Guiding Visual Surveillance by Tracking Human Attention

Ben Benfold and Ian Reid Department of Engineering Science University of Oxford

UNIVERSITY OF OXFORD

Introduction

Problem: How can we identify objects and events that might be of interest to an observer?

We would like to detect events such as road accidents, fighting and people collapsing. The performance of current automatic detection systems is far below that of a human.

Solution: Pedestrians know what is interesting so work out what they are looking at!

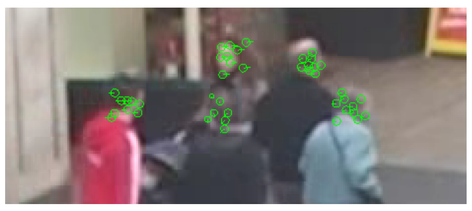
Monitored Scenes usually contain a number of people, who will naturally be interested in unusual occurrences around them. By estimating the directions in which they look we can measure the amount of interest in different areas of a scene.

Head Tracker

Predictions → ← **Observations**

Feature Tracking

The best features for estimating the head velocity are identified by combining observations using a dynamic Bayesian network and used to make robust velocity estimations.

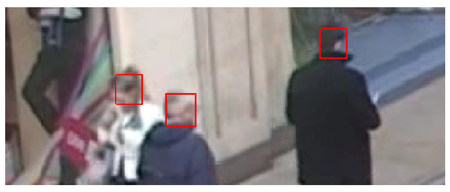


Kalman Filter

We replace the usual physics based prediction model with velocity estimates from the feature tracking. Observations from infrequent HOG detections prevent drift.

HOG Detections

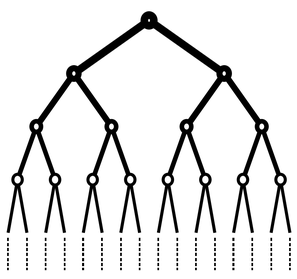
Using the same HOG+SVM algorithm that Dalal and Triggs used to detect pedestrians, we detect heads in a constrained region around the expected head position to provide absolute location estimates.

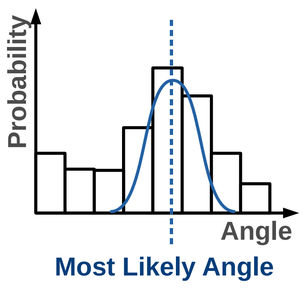


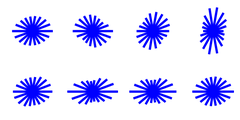
Gaze Direction Estimation

Randomised Ferns

The gaze directions for the tracked head regions are estimated using randomised ferns, a type of randomised tree classifier. Two decision types for the branches take advantage of different aspects of the image. Each leaf has a histogram representing the probability distribution over eight direction classes.

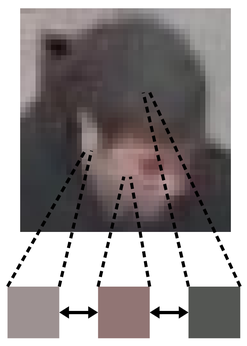






HOG Decisions

The magnitudes of two different gradient orientation histogram bins are compared



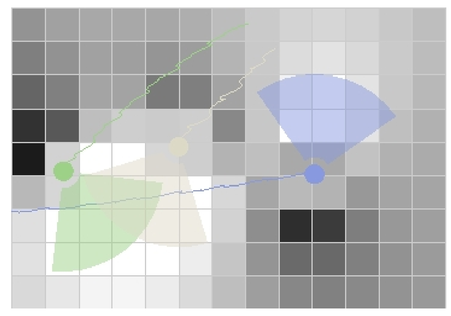
Colour Triplet Comparisons

One colour sample is compared to two others and the decision is made based on which is most similar.

Attention Maps

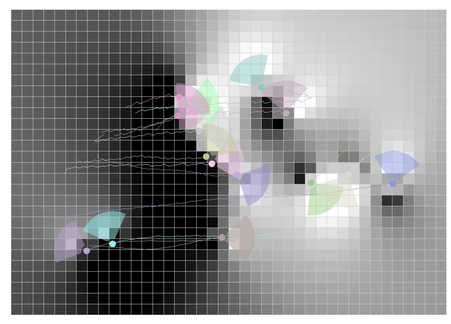
Estimating Attention

The gaze estimates for the tracked people are projected onto each square metre of the ground plane. Gazes away from the direction of motion are more significant so provide a larger response.

