A Comparative Study of Fixation Density Maps

Ulrich Engelke, Hantao Liu, Junle Wang, Patrick Le Callet, Ingrid Heynderickx, Hans-Jürgen Zepernick, Anthony Maeder

A Comparative Study of Fixation Density Maps

Ulrich Engelke*, *Member, IEEE*, Hantao Liu, *Member, IEEE*, Junle Wang, Patrick Le Callet, *Member, IEEE*, Ingrid Heynderickx, Hans-Jürgen Zepernick, *Senior Member, IEEE*, and Anthony Maeder, *Member, IEEE*

***Abstract*— Fixation density maps (FDM) created from eye tracking experiments are widely used in image processing applications. The FDM are assumed to be reliable ground truths of human visual attention and as such one expects high similarity between FDM created in different laboratories. So far, no studies have analysed the degree of similarity between FDM from independent laboratories and the related impact on the applications. In this paper, we perform a thorough comparison of FDM from three independently conducted eye tracking experiments. We focus on the effect of presentation time and image content and evaluate the impact of the FDM differences on three applications: visual saliency modelling, image quality assessment, and image retargeting. It is shown that the FDM are very similar and that their impact on the applications is low. The individual experiment comparisons, however, are found to be significantly different, showing that inter-laboratory differences strongly depend on the experimental conditions of the laboratories. The FDM are publicly available to the research community.**

***Index Terms*— Eye tracking, visual attention, fixation density maps, inter-laboratory differences.**

I. INTRODUCTION

THE human visual system (HVS) receives and processes an abundant amount of information at any instant in time. To reduce the complexity of scene analysis, several bottom-up and top-down visual attention (VA) [1], [2] mechanisms are deployed. The former is fast, signal driven, and independent of a particular viewing task. The latter mechanism is slower as it requires a voluntary gaze shift that is strongly dependent on the viewing task and semantic information in the visual scene. The two mechanisms together achieve that the most relevant visual information is favoured in any given context.

Various image and video processing applications, including source coding [3], retargeting [4], retrieval [5], and quality assessment [6], integrate VA mechanisms with the aim to improve system performance. To fully exploit the benefits of VA-based processing systems, the visual locations that attract an observers attention need to be determined using computational VA or saliency models [7]–[10]. These models are often developed using a ground truth recorded in subjective experiments. In some cases, such experiments require the observers to manually label the interesting [11] or important [12], [13] regions. Given the strong link between overt VA and eye movements [14], [15], more frequently such a ground truth is obtained through eye tracking experiments [16]. The resulting gaze patterns can be post-processed into fixation density maps (FDM) and the average FDM over all observers are then considered to be reliable ground truths of overt VA.

There is a strong demand for publicly available image and video eye tracking databases [10], [17]–[20] providing FDM. Having a reliable ground truth for computational modelling is common to other image processing applications, such as image quality assessment, where standardised procedures exist to perform subjective quality experiments. However, no standardised methodologies exist for eye tracking experiments. Instead, researchers usually follow best-practice guidelines [21]. The experiment outcomes hence depend on several factors related to the observer panel and the experimental design. The observers differ with respect to their cultural background, age, gender, interest, and expectations. These variations are the main reason why averaged FDM instead of individual FDM are used. Environmental aspects further affect the final FDM, such as the experimental procedures, the eye tracker hardware, and the viewing conditions. The lack of agreement on these experimental methodologies may lead to considerable differences in the resulting FDM. To identify the reliability of the FDM as a ground truth for image processing applications it is thus crucial to evaluate the similarity of FDM obtained from independent laboratories. Such inter-laboratory comparisons are common in research disciplines related to natural sciences and medical sciences. However, they are less common in computer science where researchers often restrict themselves to assessing the differences amongst observers within an experiment [22], [23].

To the best of our knowledge, there are no comprehensive studies on the differences of eye tracking data between laboratories and the implications for image processing applications. In this article, we therefore study the degree to which FDM of images differ between three experiments. These experiments were not conducted conjointly for the purpose of FDM comparison, but they were performed independently with each experiment considering their FDM to be solid ground truths for image processing. The goal here is therefore not to compare FDM that were created using exactly the same setup, but rather to analyse the differences amongst FDM and to estimate the impact on the performance of image processing applications. We further focus on the influence of

U. Engelke is with Philips Research Laboratories, 5656 AE Eindhoven, The Netherlands (e-mail: ulrich.engelke@philips.com, phone: +31-40-2748160, fax: +31-40-2746321).

H. Liu is with the University of Hull, Hull, HU6 7RX, United Kingdom (e-mail: hantao.liu@hull.ac.uk).

J. Wang and P. Le Callet are with *LUNAM Université, Université de Nantes, IRCCyN UMR CNRS 6597 (Institut de Recherche en Communications et Cybernétique de Nantes), Polytech Nantes, France (e-mail: {junle.wang,patrick.lecallet}@univ-nantes.fr).

I. Heynderickx is jointly affiliated with Philips Research Laboratories, 5656 AE Eindhoven, The Netherlands, and with the Delft University of Technology, 2628 CD Delft, The Netherlands (e-mail: ingrid.heynderickx@philips.com).

H.-J. Zepernick is with the Blekinge Institute of Technology, 37179 Karlskrona, Sweden (e-mail: hans-jurgen.zepernick@bth.se).

A. Maeder is with the University of Western Sydney, Locked Bag 1797, Penrith South DC, NSW 1797, Australia (e-mail: a.maeder@uws.edu.au).

TABLE I
REFERENCE IMAGES IN THE LIVE IMAGE QUALITY DATABASE [26].

#	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		

two factors: visual content and image presentation time. The former factor can be assumed to have a strong impact since the agreement may vary with the degree to which objects in the scene attract the viewers' attention. The latter factor is of interest because the similarity between FDM may vary with respect to the duration that the images are viewed. We presented initial results on these issues in [24], based on two experiments. In this article, we extend this work by a third experiment and by providing a considerably more detailed discussion and analysis using different similarity measures. We further address the impact of FDM similarity on three contemporary applications: visual saliency modelling, image quality assessment, and image retargeting.

The remainder of the article is organised as follows. Section II introduces the eye tracking experiments and Section III describes the FDM similarity measures. Section IV provides a detailed comparison of the FDM from the three experiments. The influence of the FDM on three applications is discussed in Section V. A discussion of the results is provided in Section VI. Finally, conclusions are drawn in Section VII.

II. EYE TRACKING EXPERIMENTS

The eye tracking experiments were conducted in three independent laboratories, i.e. at the School of Computing and Mathematics at the University of Western Sydney (UWS), Australia [18], the Man-Machine Interaction group of Delft University of Technology (TUD), The Netherlands [19], and the Image and Video Communications Group at IRCCyN of the University of Nantes (UN), France [25].

A. Test images

The stimuli presented in all three experiments were the 29 original (reference) images from the LIVE image quality database [26]. These images cover a wide range of content, including, natural scenes, buildings, boats, humans, animals, and written text. A list of all images is given in Table I with the original names from the LIVE database.

B. Comparison of experimental procedures

An overview of the three experiments is presented in Table II. We do not repeat all details here but instead highlight differences amongst the experiments that can be expected to be a dominant source of variability in the recorded data.

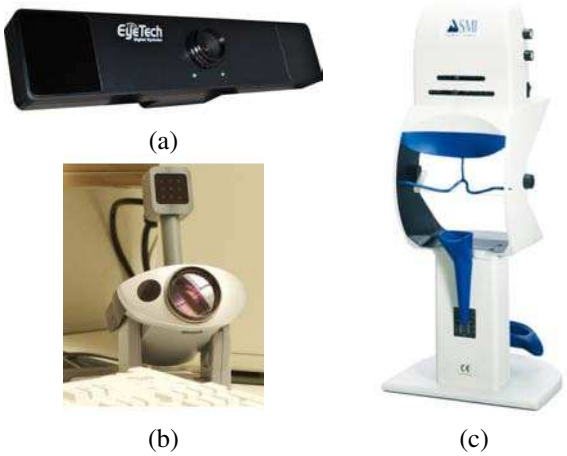

Fig. 1. Eye trackers: (a) EyeTech TM3 (UWS) [30], (b) SMI iView X RED (TUD) [28], and (c) SMI iView X Hi-Speed (UN) [29].

Amongst the major differences in the observer panel is the number of participants, which ranges from 15 to 21. Generally, eye tracking data averaged over an observer population becomes more stable with an increased number of participants [21]. The average age between UWS and UN is considerably different and may have an impact on the viewing behaviour as people of different ages have different interests. No ages were recorded in experiment TUD, however, given that all participants were students the average age is expected to be the lowest amongst the experiments. Finally, as the experiments were conducted in three countries, cultural differences between the observer panels may have an influence.

