TurkerGaze: Crowdsourcing Saliency with Webcam based Eye Tracking

Anonymous ICCV submission
Paper ID 80

In the supplementary material, we demonstrated more examples collected by TurkerGaze of:

- gaze prediction error during image free viewing
- gaze prediction accuracy evaluated by a 33-point chart
- meanshift fixation detection results
- image saliency for fully annotated images
- center-bias-free saliency map for a panoramic image
- video saliency

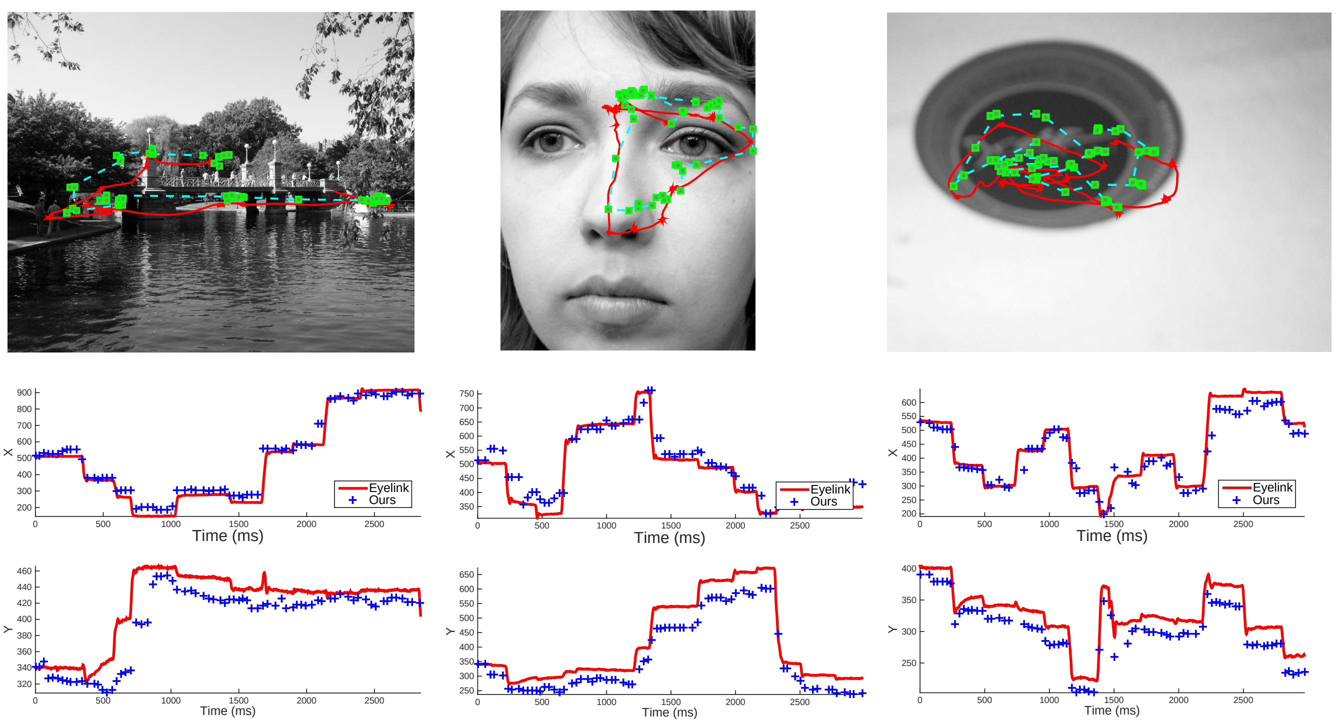

Figure 1: **Gaze prediction error during image free viewing.** On the top, the red curve shows the eye position recorded by the Eyelink eye tracker operating at 1000 Hz, and the green squares indicate the predictions of our system with a 30 Hz sample rate. On the bottom is position error in the x and y direction over time.

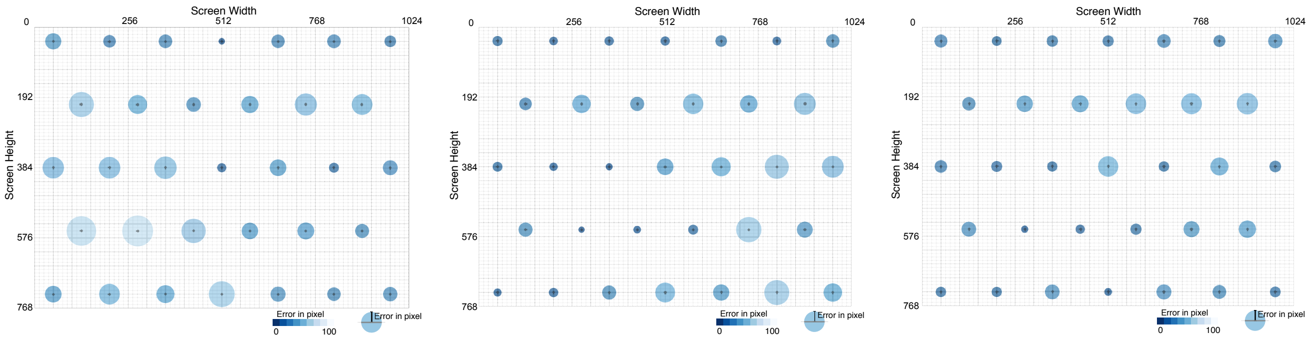

Figure 2: **Gaze prediction accuracy evaluated by a 33-point chart.** The radius of each dot indicates the median error of the corresponding testing position. Each plot corresponds to one experiment for one subject.

