracking. In: International Conference on Neural Information Processing Systems, pp. 809–817 (2013)