Impact of Ultra High Definition on Visual Attention

Hiromi Nemoto
MMSPG, EPFL
hiromi.nemoto@epfl.ch

Philippe Hanhart
MMSPG, EPFL
philippe.hanhart@epfl.ch

Pavel Korshunov
MMSPG, EPFL
pavel.korshunov@epfl.ch

Touradj Ebrahimi
MMSPG, EPFL
touradj.ebrahimi@epfl.ch

ABSTRACT

Ultra high definition (UHD) TV is rapidly replacing high definition (HD) TV but little is known of its effects on human visual attention. However, a clear understanding of this effect is important, since accurate models, evaluation methodologies, and metrics for visual attention are essential in many areas, including image and video compression, camera and displays manufacturing, artistic content creation, and advertisement. In this paper, we address this problem by creating a dataset of UHD resolution images with corresponding eye-tracking data, and we show that there is a statistically significant difference between viewing strategies when watching UHD and HD contents. Furthermore, by evaluating five representative computational models of visual saliency, we demonstrate the decrease in models' accuracies on UHD contents when compared to HD contents. Therefore, to improve the accuracy of computational models for higher resolutions, we propose a segmentation-based resolution-adaptive weighting scheme. Our approach demonstrates that taking into account information about resolution of the images improves the performance of computational models.

Categories and Subject Descriptors

I.2.10 [**Artificial Intelligence**]: Vision and Scene Understanding—*perceptual reasoning, representations, data structures, and transforms*; H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems—*evaluation/methodology, video*

Keywords

Visual attention; ultra high definition; saliency map; subjective evaluations; eye-tracking

1. INTRODUCTION

Ultra high definition (UHD), a rapidly emerging immersive video technology, is expected to replace high definition (HD) as the next standard video format of digital TV. Most of the TV displays and video cameras manufacturers, as well as broadcasting companies, strongly promote UHD video content, since this technology enhances sensation of presence and provides better viewing experience [15, 24]. The increased resolution of UHD TV typically leads to larger display sizes and, hence, for the full enjoyment of UHD content, ITU standardization body recommends certain viewing conditions [18], as illustrated in Figure 1. The figure demonstrates the difference in viewing conditions between HD and UHD, suggesting that there might be also large differences in viewing strategies and in visual attention patterns of people watching HD and UHD TVs.

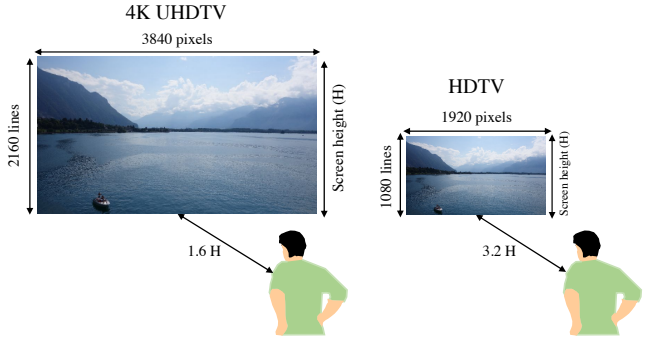


Figure 1: Recommended viewing conditions for 4K UHD and HD.

Visual attention is a widely studied topic and its practical applications include gaze-adaptive image and video compression [16, 6], objective image quality metrics [26], image retargeting [30], and image retrieval [29]. It even reaches beyond computing, proving useful in areas such as attention-based advertising, art, and cinema. To take advantage of visual attention information in practical applications, *salient regions* in images, i.e., regions that attract most of the attention, are either detected using an eye-tracking device or predicted using computational models of visual attention. One of the first computational model was proposed by Itti and Koch [17] in 1998, which uses image features such as luminance intensity, color, and orientation to construct a *saliency map*, i.e., a map predicting visual attention of a corresponding visual scene. The practical usefulness of computational models fueled the research for many years, resulting in many visual attention models, creation of different evaluation datasets with ground truth eye-tracking data, and various evaluation methodologies and metrics.

Although a significant number of public image and video datasets for visual attention exist [31], no dataset with eye-tracking data is available for UHD content. However, without this subjective data, it is hard to understand what is the impact of UHD on visual attention and whether it is significant for practical applications. In addition, computational models of visual attention are also evaluated using existing public databases, featuring images with resolution of HD or less, which means that little is known about the accuracy of these models when used for UHD content.

MM'14, November 3–7, 2014, Orlando, Florida, USA.

http://dx.doi.org/10.1145/2647868.2654917.

Since UHD has the ability to provide more details and requires higher data rate when compared to HD, understanding human attention patterns and viewing strategies for UHD content is important for developing efficient data compression algorithms and accurate objective quality metrics. The knowledge of visual attention for UHD can also help electronics manufactures to create better acquisition and display devices and content creators, such as photographers, movie and TV makers, to create images and video sequences with higher appeal value.

Therefore, this paper investigates the impact of UHD content on visual attention and on existing computational models of visual attention. To answer this question, a dataset is created first, which consists of 45 UHD 4K images with high variety of content (see examples in Figure 3 and refer to [25] for more details). Then, a subjective experiment involving 20 naïve subjects is conducted to collect eye-tracking data for these images using a professional eye-tracking system (Smart Eye Pro 5.8) and a professional reference monitor (Sony Trimaster SRM-L560). A similar experiment is also conducted with images resized to HD resolution for comparison with UHD. Fixation density maps (FDM) computed from the eye-tracking data for UHD and HD resolutions are compared using three metrics: attentional focus [20], similarity score [22], and Kullback-Leibler divergence (KLD) [10, 2]. Five representative models of visual attention were selected: 'Judd' by Judd *et al.* [23], 'GBVS' by Harel *et al.* [14], 'Itti' by Itti *et al.* [17], 'AIM' by Bruce [4], and 'Context-Aware' by Goferman *et al.* [12]. Their saliency maps were computed for UHD and HD contents and compared with corresponding subjective eye-tracking data using three metrics: similarity score, KLD, and area under the curve (AUC) [13].

The aim of this work is not only to evaluate the behavior of visual attention on UHD content, but also to investigate the impact on current computational models of visual attention and, if possible, to improve these models by making them adaptive to the content resolution. In summary, the main contributions of the paper are:

1. Dataset of UHD and HD images with corresponding subjective eye-tracking data;
2. Similarity analysis of FDMs for HD and UHD contents to understand if there is a difference in visual attention between UHD and HD resolutions;
3. Evaluation of five existing computational models of visual attention computed for UHD images to see if their performance is degraded when compared to HD;
4. Proposed resolution-adaptive weighting scheme to improve the accuracy of computational models of visual attention for UHD contents.

The rest of this paper is organized as follows. Section 2 presents related works including computation of FDM and several metrics for comparison of FDMs or FDM and saliency map. Section 3 describes eye tracking experiments, while Section 4 presents the results of subjective experiments. Sections 5 compares the performances of computational models of visual attention and Section 6 proposes the improvement of their performance for UHD content. Section 7 concludes the paper.

