Object Tracking with Kernelized Correlation Filters

--CS585 Project Report

Name: Nan Zhou
Name: Lingzhuo Zhao
Name: Yanjie Chen

Abstract

This project ends to study a traditional method, Kernelized Correlation Filters(KCF) to achieve high-speed tracking. KCF is an improvement for the circulant structure tracker (CST) [1] by adding in the Histogram of Oriented Gradients(HOG) feature. Another tracking algorithm that exploits dese Spatio-temporal context (STC) information will be introduced to help visualise the KCF framework. We aim to apply the KCF tracker to some of the benchmark datasets and analyse the performance in comparison with other tracking algorithms.

Problem Definition

Given a series of video frames, it is required to track at least one object in the video. We are required to develop one algorithm and test it in different datasets.

Related Work

Tracking-by-detection

Visual tracking is a popular research problem in computer vision due to the applications in different areas such as activity recognition, human-computer interaction and motion detection. Numerous tracking algorithms have been proposed which allow for some assumptions about the target object. It is ideal to track an object with little information about it. A very successful method is tracking-by-detection.

Most tracking algorithms can be categorized to two types: generative and discriminative based on their appearance models. A generative tracking method learns an appearance model to represent the target and searches for image regions with best matching scores as the results[6]. A discriminative classifier is the key component of modern trackers in general. It aims to distinguish the target and its surroundings. Some examples of the related tackers are Support Vector machines(SVM), random Forest classifiers, boosting variants and Naïve Bayes classifiers proposed by Zhang et al. [3], which can predict the target's location directly rather than its presence in an image patch. This paper will focus on two discriminative methods. One is proposed by Kalal et al[4], and the other is proposed by Bolme et al. [5].

The approach proposed by Kalal et al[4] can cut down the number of training samples as it uses a list of structural constraints to lead the sampling process of a boosting classifier. However, there are some limitations such as pricey training process, limited training features that can be used and tedious structural heuristics tuning.

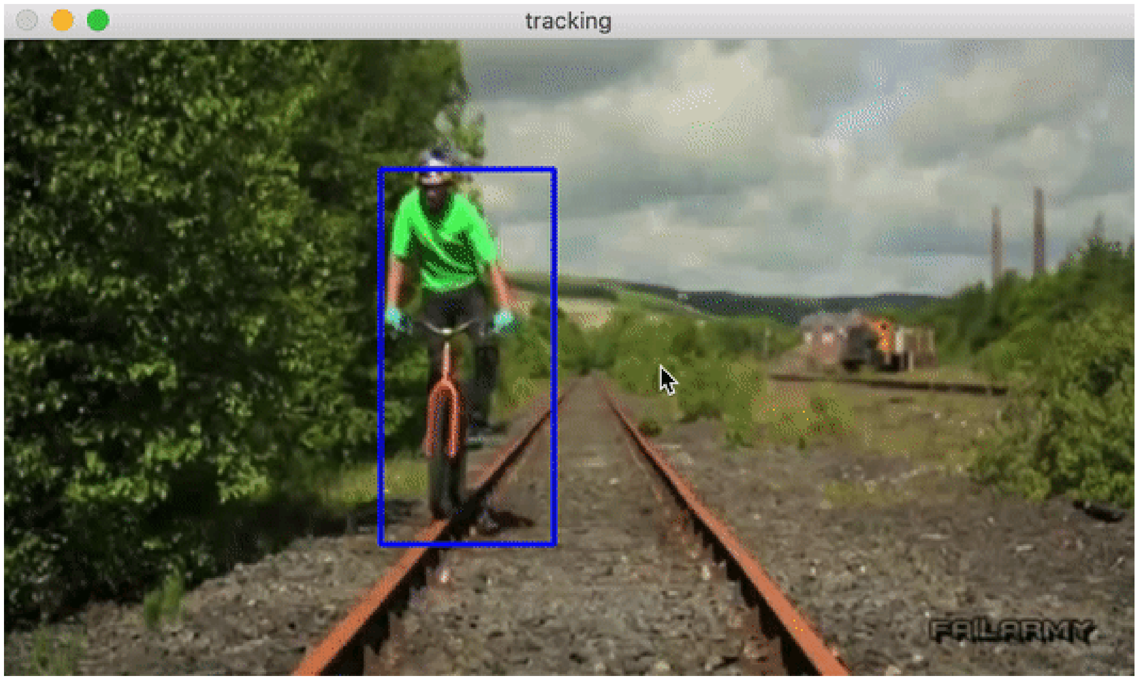
The correlation filters [5] only use partial computation power although it is more sophisticated. The trick is the convolution of two patches is the same as an element-wise product in the Fourier domain, which makes the process of translation and image shifts by linear classifier more efficient. But, it can be limiting as well.

