Comparing Two Eye-Tracking Databases: The Effect Of Experimental Setup And Image Presentation Time On The Creation Of Saliency Maps

Ulrich Engelke, Hantao Liu, Hans-Jürgen Zepernick, Ingrid Heynderickx, Anthony Maeder

International Picture Coding Symposium

2010 Nagoya

28th Picture Coding Symposium, PCS2010, December 8-10, 2010, Nagoya, Japan

COMPARING TWO EYE-TRACKING DATABASES: THE EFFECT OF EXPERIMENTAL SETUP AND IMAGE PRESENTATION TIME ON THE CREATION OF SALIENCY MAPS

Ulrich Engelke[1], *Hantao Liu*[2], *Hans-Jürgen Zepernick*[1], *Ingrid Heynderickx*[2,3], *and Anthony Maeder*[4]

[1]Blekinge Institute of Technology, 371 79 Karlskrona, Sweden, E-mail: uen@bth.se
[2]Department of Mediamatics, Delft University of Technology, Delft, The Netherlands
[3]Group Visual Experiences, Philips Research Laboratories, Eindhoven, The Netherlands
[4]University of Western Sydney, Locked Bag 1797, Penrith South DC, NSW 1797, Australia

ABSTRACT

Visual attention models are typically designed based on human gaze patterns recorded through eye tracking. In this paper, two similar eye tracking experiments from independent laboratories are presented, in which humans observed natural images under task-free condition. The resulting saliency maps are analysed with respect to two criteria; the consistency between the experiments and the impact of the image presentation time. It is shown, that the saliency maps between the experiments are strongly correlated independent of presentation time. It is further revealed that the presentation time can be reduced without substantially sacrificing the accuracy of the convergent saliency map. The results provide valuable insight into the similarity of saliency maps from independent laboratories and are highly beneficial for the creation of converging saliency maps at reduced experimental time and cost.

Index Terms— Eye tracking experiments, visual saliency, correlation analysis, natural images.

1. INTRODUCTION

The human visual system (HVS) processes an enormous amount of information, which is considerably reduced by an integral property of the HVS known as visual attention (VA) [1]. The importance of VA models in digital image and video processing systems has increasingly been realised in past years [2, 3]. The range of applications for VA models is wide and includes video surveillance, object tracking, and visual quality assessment. Especially the usefulness of VA for quality assessment has recently received great interest [4–6]. The ground truth for VA model design is typically obtained from eye tracking experiments, where the gaze of human observers is recorded while viewing the presented stimuli. The gaze patterns are further processed into saliency maps (SM), which are then used for the design and validation of the VA models. Several problems arise in this context that have not been sufficiently addressed in the literature.

Firstly, little is known about SM being independently created from eye tracking data of different laboratories. As such, SM based on eye tracking data from one laboratory might constitute a largely different ground truth as compared to SM from another laboratory. The lack of publicly available eye tracking data, and thus, the deficit of comparison between eye tracking experiments from different laboratories further aggravates this shortcoming. Secondly, in eye tracking experiments with still images, the duration of the presentation has not been extensively studied and there is no common agreement as to how long an image should be presented to obtain valid SM. If the presentation time is too short, then valuable information regarding the attended regions might be missing in the SM. If the presentation time is too long, redundant information may be present in the SM, thus, unnecessarily prolonging tedious and expensive eye tracking experiments.

In this paper, we take a first step to address these problems. We present two similar eye tracking experiments conducted in independent laboratories using as stimuli the reference images from the LIVE image quality database [7]. Thus, the results may be of particular interest for the image quality research community, but also for anyone working in related fields that involve eye tracking. We conduct a correlation analysis on the SM to address questions like: Are SM consistent between eye tracking experiments from two independent laboratories? How long should images be presented as a compromise between SM convergence and experiment duration?

The paper is organised as follows. In Section 2, we introduce the eye tracking experiments. In Section 3, we provide a detailed correlation analysis of the SM from the eye tracking experiments. Finally, conclusions are drawn in Section 4.