Three different eye trackers were used in the experiments, which are illustrated in Fig. 1. The considerably higher frequency of the eye tracker in experiment UN is instrumental for the analysis of saccadic eye movements, for which frequencies of below 50 Hz are not sufficient. Since we focus on the analysis of fixations, the recording frequency of the eye trackers is not expected to have a strong impact on the results. More importantly, the head rest and comparably higher accuracy of the eye tracker in UN may result in more accurate gaze data compared to the other eye trackers.

The image presentation differs mainly in three factors. Firstly, the duration ranges from 10 s to 15 s, which means that for the purpose of comparing the three experiments, we are limited to the first 10 s of each experiment. Secondly, the visual angle of the displayed images differs due to varying viewing distance and screen resolution. Finally, experiment UWS utilised a central fixation point during the grey screen presented between images to ensure that image viewing is started from the same location. This factor can be expected to have an impact especially on the early fixations.

C. Creation of fixation density maps

The recorded gaze patterns are post-processed into FDM, which are normalised intensity maps with values between 0 and 1. The magnitudes within the FDM represent the amount of overt attention at certain locations, but they do not account for the timely order of the fixations. However, fixation order

TABLE II
OVERVIEW OF THE EYE TRACKING EXPERIMENTS.

Category	Details	UWS	TUD	UN
Participants	Number	15	18	21
	Age range (average age)	20-60 (42)	-	18-42 (26)
	Male/female	9/6	11/7	11/10
	Non-experts/experts	12/3	18/0	21/0
	Occupation	University staff/students	University students	University staff/students
	Compensated	No	No	Yes
Viewing conditions	Environment	Laboratory		
	Illumination	Low		
	Viewing distance	60 cm	70 cm	70 cm
	Task	Free-viewing: the observers were not instructed with any particular task but to view the images		
Display	Make	Samsung SyncMaster	iiyama	DELL
	Type	LCD	CRT	LCD
	Size	19”		
	Resolution [pixels]	1280×1024	1024×768	1280×1024
Eye tracker	Make	EyeTech TM3 [27]	SMI iView X RED [28]	SMI iView X Hi-Speed [29]
	Type	Infrared video-based		
	Frequency	45 GP/s	50 GP/s	500 GP/s
	Accuracy	$<$ 1 dva	0.5-1 dva	0.25-0.5 dva
	Mounting	Under the display	Under the display	Tower with head rest
	Calibration	16 point screen	9 point screen	9 point screen
Image presentation	Order	Random		
	Image duration	12 s	10 s	15 s
	Grey-screen duration	3 s		
	Max. visual angle [pixels/deg]	36	32.8	41.8
	Central fixation point	Yes	No	No

cannot easily be predicted using a computational model [31], for which reason FDM are typically used. The conversion into FDM is conducted for UWS using the implementation that is explained in [18]. Experiments TUD and UN utilised the SMI Begaze software that accompanied the eye tracker. Despite the different software used, the underlying process comprised of the same steps. Firstly, gaze points (GP) belonging to saccades were removed since vision is greatly suppressed during these fast eye movements. The remaining GP were clustered into fixations, with the magnitudes of the fixations corresponding to the fixation lengths. Finally, the fixation map was filtered using a Gaussian kernel to account for eye tracker inaccuracies as well as the decrease in visual accuracy with increasing eccentricity from the fovea. All three experiments assumed a minimum fixation length of 100 ms and a foveal coverage of approximately 2 degrees visual angle (dva).

We created FDM based on a range of presentation times $t \in \{0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ s to allow for FDM similarity analysis for different viewing durations. The FDM for image i, created from a particular presentation time t, and belonging to one of the three experiments UWS, TUD, and UN, are denoted as $M^{(t)}_{UWS}(i)$, $M^{(t)}_{TUD}(i)$, and $M^{(t)}_{UN}(i)$, respectively. All FDM are presented in Fig. 17 and Fig. 18 (last page of this article) for a presentation time $t = 10$ s. In addition to the experimental FDM, we created random FDM for each image that serve as a lower limit, both for the FDM similarity evaluation as well as the performance evaluation of the applications in Section V. The random FDM were created by randomly substituting the FDM between images of the same or similar size within the same database. For better comparability, the same random substitution was used for all databases. The random FDM are in the following denoted as $M^{(t)}_{RND}$.

III. SIMILARITY MEASURES

There are no standardised measures to compare the similarity between two FDM or between FDM and saliency maps from computational models. However, there is a range of measures that are widely used to perform this task: correlation coefficient [8], [9], [32], Kullback-Leibler divergence [8], [10], receiver operating characteristics (ROC) analysis [32], [33], and normalised scanpath saliency (NSS) [23], [33]–[35]. The former three are directly applicable to saliency maps and FDM, whereas NSS compares the actual fixations to a saliency map. We utilise two similarity measures: the Pearson linear correlation coefficient and the area under the ROC curve. For the purpose of computing these measures, the original FDM values are linearly transformed from the range [0,1] to the range [0,255]. Visual comparison of the FDM is additionally facilitated by the FDM provided in Fig. 17 and Fig. 18.

A. Pearson linear correlation coefficient (PLCC)

The Pearson linear correlation coefficient (PLCC) [36] measures the strength and direction of a linear relationship between two variables. We compute it between two FDM, $M^{(i)}$ and $M^{(j)}$, as follows

$$\rho_P(M^{(i)}, M^{(j)}) = \frac{\sum_k \sum_l (M^{(i)}_{kl} - \mu^{(i)})(M^{(j)}_{kl} - \mu^{(j)})}{\sqrt{\sum_k \sum_l (M^{(i)}_{kl} - \mu^{(i)})^2} \sqrt{\sum_k \sum_l (M^{(j)}_{kl} - \mu^{(j)})^2}} \quad (1)$$

where $k \in [1, K]$ and $l \in [1, L]$, respectively, are the horizontal and vertical pixel coordinates, and $\mu^{(i)}$ and $\mu^{(j)}$ are the mean pixel values of the FDM.

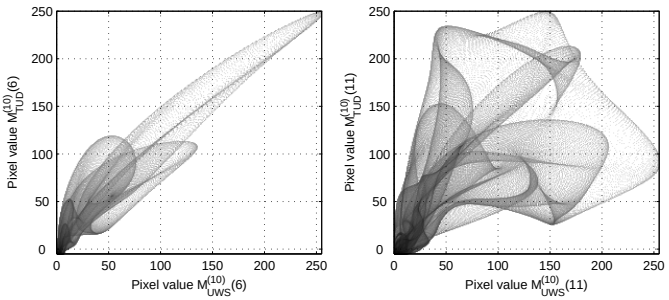

Fig. 2. Scatter-like plot of the conjoint pixel values between two FDM for (a) highly similar FDM ($M_{UWS}^{(10)}(6)$ and $M_{TUD}^{(10)}(6)$) and (b) highly dissimilar FDM ($M_{UWS}^{(10)}(11)$ and $M_{TUD}^{(10)}(11)$).

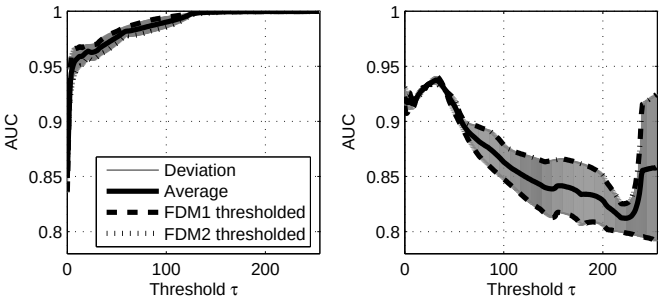

Fig. 3. AUC for 255 thresholds τ between for (a) highly similar FDM ($M_{UWS}^{(10)}(6)$ and $M_{TUD}^{(10)}(6)$) and (b) highly dissimilar FDM ($M_{UWS}^{(10)}(11)$ and $M_{TUD}^{(10)}(11)$).

For illustration purposes, scatter-like plots are presented in Fig. 2(a) for two highly correlated FDM ($M_{UWS}^{(10)}(6)$ and $M_{TUD}^{(10)}(6)$) with $\rho_P = 0.933$ and in Fig. 2(b) for two lowly correlated FDM ($M_{UWS}^{(10)}(11)$ and $M_{TUD}^{(10)}(11)$) with $\rho_P = 0.637$. Naturally, the highly correlated FDM exhibit values much closer to the main diagonal. There are, however, also very distinct structures in the plots, which inherently result from the structures contained in the actual FDM (see Fig. 17 and Fig. 18). The PLCC does not account for these structural differences between the FDM and also cannot distinguish whether differences amongst FDM are caused mainly from high magnitude pixels or low magnitude pixels. The area under the ROC curve accounts for these missing aspects of PLCC.