2. BACKGROUND AND RELATED WORK

A lot of work have been done over the years on fixation density map and computational visual attention modeling. In this section, we first give a brief overview of state of the art in visual attention, then describe how fixation density map (FDM) is computed and follow up with different approaches to measure the similarity of two FDMs or a FDM and a saliency map.

2.1 Related Work

A number of eye tracking experiments have been conducted for the purpose of investigating human visual attention mechanism. Previous research has suggested that a variety of factors influence human fixation patterns. The Delft Image Quality Lab carried out two eye tracking experiments in free-viewing task and in quality assessments task conditions [1] and found that there are significant differences on eye fixation patterns between these two different task conditions. In [27], the authors also reported that image distortion affects visual attention, especially for low quality images. The authors of [5] demonstrated that human faces are significant attentive regions in both free-viewing and searching conditions.

MIT CSAIL Saliency Database [21] provides a saliency database of low resolution images and tests the effect of resolution on fixation patterns. The results of the study showed that lower image resolutions contribute to the consistency of eye fixations and also noted that humans have a tendency to look at the image center for all resolutions. Image appeal is another feature that affects human fixation pattens. According to [11], more appealing images grab more attention than less appealing images. It also has been reported that current visual attention models are not able to capture this difference. The work by Engelke *et al.* [7] is the first comparative study to investigate the similarity between FDMs from independent experimental laboratory. Three eye tracking experiments were conducted independently using the same contents, though, they used somewhat different experimental conditions, e.g., viewing distance and image presentation time. It was shown that FDMs are very similar with each other and the impact of the dissimilarity of FDMs is not significant on practical applications such as visual attention modeling and image quality metrics.

Despite the large body of work on eye tracking and visual attention, to the best of our knowledge, no study have been reported on the influence of higher resolution images such as UHD resolution on human fixations or visual attention. Therefore, in this paper, we conduct subjective experiments with an eye tracking system under UHD viewing condition and perform rigorous analysis of visual attention using FDMs obtained in the subjective experiments.

2.2 Computation of fixation density maps

Fixation density maps (FDMs) are computed by convolving the recorded gaze points with a Gaussian filter, and then normalizing the result to values between 0 and 1. Only gaze points corresponding to fixation points are used to compute an FDM. Gaze points associated with saccades are not used in the computation. The eye tracking system used in our experiments (see Section 3.3) automatically discriminates between saccades and fixations based on the gaze velocity information. More specifically, during a time frame, all gaze points associated with gaze velocity below a fixation threshold are classified as fixation points, while saccades are detected when the gaze velocity lies above the fixation threshold. Blinks are also detected automatically by the eye tracking system based on the distance between the two eyelids of each eye. All detected saccades and blinks are excluded from the experimental data by the eye tracker and only the gaze points classified as fixation points are used further. These points are then filtered with a Gaussian kernel to compensate the eye tracker inaccuracies and to simulate the foveal point spread function of the human eye. As suggested in the state of the art [8, 22], the standard deviation of the Gaussian filter used for computing the FDMs is set to 1 degree of visual angle, which corresponds to $\sigma = 60$ pixels in our experiments for both HD and UHD. This standard deviation value is based on the assumption that the fovea of the human eye covers approximately 2 degrees of visual angle.

2.3 Assessment measures

Although several metrics have been proposed to measure the similarity between two FDMs or between a FDM and a saliency map computed by visual attention models, there is no standardized procedure. In [28], 12 similarity metrics have been compared using Kendall's W coefficient with a conclusion that a lot of these metrics are redundant. To avoid redundancies, the authors suggested to use three metrics, area under the curve (AUC) [3], Kullback-Leibler divergence (KLD) [10, 2], and one other optional metric (we selected the similarity score [22]) to capture different aspects of saliency maps. Since AUC is mostly used for comparison between computational saliency maps and the distributions of humans' fixation points, we used only KLD and similarity score metrics to analyze the similarities between two FDMs computed from our HD and UHD experiments (see Section 3 for more details). Additionally, the attentional focus measure [20] was used to characterize the visual focus of the subjects.

2.3.1 Attentional focus

Attentional focus [20] is defined as the number of objects that are viewed by the subjects during image observation. The rationale is to distinguish between cases where subjects look at few objects versus cases where they look more or less uniformly at several objects. To compute attentional focus, the FDM was first partitioned into blocks of $N \times N$ pixels. Then, the average intensity was computed for each block. Finally, the attentional focus was computed as the entropy of the normalized intensity across different blocks. Low entropy indicates high attentional focus while high entropy indicates low attentional focus. Figure 2 shows a schematic representation of this concept. The size of the blocks was determined so as to match the size of fovea, corresponding to 2 degrees of visual angle, which corresponds to 120×120 pixels in our experiments.

2.3.2 Similarity score

The similarity score is a distribution-based metric of how similar two saliency maps are. The similarity score S between two normalized maps P and Q is

$$S = \sum_{i,j} min(P_{i,j}, Q_{i,j}),\ where \sum_{i,j} P_{i,j} = \sum_{i,j} Q_{i,j} = 1 \quad (1)$$

If a similarity score is one, the two saliency maps are the same, if it is zero, the maps do not overlap at all.

2.3.3 Kullback-Leibler divergence

The Kullback-Leibler divergence (KLD) is usually used to estimate the dissimilarity between two probability distributions. In context of saliency maps or fixation density maps, this is a measure of dissimilarity between two histograms. If, in the corresponding histograms, $p(x)$ and $h(x)$ represent the probabilities of a pixel to have value x, the symmetric KLD is

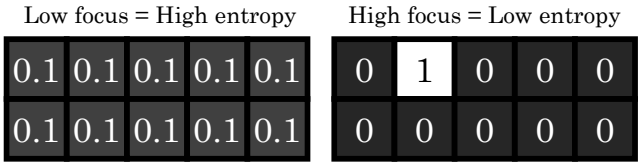
$$KLD = \frac{1}{2} \sum_{x} \left[p(x) log \frac{p(x)}{h(x)} + h(x) log \frac{h(x)}{p(x)} \right] \quad (2)$$

When two probability distributions are strictly equal, KLD value is zero, and when histograms do not overlap at all, it tends to infinity.

2.3.4 Area under the curve

Area under the curve (AUC) metric is the area under the receiver operating characteristics (ROC) curve [13]. In this metric, the human fixation points are treated as the positive set and the same number of points are randomly extracted from the image to form a negative set. Then, the saliency map is used as a classifier to separate the positive points from the negatives points. For a given threshold, all points in the saliency map having a value above this threshold are considered as fixation points. The fixation points in the positive and negative sets are labeled as true positive and false positive, respectively. By varying the threshold, the ROC curve is constructed from the corresponding false positive and true positive rates. The area under this curve indicates how well the saliency map estimates human fixations. An AUC value of 1 corresponds to perfect prediction and a value of 0.5 corresponds to random prediction.

