Supplementary Material for 'Using body-anchored priors for identifying actions in single images'

This manuscript contains the supplementary material for the paper: 'Using body-anchored priors for identifying actions in single images'. Section 1 describes the movies, provided within the current supplementary. The movies show examples of deformable part detection, namely the face, the hand and the elbow, and also examples of static action recognition (performed on single images). The static action recognition method was applied on two datasets: '12/10' dataset and Gupta *et al* dataset [10]. The aforementioned datasets are described in the paper. In addition, figure 1 depicts several examples of interesting action recognition failures.

1 Movies

Three movies are provided together with this supplementary material. The first two movies contain sample results of action recognition from static images (frames) from two datasets (only successfully recognized frames are shown). The first dataset, titled '12/10', consists of 120 videos of 10 people performing 12 similar transitive actions, namely drinking, talking on the phone, scratching, toasting, waving, brushing teeth, smoking, wearing glasses, eating with a spoon, singing to a microphone, and also drinking without a cup and making a phone call with a bottle. The second dataset, provided by Gupta *et al* [10], consists of 46 videos of 9 people performing 6 transitive actions, namely drinking, spraying, answering the phone, making a call, pouring from a pitcher and lighting a flashlight. The third movie contains examples of detection of the hand, the face, and the elbow on the '12/10' dataset.

All the action recognition experiments conducted are performed in leave-one-out manner, where all the frames that contain a test person are excluded from the training. The details of the experiments and the quantitative evaluation of the results can be found in the paper.

To view the movies it is necessary to have the Xvid codec installed. The codec can be found within K-lite codec pack:

http://www.codecguide.com/download_kl.htm

The movies illustrating the results of static action recognition experiments performed on the two datasets are:

1. **'12_actions_10_people_Xvid.avi'** - shows sample successful static action recognition results on '12/10' dataset.

1

2. **'6_actions_9_people_Gupta_et_al_Xvid.avi'** - shows sample successful static action recognition results on Gupta *et al* dataset [10].

The movie containing examples of the parts detection is:

1. **'example_part_detections_12_10_Xvid.avi'** - shows example detections of the hand, the face, and the elbow on frames from '12/10' dataset.

In each frame of all the movies the detected location of the face of the person is marked by a magenta circle filled in by light blue. In the current method only the right hand of the person is being detected. A green star with red color inside marks the detected location of the right hand. The detected location of the elbow is marked by a yellow square.

In movies showing action recognition examples, the bar graph on the right shows the log-likelihood estimates obtained by the proposed method for all the considered actions.

2
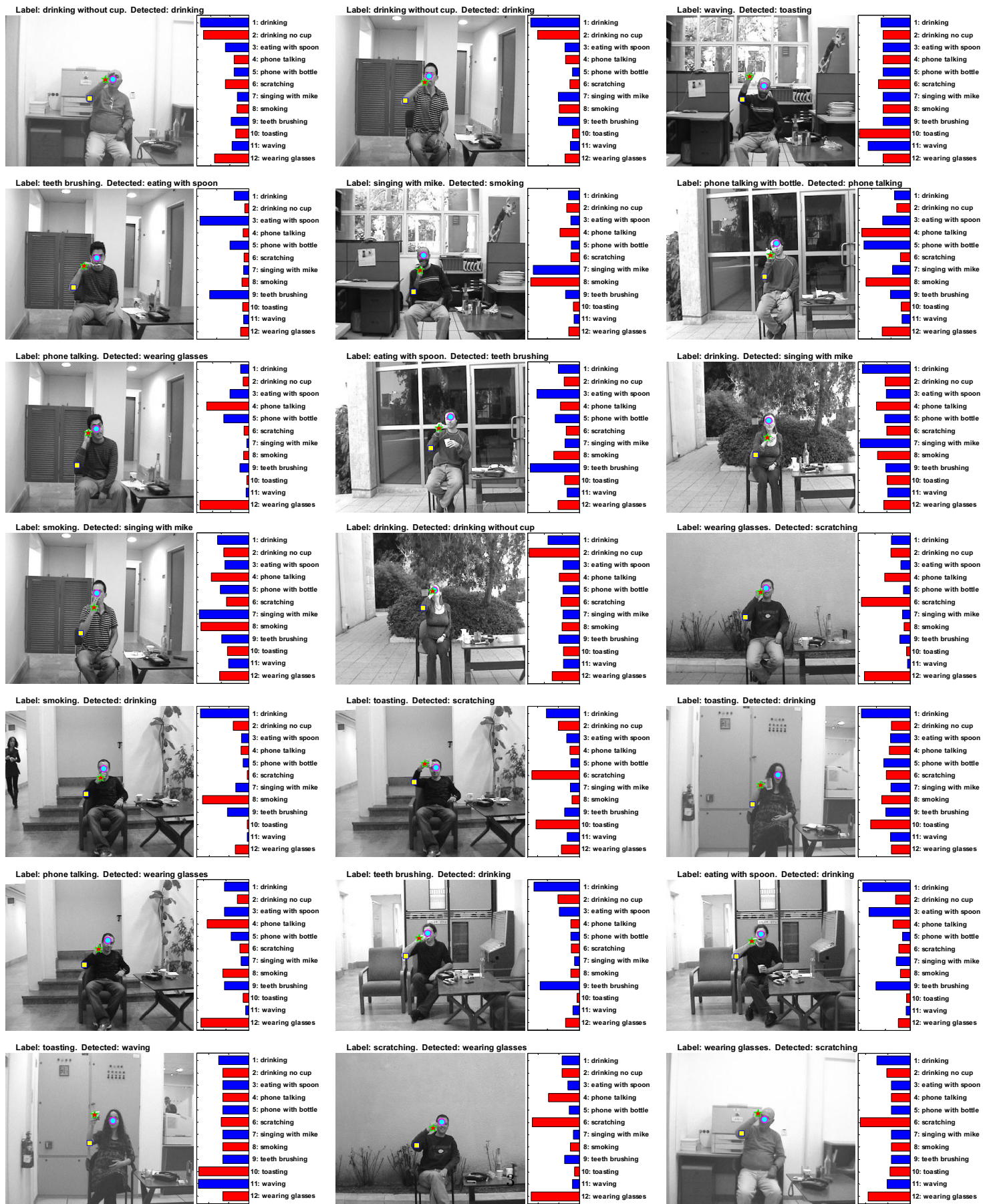
Figure 1: Some examples of interesting static action recognition failures. In order to see the fine details, please zoom in.