B. Area under the ROC curve (AUC)

To facilitate the use of the area under the ROC curve (AUC) [37], [38] for measuring FDM similarity, one of the two FDM has to be thresholded into a binary map as

$$M_{bin,DB}^{(t)}(i) = \begin{cases} 1 & \text{for} \quad M_{DB}^{(t)}(i) \geq \tau \\ 0 & \text{for} \quad M_{DB}^{(t)}(i) < \tau \end{cases} \tag{2}$$

with $\tau \in [0 \ldots 254]$ and $DB \in \{UWS, TUD, UN\}$. ROC analysis is non-symmetrical and depending on which FDM is used to create the binary map, the value of the resulting AUC can vary. We therefore compute the average over the two non-symmetrical AUC. Depending on the threshold τ chosen, different properties of the FDM are analysed. For a low threshold the binary map covers a larger area than for a large threshold. Hence, for low values of τ the AUC accounts for coverage similarity between the FDM whereas for large values it identifies the similarity between the peaks.

For illustration of this behaviour, Fig. 3 presents both non-symmetrical AUC computations between two FDM, along with their mean for all 255 thresholds τ. For highly similar FDM (Fig. 3(a)), the AUC rises fast towards the maximum level and the difference between the AUC is small. For highly dissimilar FDM (Fig. 3(b)), the AUC is low and in this case even decreases with an increasing threshold. These lower AUC for large τ quantify that FDM $M_{UWS}^{(10)}(11)$ and $M_{TUD}^{(10)}(11)$ have different peaks, as can be visually observed from Fig. 17.

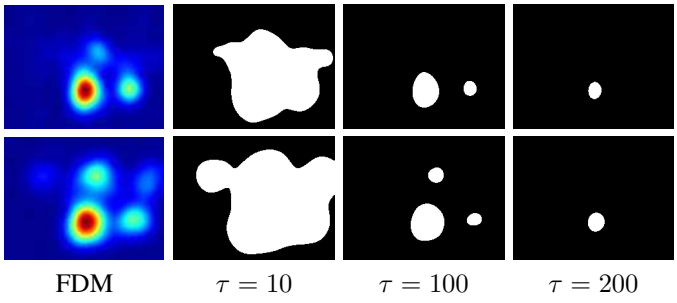

Fig. 4. Binary maps for highly similar FDM ($M_{UWS}^{(10)}(6)$ and $M_{TUD}^{(10)}(6)$) after thresholding with $\tau = 10$, $\tau = 100$, and $\tau = 200$.

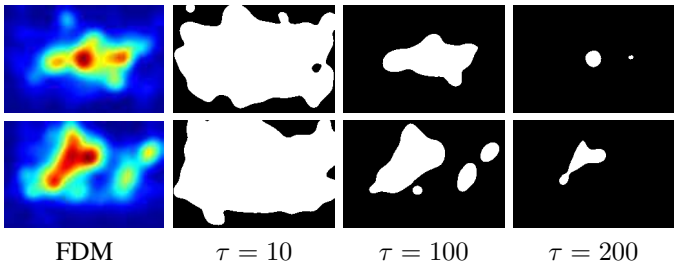

Fig. 5. Binary maps for highly dissimilar FDM ($M_{UWS}^{(10)}(11)$ and $M_{TUD}^{(10)}(11)$) after thresholding with $\tau = 10$, $\tau = 100$, and $\tau = 200$.

To capture different properties of FDM, in terms of coverage and peak similarity, we consider in the following a low threshold $\tau = 10$, a high threshold $\tau = 200$, and also an intermediate threshold $\tau = 100$ to account for lower order peaks. Fig. 4 and Fig. 5 illustrate the binary maps resulting from the thresholding for the FDM presented in Fig. 3(a) and Fig. 3(b), respectively. The similarity between the binary maps in Fig. 4 reflects well the increase in AUC as presented in Fig. 3(a). Similarly, the decrease in AUC in Fig. 3(b) is also reflected in the visual inspection of the binary maps in Fig. 5.

C. Monotonicity between PLCC and AUC

Despite the different purposes for the PLCC and AUC, they are expected to vary conjointly to some degree. To identify the degree to which the two measures interrelate to each other, we compute the Spearman rank order correlation

TABLE III
SPEARMAN RANK ORDER CORRELATION BETWEEN PLCC AND AUC.

	UWS vs. TUD	UWS vs. UN	TUD vs. UN
$\tau = 10$	0.317	0.474	0.409
$\tau = 100$	0.836	0.784	0.762
$\tau = 200$	0.691	0.64	0.762

coefficient (SRCC). The SRCC computed over all images and presentation times t ($N = 29 \times 11 = 319$) are presented in Table III. The threshold τ has a strong impact on the similarity between the ranks of PLCC and AUC. The higher SRCC for $\tau = 100$ and $\tau = 200$ as compared to $\tau = 10$ can be attributed to the fact that the PLCC is only high if also the large magnitudes in the FDM (the peaks) agree with each other.

IV. INTER-LABORATORY COMPARISON

In the following sections, the FDM similarity is evaluated using the PLCC and AUC measures.

A. Inter-laboratory differences

The main objective of this article is the evaluation of inter-laboratory differences between FDM. For this purpose, we present in Fig. 6 and Fig. 7 the PLCC and AUC, respectively, as a function of presentation time t for the 3 inter-laboratory comparisons and the comparison to the random FDM. Each figure shows the means along with their standard errors over all 29 images. Figure 6 illustrates that the progression of the mean PLCC with presentation time is similar between the three experiments. The increase quickly flattens out and the PLCC only marginally depends on the presentation time for $t \geq 2\,\text{s}$. For TUD-UN this observation already holds for $t \geq 1\,\text{s}$.

Despite similar progression of the PLCC, the overall magnitudes between the three comparisons differ to some degree, with TUD-UN having the highest correlations, followed by UWS-UN and UWS-TUD. Hence, the FDM of experiments TUD and UN appear to be most similar, whereas the respective similarities to experiment UWS are to some degree lower. In addition to the lower mean PLCC, it can also be observed that the standard errors are larger for UWS-TUD and UWS-UN compared to TUD-UN, which indicates that there is a larger variance of the PLCC with respect to the image content. The significantly higher PLCC for $t \geq 1\,\text{s}$ between the experimental FDM, as compared to the random FDM, emphasize high similarity between the experimental FDM.

Similar observations as for the PLCC also hold for the AUC presented in Fig. 7. The AUC plots confirm the order of similarity between the experiments and also the increasing similarity between experiments with an increase in presentation time. It is interesting to note the difference of the AUC values for the three different thresholds $\tau = 10$, $\tau = 100$, and $\tau = 200$. The AUC increases with the threshold, indicating that the FDM similarity is generally higher for the strongly salient regions as compared to the remainder of the images. The significantly lower AUC between the experimental FDM and random FDM confirm the observations on the PLCC.

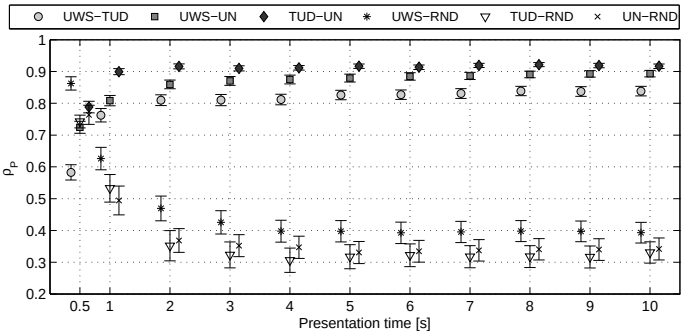

Fig. 6. Mean PLCC and standard errors over all images for all t.

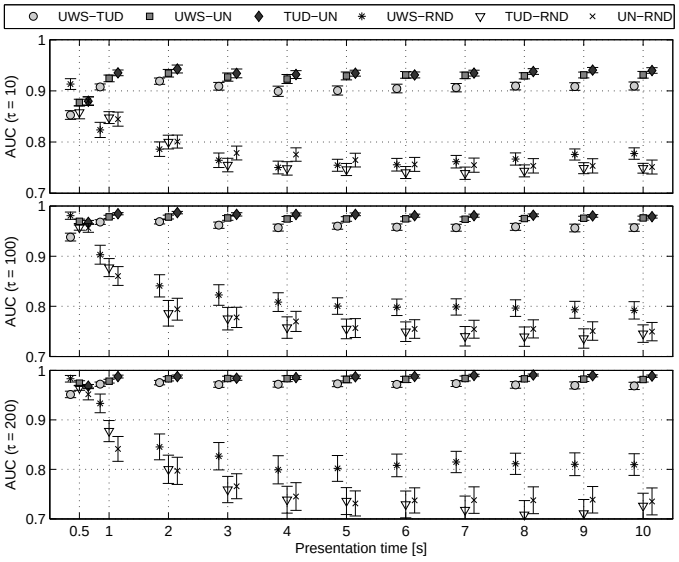

Fig. 7. Mean AUC and standard errors over all images for all t and for $\tau = 10$ (top), $\tau = 100$ (middle), and $\tau = 200$ (bottom).