The classical AUC method is often criticized, because its value can be artificially increased if a computational model considers the center bias technic, i.e., considering that human eye fixations are rarely located near the borders of typical test images. To overcome this drawback, the shuffled AUC has been proposed for better saliency maps validation [32]. In the shuffled AUC metric, to moderate center-bias effects, human fixations of other images from the same dataset are used to create the negative set instead of uniformly random points. To fairly compare computational models, the shuffled AUC was used in this paper.



*The values in the windows indicate normalized average intensity of FDM.

Figure 2: Illustration of attentional focus metric.

3. EYE TRACKING EXPERIMENTS

To investigate how different are visual attention and viewing strategies for HD and UHD resolutions, we conducted extensive experiments to acquire eye movements while a set of still images was shown in both HD and UHD resolutions to the subjects.

3.1 Test images

Since there is no publicly available standard dataset, at least to our knowledge, of UHD resolution images suitable for visual attention modeling, we constructed such dataset. For the dataset, we used still images with native resolution higher than UHD acquired by some of the latest digital cameras, including Sony DSC-RX100 II, Sony NEX-5N, FUJIFILM XF1, Olympus E-PL2, and RED SCARLET-X. Additionally, some high resolution painting images were obtained from the Europeana internet portal[1]. A total of 45 images were selected to cover a wide variety of content, e.g., natural scenes (both indoor and outdoor), humans, ships, animals, music gigs, historical scenes, etc. For the dataset, all images were cropped to 3840×2160 pixels for UHD resolution and then downsampled to 1920×1080 pixels for HD resolution using Lanczos resampling. Figure 3 shows some examples of the images and more details about the dataset can be found in [25].

3.2 Participants

A total of 20 naïve subjects (7 females and 13 males) took part in the experiments. Subjects were between 18 and 28 years old with an average of 23.8 years of age. Before the experiment, a consent form was handed to subjects for signature. All subjects were screened for correct visual acuity and color vision using Snellen and Ishiara charts respectively.

[1]Europeana: think culture, http://www.europeana.eu

Figure 3: Examples of images from our UHD dataset. The dataset is publicly accessible.[2]

3.3 Test environment

The experiments were conducted at the MMSPG quality test laboratory, which fulfills the recommendations for the subjective evaluation of visual data issued by ITU-R [19]. The test room was equipped with a controlled lighting system with a 6500 K color temperature and an ambient luminance at 15% of the maximum screen luminance, whereas the color of all the background walls and curtains present in the test area were in mid grey. The test room was separated in two by a curtain to isolate the subject and equipment from the test operators, which were present during the test session to supervise the recording of the eye tracking data. The laboratory setup was intended to ensure the reproducibility of the results and to avoid unintended influence of external factors.

Test stimuli were displayed on a professional high-performance 4K/QFHD 56" LCD reference monitor Sony Trimaster SRM-L560. A Smart Eye Pro 5.8 remote eye tracking system was employed to determine the gaze position on the screen of the left and right eyes independently. The system was equipped with three Sony HR-50 cameras at a frame rate of 60 fps and two infrared flashes, which enabled us to measure the gaze position with under 0.5 visual degrees error, while an accurate gaze output was available for at least ± 45 degrees of head rotation. All measurements from the eye tracker were recorded on a separate computer.

The experiment involved one subject per test session. The subject was seated in line with the center of the monitor at the distance of 3.2 and 1.6 times the image height for HD and UHD contents, respectively, as suggested in [18] as optimal viewing distance, which corresponds to roughly 1.1 meters from the monitor in both cases. The eye tracking system was placed at 0.7 meters from the monitor such that the face was well captured by the cameras. Figure 4 depicts the conditions of the experiments.

At the beginning of the test, the aperture and focus settings of the eye tracker cameras were adjusted for optimal conditions and a full camera calibration was performed to maximize the accuracy of the measurements. For each subject, a personal profile was created by recording several head poses and gaze calibrations using four calibration points close to the screen corners and one at the center of the screen. To ensure the accuracy of the eye tracking data, subjects were instructed to hold their head still while watching the images, and test operators made sure that all features were correctly detected by at least two out of three cameras during the experiment.

3.4 Experimental protocol

The experiment was separated into two different sessions to avoid inter-resolution comparison: one sessions was dedicated to UHD resolution only and another session to HD resolution only. To reduce the influence of potential memory effects on visual attention from viewing the same contents twice, the participants were divided into two groups of ten subjects each: the first group watched the images in UHD resolution first and then in HD resolution, while the reverse order was considered for the second group. Table 1 depicts the arrangement of the test sessions. To reduce contextual effects, the stimuli orders of display were randomized by applying different permutation for each subject. To reduce fatigue effects, each subject took a 15 minutes break between the two sessions.

[2]Ultra-Eye dataset, http://mmspg.epfl.ch/ultra-eye

Figure 4: Experimental setup.

Table 1: Arrangement of test sessions for HD and UHD resolutions.

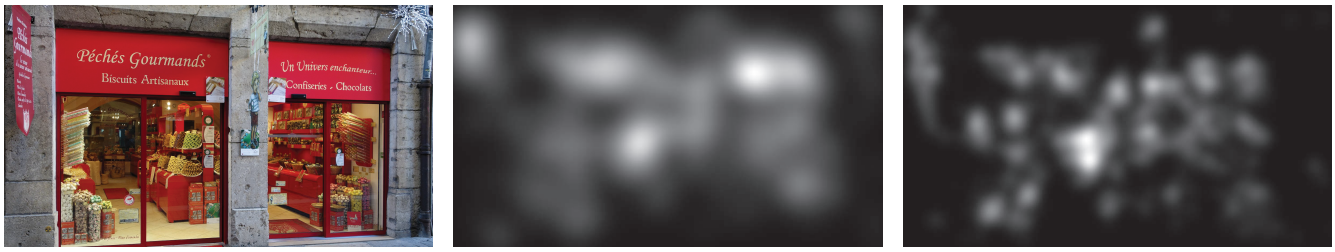
	Group #1	Group #2
First session	UHD resolution	HD resolution
Second session	HD resolution	UHD resolution

According to [7], the FDM is almost saturated at about four seconds presentation time. However, since the images used in our experiments were about four times larger than the ones used in [7], it is possible that the subjects are not able to watch all salient regions in the image if the presentation time is too short. Therefore, each image was shown for 15 seconds in our experiments. Additionally, a two seconds mid-grey background was displayed prior to the presentation of each test stimuli to reset subject's attention. With this timing, each session was approximately 15 minutes long.

Since the purpose of these experiments was to investigate the difference in visual attention and viewing strategies for HD and UHD resolutions, subjects were instructed to watch the images in a free-viewing scenario. Additionally, a training session was organized to allow subjects to familiarize with the procedure. The training materials were presented to subjects exactly as for the test materials.

To understand the influence of the memory effect on the subjective data, the following categories of FDMs were analyzed separately:

1. **UHD-First:** Group #1, first session (10 subjects).
2. **HD-First:** Group #2, first session (10 subjects).
3. **UHD-Second:** Group #2, second session (10 subjects). They watched UHD contents after watching the same images with HD resolution, followed by a 15 minutes resting phase.
4. **HD-Second:** Group #1, second session (10 subjects). They watched HD contents after watching the same images with UHD resolution, followed by a 15 minutes resting phase.
5. **UHD-All:** All 20 subjects.
6. **HD-All:** All 20 subjects.