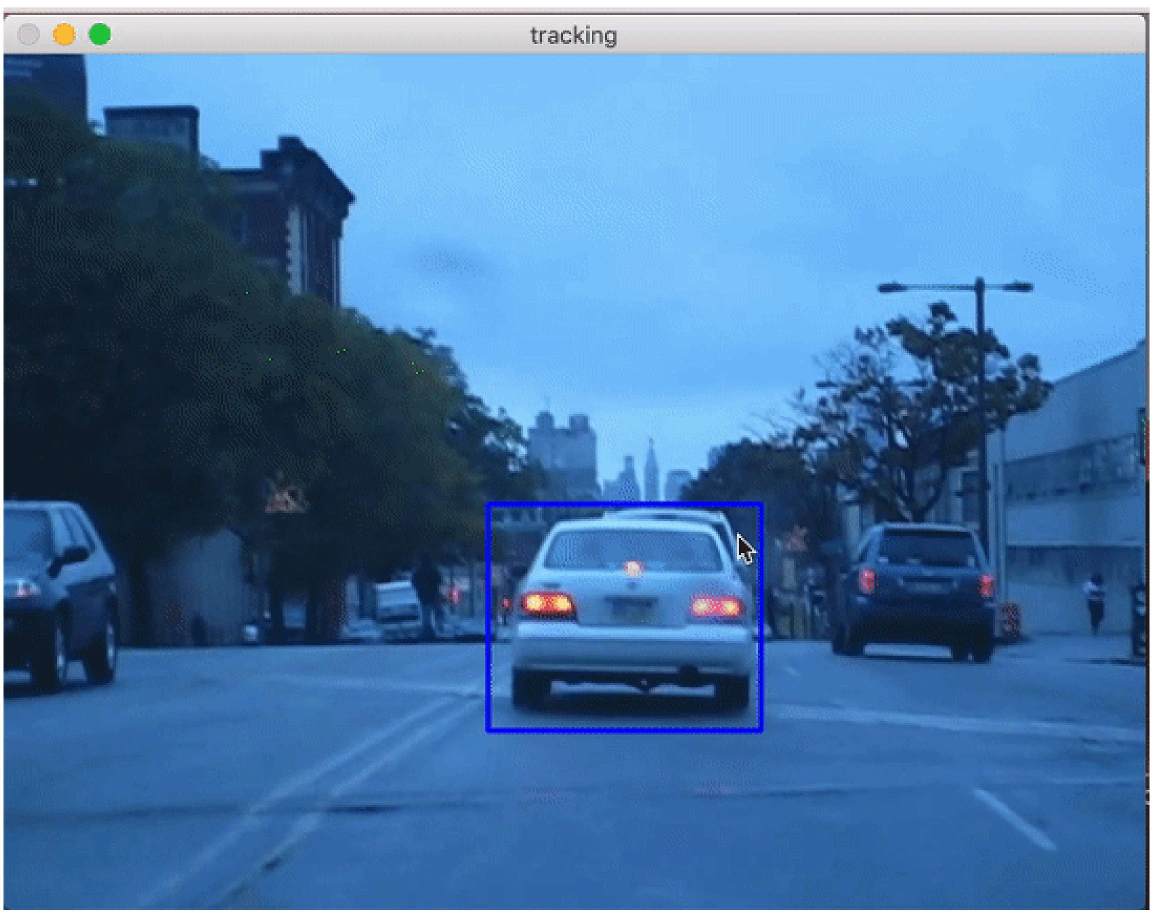
Method

KCF is a paper published in 15 years and is a more representative discriminant tracking algorithm. On the basis of the original CSK algorithm, the main feature is to use the circulant matrix Fourier space diagonalization to simplify the calculation. Before learning KCF, you can read the other two papers, namely the 10-year MOSSE tracking algorithm and the CSK algorithm. The MOSSE algorithm applied convolution to the tracking algorithm for the first time, which greatly improved the algorithm speed. Later, CSK on this basis, the algorithm adds a regular term to avoid over-fitting, and at the same time introduces a circulant matrix and a kernel function to improve computational efficiency, and the performance is greatly improved.