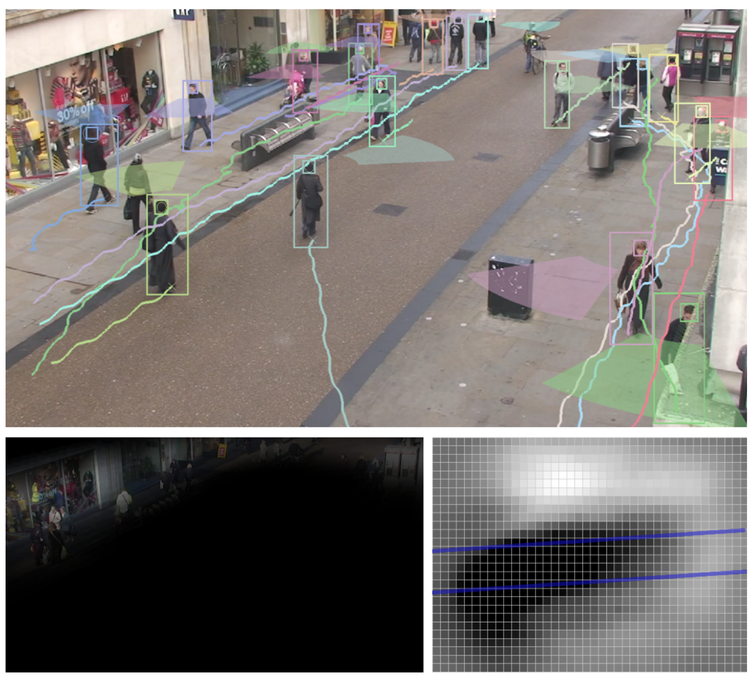


Building Maps

For the detection of static objects, gaze estimates are simply accumulated over time. To detect transient objects, intersections between gaze directions are found.



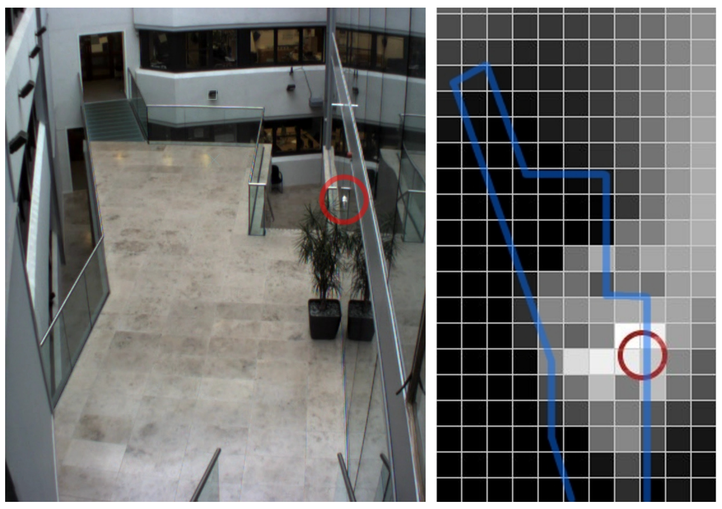
Experiment 1



Background Attention

In the first experiment pedestrians were tracked in a busy town centre street. Up to thirty pedestrians were tracked simultaneously and had their gaze directions estimated. A gaze map was built up over the full twenty-two minute video sequence, covering approximately 2200 people. When projected, the attention map identifies the shops on the left of the view as a common subject of attention.

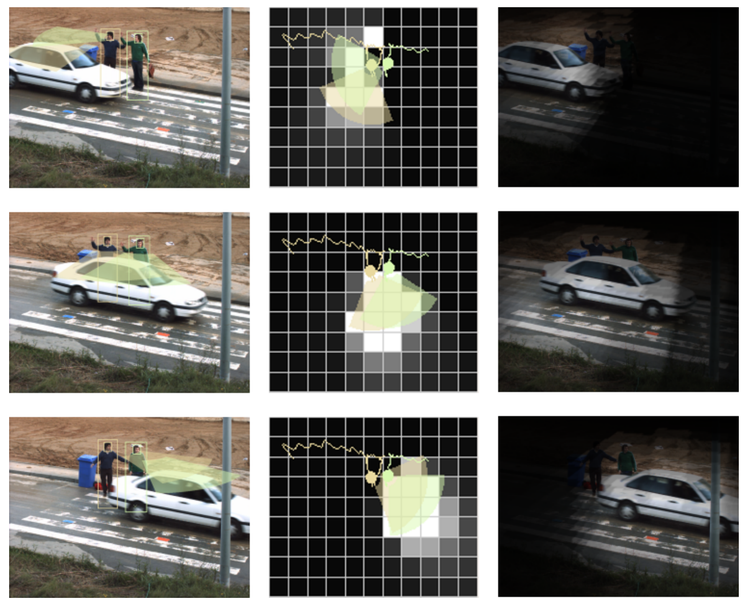
Experiment 2



Artificial Stimulus

The purpose of the second experiment was to artificially draw the attention of pedestrians to a particular location. To achieve this, a bright light was attached to the wall at the location indicated by the red circle in the above images. The blue lines show the outline of the floor. For this experiment the attention map was generated by taking the difference between the attention received both with and without the light stimulus to correct for the stimuli normally present in the scene. A total of 200 minutes of video were analysed and 477 people were tracked.

Experiment 3



Transient Object

The aim of the third experiment was to identify a transient subject of attention. To resolve the ambiguities caused by not knowing the distance between pedestrians and the subject of their attention, the gaze estimates from both people were multiplied and combined over a sliding window of three frames. The resulting intersection correctly identifies the car as the subject of attention when projected back onto the video.

Ben Benfold (bbenfold@robots.ox.ac.uk)
Ian Reid (ian@robots.ox.ac.uk)
Department of Engineering Science, University of Oxford, OX1 3PJ, Oxford, UK
http://www.robots.ox.ac.uk/ActiveVision/

This research was supported by Oxford Risk and the EU project HERMES (IST-027110)