

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

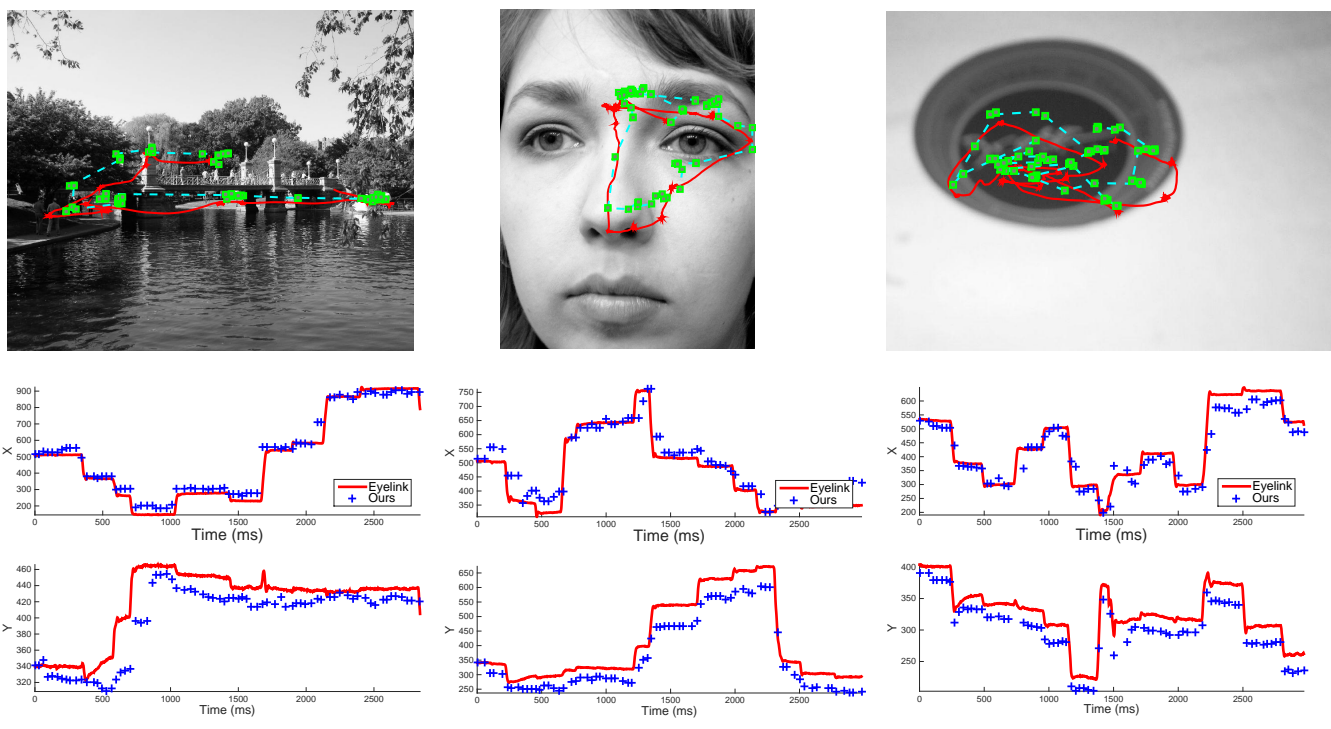


Figure 1: **Gaze prediction error during image free viewing.** On the top, the red curve shows the eye position recorded by the Eyelink eye tracker operating at 1000 Hz, and the green squares indicate the predictions of our system with a 30 Hz sample rate. On the bottom is position error in the x and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

108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161

162
163
164
165
166
167
168
1169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215

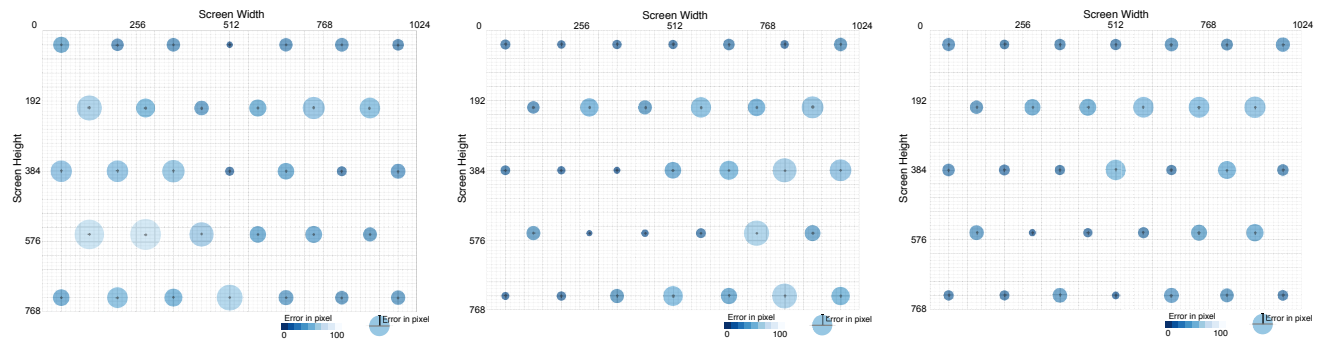


Figure 2: **Gaze prediction accuracy evaluated by a 33-point chart.** The radius of each dot indicates the median error of the corresponding testing position. Each plot corresponds to one experiment for one subject.



