Signboard Saliency Detection in Street Videos

Onkar Krishna[1], Kiyoharu Aizawa[2], Saskia Reimerth[3]

1 NTT Communication Science Laboratories, Japan, 2 The University of Tokyo, Japan, 3 Technische Universitt Wien, Austria

Abstract

Our main contribution is three-folded:

- ❑ We introduce a new eye gaze dataset collected over 30 observers viewing
- ❑ two street videos in free viewing and task viewing scenarios, the tasks being to look for a place to have either lunch
- ❑ we propose a metric for quantitative analysis of the collected eye gaze data to find differences in tendencies of gaze distribution for "signboards" in street videos during the free viewing and task viewing.
- ❑ Finally, we propose a modification to an existing video saliency algorithm, which can more accurately predict the relative ranking of signboards based saliencies during free viewing and task viewing.

Materials and Methods

Participants and Stimuli:

- ❑ A total of 30 participants attended the experiment including 15 university students (3 female, 12 male, age range 21-31, mean age 24.1) and 15 elderly subjects (4 female, 11 male, age range 66-80, mean age 73.1)
- ❑ All the subjects reported normal or corrected to normal vision.
- ❑ Two different street videos full of restaurants, each with a duration of two minutes thirty seconds, were used in the experiment.
- ❑ Tobii x2-60 eye tracker was used for recording the eye-gaze data, whereas the fixations and saccades were detected by the default Tobii fixation filter.

Procedure and Task:

- ❑ The 30 participants who took part in the study were divided into two groups of 15 participants each.
- ❑ In order to avoid repeating the same video in free viewing and task viewing mode for any one participant, one group of the participants watched the video in free viewing mode, whereas the another group watched the same video with the given task of finding a place to have lunch.
- ❑ Before the video stimuli began, participants were instructed to either view freely or to fulfill the task.

Figure: Illustration of the experiment setup, half of the participants were recruited from retirement job centers in Tokyo and rest of them were graduate students of Tokyo University.

Analysis

Preprocessing:

- ❑ To analyze different tendency of gaze landings during free viewing and task viewing we manually labelled the signboard.
- ❑ To perform the manual labelling, first, we labelled two instances of each signboard and then interpolated the label for the rest of the frames containing the same signboard.

Figure: Few instances of manually labelled signboard (only restaurants).

Three approaches were adopted to analyze the gaze landings around the signboards labeled in the previous step:

- ❑ First, we simply measured the total number of gazes that have landed on each signboard during the free viewing and the task viewing scenario for the whole duration of the signboard's appearance.
- ❑ Second, entropy-based metrics were developed to measure the explorativeness during free viewing and task viewing. We have generated and measured the entropy of 149 saliency maps (total 4473 frames divided by 30), showing the area explored during each second of the video in a single frame.

$$H(I_j) = \sum_l h_{I_j}(l) * \log(L / h_{I_j}(l))$$

Where I_j is the saliency map of the total gazes recorded during one second of viewing for which entropy is calculated and $h_{Ij}(l)$ is the histogram entry of intensity value l in image.

- ❑ The center bias for two different viewing modes can further be measured by measuring the Euclidean distance between the centroid of the average maps and the center pixel of the image.

