RHODES UNIVERSITY
Where leaders learn

Real Time Detection and Tracking of People in Crowds

J L Gouws
Supervisor: Mr. J Connan
Computer Science Department, Rhodes University

April 10, 2022

Abstract

Tracking of objects in videos streams is a powerfull tool in the field of computer vision. It is, however, a relatively unexplored field. I propose to use methods of machine learning and image processing to further explore object tracking and detection. The specific focus of my research will be the detection and tracking of people in crowds. Another concern of the research is performance of the implemented system. A large amount of computation power is often needed to analyze videos of crowds, investigation into reducing the required computation power will be carried out.

1 Introduction

Analysis of videos of crowds of people has many practical applications. Most real world situations that involve people have people that are moving. On top of that most often people are not isolated, but instead form groups of people.

For these reasons it makes sense to investigate ways to develop a system that can efficiently and accurately detect and track multiple people in a video stream. Tracking targets in a video stream is by and large an understudied field. The literature is relatively limited in comparison with the number of real world applications of tracking systems.

2 Reseach Statement

The TLD frame work can be used to develop a system that is capable of detecting and tracking people in a crowd. The system can be designed so as to start with minimal initial offline data and training, and require minimal compute power.

3 Research Objectives

- Using the TLD framework to build a system for idenitifying and tracking people in crowded scenes.
- Investingating ways to improve tracking and classification performance.
- Investigating ways to decrease the computation power required by the system.
- Developing ways to store the system state so that it can maintain it's learned parameters between instances.
- Improving the scalability of the system so that new tracking targets can be added easily.

4 Related Works

4.1 Closely related works

In 2011 Zdenek Kalal invented the Tracking, Learning and Detection(TLD) framework for the longterm tracking of objects in a video stream[1]. Kalal's original implementation uses a median flow tracking stage, P-N learning, and a random forrest based detector. The system requires online(learning as data becomes available) learning in order for the system to work, Kalal developed the P-N Learning paradigm [2], a semi-supervised bootstrapping model, tailored to the needs of TLD.

There have been improvements in tracking methods since his development of TLD. Most notably Kernelized Correlation Filters(KCF) being applied on Histogram of Oriented Gradient features[3]. KCFs, however, do not have the ability to detect, and so if they fail they are unlikely to recover. KCFs also require a stage that will start them tracking. KCFs have great potential to be used as the tracking stage for a TLD tracker.

Kalal uses Random Forest for detection in his implementation of TLD. A more modern approach would be to use Extreme Gradient Boosting(XGB) for the detector. XGB generally offers similar if not better classification performance than Random Forest and requires significantly less time to train [4].

There has also been investigation into the use of Convolutional Neural Networks in tracking [5]. This offers high performance tracking of generic objects. It will be investigated for extension to tracking and detecting people in crowds. It will not be a main point of research unless it turns out to be fruitful– seeing as CNNs tend to require large amounts of training time, data, and memory space.

4.2 Less related work

There has also been relevant work done on correlation tracking by [6]. Ma. et al investigated the problem of long term tracking where the target undergoes abrupt motion, heavy occlusions and disapearing from view. There work will probably integrate well with Kernelized Correlation filters, and it has very practical advantages. It is not the forefront of this research to handle occlusions, but it is a possible extension.

The P-N learning system is not completely unrelated to Reinforcement Learning Techniques. There have been promising recent results in online reiforcement learning [7]. The work of [7] allows for variable data budgets, which should integrate well with a constant flux of input from a video stream.

TLD may not be compatible with the methods proposed in [7], but the feaibility of the ideas will be explored in the paper.

5 Limitations

It is difficult to accurately identify people from the rear. In this case it might be good to try and prevent the system from learning the appearance of the back of a person's head. A solution to this problem is also developing a system that can be distributed so as to be able to track people from a surrounded view.

It is possible to track semi-ocluded targets, but detecting them is difficult. This will cause problems for the initialization and re-initialization of the tracker. It might be possible to impute some facial features before inputing the stream into the detection system. If a face is fully ocluded, however, there is not much that the system will be able to do, but a human will not be able to do very much.

The system will require relatively good resolution cameras in order to identify people accurately from a long distance. In most situations involving crowds, a camera capturing a video stream will need to be placed a significant distance from the targets. Hence, the best we can do in software is try to use image processing techniques to artificially enlarge the target, or work on detectors that can work on minimal resolution.

6 Timeline

Time	Deliverable
30 March 2022	Seminar 1: Presentation of project
11 April 2022	Draft proposal
14 April 2022	Final proposal
18 April 2022	Functional re-implementation of TLD
25 April 2022	Using KCF as Tracking stage of tracker
29 April 2022	Literature review
11-13 July 2022	Seminar 2: Progress Presentation
19 August 2022	Progress Report
3 October 2022	First Draft of thesis
14 October 2022	Short ACM-style paper
17-19 October 2022	Seminar 3: Final Oral presentation
28 October 2022	Final project submission

References

[1] Z. Kalal, K. Mikolajczyk, and J. Matas, "Tracking-learning-detection," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, pp. 1409–1422, 7 2011.

[2] Z. Kalal, J. Matas, and K. Mikolajczyk, "P-n learning: Bootstrapping binary classifiers by structural constraints," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010, pp. 49–56. DOI: 10.1109/CVPR.2010.5540231.

[3] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, pp. 583–596, 3 2014.

[4] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, "A comparative analysis of gradient boosting algorithms," *Artificial Intelligence Review*, vol. 54, no. 3, pp. 1937–1967, 2021.

[5] H. Nam and B. Han, “Learning multi-domain convolutional neural networks for visual tracking,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016.

[6] C. Ma, X. Yang, C. Zhang, and M.-H. Yang, “Long-term correlation tracking,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2015.

[7] J. Schrittwieser, T. Hubert, A. Mandhane, M. Barekatain, I. Antonoglou, and D. Silver, “Online and offline reinforcement learning by planning with a learned model,” in *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, Eds., vol. 34, Curran Associates, Inc., 2021, pp. 27 580–27 591. [Online]. Available: `https://proceedings.neurips.cc/paper/2021/file/e8258e5140317ff36c7f8225a3bf9590-Paper.pdf`.