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.

216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269

270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323

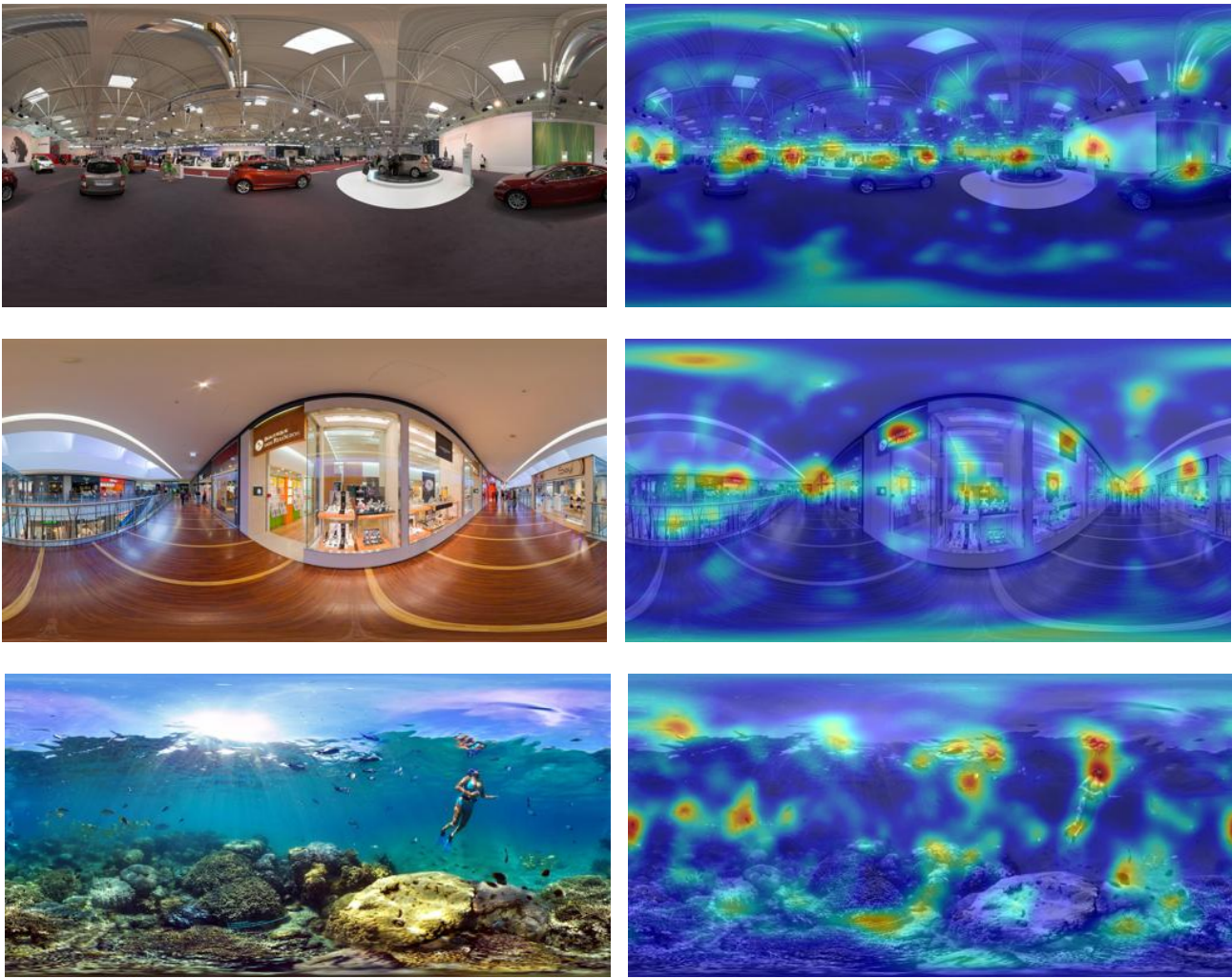


Figure 4: **Examples of center-bias-free saliency maps for panoramic images collected by our system.** We uniformly sampled overlapping views from the panorama, projecting each to a regular photo perspective, and collected gaze data for each projection. We then projected the saliency maps back onto the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

324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377

378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431



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.