Figure 3: **Meanshift fixation detection results on two Judd images.** On left is the noisy, subsampled gaze data used for testing. On right are the detected fixations. Red circles indicate ground truth (detected in 240Hz data using code provided by Judd *et al*.) and yellow crosses indicate fixations detected in the noisy 30Hz data using our meanshift approach.

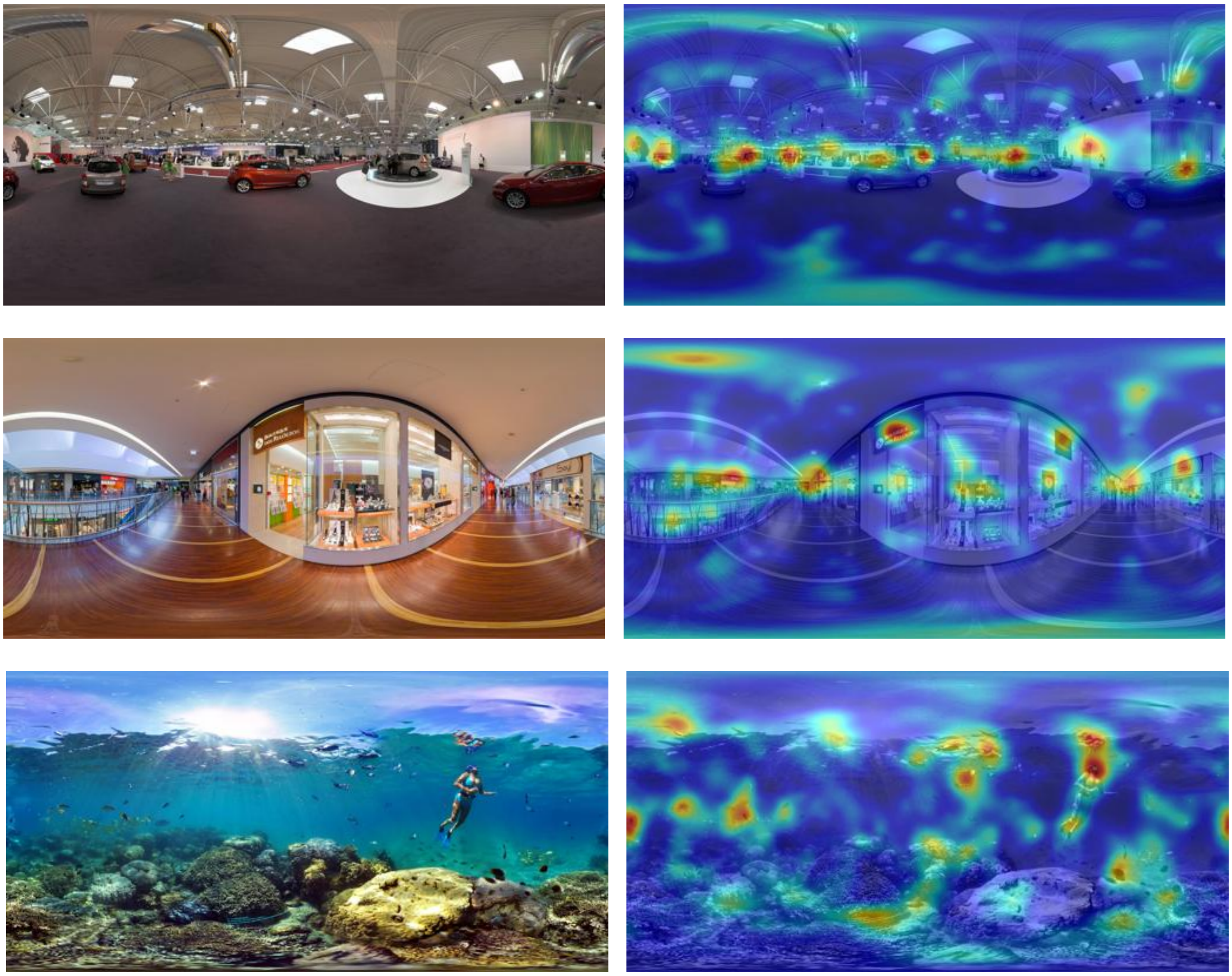

Figure 4: **Examples of center-bias-free saliency maps for panoramic images collected by our system.** We uniformly sampled overlapping views from the panorama, projecting each to a regular photo perspective, and collected gaze data for each projection. We then projected the saliency maps back onto the panorama and averaged saliency for overlapping areas. This gives a panoramic saliency map with no photographer bias.

Figure 5: **Examples of the image saliency data obtained by our system.** In a free-viewing task, users were shown selected images from the SUN database with fully annotated objects. From the raw eye tracking data, we used the proposed algorithm to estimate fixations and then computed the saliency map. This map could be used to evaluate the saliency of individual objects in the image. The 1st row: the original image; the 2nd row: the raw eye tracking data; the 3rd row: the fixations detected by the meanshift algorithm; the 4th row: the saliency map; the 5th row: the semantic object segments; the 6th row: salient objects.