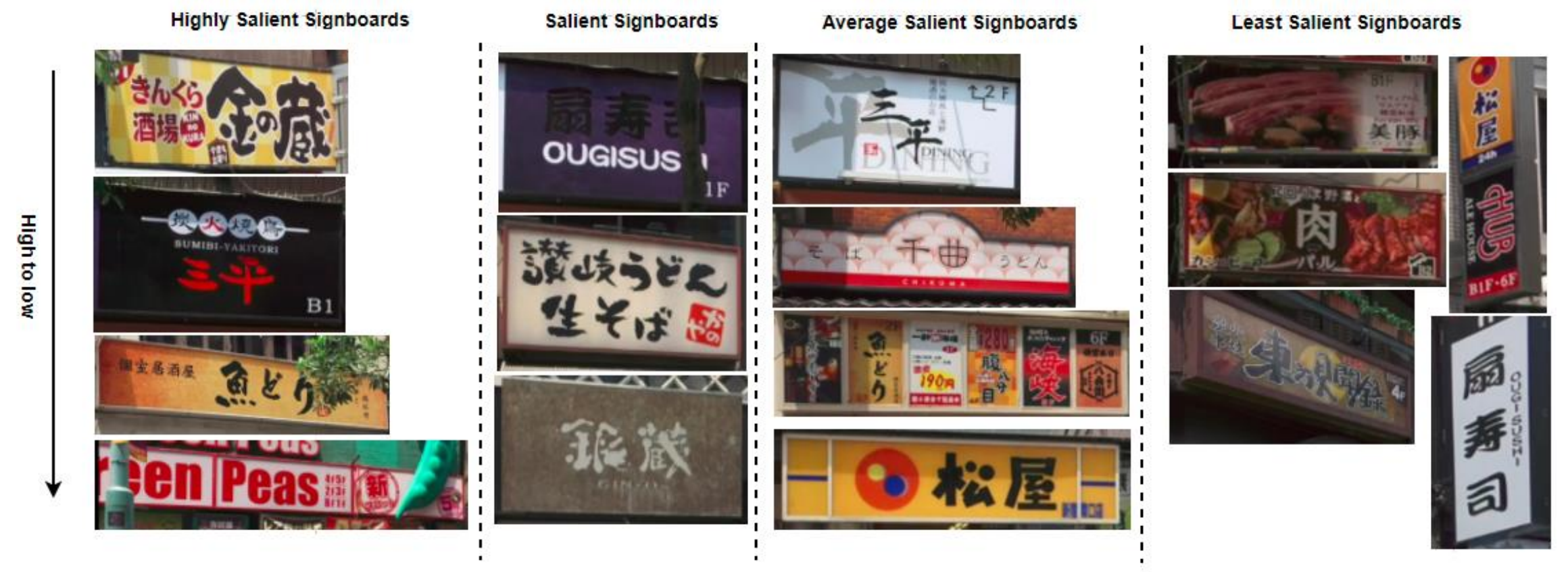


Figure: Ground Truth Ranking: Segmented signboards of restaurants in the street video and their corresponding saliency rankings during free viewing, generated from measuring total number of gaze landings (task viewing ranking: board 13, 5, 14, 11, 15, 12, 2, 1, 6, 7, 10, 16, 3, 8, 4, 9.)

Results

Explorativeness:

- ❑ The average score of the entropy value suggests higher explorativeness during task viewing than free viewing (task viewing - 1.94, free viewing - 1.40).
- ❑ The one-way ANOVA showed an effect of a task in scene exploration tendencies, F(1, 148) = 22.13; p < 0.001.