2. EYE TRACKING EXPERIMENTS

The first experiment, referred to as E1, was conducted at the University of Western Sydney (UWS), Australia [5] and the second experiment, referred to as E2, was conducted at Delft University of Technology (TUD), The Netherlands [6]. The stimuli in E1 and E2 were the 29 publicly available reference images from the LIVE image quality database [7] which are listed in Table 1 with their original names.

978-1-4244-7135-5/10/$26.00 ©2010 IEEE

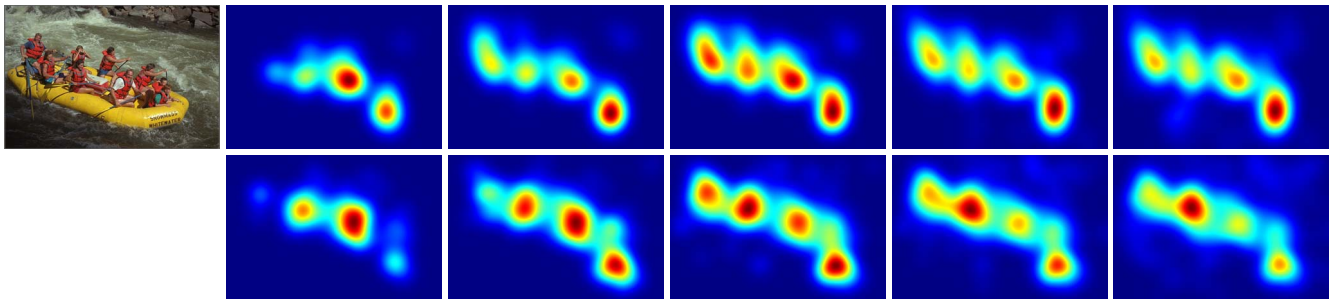

Fig. 1. SM for image 20 ('rapids') with increasing t from left to right and for experiments E1 (top row) and E2 (bottom row).

Table 1. Numbers and names of the LIVE reference images.

#	Name	#	Name	#	Name
1	bikes	11	house	21	sailing1
2	building2	12	lighthouse	22	sailing2
3	buildings	13	lighthouse2	23	sailing3
4	caps	14	manfishing	24	sailing4
5	carnivaldolls	15	monarch	25	statue
6	cemetry	16	ocean	26	stream
7	churchandcapitol	17	paintedhouse	27	studentsculpture
8	coinsinfountain	18	parrots	28	woman
9	dancers	19	plane	29	womanhat
10	flowersonih35	20	rapids		

2.1. Details of experiment E1

Fifteen non-expert staff and students of UWS participated in experiment E1, of which nine were male and six were female. The images were presented on a Samsung SyncMaster LCD monitor of size 19", with a native resolution of 1280×1024 pixels. An infrared video-based EyeTech TM3 eye tracker recorded the gaze patterns of the observers while viewing the images. The recording frequency was 45-50 gaze points per second (GP/s) at an accuracy of <1 degree of visual angle (dva). A 16-point calibration screen was presented before each session, to calibrate the eye tracker to each observer.

Before commencing the experiment, the participants were tested for visual acuity using the Snellen chart. The images were presented in random order at a viewing distance of approximately 60 cm. Each image was presented for 12 s with a mid-grey screen being shown in between for 3 s. The eye tracking experiment was conducted under task-free condition, meaning, that no particular instructions were given to the participants, other than to just view the images.

2.2. Details of experiment E2

Twenty non-expert students of TUD, twelve being male and eight being female, were recruited as participants for experiment E2. The stimuli were displayed on an Iiyama 19" CRT monitor with a resolution of 1024×768 pixels. The eye movements were recorded with an infrared video-based eye tracking system, the iView X RED by SensoMotoric Instruments. The iView X RED has a sampling rate of 50 GP/s and a gaze position accuracy of 0.5-1 dva. A chin rest was deployed to reduce head movements and ensure a constant viewing distance. A 9-point calibration screen was used for eye tracker calibration to the individual observers.

The images were displayed in random order at a viewing distance of approx. 70 cm. Each stimulus was shown for 10 s, with a mid-grey screen presented for 3 s between the images. The eye tracking experiment was conducted under task-free condition with the participants being asked to look at the images in a natural way, as they usually would look at them.

