University Defence Research Centre (UDRC)
In Signal Processing

Sponsored by the UK MOD

[C5] Real-Time Multi-Modal Person Tracking
Theme: Detection, Localisation & Tracking Theme

PI: Tania Stathaki, Imperial College

Researchers: A. Dulai

Project Objectives

•The aim of this project is to create a **real-time person tracking system**.
•Tracking people in real-time using video presents a variety of challenges, due to variability in lighting, appearance, visibility and scale.
•Existing approaches to person detection are slow, and most existing methods for tracking require manual initialisation.
•This project will also combine video information with information from other modalities.

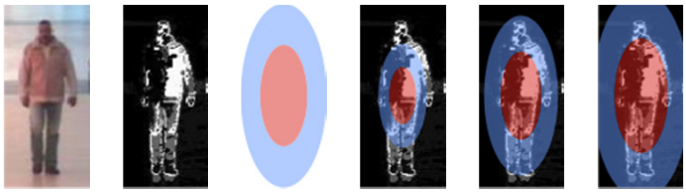
Technical Work

Mean-shift is a very general iterative technique for finding the local mode of a distribution.

It can be adapted for use in video tracking. An object can be identified by its **colour histogram**, and mean-shift can be used to find the best match in the next frame without performing an exhaustive search.

Mean-shift tracking algorithms have been shown to be suitable for tracking a wide variety of objects, such as faces, people and cars. This flexibility is due to the fact that a target is characterised by its global colour properties, which tend to remain unaffected by changes in a target's pose and orientation.

Mean-shift tracking through scale is non-trivial. In the course of this project, a novel method for scale adaptive mean-shift tracking has been developed. It relies on masking a back-projected image with a novel kernel function. The figure below shows (from left to right) (a) an original image, (b) the back-projected version, (c) a schematic representation of the kernel where the **red** area represents negative values, and the **blue** area represents positive values (d) a back projected image that is masked by a kernel that is too small, (e) masked by a kernel that is the correct size, and (f) masked by a kernel that is too large. The mask can be used to calculate the correct scale.



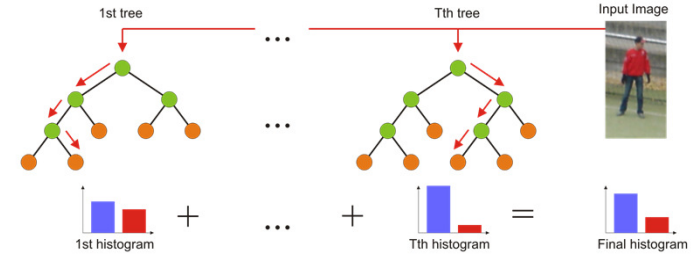
(a) (b) (c) (d) (e) (f)

In this project, the use of **randomised decision forests** for **people detection** has been investigated. Randomised decision forests are ensembles of decision trees that have been trained using **randomly selected features** on **random subsets** of the training data. They have been used for a variety of tasks in the field of computer vision, and have the advantage of **short training times** and **rapid execution speeds**.

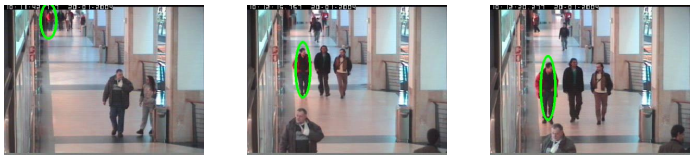
Randomised decision forests are trained via supervised learning, and for training a person detector, labelled images of people and 'non-person' images are required. **Edge orientation features** have been used, as edge features have been shown to be effective for person detection in past research.

A tree consists of split nodes (shown in the diagram below in **green**) and leaf nodes (shown in the diagram in **orange**). Each split node extracts a feature from the image window and tests it against a learned threshold. If the feature value is less than the threshold, then the left branch of the tree is chosen, otherwise the right branch is chosen. The tree is traversed in this fashion until a leaf node is reached, which returns a probability distribution across the object classes.

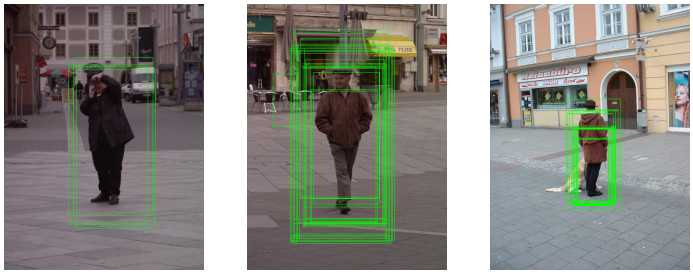
An image window is evaluated on every tree in the forest, and the final probability distribution that is returned is the average across all the trees.



The images below show a person being tracked through scale using the proposed method.



Some initial results for the random decision forest detector (without non-maxima suppression) are shown below. Initial results on the INRIA positive test data set indicate an accuracy of 77%.



Summary

We have proposed a method for mean-shift tracking through scale that is considerably more efficient than existing methods. This technique is able to track people as they move through scale in real-time.

We have explored the use of randomised decision forests for people detection. Edge orientation features have been used to discriminate between person and non-person images. Our results are promising.

Exploitation

The research produced by this project has a variety of potential applications:

•Surveillance technology would benefit greatly from the ability to recognise targets of interest.
•The approach could be extended to other target categories.
•The ability to integrate information from multiple modalities can yield better information than that of a single modality.

[dstl]

MINISTRY OF DEFENCE

EPSRC
Engineering and Physical Sciences Research Council

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) and the MOD University Defence Research Centre on Signal Processing (UDRC).