B. Content dependency

The standard errors in Fig. 6 and Fig. 7 show that the FDM similarity is to some degree content dependent. We therefore analyse here the similarity between the FDM in relation to the content of the images. Given that the PLCC and AUC are very similar for $t \geq 2\,\text{s}$, we consider two presentation times: $t = 1\,\text{s}$ and $t = 10\,\text{s}$. These presentation times allow us to compare the impact of image content on the early fixations ($t = 1\,\text{s}$) and on a more exhaustive viewing of the images ($t = 10\,\text{s}$). The PLCC and AUC for all 29 images are presented in Fig. 8 and Fig. 9, respectively. The PLCC and AUC are presented for the individual comparisons amongst experiments (UWS-TUD, UWS-UN, TUD-UN) as well as for the average over the three comparisons. All presented values are sorted with respect to the decreasing average measures.

The similarity amongst FDM strongly depends on the image content, both for $t = 1\,\text{s}$ and $t = 10\,\text{s}$. For high average PLCC and AUC, the values from the individual experimental comparisons are located closely together, whereas for low average PLCC and AUC the deviation of the individual values is considerably higher. Furthermore, the PLCC and AUC of

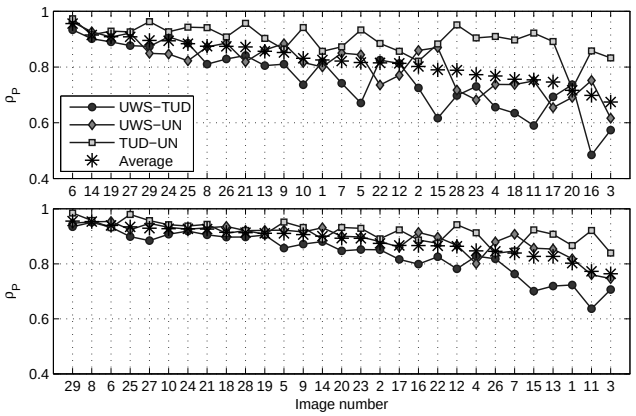

Fig. 8. Impact of the image content measured using PLCC for $t = 1\,\mathrm{s}$ (top) and $t = 10\,\mathrm{s}$ (bottom).

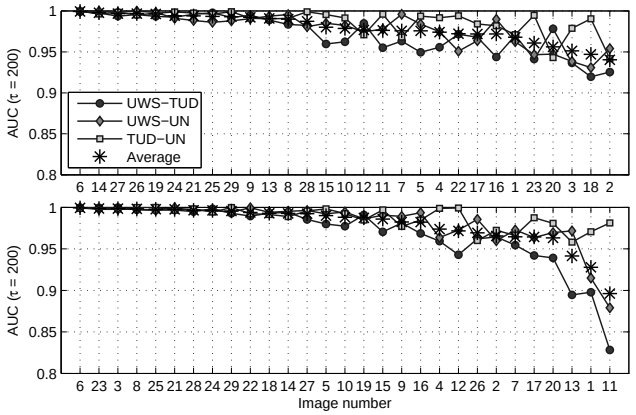

Fig. 9. Impact of the image content measured using AUC ($\tau = 200$) for $t = 1\,\mathrm{s}$ (top) and $t = 10\,\mathrm{s}$ (bottom).

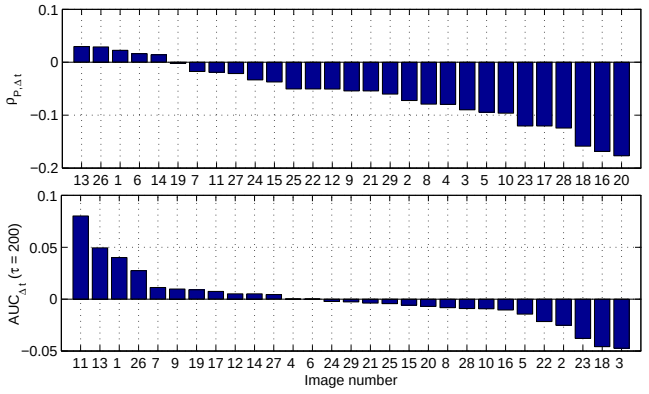

Fig. 10. Differences between $t = 1\,\mathrm{s}$ and $t = 10\,\mathrm{s}$ for PLCC (top), $\rho_{P,\Delta t}$, and AUC with $\tau = 200$ (bottom), $\mathrm{AUC}_{\Delta t}$.

the same images can be different between the two presentation times. This essentially means that the content dependency of the similarity between FDM is a function of time. To better illustrate the differences between the two presentation times, we present in Fig. 10 two bar plots of the PLCC and AUC difference between $t = 1\,\mathrm{s}$ and $t = 10\,\mathrm{s}$ denoted as $\rho_{P,\Delta t}$ and

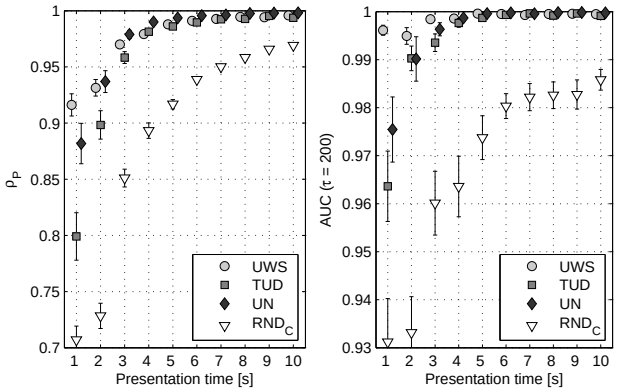

Fig. 11. PLCC (left) and AUC with $\tau = 200$ (right) between FDM of consecutive presentation times.

$\mathrm{AUC}_{\Delta t}$, respectively. Together with Fig. 8 and Fig. 9 as well as visual inspection of the FDM at all presentation times, these results allow for a detailed discussion as follows.

For both PLCC and AUC, many of the images exhibit rather small differences $\rho_{P,\Delta t}$ and $\mathrm{AUC}_{\Delta t}$. For instance, image number 6 ('cemetry') is rated very high for both $t = 1\,\mathrm{s}$ and $t = 10\,\mathrm{s}$. This image contains two plaques with written text which attracted the attention of the observers upon presentation of the image and kept the attention throughout the image presentation. Many other images have large differences between presentation times. Image number 3 ('buildings'), for instance, exhibits the largest $\mathrm{AUC}_{\Delta t}$ difference in the set due to a low AUC at $t = 1\,\mathrm{s}$ and a high AUC at $t = 10\,\mathrm{s}$. Like image number 6, this image also contains text. However, due to the high complexity of the image, the text is not as dominant and the observers needed more time to detect it. Similarly, image number 18 ('parrots') has considerably higher PLCC and AUC for $t = 10\,\mathrm{s}$ as compared to $t = 1\,\mathrm{s}$. This image contains two distinct salient regions (the parrot heads) whose attendance leveled off with increased presentation time.

C. Convergent behaviour

The previous sections revealed a strong similarity between the FDM of the three laboratories, especially for an increased presentation time ($t \geq 2\,\mathrm{s}$). This increased similarity suggests that the FDM become more stable with longer presentation times. In applications that require a converged FDM, verification of the convergence speed of FDM may aid in reducing experimental time and cost while sacrificing only marginally the accuracy of a final FDM. We therefore analyse in the following the PLCC and AUC ($\tau = 200$) between FDM created from two consecutive presentation times for each experiment individually. These PLCC and AUC values are presented as a function of presentation time in Fig. 11 for the three experiments. The labels on the abscissa indicate the higher presentation time, for instance, $t = 1\,\mathrm{s}$ relates to the similarity between FDM based on $t = 500\,\mathrm{ms}$ and $t = 1\,\mathrm{s}$.

