

CS224W Course Project Proposal: A survey of Network Alignment

Danqi Chen
Stanford University
danqi@stanford.edu

Botao Hu
Stanford University
botaohu@stanford.edu

Shuo Xie
Stanford University
shuoxie@stanford.edu

October 22, 2012

Abstract

This document is a project proposal for the CS231A open course project. It details our plans for contributing to current research in real-time object tracking. Possible datasets, algorithms, readings and evaluation methods are reviewed.

1 Problem Statement

Several attempts have recently been made to improve real-time object tracking in a sequence of frames by using a detector in addition to the tracker (Kalal [1], Pernici [4], Nebehay [3]). The main goals of the detector are to prevent the tracker from drifting away from the object, and recover tracking after an occlusion. Since the only prior knowledge about the object is a bounding box in the initial frame, the detector must be trained online. In order to build such a system, two critical challenges must be addressed:

1. finding an efficient feature-extraction algorithm to perform detection on thousands of sub-windows in each frame
2. using a powerful learning strategy to update the template used by the detector

In addition, the solutions to these two problems are dependent on each other, and as such, they must be designed so as to fit into a single system.

The goal of this project is to investigate new algorithms for 1. and 2. and try to find improvements in terms of:

- robustness of tracking (good performance with a wide range of objects, tolerance to poor video quality such as camera blur, low resolution, low frame rate, etc)
- efficiency (time and space complexity)

For demonstration purposes, some additional features could be introduced, such as simultaneous tracking of multiple objects.

2 Algorithms

Several feature extraction algorithms can be used for template matching. Feature descriptors such as FREAK (Vandergheynst [6]), BRISK (Leutenegger [2]) and ORB (Rublee [5]) could help speed up the template matching process.

We would like to investigate a deep learning approach for improving the learning step. Zou, W. [7] should be a good reference for designing our solution.

3 Readings

The papers mentioned in the references section of this document will provide context and background for this project.

References

- [1] Z. Kalal, K. Mikolajczyk, and J. Matas. Tracking-learning-detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(7):1409–1422, 2012.
- [2] S. Leutenegger, M. Chli, and R. Siegwart. Brisk: Binary robust invariant scalable keypoints. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2548–2555. IEEE, 2011.
- [3] G. Nebehay, B. Matusik, C. Picus, and R. Pflugfelder. Evaluation of an online learning approach for robust object tracking. 2011.
- [4] F. Pernici. Facehugger: The alien tracker applied to faces.
- [5] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. Orb: an efficient alternative to sift or surf. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2564–2571. IEEE, 2011.
- [6] P. Vanderghenst, R. Ortiz, and A. Alahi. Freak: Fast retina keypoint. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 510–517. IEEE, 2012.
- [7] W. Zou, S. Zhu, A. Ng, and K. Yu. Deep learning of invariant features via tracked video sequences. NIPS, 2012.

4 Appendix