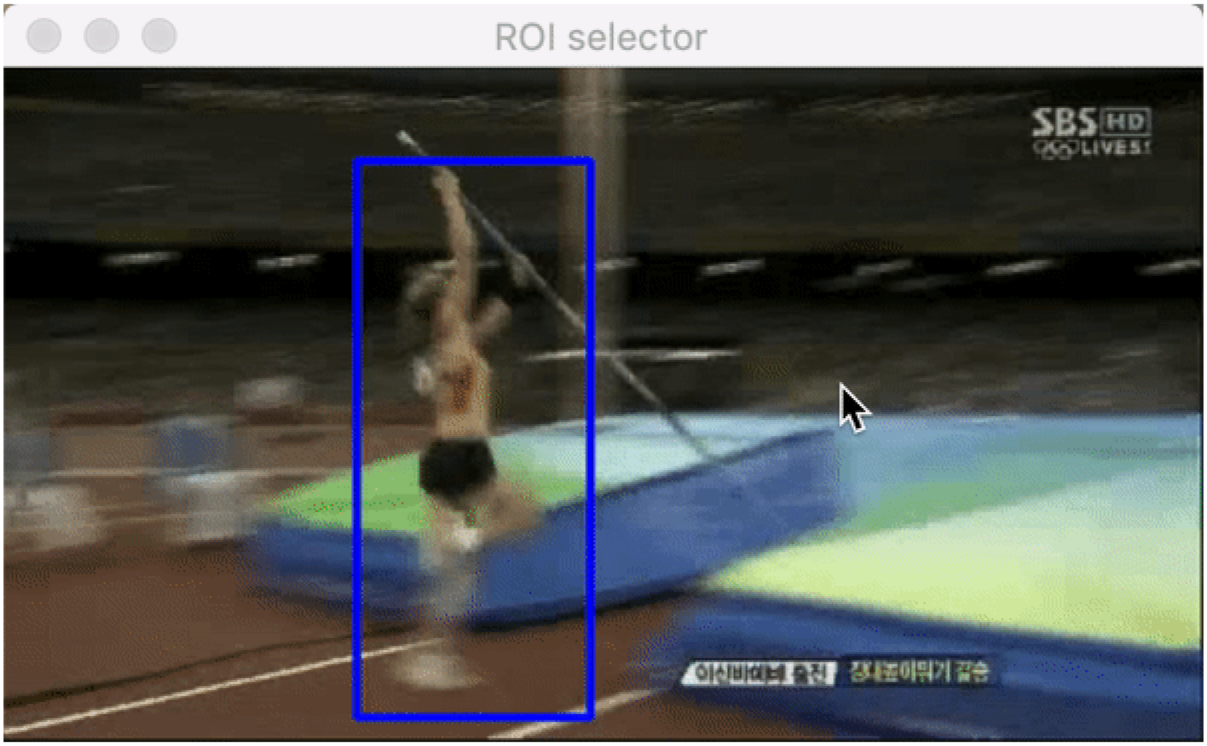
KCF is a tracking method of Tracking-By-Detection, which is the same as TLD and OAB. The tracking object is a positive sample, and the surrounding environment is a negative sample to train a discriminative classifier. It uses the circulant matrix of the expanded area of the target to collect positive and negative samples, then use ridge regression to train the detector, and successfully use the diagonalization property of the circulant matrix in Fourier space to convert complex matrix operations into vector dot multiplication, which greatly reduces the amount of calculations improves the efficiency of tracking. The ridge regression of the linear space is mapped to the higher-dimensional nonlinear space through the kernel function. In the nonlinear space, by solving a dual problem and some common constraints, it is also possible to use the circulant matrix Fourier space diagonalization to simplify the operation. Expose the limitation that can only handle single-channel features, and give a way to deal with multi-channel features.

Result





Fail to track



Future Work

There is still some future work that could be done to improve our KCF algorithm. First, we could test our algorithm in some different datasets. Based on the results of different datasets, some parameters in our algorithm would be tuned and we can get better results. Second, right now we need to manually select an object to track using a bounding box at the initial frame for our algorithm. We can improve this by making the program automatically tracking the object in the video. Also, we could develop multi-object tracking based on this algorithm.

References

1. *Henriques, J., Caseiro, R., Martins, P., Batista, J.: Exploiting the circulant structure of tracking-by-detection with kernels. In: ECCV. pp. 702–715 (2012)*
2. *J. F. Henriques, R. Caseiro, P. Martins and J. Batista, "High-Speed Tracking with Kernelized Correlation Filters," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 3, pp. 583-596, 1 March 2015, doi: 10.1109/TPAMI.2014.2345390.*

3. *K. Zhang, L. Zhang, and M.-H. Yang, "Real-time compressive tracking," in ECCV, 2012.*

4. *Z. Kalal, K. Mikolajczyk, and J. Matas, "Tracking-learning- detection," TPAMI, 2012.*

5. D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, "Visual object tracking using adaptive correlation filters," in CVPR, 2010, pp. 2544-2550.

6. K. Zhang, L. Zhang, Q. Liu, D. Zhang, and M.-H. Yang. Fast visual tracking via dense spatio-temporal context learning. In Proceedings of the European Conference on Computer Vision, 2014.