The average over all contents illustrates that there is indeed a strong convergent behaviour of the FDM similarity with presentation time. The PLCC and AUC curves generally

TABLE IV

PLCC BETWEEN THE SALIENCY MAPS AND FDM.

t	Data-base	Visual attention models					σ_{SAL}
		Itti [7]	Rajashekar [9]	Bruce [10]	Achanta [39]	Hou [40]	
1 s	UWS	0.096	0.288	0.218	0.202	0.241	0.071
	TUD	0.097	0.348	0.244	0.242	0.3	0.094
	UN	0.099	0.372	0.272	0.254	0.32	0.103
	σ_{DB}	0.001	0.044	0.027	0.027	0.041	—
	RND	0.04	0.282	0.164	0.12	0.152	0.087
10 s	UWS	0.147	0.435	0.371	0.312	0.384	0.111
	TUD	0.152	0.448	0.369	0.333	0.415	0.115
	UN	0.15	0.449	0.376	0.335	0.421	0.118
	σ_{DB}	0.003	0.008	0.004	0.013	0.02	—
	RND	0.072	0.297	0.208	0.135	0.171	0.084

TABLE V

AUC ($\tau = 200$) BETWEEN THE SALIENCY MAPS AND FDM.

t	Data-base	Visual saliency models					σ_{SAL}
		Itti [7]	Rajashekar [9]	Bruce [10]	Achanta [39]	Hou [40]	
1 s	UWS	0.624	0.733	0.683	0.651	0.676	0.04
	TUD	0.62	0.786	0.713	0.693	0.748	0.062
	UN	0.648	0.801	0.749	0.681	0.758	0.061
	σ_{DB}	0.015	0.036	0.033	0.022	0.044	—
	RND	0.621	0.717	0.633	0.603	0.65	0.044
10 s	UWS	0.66	0.797	0.758	0.687	0.737	0.055
	TUD	0.671	0.803	0.772	0.692	0.771	0.057
	UN	0.658	0.802	0.773	0.69	0.76	0.06
	σ_{DB}	0.007	0.003	0.008	0.002	0.017	—
	RND	0.616	0.682	0.624	0.554	0.606	0.046

follow a very similar progression for all three experiments, with exception of the PLCC and AUC between presentation times $t = 500\,\text{ms}$ and $t = 1\,\text{s}$. For presentation times larger than $t = 4\,\text{s}$ the average PLCC and AUC are well above 0.95 and 0.995, respectively. This observation holds for all three experiments and for a wide range of natural image content. The convergence is partly a result of the effect that the number of new fixations relative to the total number of fixations decreases with an increase of presentation time. To visualise this effect, we created through simulation FDM containing randomly distributed fixations, with an averaged fixation length of $250\,\text{ms}$ that was estimated from the eye tracking data. These random FDM, denoted as RND_C in Fig. 11, also exhibit a convergent behaviour of PLCC and AUC over time. However, the convergence is considerably slower, providing further evidence of the strong convergent behaviour of the experimental FDM.

V. APPLICATIONS

The similarity measures indicate high similarity between the FDM but they do not provide direct insight into the reliability of the FDM as a ground truth for image processing applications. In this section, we therefore identify the sensitivity of three applications to the FDM used: visual saliency modelling, image quality assessment, and image retargeting.

TABLE VI

IMAGE QUALITY PREDICTION PERFORMANCE GAIN BASED ON PLCC.

		JPEG 1	JPEG 2	J2K 1	J2K 2	Gaussian blur	White noise	Fast fading	Average
PSNR	UWS	0.004	0.028	0.002	0.037	0.006	0	0.009	0.012
	TUD	0.008	0.031	0.006	0.037	0.02	0	0.016	0.017
	UN	0.006	0.029	0.003	0.036	0.018	0	0.015	0.015
	RND	0	-0.006	-0.001	0.005	-0.02	0.014	0.001	-0.008
SSIM	UWS	0.017	0.041	0.022	0.041	0.075	0.008	0.029	0.033
	TUD	0.019	0.039	0.019	0.038	0.07	0.008	0.023	0.031
	UN	0.014	0.036	0.019	0.037	0.07	0.009	0.025	0.03
	RND	0.002	-0.004	0.002	0	0.064	-0.002	0.023	0.012
VIF	UWS	0.022	0.008	0	0.007	0.017	0.01	0.012	0.011
	TUD	0.022	0.008	0	0.009	0.017	0.012	0.009	0.011
	UN	0.026	0.008	0.004	0.009	0.021	0.01	0.008	0.012
	RND	0.018	0.007	0	0.007	0.013	0.008	0.003	0.008

A. Visual saliency models

FDM obtained from eye tracking experiments are typically used for the training and validation of visual saliency models. We analyse here to what degree the validation of saliency models depends on the ground truth, the FDM. We consider in the following the well known saliency model by Itti et al. [7] as well as the models by Rajashekar et al. [9], Bruce et al. [10], Achanta et al. [39], and Hou et al. [40]. We compute the saliency maps for all images using these models and compute the similarity between them and the FDM based on presentation times $t = 1\,\text{s}$ and $t = 10\,\text{s}$. The results for PLCC and AUC ($\tau = 200$) are presented in Table IV and V, respectively. In addition, the standard deviations over the PLCC and AUC are given over the three databases, σ_{DB}, and the five saliency models, σ_{SAL}.

The results show that both similarity measures differ considerably more between the visual saliency models than between the FDM. This observation holds for both presentation times $t = 1\,\text{s}$ and $t = 10\,\text{s}$. Interestingly, all saliency models perform better on the FDM with $t = 10\,\text{s}$, even though these models aim to predict salient locations that are widely known to drive mainly rapid bottom-up VA mechanisms. For the PLCC, this higher performance might be influenced to some degree by the larger number and thus a wider spread of fixations for $t = 10\,\text{s}$ compared to $t = 1\,\text{s}$. The consistently higher performance of all saliency models on the experimental FDM compared to the random FDM further illustrates that the models predict saliency with an accuracy above chance.

B. Quality prediction models

Saliency maps and FDM are often integrated into image quality models with the aim to improve quality prediction performance [19], [41]. We analyse to what degree the improvement of three quality prediction models, the Peak Signal-to-Noise Ratio (PSNR), Structural Similarity (SSIM) Index [42], and Visual Information Fidelity (VIF) criterion [43], varies with the FDM used. Following the procedure in [44], we integrate the FDM based on a presentation time of $t = 10\,\text{s}$ into the quality models by local, multiplicative weighting of the respective distortion map. As the images we used in our

eye tracking experiments are taken from the LIVE image quality database, we have a large set of distorted images and their respective mean opinion scores (MOS) available for the design and validation of the quality models. To analyse the prediction performance, we computed PSNR, SSIM, and VIF on all distorted images of the LIVE database before and after incorporation of the FDM. The model predictions were then compared to the respective MOS by computing the PLCC. The analysis is conducted independently for the different distortion classes contained in the LIVE database. The performance gain through incorporation of the FDM is presented in Table VI.

The results show that for all distortion classes and the related averages, the improvements are very similar between the three experiments. The improvement, however, differs considerably between the quality prediction models and between the different distortion classes. The consistency between FDM is thus better than the consistency between quality prediction models and distortion classes.

C. Saliency-based image retargeting

Image retargeting algorithms [45] resize images by cutting out vertical seams of lowest energy, thus preserving the most important regions in the images. Saliency-based image retargeting algorithms allow for an additional importance weighting based on the visual saliency in the scene. We used the FDM based on $t = 1\,\mathrm{s}$ and $t = 10\,\mathrm{s}$ in the saliency-based image retargeting algorithm by Wang et al. [4] to determine the similarity of the resulting retargeted images. Examples are presented in Fig. 12-14 for the images 13, 27, and 29 of the LIVE database.

The retargeting is generally inferior when using the randomly substituted FDM (RND) as compared to the experimental FDM. An exception is image 29 for $t = 1\,\mathrm{s}$, which looks good also for RND. The most relevant regions are generally well preserved when using the experimental FDM (UWS, TUD, UN) with the outcomes being very similar between the databases. The similarity is particularly high between TUD and UN and is somewhat lower for UWS, which confirms our earlier results on PLCC and AUC in Section IV.

VI. DISCUSSION

In Section II-B, we discussed the differences between the three experiments and how they can be expected to have an impact on the FDM similarity. Given the multitude of varying factors due to the independently conducted experiments, we could only speculate here as to what degree each of the factors influences the inter-laboratory differences. In the following, we therefore focus on a discussion of the overall differences between the databases.

A. Inter-laboratory comparisons: revisited

Both the PLCC and AUC show similar progressions with presentation time between the three databases (see Fig. 6 and Fig. 7). The absolute values, however, are not exactly the same for the individual comparisons between the databases, with TUD-UN being most similar, followed by UWS-UN and UWS-TUD. This trend is transferred to some degree to the image processing applications, where UWS typically differs somewhat more from the other databases. To see whether there are significant differences between the individual database comparisons, we performed paired t-tests for all individual

Fig. 12. Retargeted image number 13 based on FDM with $t = 1\,\mathrm{s}$ (top) and $t = 10\,\mathrm{s}$ (bottom). From left to right: UWS, TUD, UN, RND.