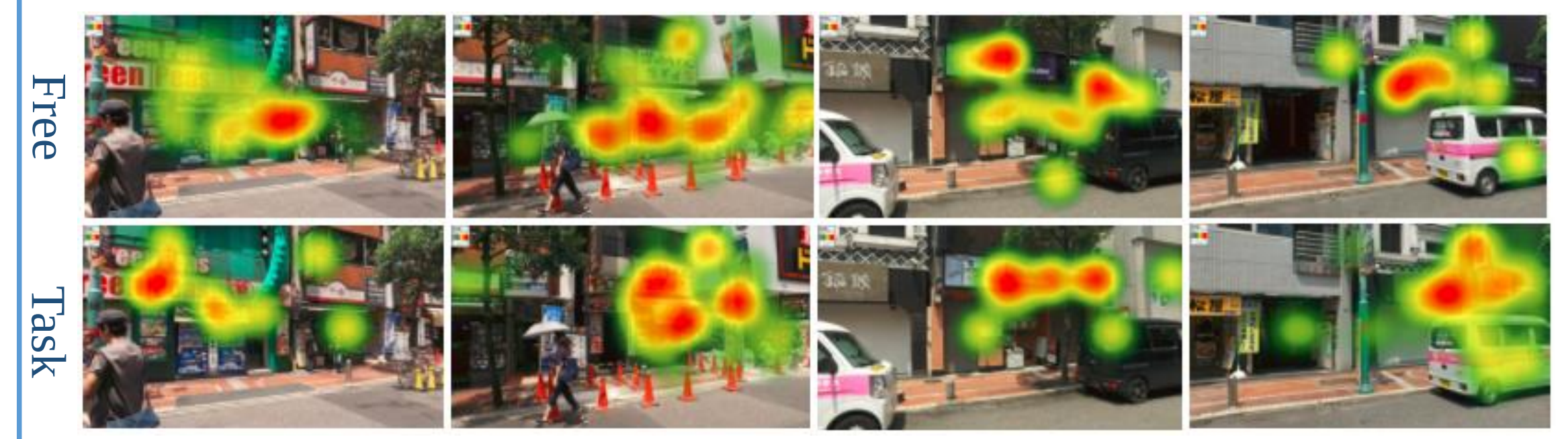
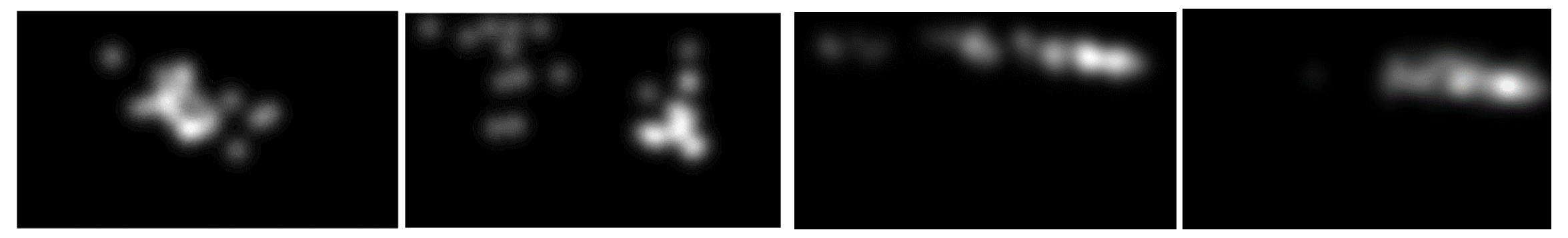


Figure: Visualization of different tendencies of explorativeness in free viewing and task viewing scenario.

Center Bias:

- ❑ The higher Euclidean distance for task viewing suggests the lower center bias in the task viewing scenario compared to the free viewing scenario (203 and 267 are the Euclidean distances in pixels for free viewing and task viewing).



(a) Free Viewing (b) Task Viewing (c) Free Viewing (d) Task Viewing

Figure 4: (a, b) Average tendency of the rate of video exploration during free viewing and task viewing. (c, d) Different tendency of explorativeness around board 14 only during free or task viewing

The proposed analysis suggests three major findings:

- ❑ We generated the ground truth rankings of the signboards based on the eye-gaze data collected in the free viewing and the task viewing scenario.
- ❑ Secondly, the explorativeness results indicate a higher exploration tendency during task viewing compared to free viewing.
- ❑ Lastly, we discovered a higher center-bias for free viewing than task viewing.

Saliency Model Selection

We first evaluated the performance of existing video saliency algorithms in predicting the gazes landed on the signboards.

Table 1: Prediction accuracy of different algorithms for the gaze over the full duration for free viewing and task viewing.

	GBVS	s_map	m_map	e_map	Itti
Free	0.7969	0.7626	0.6996	0.7429	0.7961
Task	0.7717	0.7350	0.6836	0.7045	0.7483

Proposed Saliency Model

- ❑ Motivated by the prediction accuracy of Itti's model, we applied the recommendations from the analysis' results to upgrade Itti's model with motion features [1] to predict the signboard saliencies
- ❑ To make our model adapt to differences in explorativeness during free viewing and task viewing we focused on feature scale selection.
- ❑ We identified the subsets of the feature map scales that best represented the different levels of details viewed by the observers during free viewing and task viewing.

$$Intensity = \bigoplus_{i=s}^{6} \mathcal{N}(Intensity_i)$$

$$Color = \bigoplus_{i=s}^{6} [\mathcal{N}(\mathcal{RG}_i) + \mathcal{N}(\mathcal{BY}_i)]$$

$$Orientation = \sum_{\theta \in \{0,45,90,135\}} \bigoplus_{i=s}^{6} \mathcal{N}(Orientation_i(\theta))$$

s is the starting index from where maps were taken, scale 1 (finer) to scale 6 (coarser). The experimental result shows that the following subset of the coarser scales *s* = 4, 5, 6 are suitable in free viewing, conversely the task viewing prediction accuracy improved for the finer scales *s* = 1, 2, 3.