(a) Presented image (b) FDM of HD resolution (HD-all) (c) FDM of UHD resolution (UHD-all)

Figure 5: Examples of FDMs for a presentation time of 15 s.

4. IMPACT ON VISUAL ATTENTION

Figure 5 shows examples of FDMs for HD and UHD resolutions computed from the eye-tracking data (across all subject groups). It can be noted from the figure that FDM of UHD resolution is more scattered and more 'focused' compared to FDM of HD resolution. In both cases, subjects look at various objects in the images. However, subjects watched specific objects in UHD images with higher intensity but browsed HD images in a more 'relaxed' way.

In this section, we compare FDMs for UHD and HD contents using the similarity metrics discussed in Section 2.3: attentional focus, similarity score, and KLD.

4.1 Attentional focus

Figure 6 shows the attentional focus computed separately for categories of FDMs described in Section 3.4 vs. varying presentation time. From the figure, it can be noted that, at each presentation time, the attentional focus of UHD resolution has lower value, which means that UHD has lower entropy or higher focus when compared to HD resolution, regardless of the presentation order. A possible explanation is that the higher level of details in UHD images make subjects' attention more focused and concentrated compared to HD images.

Also, attentional focus saturates faster for HD resolution than for UHD resolution, since UHD resolution images are four times bigger. To estimate the presentation time at which the FDMs are saturated, the attentional focus values were fitted using the response curve of a first order lag system according to the equation:

$$f(t) = a\left[1 - \exp\left(-\frac{t}{\tau}\right)\right] + b, \tag{3}$$

where t is the presentation time (how long the image was viewed by the subjects), a and b are the amplitude and the offset of the resulted attentional focus curve, and τ is a constant representing the time at which the attentional focus reaches 63.2% of its maximum value.

Considering that the saturation (95% of the maximum value) is achieved at 3τ, the FDMs are saturated after about 10.67 s for HD and after about 13.02 s for UHD. It means that 10 s is not enough to get a stable FDM for UHD resolution and that a longer presentation time is required.

Figure 6 shows that there is no influence of presentation order on the attentional focus and there is a difference between UHD and HD resolutions, but it does not show if these findings are statistically significant. To answer this question, we performed an analysis of variance (ANOVA) on attentional focus results at the presentation time equal to 15 seconds. ANOVA analysis was done for different pairs of FDMs with results shown in Table 2. The table shows that attentional focus is statistically significantly different for HD and UHD resolutions, while the presentation order of HD or UHD content, i.e., the order in which a subject viewed content, does not affect attentional focus in a statistically significant way. It means that even though each image was presented twice to the subjects, the influence of potential memory effects does not significantly impact attentional focus, indicating that the results from both groups of subjects can be combined.

Table 2: p-value computed for attentional focus ($t = 15$ s).

	HD-Second	UHD-First	UHD-Second
HD-First	0.074	< 0.001	< 0.001
HD-Second		< 0.001	< 0.001
UHD-First			0.11

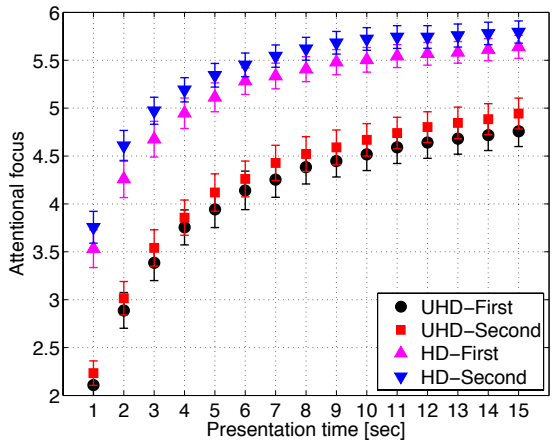
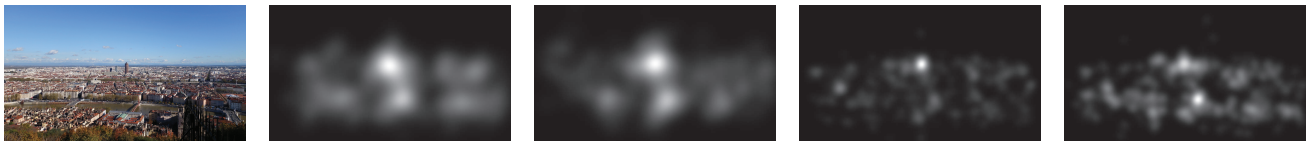


Figure 6: Attentional focus of FDMs with confidence intervals.

4.2 Similarity score

While attentional focus only measures one FDM, similarity score compares two different FDMs. We computed similarity scores for all meaningful pairs of FDMs (see Section 3.4 for explanation of different FDMs): HD-First vs. UHD-First, HD-Second vs. UHD-Second, HD-First vs. HD-Second, and UHD-First vs. UHD-Second; and the corresponding scores are shown in Figure 7 for all presentation times varying from 1 to 15 seconds. The figure demonstrates that HD-First (HD images were viewed before UHD) is more similar to HD-Second (UHD images were viewed before HD) than UHD-First and UHD-Second FDMs. This high similarity between FDMs for HD can also be noticed visually, for instance by comparing FMD of HD-First in Figure 8 (b) of a sample image in Figure 8 (a) with FDM of HD-Second in Figure 8 (c). In turn, the FDMs of UHD-First, shown in Figure 8 (d), and UHD-Second, shown in Figure 8 (e), are quite different visually too.



(a) Presented image (b) FDM of HD-First (c) FDM of HD-Second (d) FDM of UHD-First (e) FDM of UHD-Second

Figure 8: Examples of FDMs for different resolutions and viewing orders.

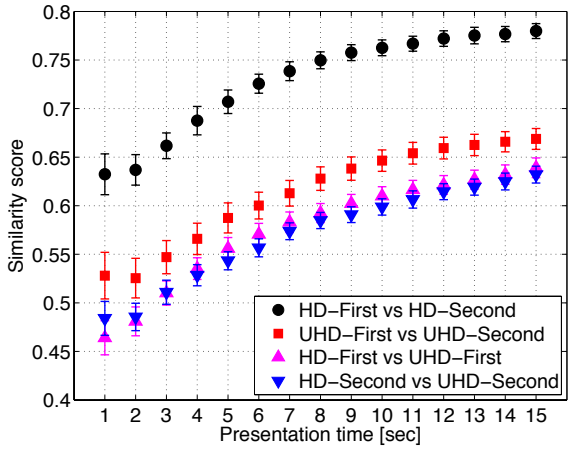


Figure 7: Similarity score of FDM pairs with confidence intervals.