Fig. 13. Retargeted image number 27 based on FDM with $t = 1\,\mathrm{s}$ (top) and $t = 10\,\mathrm{s}$ (bottom). From left to right: UWS, TUD, UN, RND.

Fig. 14. Retargeted image number 29 based on FDM with $t = 1\,\mathrm{s}$ (top) and $t = 10\,\mathrm{s}$ (bottom). From left to right: UWS, TUD, UN, RND.

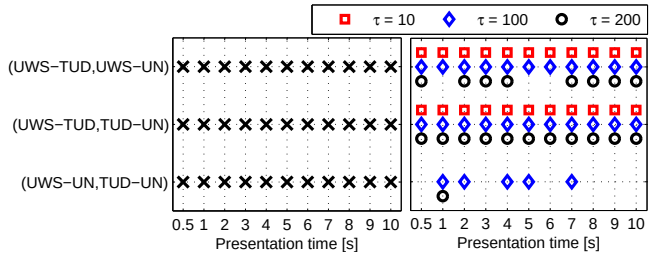

Fig. 15. Paired t-test at 95% confidence between for PLCC (left) and AUC (right). Markers indicate statistically significant difference.

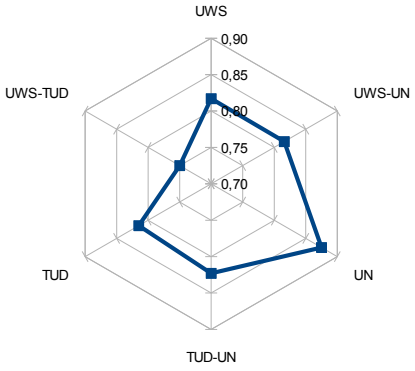

Fig. 16. Spider chart of the intra-experiment (UWS, TUD, UN) and inter-experiment (UWS-TUD, UWS-UN, TUD-UN) correlations (PLCC) between FDM based on 7 observers (values averaged over 100 random samples).

comparisons and all presentation times on the data presented in Fig. 6 and Fig. 7. The markers in Fig. 15 reveal if there are significant differences between the comparisons at 95% confidence. For PLCC all comparisons are statistically different. For AUC, only the comparisons UWS-TUD and UWS-UN are statistically the same in most cases. These results show that inter-laboratory comparisons vary significantly with the laboratories that are involved. Despite the significant differences of the similarity metrics, the experimental FDM still have a similarly positive impact on the image processing applications, as compared to using randomly chosen FDM.

B. Intra- versus inter-experiment differences

It could be argued that the differences between experiments are due to intrinsic variations amongst the observer groups. We therefore take a closer look at the variations within the experiments (intra-experiment) in comparison to the variations between the experiments (inter-experiment). We adopt the performance efficiency method [46] by repeatedly splitting the observer panel within an experiment into two sub-groups and computing the PLCC between the FDM created from these groups. These PLCC serve as an intrinsic ground truth and upper theoretical limit of the variations amongst observers within an experiment. To facilitate a fair comparison, we adapt the method in [46] by selecting the same size of the sub-groups for the three comparisons within and between the experiments. Based on these sub-groups, we create FDM for a presentation time of $t = 10\,\mathrm{s}$. For the sub-group selection, we are bound by the lowest number of 15 observers in experiment UWS. We therefore randomly select two groups of 7 observers within each of the experiments and compute the intra-experiment PLCC between the related FDM. Similarly, we select randomly 7 observers from each experiment and compute the inter-experiment PLCC. To obtain a robust estimate we repeat this process 100 times for the intra- and inter-experiment comparisons and compute the average PLCC.

All intra- and inter-experiment PLCC are presented in Fig. 16. The intra-experiment correlation for UN is approximately 5% higher than for UWS and TUD. One could speculate that the larger foveal coverage in relation to the image size in UN (see Section II-B) may enhance observers to grasp the gist of the scene. Thus, the number of possible target objects is lower and agreement between observers is higher. The superior accuracy of the eye tracker in UN could also have an impact on these results. Finally, the larger Gaussian kernel sizes relative to the image size inherently increases the PLCC to some degree. The inter-experiment PLCC are considerably higher for UWS-UN and TUD-UN than for UWS-TUD. The lower observer differences within UN may be one reason why this experiment correlates higher with the other experiments.

TABLE VII
RATIOS OF INTER- VS INTRA-LABORATORY PLCC.

Intra-laboratory	Inter-laboratory	Ratio
UWS	UWS-TUD	0.918
	UWS-UN	0.998
TUD	UWS-TUD	0.92
	TUD-UN	1.009
UN	UWS-UN	0.933
	TUD-UN	0.941

The ratios between the inter-laboratory PLCC and both corresponding intra-laboratory PLCC are presented in Table VII. The ratios that are lower than 1 suggest that there are indeed differences between the laboratories that are not accounted for by only the intrinsic observer variations. Only the ratios 0.998 and 1.009 for UWS-UN and TUD-UN with UWS and TUD as ground truths, respectively, show that the intrinsic differences within these experiments are as high as the differences between the experiments.

VII. CONCLUSIONS AND OUTLOOK

We analysed FDM similarity between three independent eye tracking experiments using two different similarity measures: PLCC and AUC. We showed that these measures capture different properties while being coherent in predicting the similarity of the FDM. Only for short presentation times ($t \leq 1\,\mathrm{s}$), PLCC was found to deviate from AUC.

Despite various differences between the experiments, the FDM were found to be very similar. The similarity, however, was dependent on the individual experimental comparisons, with UWS being more different to TUD and UN. The FDM similarity was further revealed to be highly dependent on the image content, with images that contain a distinct salient region experiencing a higher FDM similarity as compared to images with multiple or no salient regions. A similar

convergent behaviour of the FDM was identified for all three experiments. The speed of convergence may need to be verified for different numbers of observers as well as different experimental conditions. The reliability of the FDM as a ground truth was validated on three image processing applications: visual saliency modelling, image quality assessment, and image retargeting. On all applications it was shown that the difference between the experimental FDM on the outcomes was low. These findings suggest that FDM from independent eye tracking experiments can indeed be considered to be reliable ground truths for image processing applications.

Given the independency of the experiments, we could not clearly identify the degree to which the differentiating factors are impacting the FDM. It is therefore instrumental to extend this work by conducting experiments conjointly with careful variation of certain factors to evaluate their impact on the FDM. It further needs to be verified whether a larger number of participants would result in even more stable FDM and thus in higher similarity between the experiments and an even faster convergent behaviour. Thresholds need to be determined that specify the minimum number of participants in order to achieve FDM for given similarity constraints. Finally, the comparisons presented in this article hold for eye-tracking experiments under task-free condition. Different results could be expected under a variety of viewing tasks, for instance, visual search tasks. These issues are out of the scope of this article and are subject for future work. To stimulate further VA research, we made the fixation data and FDM of the three databases publicly available to the research community: UWS at [47], TUD at [48], and UN at [25].

ACKNOWLEDGEMENT

The authors would like to thank Dr. Wang et al. for providing the image retargeting algorithm and the eye tracking experiment participants for sparing their valuable time.

