ABC: Using Object Tracking to Automate Behavioural Coding

Aitor Apaolaza[1], Robert Haines[1], Amaia Aizpurua[1], Andy Brown[2], Michael Evans[2], Stephen Jolly[2], Simon Harper[1], Caroline Jay[1]

[1]University of Manchester, UK; [2]BBC R&D, Salford, UK

Video data of people interacting with devices contains rich information about human behaviour that can be used to design or improve user experience. As a first step, it must be interpreted — or coded — into a form that can be analyzed systematically. The coding process is currently performed manually, and it can be slow and difficult, and biased by subjectivity. This is particularly problematic when trying to obtain data that should be objective, such as the movements of a user in relation to a device. We describe Automated Behavioural Coding (ABC), an open source object tracking technique designed to log user and device movements, and then output positional data that can be used to model interaction. We validate the technique in a study of dual screen TV viewing, and show that the ABC tool is able to correctly classify the direction of gaze to the TV or tablet up to 95% of the time, in a fraction of the time it takes to capture this data manually.

Workflow

Traditional video recordings are employed as input.

Manual annotations of attention were used as ground truth.

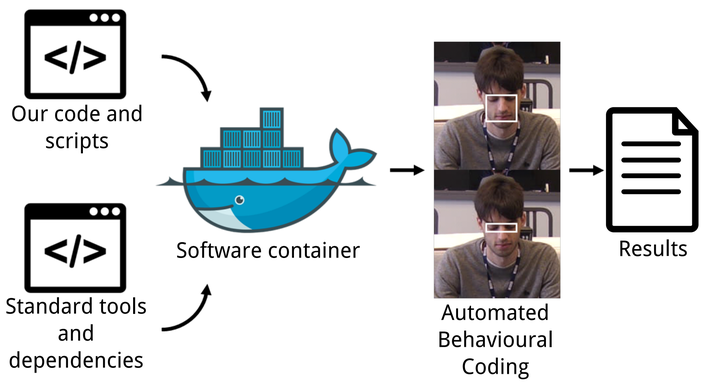
Object tracking provides position and movement of participants' face and eyes.

It was found that obstructions – e.g. touching the face – affected the stability of the tracking.

Movement data from the object tracking is used to predict location of attention.

The model was trained with between 1 and 10 minutes of manually annotated video.

Reproducibility



To make it easier for ourselves and other investigators to use our automated pipelines, we have used software containers to wrap up each stage of our method with everything needed for it to run. This guarantees that our tools will always run the same way, regardless of the environment they are running in, and ensures that any research is as reproducible as possible.

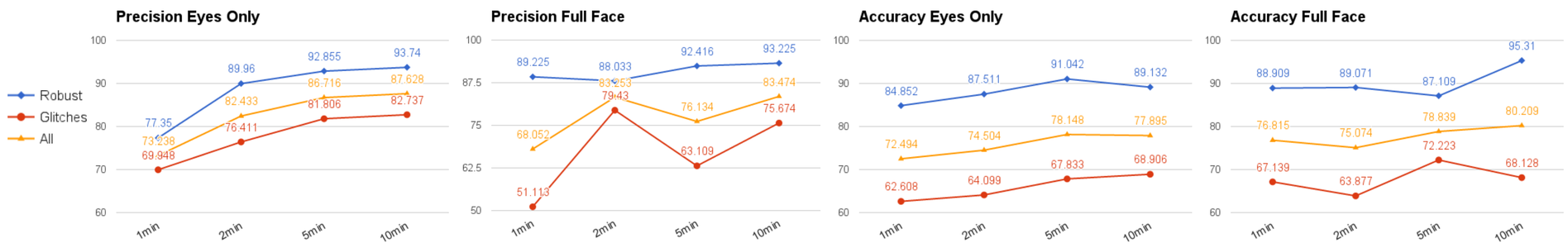
Results

Inputs have been categorised according to the stability of the object tracking:

Robust: when the object tracking showed no visible problem. The object of interest remains within the bounding box throughout the video, with the centre of the box located over it.

Glitches: periods of time where the object tracking is considered to be erratic.

All: all users.



Precision of predicting attention towards the iPad

Accuracy of attention prediction

Varying lengths of training

- The prediction model has been trained using varying lengths of annotated input, simulating partially annotated video recordings.

Observations

- Object tracking of the eyes is superior to full face tracking in instances where the face is occluded.
- 5 minutes of manually annotated videos are enough to obtain an 80% precision.
- When object tracking was robust, 2 minutes of manually annotated video was sufficient to classify attention over the full 20 minute video with nearly 90% accuracy.

Traditional approach

Requires thorough manual analysis of the entire video recording.

Cumbersome
Researchers need to watch the video at a slow playback speed.

Subjective
Manual annotations are dependent on researchers.

Coarse-grained categories
Observation can only model the interaction into qualitative categories.

Proposed approach

Supports automatic annotation of attention, requiring only a few minutes of manual annotation.

Automatic
With a significantly shorter annotation time ABC provides accurate attention predictions.

Objective quantifiable metrics
Precise quantifications of direction, velocity, and spatial relationships.

Acknowledgements
We would like to thank the EPSRC and the BBC for their funding contribution to this work via grants EP/K503782/1, EP/H500154/1, and EP/M017133/1.