2.3. Post-processing of the eye tracking data

The gaze patterns were post-processed into SM by eliminating GP belonging to saccadic eye movements and only accounting for GP belonging to visual fixations. Individual SM were created for each image, based on the gaze patterns of all observers from the respective experiments. To identify the impact of image presentation time on the SM, we created 5 different SM for each image, taking into account the GP recorded through the first 2 s, 4 s, 6 s, 8 s, and 10 s. The corresponding SM are denoted as $SM_{E1}^{(t)}$ and $SM_{E2}^{(t)}$, $t \in \{2, 4, 6, 8, 10\}$, respectively, for experiments E1 and E2.

An example of a full set of SM for image number 20, named 'rapids' in the LIVE database, is shown in Fig. 1. The top and bottom row show, respectively, the SM for experiments E1 and E2 for all 5 considered presentation times t.

3. CORRELATION ANALYSIS

We assess the similarity between two SM $SM^{(i)}$ and $SM^{(j)}$ using the Pearson linear correlation coefficient

$$\rho_P = \frac{\sum_m \sum_n (SM_{mn}^{(i)} - \mu^{(i)})(SM_{mn}^{(j)} - \mu^{(j)})}{\sqrt{\sum_m \sum_n (SM_{mn}^{(i)} - \mu^{(i)})^2} \sqrt{\sum_m \sum_n (SM_{mn}^{(j)} - \mu^{(j)})^2}} \quad (1)$$

where $m \in [1, M]$ and $n \in [1, N]$, respectively, are the horizontal and vertical pixel coordinates, and $\mu^{(i)}$ and $\mu^{(j)}$ denote

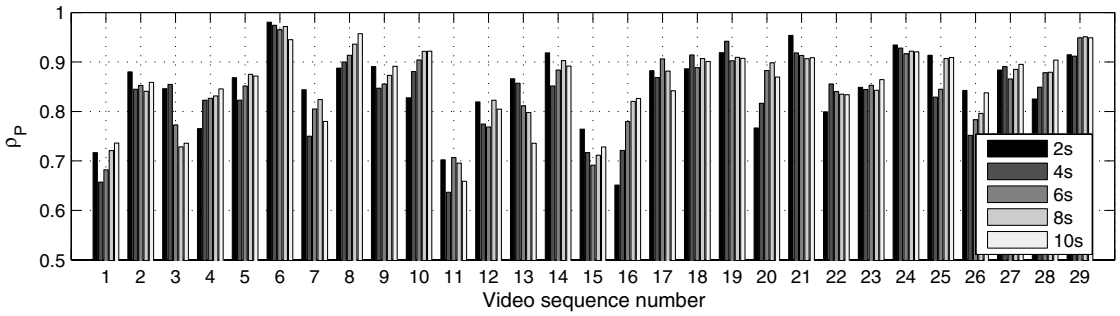

Fig. 2. Correlations ρ_P between $SM_{E1}^{(t)}$ and $SM_{E2}^{(t)}$ for all 29 images and 5 presentation times t.

the respective mean pixel values. The correlation coefficient ρ_P is computed in the range from -1 to 1, with a larger value corresponding to higher similarity between the SM.

3.1. Consistency between eye tracking experiments

From Fig. 1 one can observe that the SM between experiments E1 and E2 are fairly similar, especially for longer presentation times. To verify whether this is a general phenomenon that applies for all images, we present in Fig. 2 the correlations between $SM_{E1}^{(t)}$ and $SM_{E2}^{(t)}$ for all 29 images and 5 presentation times t. It can be seen, that the correlation between the SM of the two experiments varies, in general, more with the image content as compared to the presentation time. To provide further evidence, we present the correlations averaged over all 29 contents in Fig. 3 and averaged over the 5 presentation times t in Fig. 4. Figure 3 shows that the SM between the experiments are highly correlated with ρ_P around 0.84-0.85 for all t. The mean correlations in this figure also reveal that the correlations between $SM_{E1}^{(t)}$ and $SM_{E2}^{(t)}$ do not fluctuate much with t. The magnitudes of the corresponding standard errors indicate that the fluctuation amongst image contents is similar across t.

