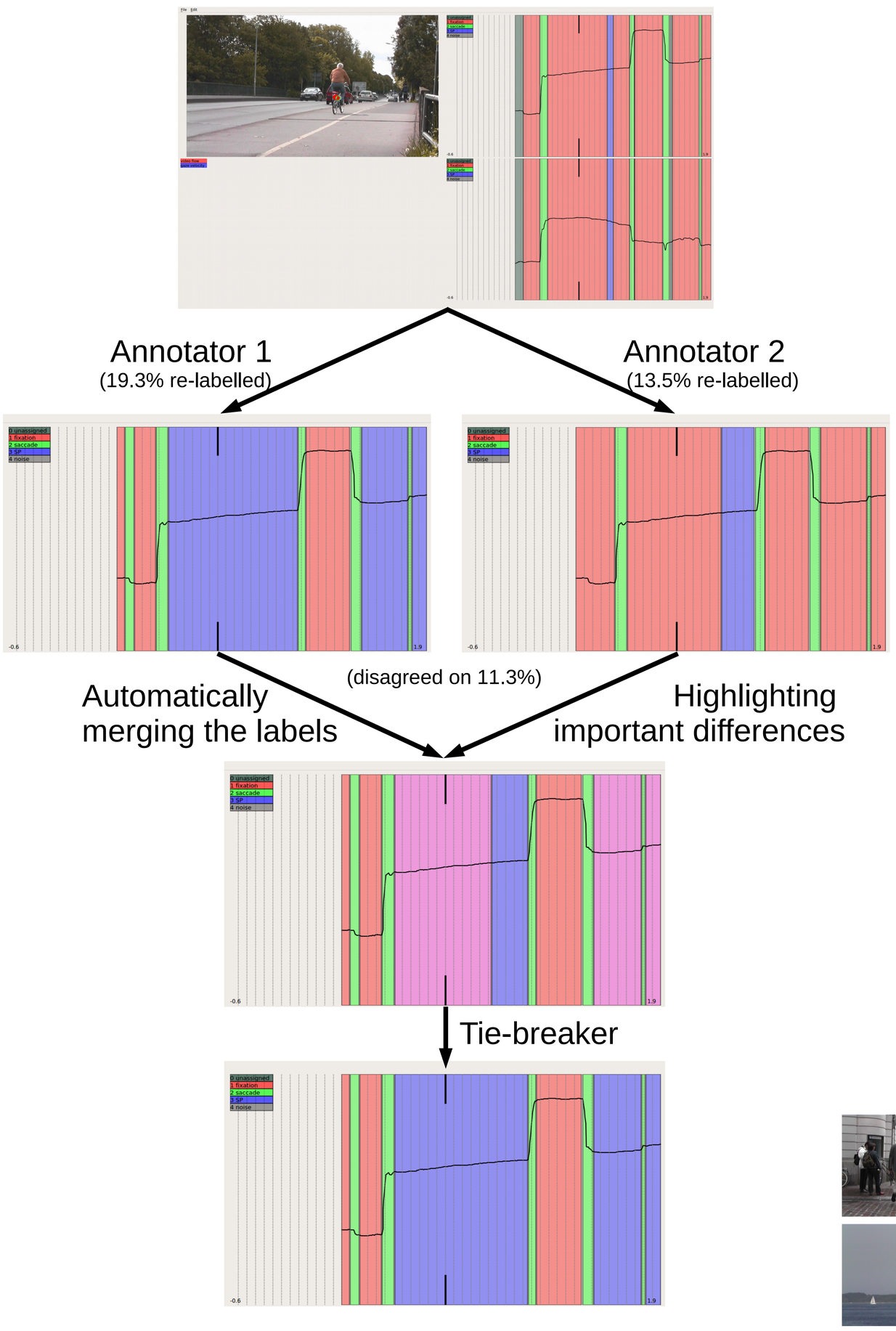
Manual & Automatic Detection of Smooth Pursuit in Dynamic Natural Scenes

Mikhail Startsev[1], Ioannis Agtzidis[2], Michael Dorr[3]
Technical University Munich
[1] *mikhail.startsev@tum.de*, [2] *ioannis.agtzidis@tum.de*, [3] *michael.dorr@tum.de*

To understand gaze behaviour, we need to abstract from the raw point-of-regard data and segment the gaze trace into eye movement types. For static stimuli, these are typically limited to fixations and saccades, but dynamic stimuli may induce smooth pursuit (SP) as well. Detecting SP on naturalistic videos is challenging because the targets and their trajectories are unknown a priori, the episodes may be short (average uninterrupted SP episode in hand-labelled data lasts 0.41s) and have speeds not much greater than both oculomotor and tracker noise around fixations. In this work we further evaluate our previously developed algorithm that uses information from several observers to address these challenges, which showed excellent performance in our preliminary evaluation. We now collected a manually annotated “ground truth” for the entire GazeCom dataset, on which our detection algorithm achieves precision and recall of 74.2% and 46.4%, respectively. As part of the pipeline, we also detect fixations, with precision and recall of 91.3% and 90.2%, respectively.

I. Hand-labelling Pipeline

To speed up the labelling process, the recordings are pre-labelled by a set of simpler approaches (92.7% labelled):



GazeCom data set:
- 18 clips
- 20s each
- over 4.5 viewing hours in total
- 47 observers per video on average
- full manual annotation of eye movements

II. Results

In this work we present the full-dataset evaluation **(over 4.5 hours of manually annotated viewing time)** of our detection algorithm against previous state-of-the-art approaches:

	Smooth Pursuit	Fixations
[Larsson et al. 2015]	F1: 0.459 Precision: 0.576 Recall: 0.382	F1: **0.912** Precision: 0.872 Recall: **0.956**
[Berg et al. 2009]	F1: 0.420 Precision: 0.509 Recall: 0.358	F1: 0.884 Precision: 0.900 Recall: 0.868
Ours	F1: **0.571** Precision: **0.742** Recall: **0.464**	F1: 0.908 Precision: **0.913** Recall: 0.902

We see similar results if we only consider the parts of the data set where the initial label was changed by the annotators (after tie-breaking, 18.5%).

III. Data set & Tools

We process data as ARFF files, relying on attributes *time, x, y* and (optionally) the tracker *confidence*. The annotated GazeCom recordings are provided in this format already. We also supply simple converters from SMI and EyeLink data to ARFF.

The data sets used in this work, the Python implementation of our detection algorithm and the Qt-based hand-labelling tool are publicly available at http://michaeldorr.de/smoothpursuit