This observation with the fact that other two pairs, HD-First vs. UHD-First and HD-Second vs. UHD-Second have almost the same similarity as UHD-First vs. UHD-Second, indicate that the fixation patterns for UHD resolution have higher diversity compared to HD resolution, i.e., different subjects look at UHD images in many different ways compared to a more unified way of viewing HD images. It also means that the presentation order does not influence the similarity score.

To analyze the statistical significance of the similarity score results, we performed an ANOVA analysis and computed p-values, comparing similarity scores between different pairs of FDMs as shown in Table 3. The table demonstrates that all results show statistically significant difference, except when comparing HD-First vs. UHD-First with HD-Second vs. UHD-Second. This analysis confirms the observation that there is a significant difference between FDMs of HD and UHD resolutions but the presentation order, in other words, memory effect, has no influence on the results, confirming the conclusions given in Section 4.1.

To better understand the dissimilarity between HD and UHD resolutions, scatter-like plot of the conjoint intensity values between two FDMs can be used [7]. Figure 9 shows such plot for the FDMs given in Figure 5 (b) and (c). In this plot, highly correlated FDM values lie closer to the main diagonal (dashed line). As it can be observed, there are several structural dissimilarities, especially for highly fixated points, which are due to the difference in the number of peaks and their respective positions in the actual FDMs.

Table 3: p-value computed for similarity score ($t = 15$ s).

	UHD-1st vs. UHD-2nd	HD-1st vs. UHD-1st	HD-2nd vs. UHD-2nd
HD-1st vs. HD-2nd	< 0.001	< 0.001	< 0.001
UHD-1st vs. UHD-2nd		< 0.001	< 0.001
HD-1st vs. UHD-1st			0.30

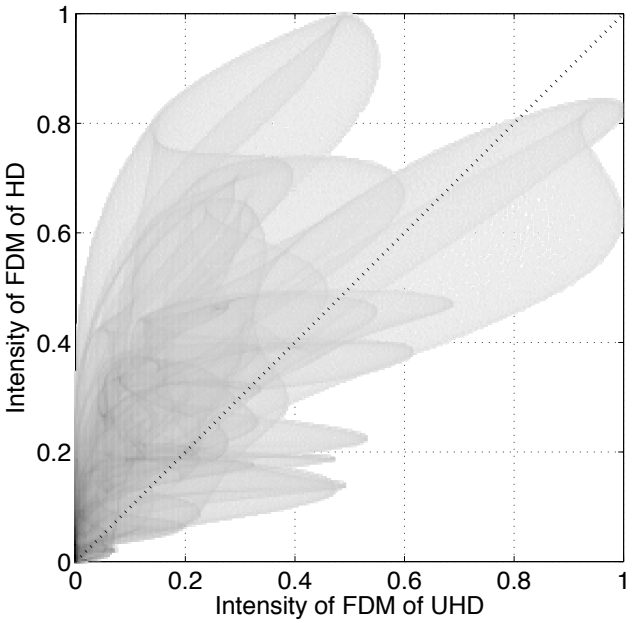


Figure 9: Scatter-like plot of the conjoint intensity values between the two FDMs of Figure 5.

Also, similarly to attentional focus, estimated τ time constant of similarity scores for HD ($\tau = 4.2$ s) is lower than for UHD resolution ($\tau = 5.1$ s).

4.3 Kullback-Leibler divergence

KLD metric was computed in the same way as similarity scores metric and results are shown in Figure 10. As it can be observed, KLD values are clearly saturated after 3 s for both HD and UHD. Since KLD measures the dissimilarity of the two histograms, this metric does not consider the spacial distribution but only evaluates the difference in the number of points of attention and their intensities. Therefore, this metric shows the difference between the viewing strategy of the subjects. Results in Figure 10 show very small KLD values when comparing FDMs of the same resolution, which suggests that the strategy to browse the images does not change much across subjects for a specific resolution. The fact that KLD values for UHD resolution (UHD-First vs. UHD-Second pair) are the lowest suggests that subjects are focusing on a fewer attentive regions in UHD compared to HD, probably, due to the higher resolution and higher level of details in UHD images. It can also be noted from the figure that KLD values for HD vs. UHD FDM pairs are much higher than for the same resolution pairs, which means that viewing strategies for HD and UHD resolutions are different.

Table 4: p-value computed for KLD ($t = 15$ s).

	UHD-1st vs. UHD-2nd	HD-1st vs. UHD-1st	HD-2nd vs. UHD-2nd
HD-1st vs. HD-2nd	< 0.001	< 0.001	< 0.001
UHD-1st vs. UHD-2nd		< 0.001	< 0.001
HD-1st vs. UHD-1st			0.53

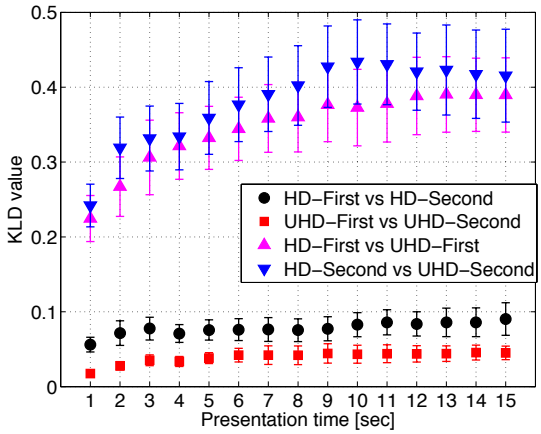


Figure 10: KLD of FDM pairs with confidence intervals.

To investigate the statistical significance of KLD results, we performed an ANOVA analysis in the same way as for similarity score metric. Table 4 shows p-values of the ANOVA analysis for KLD metric computed on different pairs of FDMs. From the table, it is clear that only HD-First vs. UHD-First is not significantly different compared to HD-Second vs. UHD-Second, which, similarly to the earlier observations, means that influence of memory effects is insignificant, but HD is significantly different from UHD.

5. IMPACT ON SALIENCY MAPS

In this section, we investigate the impact of UHD viewing on the performance of computational models of visual attention using three metrics: similarity score, KLD, and AUC.

5.1 Selection of computational models

In [22], Judd *et al.* benchmarked 10 computational models of visual attention using human fixations recorded with an eye tracker as ground truth. The authors have reported that Judd *et al.* [23] and graph-based visual saliency (GBVS) [14] models have the best and second best performance respectively. Other computational models of visual attention in the top five of their benchmarking are the Itti's [17], AIM [4], and Context-Aware [12] models. Following this benchmark, we use these five computational models of visual attention to estimate human fixations and to investigate whether using UHD content affect their performances. The selected models cover the whole spectrum of approaches. Context-Aware is top-down, Judd is hybrid, and the rest are bottom-up models. All models take image size into account when computing saliency map, therefore, we used UHD and HD images in their original sizes as inputs in the models.

5.2 Benchmarking of computational models

The performance of the computational models of visual attention is evaluated by comparing saliency maps generated by the models to the ground truth FDMs from our eye tracking experiments (see Section 3 for details). The similarity score, KLD, and AUC metrics are used as performance indexes. The FDMs of HD-All and UHD-All (see Section 3.4 for the explanation of the terms), obtained after 15 seconds presentation time, are used for the performance comparison, as our analysis showed that these FDMs are stable enough (see Section 4).