Figure 4 highlights, that the correlations between $SM_{E1}^{(t)}$ and $SM_{E2}^{(t)}$ are highly dependent on the image content, with the highest correlation being $\rho_P = 0.967$ for image 6 and the lowest correlation being $\rho_P = 0.68$ for image 11. The generally narrower standard errors for higher ρ_P indicate lower fluctuation across t. From visual inspection of the SM corresponding to all image contents, we found that the similarity of the SM between E1 and E2 is generally higher if the image contains highly salient regions, such as humans, faces, animals, and text. If no such salient region is present, the correlations were found to be lower. It was also found, that multiple salient regions, such as given by multiple humans in image 20, typically reduced the correlation to some degree.

3.2. Effect of image presentation time

The effect of image presentation time is assessed by computing ρ_P between the SM based on any two consecutive t,

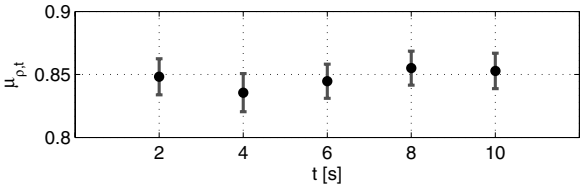

Fig. 3. Correlations between $SM_{E1}^{(t)}$ and $SM_{E2}^{(t)}$ averaged over all 29 images for the 5 presentation times t.

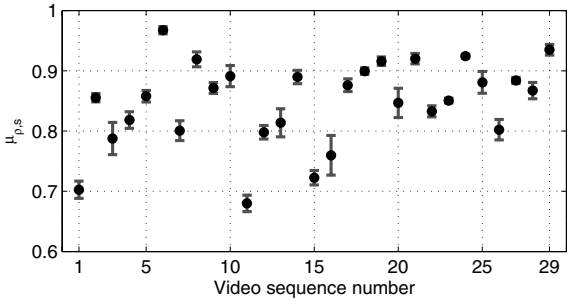

Fig. 4. Correlations between $SM_{E1}^{(t)}$ and $SM_{E2}^{(t)}$ averaged over the 5 presentation times t for all 29 images.

$SM^{(t)}$ and $SM^{(t+1)}$. These correlations are presented for all images and for E1 and E2 in Fig. 5. It can be seen, that almost exclusively for all images, the similarity between SM increases with longer presentation times t. It should be noted, that the incline in ρ_P is in most cases very similar between E1 and E2. Also, the correlation between $SM^{(2)}$ and $SM^{(4)}$ is often considerably lower than the other correlations.

The incline in correlations with t and the agreement between E1 and E2 is further highlighted by the correlations averaged over all image contents, as presented in Fig. 6. For both experiments E1 and E2 the similarity between SM clearly rises with t until some saturation effects occur, revealing, that the presentation time can be limited without sacrificing much of the accuracy of the converging SM. This result is in high agreement between the two experiments.

Figure 7 presents the average correlations between $SM^{(t)}$ and $SM^{(t+1)}$ for all images. For all contents, the correlations are very high, above 0.91. The large standard errors of some

