

Training a Tool To Recognise Behaviour From Video

David Mawdsley₁, Aitor Apaolaza₂, Robert Haines₁, Caroline Jay₂

Research IT(1) / School of Computer Science(2), University of Manchester

Introduction

Video offers a rich source of information about behaviour, but “coding” its content for further analysis is a time consuming and manual process.

We are developing an “Automated Behavioural Coding” tool [1] to combine the skills of humans and computers to transform the speed and accuracy of coding behaviours from video.

Experimental Setup

We use a “spot-the-difference” experiment to develop our approach. 19 participants had to identify differences in images shown simultaneously on a wall mounted TV and a tablet placed on the table in front of them (part 1), or which they were allowed to hold as they wished (part 2). Our aim was to determine whether or not the participants were looking at the tablet. We compare our approach to manually encoded data, which serves as our ground truth.

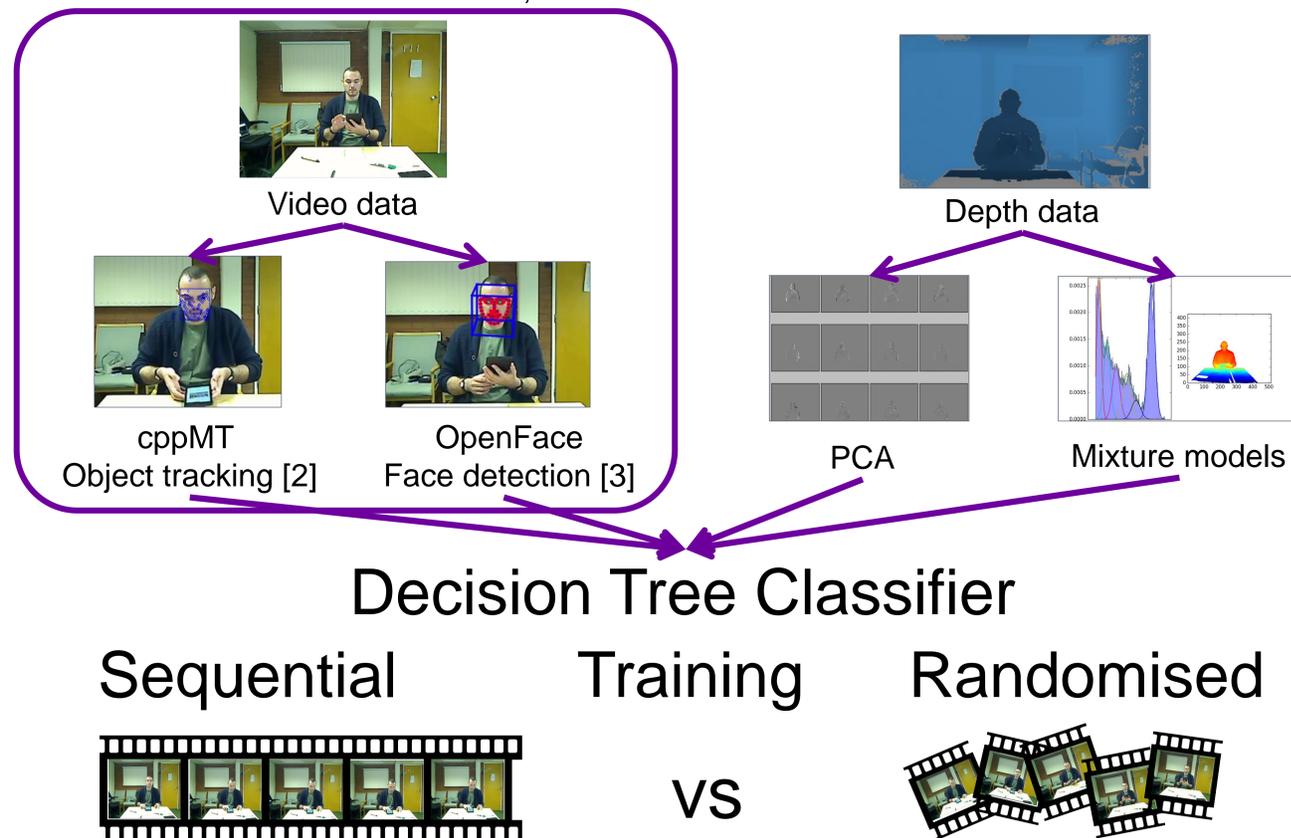
Colour video and depth data were collected, from cameras in front of and to the side of the participant.

Our analysis so far has focused on the front facing colour video.

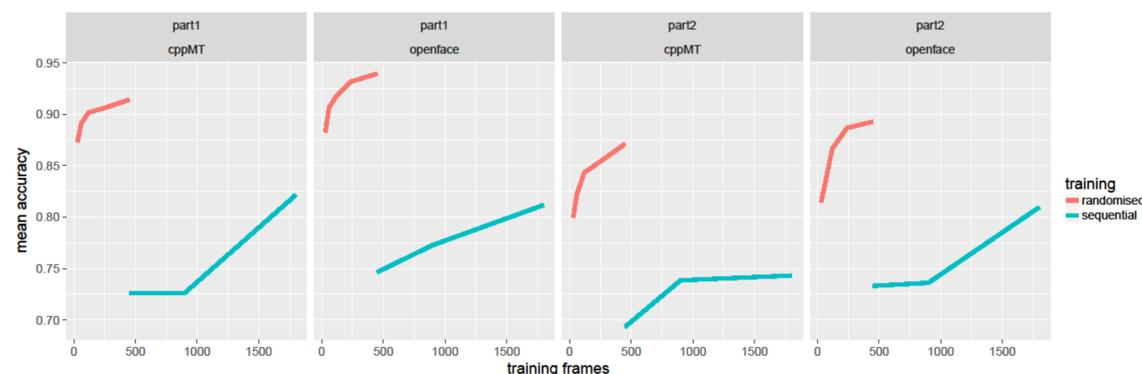


Workflow

We are exploring several data sources and dimensionality reduction techniques, which give data we can pass to a classifier to predict behaviour. The classifier can be trained using sequential frames from the start of the video, or frames chosen at random:



Results



Reproducibility

Reproducibility is an important part of our analysis and publishing pipeline:

- Each stage of the analysis runs in its own Docker container.
- We are writing a paper using Knitr [4], which combines LaTeX and R code.
- A Makefile calls the Docker containers to run the analysis steps, compiles the manuscript and handles the dependencies between them.

Discussion

Face detection and object tracking perform, on average, similarly, though this masks some substantial differences for individual participants.

Selecting training frames at random substantially improves the performance of the classifier and reduces the number of frames needed for training. Our tool can be used without ground truth data, where the user classifies a selection of randomly selected frames. We can obtain an estimate of the classifier’s performance without ground truth data using cross validation.

Work to include depth data, and to combine multiple data sources in the classifier is ongoing.

References

1. A Apaolaza et al., *ABC : Using Object Tracking to Automate Behavioural Coding*. CHI Ext. Abs. on Human Factors in Computing Systems 2766-2773, 2016
2. G Nebehay and R Pflugfelder. *Clustering of Static-Adaptive Correspondences for Deformable Object Tracking*. Comp. Vis. and Patt. Recog. June 2015
3. T Baltrusaitis, P Robinson and L-P Morency. *OpenFace: an open source facial behavior analysis toolkit*. IEEE Winter Conference on Applications of Computer Vision 2016
4. Y Xie (2016). *knitr: A General-Purpose Package for Dynamic Report Generation in R*. R package version 1.15.1.