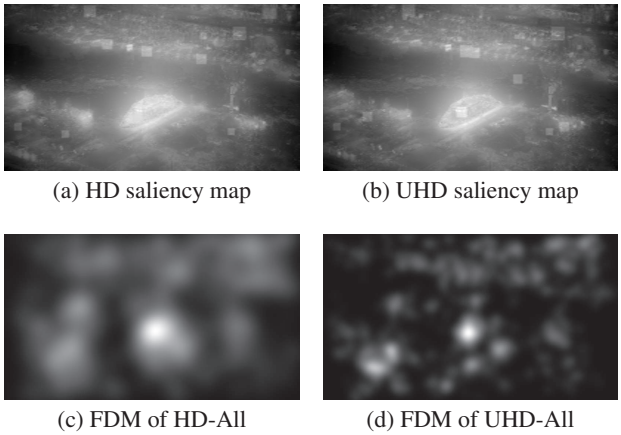
Table 5 reports the similarity scores, KLD, and AUC computed between saliency maps and FDMs. In case of UHD resolution, the performance of the models is very close to random (values are near 0.5) according to the similarity score metric. In particular, the Judd and GBVS models, which are reported to be the best models according to [22], show a drastic drop in performance between HD and UHD. ANOVA analysis was performed to determine whether the difference between the results of HD and UHD was statistically significant. The p-values are reported in Table 5. As it can be observed, the performances of all computational models of visual attention significantly decrease under UHD viewing condition when compared to HD viewing condition, with $p < 0.001$ in most cases, except for the AUC metric. Unlike the other metrics, according to the AUC metric, there is no significant difference between HD and UHD resolutions with regards to the performance of computational models of visual attention. This is most likely because AUC metric considers only the location of the fixation points, which are similar for HD and UHD, but it does not capture their intensity [33].

For KLD metric, the performance of computational models of visual attention was significantly lower in case of UHD resolution when compared to HD resolution, except for the AIM model. As the AIM model generates saliency maps having narrower probability distribution when compared to the other models, its performance is barely affected by the difference of the saliency map's statistical distribution between HD and UHD when benchmarked using the KLD metric. According to the KLD metric, the performance of Judd's model drops drastically. Looking at the saliency maps of this model (see Figure 11 (a) and (b)) show that the visual attention predicted by this model is quite spread, which is somewhat comparable to the FDM of HD resolution (see Figure 11 (c)), whereas this property does not match well with the structure of the FDM of UHD resolution (see Figure 11 (d)).

Overall, the results of the similarity and KLD metrics show that the current computational models of visual attention are not reliable to predict visual attention of the UHD content.

Table 5: Performance assessment of computational saliency maps compared to the FDMs of both HD and UHD.

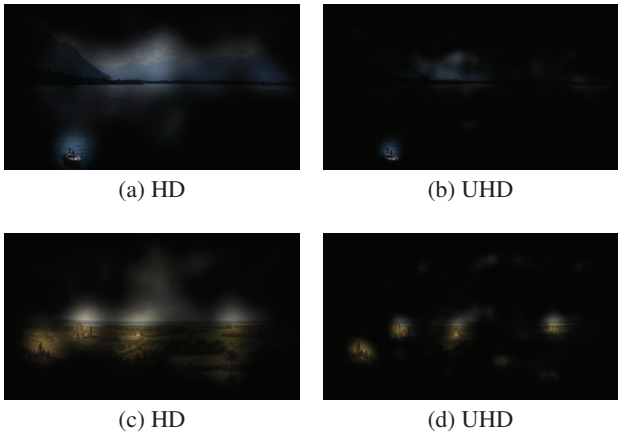
	Judd [23]		GBVS [14]		Itti [17]		AIM [4]		Context-Aware [12]	
	HD	UHD	HD	UHD	HD	UHD	HD	UHD	HD	UHD
Similarity score	0.64	0.51	0.69	0.54	0.63	0.52	0.58	0.47	0.62	0.52
p-value	< 0.001		< 0.001		< 0.001		< 0.001		< 0.001	
KLD	4.48	8.12	0.69	2.40	1.99	5.73	0.93	1.03	0.99	2.59
p-value	< 0.001		< 0.001		< 0.001		0.29		< 0.001	
AUC	0.61	0.61	0.61	0.61	0.63	0.65	0.62	0.62	0.64	0.65
p-value	0.87		0.97		0.16		0.51		0.58	



(a) HD saliency map (b) UHD saliency map

(c) FDM of HD-All (d) FDM of UHD-All

Figure 11: Performance of Judd model's saliency map vs. FDMs.



(a) HD (b) UHD

(c) HD (d) UHD

Figure 12: FDMs of HD-All and UHD-All overlaid on the images.

6. RESOLUTION-BASED WEIGHTING

The decrease in the performance of computational visual attention models for UHD resolution demonstrated in Section 5.2 indicates that these models should take viewing condition associated with image resolution into account. To verify this hypothesis, we propose a simple resolution-weighting scheme and show that it can be used to improve the performance of computational models when the higher resolution content is used.

6.1 Proposed method

By studying FDMs of UHD content, we noticed that subjects seem to look at smaller objects more when watching UHD images than when watching HD images. For example, in Figure 12 (b), subjects looked at the small boat in a focused way in UHD version (see the original image in Figure 3), while they browse other areas equally in HD (see Figure 12 (b)). Also, in Figure 12 (d) the subjects focused at the small house, the stone architecture, and humans (refer to Figure 3 for the original image) in a more concentrated manner when viewing UHD than when viewing HD (see Figure 12 (c)). Following this observation, we propose to tune saliency maps according to the sizes of objects in images. To automatically estimate these sizes, we use a graph-based image segmentation [9]. This particular segmentation technique was chosen because it ignores details in high-variability regions, which matches our purpose of estimating the sizes of main objects only. An example of

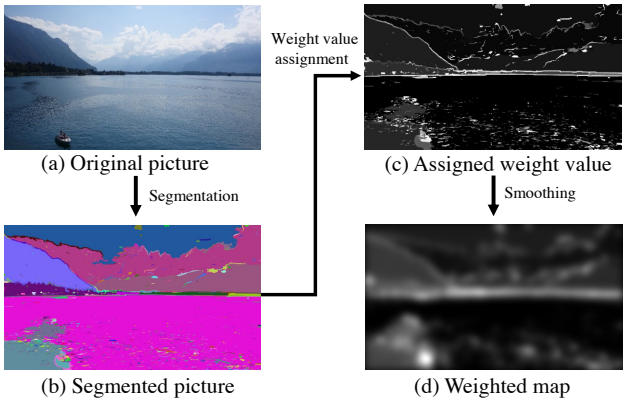


Figure 13: Computation of weighted map.

the segmentation is shown in Figure 13 (b). The authors of the segmentation use certain parameters to adjust the behaviors of the algorithm, which we set according to values in Table 6. These values may not work for other resolutions than UHD but, in this paper, we only want to demonstrate the possibility of improving saliency maps for UHD resolutions. The development of computational model of visual attention that is automatically adaptive to changing resolution is out of the scope of this paper.