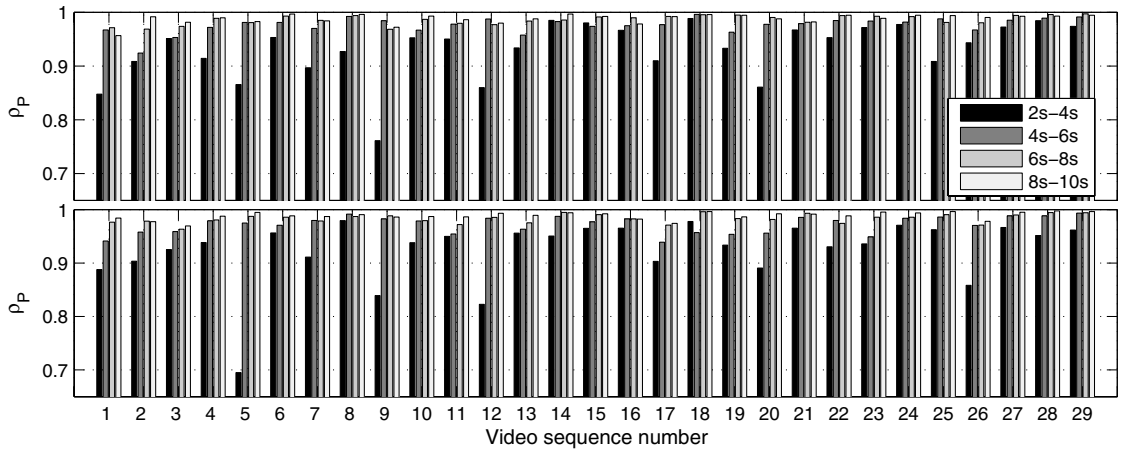

Fig. 5. Correlations ρ_P between the SM of two consecutive presentation times t for experiments E1 (top) and E2 (bottom).

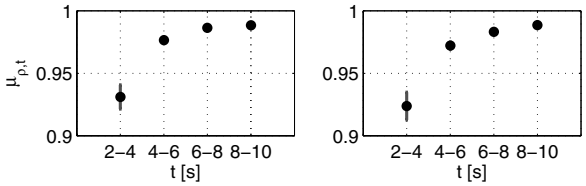

Fig. 6. Correlations between SM of two consecutive times t: average over all 29 images for E1 (left) and E2 (right).

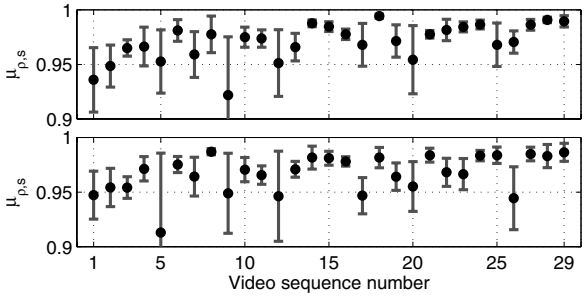

Fig. 7. Correlations between SM of two consecutive times t: average over all times t for E1 (top) and E2 (bottom).

images, e.g. images 5, 9, and 12, mostly identify the images which have a particularly low ρ_P between $SM^{(2)}$ and $SM^{(4)}$. Finally, the results in Fig. 7 show again a strong agreement between the two experiments E1 and E2.

4. CONCLUSIONS

We analysed SM created from two independent eye tracking experiments with respect to the consistency between the experiments and the effect of presentation time. The SM were shown to be highly correlated between the experiments, largely independent of presentation time. The change of the SM was further found to decrease with time, suggesting, that the presentation time can be limited to reduce the duration of eye tracking experiments. These results are particularly worth noting, as the experiments were not performed conjointly, but were conducted independently.

The eye tracking data obtained from experiments E1 and E2 are made publicly available to the research community at [8] and at [9], respectively.

5. REFERENCES

[1] J. Wolfe, "Visual attention," in *Seeing*, K. K. De Valois, Ed., pp. 335–386. Academic Press, 2000.

[2] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. PAMI*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.

[3] O. Le Meur, P. Le Callet, D. Barba, and D. Thoreau, "A coherent computational approach to model bottom-up visual attention," *IEEE Trans. PAMI*, vol. 28, no. 5, pp. 802–817, May 2006.

[4] A. K. Moorthy and A. C. Bovik, "Visual importance pooling for image quality assessment," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 2, pp. 193–201, Apr. 2009.

[5] U. Engelke, A. J. Maeder, and H.-J. Zepernick, "Visual attention modelling for subjective image quality databases," in *Proc. of IEEE MMSP*, Oct. 2009.

[6] H. Liu and I. Heynderickx, "Studying the added value of visual attention in objective image quality metrics based on eye movement data," in *Proc. of IEEE ICIP*, Nov. 2009, pp. 3097–3100.

[7] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "LIVE image quality assessment database release 2," http://live.ece.utexas.edu/research/quality, 2005.

[8] U. Engelke, A. Maeder, and H.-J. Zepernick, "VAIQ: The Visual Attention for Image Quality database," http://www.bth.se/tek/rcg.nsf/pages/vaiq-db, 2009.

[9] H. Liu and I. Heynderickx, "TUD image quality database: Eye-tracking release 1," http://mmi.tudelft.nl/ ingrid/eyetracking1.html, 2010.