# CS231A Course Project Proposal

Martin Raison
Stanford University
mraison@stanford.edu

Botao Hu
Stanford University
botaohu@stanford.edu

## 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.*

## Future Distribution Permission

The author(s) of this report give permission for this document to be distributed to Stanford-affiliated students taking future courses.

## 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 *et al.* [**?**], Pernici *et al.* [**?**], Nebehay *et al.* [**?**]). 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 subwindows 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 *et al.* [**?**]), BRISK (Leutenegger *et al.* [**?**]) and ORB (Rublee *et al.* [**?**]) could help speed up the template matching process.

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

## 3. Data & Evaluation

We plan on comparing our implementation with state-of-the-art methods such as the TLD framework [**?**] (marketed as Predator) and ALIEN [**?**] by using the same publicly available datasets, such as PETS2009[1] or the David Indoor sequence[2].

[1] http://www.cvg.rdg.ac.uk/PETS2009/a.html

[2] http://www.cs.toronto.edu/ dross/ivt/

Qualitative results will be provided through real-time tests with video sequences. Metrics such as precision or recall will be used for quantitative comparison purposes.

Apart from the above-mentioned datasets, other video sources (TV programmes, webcams[3], etc) can be used for experimenting the system with different kinds of content and video quality.

## 4. Readings

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

## 5. Acknowledgements

We would like to express our gratitude to Alexandre Alahi (Post-doc at the Stanford Computer Vision lab), who accepted to mentor our project.

## 6. Appendix

a) This project is shared with the CS229 course for all members of the team.

b) Since the two of us are enrolled in both CS231A and CS229, we will both contribute to the computer vision and the machine learning parts of the project

c) The portion of the project that is being counted for CS231A is the first of the two challenges highlighted in the Problem Statement section. More specifically, it consists in experimenting binary descriptors such as FREAK, BRISK or ORB to replace feature-extraction algorithms used in state-of-the-art methods.

[3] http://www.earthcam.com