REFERENCES

[1] B. A. Wandell, *Foundations of Vision*. Sinauer Associates, Inc., 1995.
[2] J. Wolfe, "Visual attention," in *Seeing*, K. K. D. Valois, Ed. Academic Press, 2000, pp. 335–386.
[3] K. H. Park and H. W. Park, "Region-of-interest coding based on set partitioning in hierarchical trees," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 12, no. 2, pp. 106–113, Feb. 2002.
[4] D. Wang, G. Li, W. Jia, and X. Luo, "Saliency-driven scaling optimization for image retargeting," *The Visual Computer*, Mar. 2011.
[5] K. Vu, K. A. Hua, and W. Tavanapong, "Image retrieval based on regions of interest," *IEEE Trans. on Knowledge and Data Mining*, vol. 15, no. 4, pp. 1045–1049, Jul. 2003.
[6] U. Engelke, H. Kaprykowsky, H.-J. Zepernick, and P. Ndjiki-Nya, "Visual attention in quality assessment: Theory, advances, and challenges," *IEEE Signal Processing Magazine*, vol. 28, no. 6, pp. 50–59, Nov. 2011.
[7] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.
[8] O. Le Meur, P. Le Callet, D. Barba, and D. Thoreau, "A coherent computational approach to model bottom-up visual attention," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 28, no. 5, pp. 802–817, May 2006.
[9] U. Rajashekar, I. van der Linde, A. C. Bovik, and L. K. Cormack, "GAFFE: A gaze-attentive fixation finding engine," *IEEE Trans. on Image Processing*, vol. 17, no. 4, pp. 564–573, Apr. 2008.
[10] N. D. B. Bruce and J. K. Tsotsos, "Saliency, attention, and visual search: An information theoretic approach," *Journal of Vision*, vol. 9, no. 3:5, pp. 1–24, 2009. [Online]. Available: http://journalofvision.org/9/3/5/
[11] U. Engelke and H.-J. Zepernick, "A framework for optimal region-of-interest based quality assessment in wireless imaging," *Journal of Electronic Imaging*, vol. 19, no. 1, 011005, Jan. 2010.
[12] A. J. Maeder, "The image importance approach to human vision based image quality characterization," *Pattern Recognition Letters*, vol. 26, no. 3, pp. 347–354, Feb. 2005.
[13] J. Wang, D. M. Chandler, and P. Le Callet, "Quantifying the relationship between visual salience and visual importance," in *Proc. of IS&T/SPIE Human Vision and Electronic Imaging XV*, vol. 7527, Jan. 2010.
[14] L. Itti and C. Koch, "Computational modelling of visual attention," *Nature Reviews Neuroscience*, vol. 2, pp. 192–203, Mar. 2001.
[15] J. M. Wolfe and T. S. Horowitz, "What attributes guide the deployment of visual attention and how do they do it?" *Nature Reviews Neuroscience*, vol. 5, pp. 1–7, Jun. 2004.
[16] A. T. Duchowski, *Eye Tracking Methodology: Theory and Practice*. Springer-Verlag New York, Inc., 2007.
[17] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," in *Proc. of IEEE Int. Conf. on Computer Vision*, 2009, pp. 2106–2113.
[18] U. Engelke, A. J. Maeder, and H.-J. Zepernick, "Visual attention modelling for subjective image quality databases," in *Proc. of IEEE Int. Workshop on Multimedia Signal Processing*, Oct. 2009, pp. 1–6.
[19] 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 Int. Conf. on Image Processing*, Nov. 2009, pp. 3097–3100.
[20] M. Cerf, E. P. Frady, and C. Koch, "Faces and text attract gaze independent of the task: Experimental data and computer model," *Journal of Vision*, vol. 9, no. 12:10, pp. 1–15, 2009. [Online]. Available: http://journalofvision.org/9/12/10/
[21] A. T. Duchowski, "A breadth-first survey of eye-tracking applications," *Behavior Research Methods*, vol. 34, no. 4, pp. 455–470, 2002.
[22] U. Engelke, A. Maeder, and H.-J. Zepernick, "Analysing inter-observer saliency variations in task-free viewing of natural images," in *Proc. of IEEE Int. Conf. on Image Processing*, Sep. 2010.
[23] M. Dorr, T. Martinetz, K. R. Gegenfurtner, and E. Barth, "Variability of eye movements when viewing dynamic natural scenes," *Journal of Vision*, vol. 10, no. 10:28, pp. 1–17, 2010. [Online]. Available: http://journalofvision.org/10/10/28/
[24] U. Engelke, H. Liu, H.-J. Zepernick, I. Heynderickx, and A. Maeder, "Comparing two eye-tracking databases: The effect of experimental setup and image presentation time on the creation of saliency maps," in *Proc. of IEEE Picture Coding Symposium*, Dec. 2010.
[25] J. Wang, R. Pepion, and P. Le Callet, "IRCCyN/IVC eyetracker images LIVE database," http://www.irccyn.ec-nantes.fr/spip.php?article839&lang=en, 2011.
[26] 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.
[27] EyeTech Digital Systems, "TM3 eye tracker," http://www.eyetechds.com/, 2009.
[28] SensoMotoric Instruments, "RED/RED250/RED500," http://www.smivision.com/en/eye-gaze-tracking-systems/products/red-red250-red-500.html, 2011.
[29] ——, "iView X Hi-Speed," http://www.smivision.com/en/eye-gaze-tracking-systems/products/iview-x-hi-speed.html, 2011.
[30] EyeTech Digital Systems, "TM3 eye tracker," http://www.eyetechds.com/research/tm3-qc, 2011.
[31] C. M. Privitera and L. W. Stark, "Algorithms for defining visual regions-of-interest: Comparison with eye fixations," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2000.
[32] G. Kootstra, B. Boer, and L. R. B. Schomaker, "Predicting eye fixations on complex visual stimuli using local symmetry," *Cognitive Computation*, vol. 3, no. 1, pp. 223–240, 2011.
[33] Q. Zhao and C. Koch, "Learning a saliency map using fixated locations in natural scenes," *Journal of Vision*, vol. 3, no. 3:9, pp. 1–15, 2011. [Online]. Available: http://journalofvision.org/11/3/9/
[34] R. J. Peters, A. Iyer, L. Itti, and C. Koch, "Components of bottom-up gaze allocation in natural images," *Vision Research*, vol. 45, no. 18, pp. 2397–2416, Aug. 2005.
[35] S. Marat, T. H. Phuoc, L. Granjon, N. Guyader, D. Pellerin, and A. Guérin-Dugué, "Modelling spatio-temporal saliency to predict gaze direction for short videos," *Int. Journal of Computer Vision*, vol. 82, no. 3, pp. 231–243, May 2009.
[36] J. Cohen, P. Cohen, S. G. West, and L. S. Aiken, *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, 3rd ed. Lawrence Erlbaum Associates, 2003.

[37] D. Green and J. Swets, *Signal Detection Theory and Psychophysics*. New York: John Wiley, 1966.
[38] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.
[39] R. Achanta, S. Hemami, F. Estrada, and S. Süsstrunk, "Frequency-tuned salient region detection," in *Proc. of IEEE Int. Conf on Computer Vision and Pattern Recognition*, Jun. 2009, pp. 1597–1604.
[40] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition*, Jun. 2007, pp. 1–8.
[41] A. Ninassi, O. L. Meur, P. Le Callet, and D. Barba, "Does where you gaze on an image affect your perception of quality? Applying visual attention to image quality metric," in *Proc. of IEEE Int. Conf. on Image Processing*, vol. 2, Oct. 2007, pp. 169–172.
[42] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[43] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. on Image Processing*, vol. 15, no. 2, pp. 430–444, Feb. 2006.
[44] H. Liu and I. Heynderickx, "Visual attention in objective image quality assessment: Based on eye-tracking data," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 21, no. 7, pp. 971–982, Jul. 2011.
[45] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," *ACM Transactions on Graphics*, vol. 26, no. 3, Jul. 2007.
[46] J. Stankiewicz, N. J. Anderson, and R. J. Moore, "Using performance efficiency for testing and optimization of visual attention models," in *Proc. of IS&T/SPIE Human Vision and Electronic Imaging XVI*, vol. 7865, Jan. 2011.
[47] 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.
[48] H. Liu and I. Heynderickx, "TUD image quality database: Eye-tracking release 1," http://mmi.tudelft.nl/ ingrid/eyetracking1.html, 2010.

Ulrich Engelke (S'06-M'11) received the Dipl.-Ing. degree in Electrical Engineering from RWTH Aachen University, Germany, and the Ph.D. degree in Telecommunications from the Blekinge Institute of Technology, Sweden. Currently he is with the Visual Experiences Group at Philips Research, The Netherlands, working with perception in lighting applications. His research interests include visual scene understanding, human visual perception, psychophysical experimentation, and signal processing.

Hantao Liu (S'07–M'11) received the M.Sc. in signal processing and communications from The University of Edinburgh (United Kingdom) in 2005, and the PhD degree in Computer Science from the Delft University of Technology (The Netherlands) in 2011. He is currently a Lecturer (Assistant Professor) at the Department of Computer Science at the University of Hull (United Kingdom). His research interests include image analysis, multimedia signal processing, and human visual perception.

Junle Wang received the double-M.S. degree in Signal processing (South China University of Technology, China), and Electronic Engineering (University of Nantes, France) in 2009, and the PhD degree in computer science from University of Nantes in 2012. He is currently an ATER (Assistant Professor) at the Department of Electronic and Digital Technologies. His research focuses on visual attention, depth perception and quality of experience of stereoscopic-3D content. His research interests also include human visual perception and psychophysical experimentation.

Patrick Le Callet received M.Sc. degree PhD degree in image processing from Ecole polytechnique de l'université de Nantes. He was also student at the Ecole Normale Superieure de Cachan where he got the "Aggrégation" (credentialing exam) in electronics of the French National Education. He has working as an Assistant professor from 1997 to 1999 and as a full time lecturer from 1999 to 2003 at the department of Electrical engineering of Technical Institute of University of Nantes (IUT). Since 2003 he is teaching at Ecole polytechnique de l'université de Nantes (Engineer School) in the Electrical Engineering and the Computer Science department where he is now Full Professor. Since 2006, he is the head of the Image and Video Communication lab at CNRS IRCCyN, a group of more than 35 researchers. He is mostly engaged in research dealing with the application of human vision modeling in image and video processing. His current centers of interest are 3D image and video quality assessment, watermarking techniques and visual attention modeling and applications. He is co-author of more than 140 publications and communications and co-inventor of 13 international patents on these topics. He is co-chairing within VQEG (Video Quality Expert Group) the "Joint-Effort Group" and "3DTV" activities. He is currently serving as associate editor for IEEE Transactions on Circuit System and Video Technology and SPRINGER EURASIP Journal on Image and Video Processing.