Table 6: Parameters for the segmentation.

Parameter	Value	Explanation
$sigma$	1	Smoothing coefficient
k	100	Value for the threshold function
min	20	Minimum component size

Computation of the proposed resolution-adaptive weighting map is illustrated by Figure 13. For each segmentation component in the segmented image, a weight value is assigned depending on its size in the range between 0 and 1. Since subjects tend to look at smaller objects in UHD images, the weight value should be high when the area of the segmentation component is small. However, with such approach many small segmentation fragments, which mainly come from trivial textures in larger objects, will receive a high weight value, leading to an incorrect weighting. To avoid this, for each segmentation component C_n, we compute its weight value W_n using the following equation:

$$W_n = \begin{cases} n(1-A)/B + A & \text{if } n \leq B \\ (B-n)/(N-1-B) + 1 & \text{if } n > B \end{cases} \quad (4)$$

where A and B are tuning parameters that depend on resolution of the images and $n \in \{0, ..., N-1\}$ is the index of the list of N segmentation components sorted by size. For a fixed resolution (in our case UHD), one can fine-tune these parameters, so that the small fragments and large components receive small weights, while medium-sized components of interest receive large weights.

An example of the map with assigned weight values using this approach is shown in Figure 13 (c). This map is smoothed with a Gaussian filter (see Figure 13 (d)) to avoid sharp transitions. The standard deviation of the filter kernel is the same $\sigma = 60$ as it was used for computation of FDMs (see Section 2.2 for details). The saliency map of a given computational model of visual attention is then multiplied by this weighted map resulting in a new saliency map weighted according to the object sizes in the image.

Table 7: Comparison of saliency maps of original computational models (Original) with weighted models (Proposed) for UHD.

	Judd [23]		GBVS [14]		Itti [17]		AIM [4]		Context-Aware [12]	
	Original	Proposed	Original	Proposed	Original	Proposed	Original	Proposed	Original	Proposed
Similarity score	0.51	0.55	0.54	0.56	0.52	0.52	0.47	0.51	0.52	0.51
p-value	< 0.001		0.36		0.82		0.02		0.58	
KLD	8.12	1.66	2.40	0.53	5.73	0.96	1.03	2.0	2.59	0.76
p-value	< 0.001		< 0.001		< 0.001		< 0.001		< 0.001	
AUC	0.61	0.68	0.61	0.66	0.65	0.68	0.62	0.68	0.65	0.67
p-value	< 0.001		< 0.001		0.1		< 0.001		0.2	

This simple approach allows to indirectly embed a resolution information into the saliency map computation. The examples of the new saliency maps are shown in Figure 14 side-by-side with saliency maps obtained using the original computational models as presented in Section 5.1.

6.2 Assessment of the proposed method

To test the effectiveness of these new weighted maps, we compared them with the original saliency maps by computing similarity score, KLD, and AUC metrics, similar to how we evaluated the performance of the original saliency maps for HD and UHD resolutions in Section 5.2. The results shown in Table 7 demonstrate that the new weighted saliency maps improve the performance of the most of the computational models. Only in case of AIM model and KLD metric, the new weighted saliency map shows decrease in the performance. It means that by using a rather simple approach, taking into account resolution of the images and their content structure, we could improve the performance of the existing models on higher resolution images.

7. CONCLUSION

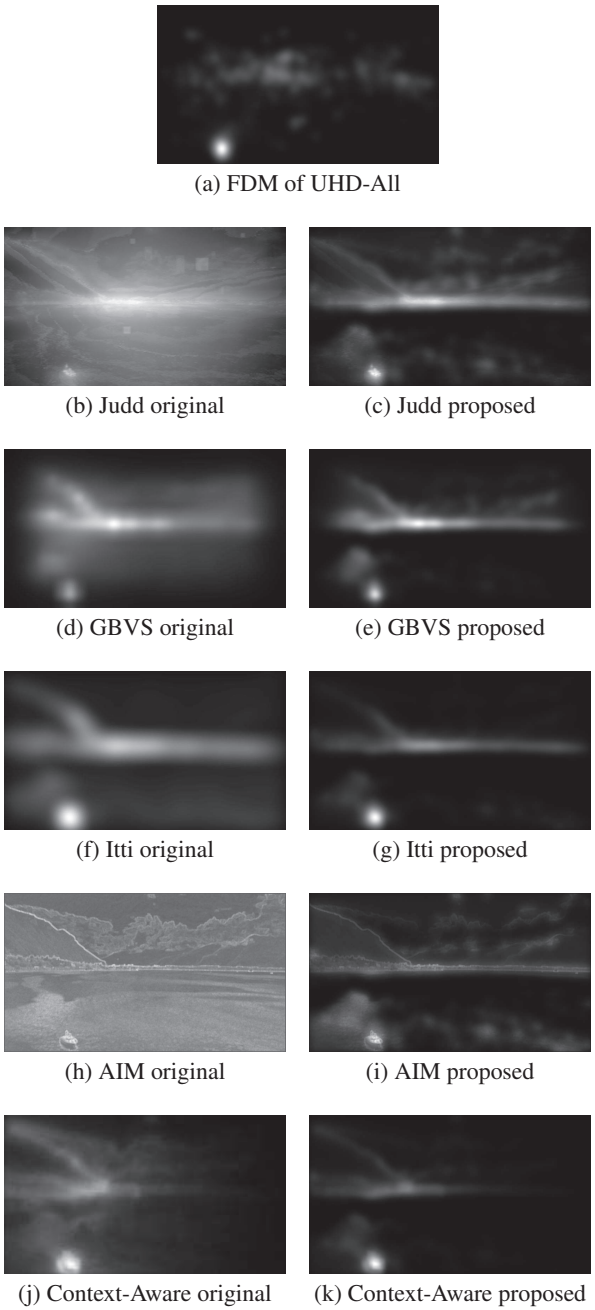
In this paper, we studied the influence of UHD resolution on human visual attention and attention models. We conducted subjective eye tracking experiments with both HD and UHD resolution images covering wide variety of scenes. We then created the fixation density maps for HD and UHD images and evaluated them using three different objective metrics: attentional focus, similarity score, and KLD, effectively creating a visual attention dataset for HD and UHD contents.

The assessment results demonstrated that (i) UHD resolution images can grab the focus of attention more than HD images; (ii) humans tend to look at a few attentive regions in the images with more intent when viewing UHD; and (iii) viewing strategy is different for HD and UHD.

We also compared five different computational models of visual attention when applying to HD and UHD images, showing that models' performance degrade with the increase in resolution. The evaluation suggests that an image resolution and the structural content information of an image should be taken into account when developing a resolution-independent computational model. Adding such information into several existing computational models via a simple segmentation-based weighting scheme increased the performance of these models.