Table 2: The signboard saliency scores (AUC score) for the lowest salient signboards (determined by gaze data) generated by different algorithms in free viewing and task viewing.

Board Ranking (Task)		12	15	13	11	Avg.
Board Ranking (Free)		14	16	13	12	
Board Name		3	8	9	16	
GBVS[2]	Free	0.73	0.53	0.65	0.70	0.65
	Task	0.76	0.63	0.52	0.70	0.65
Itti's[1]	Free	0.79	0.61	0.56	0.63	0.64
	Task	0.71	0.72	0.57	0.71	0.67
s_map[18]	Free	0.67	0.61	0.68	0.60	0.64
	Task	0.69	0.65	0.58	0.77	0.67
m_map[19]	Free	0.61	0.65	0.61	0.61	0.62
	Task	0.63	0.62	0.54	0.56	0.58
e_map[6]	Free	0.70	0.62	0.54	0.61	0.61
	Task	0.65	0.56	0.48	0.60	0.57
Ours	Free	0.70	0.51	0.49	0.63	0.58
	Task	0.71	0.55	0.45	0.56	0.56

Table 3: The signboard saliency scores (AUC score) for the highest salient signboards (determined by gaze data) generated by different algorithms in free viewing and task viewing (higher score is better).

Board ranking (Task)		8	7	1	4	
Board ranking (Free)		4	5	3	1	Avg.
Board Name		1	2	13	14	
GBVS[2]	Free	0.79	0.87	0.75	0.76	0.79
	Task	0.83	0.82	0.64	0.64	0.73
Itti's[1]	Free	0.82	0.79	0.88	0.84	0.83
	Task	0.82	0.75	0.71	0.71	0.74
s_map[18]	Free	0.81	0.74	0.88	0.79	0.80
	Task	0.75	0.72	0.66	0.68	0.70
m_map[19]	Free	0.66	0.67	0.57	0.64	0.63
	Task	0.69	0.62	0.61	0.57	0.62
e_map[6]	Free	0.76	0.72	0.75	0.74	0.74
	Task	0.77	0.72	0.61	0.61	0.67
Ours	Free	0.81	0.86	0.79	0.81	0.81
	Task	0.83	0.84	0.73	0.75	0.78

Conclusion

This paper presents a novel application of video saliency detection for ranking signboards within a street video based on the relative signboard saliencies.

The main contribution of this work is as following:

- ❑ Collection of eye-gaze data for 2 street videos for both free viewing and task viewing scenarios.
- ❑ Further, the proposal of a quantitative analysis method based on the rate of the explorativeness and center bias metrics.
- ❑ Finally those results were used in upgrading the basic saliency model for predicting signboard saliencies more accurately for free viewing and task viewing.

References

1. Laurent Itti, Christof Koch, and Ernst Niebur. 1998. A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on pattern analysis and machine intelligence 20, 11 (1998), 12541259.
2. Jonathan Harel, Christof Koch, and Pietro Perona. 2007. Graph-based visual saliency. In Advances in neural information processing systems. 545552.
3. Yuming Fang, Zhenzhong Chen, Weisi Lin, Chia-Wen Lin: Saliency Detection in the Compressed Domain for Adaptive Image Retargeting. IEEE Transactions on Image Processing 21(9): 3888-3901 (2012)
4. Yuming Fang, Weisi Lin, Zhenzhong Chen, Chia-Ming Tsai, and Chia-Wen Lin, A Video Saliency Detection Model in Compressed Domain. IEEE Trans. Circuits Syst. Video Techn. 24(1): 27-38, 2014.
5. Fang, Y., Wang, Z., Lin, W., and Fang, Z. (2014). Video saliency incorporating spatiotemporal cues and uncertainty weighting. IEEE Transactions on Image Processing, 23(9), 3910-3921.

Acknowledgements and Contact

We are thankful to VTEC laboratories Inc. for providing the video used in this study. We are also thankful to Dr. Go Irie, NTT Corporation for providing valuable insights.

Contact:

Onkar Krishna, Ph.D.

NTT Communication Science Laboratories, Japan

Email: o.krishna@lab.ntt.co.jp