Ingrid Heynderickx received her PhD degree in physics at the University of Antwerp (Belgium) in December 1986. In 1987 she joined the Philips Research Laboratories in Eindhoven (The Netherlands), and meanwhile worked in different areas of research: optical design of displays, processing of liquid crystalline polymers and functionality of personal care devices. Since 1999 she is head of the research activities on Visual Perception of Display and Lighting Systems and in 2005 she is appointed research fellow in the group Visual Experiences. She is member of the Society for Information Displays (SID), and for the SID, she was chairman of the Applied Vision subcommittee from 2002 till 2007. In 2008, she became Fellow of the SID and chairman of the European Committee of the SID. In 2005, she is appointed Guest Research Professor at the Southeast University of Nanjing (China) and Part-time Full Professor at the University of Technology in Delft (the Netherlands).

Hans-Jürgen Zepernick (M'94-SM'11) received the Dipl.-Ing. degree from the University of Siegen in 1987 and the Dr.-Ing. degree from the University of Hagen in 1994. From 1987 to 1989, he was with Siemens AG, Germany. He is currently a Professor of radio communications at the Blekinge Institute of Technology, Sweden. Prior to this appointment, he held the positions of Professor of wireless communications at Curtin University of Technology; Deputy Director of the Australian Telecommunications Research Institute; and Associate Director of the Australian Telecommunications Cooperative Research Centre. His research interests include advanced radio communications, mobile multimedia, and perceptual quality assessment.

Anthony Maeder (M') is Professor of Health Informatics at University of Western Sydney in Australia and research leader of the Telehealth Research and Innovation Laboratory there. He previously held the positions of Research Director of the CSIRO eHealth Research Centre in Brisbane, Professor and Head of School in Engineering at University of Ballarat and in Electrical and Electronic Systems Engineering at Queensland University of Technology. Prior to that he undertook his PhD in Software Engineering at Monash University. He was the founding President of the Australian Pattern Recognition Society. His research interests are in human visual understanding for digital images, and various topics in eHealth.

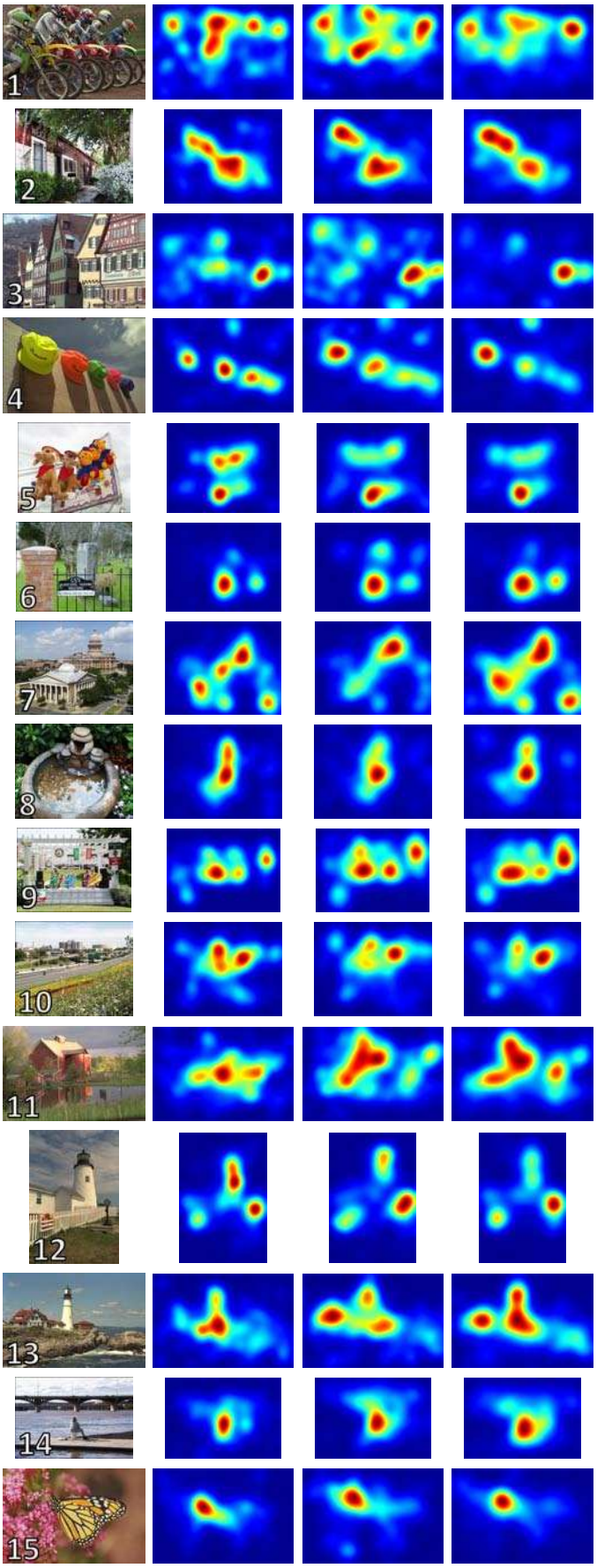

Fig. 17. Example FDM for images 1-15 and for a presentation time of 10 s. Left to right: original, UWS, TUD, UN.

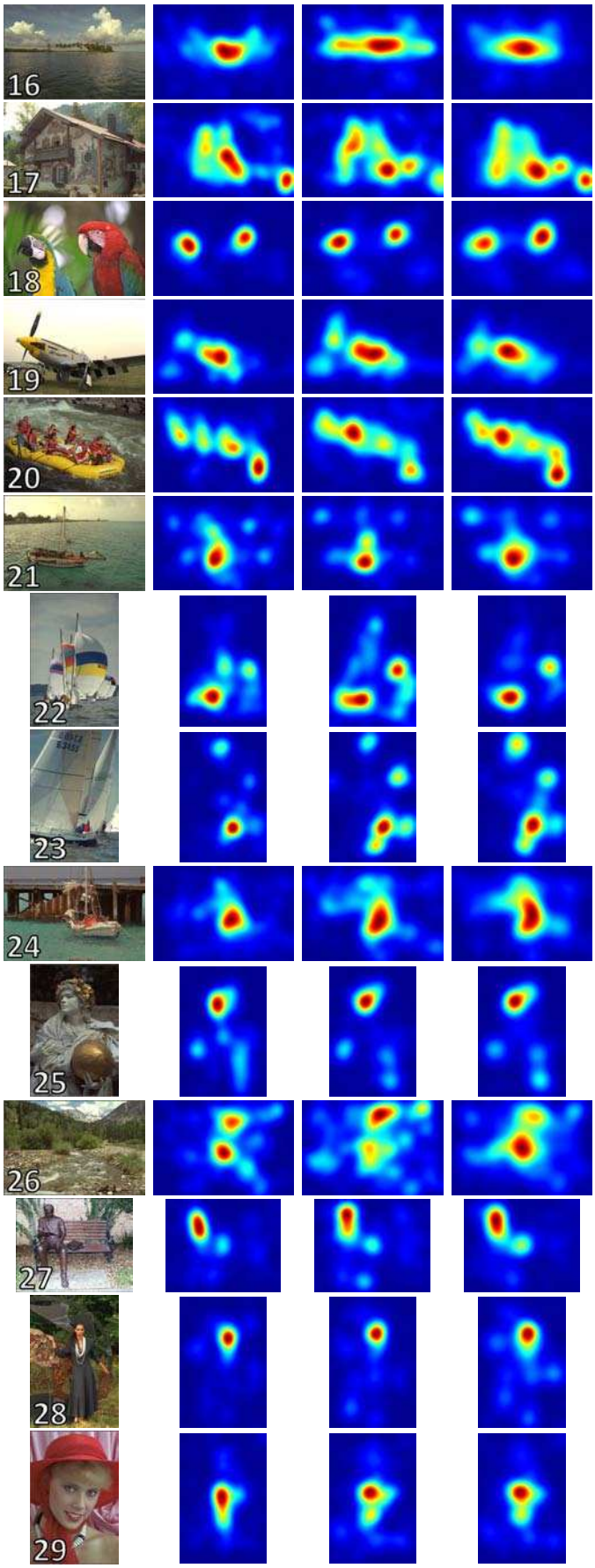

Fig. 18. Example FDM for images 16-29 and for a presentation time of 10 s. Left to right: original, UWS, TUD, UN.