Acknowledgments

This work has been conducted in the framework of the Swiss National Foundation for Scientific Research (FN 200021-143696-1), EC funded Network of Excellence VideoSense, and COST IC1003 European Network on Quality of Experience in Multimedia Systems and Services QUALINET.



(a) FDM of UHD-All

(b) Judd original (c) Judd proposed

(d) GBVS original (e) GBVS proposed

(f) Itti original (g) Itti proposed

(h) AIM original (i) AIM proposed

(j) Context-Aware original (k) Context-Aware proposed

Figure 14: Left column: original saliency maps. Right column: improved saliency maps with resolution adaptive weighting.

8. REFERENCES

[1] H. Alers, L. Bos, and I. Heynderickx. How the task of evaluating image quality influences viewing behavior. In *International Workshop on Quality of Multimedia Experience (QoMEX)*, pages 167–172, Sept. 2011.

[2] A. Borji and L. Itti. State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):185–207, Jan. 2013.

[3] A. Borji, D. N. Sihite, and L. Itti. Quantitative Analysis of Human-Model Agreement in Visual Saliency Modeling: A Comparative Study. *IEEE Transactions on Image Processing*, 22(1):55–69, 2013.

[4] N. D. B. Bruce. *Saliency, Attention and Visual Search: An Information Theoretic Approach*. PhD thesis, York University, Canada, 2008. AAINR45988.

[5] M. Cerf, J. Harel, W. Einhaeuser, and C. Koch. Predicting human gaze using low-level saliency combined with face detection. In J. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems 20*, pages 241–248. Curran Associates, Inc., 2008.

[6] Z. Chen, W. Lin, and K. N. Ngan. Perceptual video coding: Challenges and approaches. In *IEEE International Conference on Multimedia and Expo (ICME)*, pages 784–789, July 2010.

[7] U. Engelke, H. Liu, J. Wang, P. L. Callet, I. Heynderickx, H.-J. Zepernick, and A. J. Maeder. Comparative Study of Fixation Density Maps. *IEEE Transactions on Image Processing*, 22(3):1121–1133, 2013.

[8] U. Engelke, A. Maeder, and H. Zepernick. Visual attention modelling for subjective image quality databases. In *IEEE International Workshop on Multimedia Signal Processing (MMSP)*, pages 1–6, Oct. 2009.

[9] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient Graph-Based Image Segmentation. *International Journal of Computer Vision*, 59(2):167–181, Sept. 2004.

[10] M. S. Gide and L. J. Karam. Comparative evaluation of visual saliency models for quality assessment task. In *International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)*, pages 37–40, Jan. 2012.

[11] S. Gilani, R. Subramanian, H. Hua, S. Winkler, and S.-C. Yen. Impact of image appeal on visual attention during photo triaging. In *IEEE International Conference on Image Processing (ICIP)*, pages 231–235, Sept. 2013.

[12] S. Goferman, L. Zelnik-Manor, and A. Tal. Context-aware saliency detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2376–2383, June 2010.

[13] D. M. Green and J. A. Swets. *Signal Detection Theory and Psychophysics*. Wiley, New York, 1966.

[14] J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. In *Advances in neural information processing systems*, volume 19, pages 545–552. MIT Press, 2007.

[15] T. Ito. Future television — Super Hi-Vision and beyond. In *IEEE Asian Solid State Circuits Conference (A-SSCC)*, pages 1–4, Nov. 2010.

[16] L. Itti. Automatic foveation for video compression using a neurobiological model of visual attention. *IEEE Transactions on Image Processing*, 13(10):1304–1318, Oct. 2004.

[17] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, Nov. 1998.

[18] ITU-R BT.2022. General viewing conditions for subjective assessment of quality of SDTV and HDTV television pictures on flat panel displays. International Telecommunication Union, Aug. 2012.

[19] ITU-R BT.500-13. Methodology for the subjective assessment of the quality of television pictures. International Telecommunication Union, Jan. 2012.

[20] P. Jermann, M.-A. Nuessli, and K. Sharma. Attentional Episodes and Focus. In *Dual Eye-Tracking workshop in ACM Conference on Computer Supported Cooperative Work*, Seattle, Washington, USA, Feb. 2012.

[21] T. Judd, F. Durand, and A. Torralba. A model of saliency-based visual attention for rapid scene analysis. *Journal of Vision 2011*, 11(4):1254–1259, Apr. 2011.

[22] T. Judd, F. Durand, and A. Torralba. A Benchmark of Computational Models of Saliency to Predict Human Fixations. Technical Report MIT-CSAIL-TR-2012-001, CSAIL, MIT, Jan. 2012.

[23] T. Judd, K. Ehinger, F. Durand, and A. Torralba. Learning to predict where humans look. In *IEEE International Conference on Computer Vision (ICCV)*, pages 2106–2113, Sept. 2009.

[24] K. Masaoka, M. Emoto, M. Sugawara, and Y. Nojiri. Contrast effect in evaluating the sense of presence for wide displays. *Journal of the Society for Information Display*, 14(9):785–791, 2006.

[25] H. Nemoto, P. Hanhart, P. Korshunov, and T. Ebrahimi. Ultra-Eye: UHD and HD images eye tracking dataset. In *International Workshop on Quality of Multimedia Experience (QoMEX)*, Sept. 2014.

[26] J. Redi, H. Liu, P. Gastaldo, R. Zunino, and I. Heynderickx. How to apply spatial saliency into objective metrics for JPEG compressed images? In *IEEE International Conference on Image Processing (ICIP)*, pages 961–964, Nov. 2009.

[27] J. Redi, H. Liu, R. Zunino, and I. Heynderickx. Interactions of visual attention and quality perception. In *Proc. SPIE*, volume 7865, pages 78650S–78650S–11, Feb. 2011.

[28] N. Riche, M. Duvinage, M. Mancas, B. Gosselin, and T. Dutoit. Saliency and Human Fixations: State-of-the-Art and Study of Comparison Metrics. In *IEEE International Conference on Computer Vision (ICCV)*, pages 1153–1160, Dec. 2013.

[29] K. Vu, K. A. Hua, S. Member, and W. Tavanapong. Image Retrieval Based on Regions of Interest. *IEEE Transactions on Knowledge and Data Engineering*, 15:1045–1049, 2003.

[30] D. Wang, G. Li, W. Jia, and X. Luo. Saliency-driven Scaling Optimization for Image Retargeting. *Visual Computing*, 27(9):853–860, Sept. 2011.

[31] S. Winkler and R. Subramanian. Overview of Eye tracking Datasets. In *International Workshop on Quality of Multimedia Experience (QoMEX)*, pages 212–217, July 2013.

[32] L. Zhang, M. H. Tong, T. K. Marks, H. Shan, and G. W. Cottrell. SUN: A Bayesian framework for saliency using natural statistics. *Journal of vision*, 8(7):32.1–20, 2008.

[33] Q. Zhao and C. Koch. Learning a saliency map using fixated locations in natural scenes. *Journal of vision*, 11(3):1–15, Mar. 2011.