A new HD and UHD video eye tracking dataset

Toinon Vigier
IRCCyN UMR CNRS
Université de Nantes
Nantes, France
toinon.vigier@univ-nantes.fr

Josselin Rousseau
IRCCyN UMR CNRS
Université de Nantes
Nantes, France
josselin.rousseau@univ-nantes.fr

Matthieu Perreira Da Silva
IRCCyN UMR CNRS
Université de Nantes
Nantes, France
matthieu.perreiradasilva@univ-nantes.fr

Patrick Le Callet
IRCCyN UMR CNRS
Université de Nantes
Nantes, France
patrick.lecallet@univ-nantes.fr

ABSTRACT

The emergence of UHD video format induces larger screens and involves a wider stimulated visual angle. Therefore, its effect on visual attention can be questioned since it can impact quality assessment, metrics but also the whole chain of video processing and creation. Moreover, changes in visual attention from different viewing conditions challenge visual attention models. In this paper, we present a new HD and UHD video eye tracking dataset composed of 37 high quality videos observed by more than 35 naive observers. This dataset can be used to compare viewing behavior and visual saliency in HD and UHD, as well as for any study on dynamic visual attention in videos. It is available at http://ivc.univ-nantes.fr/en/databases/HD_UHD_Eyetracking_Videos/.

CCS Concepts

•**Information systems** → **Multimedia databases;** •**General and reference** → *Evaluation;*

Keywords

Eye tracking ; video ; UHD.

1. INTRODUCTION

UHD TV standard defines new video technologies as an increasing resolution from HD (1920 × 1080) to UHD, i.e. 4K (3840 × 2160) or 8K (7680 × 4320). The emergence of UHD potentially provides a better immersion of the user thanks to a wider visual angle with appropriate larger screens [4]. Indeed, ITU defines the optimal viewing distance as the distance at which scanning lines just cannot be perceived with visual acuity of 1'. It is thus set to 3H for HD and 1.5H for 4K-UHD where H is the height of the screen [2]. Figure 1 shows the increase of stimulated visual angle along with a better resolution. This increase of resolution and stimulated visual angle can modify visual attention deployment and visual patterns of people looking at HD and UHD videos.
Visual attention is a widely developed topic for many years, which finds a variety of applications as image and video compression, image and video objective quality metrics, computer vision and robotics, eye controlled display, attention-based video content creation, etc. In these applications, visual attention can be directly studied from gaze data tracked in subjective experiments or predicted using visual saliency models based on top-down or bottom-up factors. However, these prediction models most often elude viewing conditions. Therefore, the changes of viewing conditions in the transition from HD to UHD raise several issues on the impact of visual attention deployment and viewing behavior in videos, and on the performance of visual saliency models. Eye tracking experiments can provide very useful information to tackle these issues.

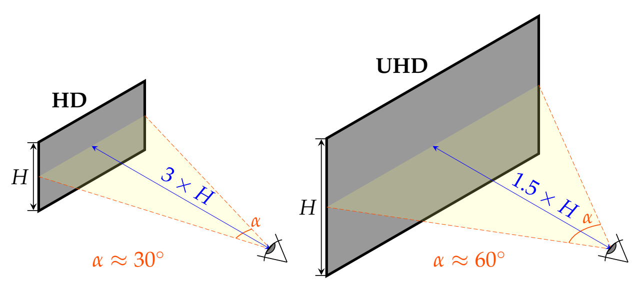


Figure 1: The increase of stimulated visual angle from HD to UHD.

In this paper, we propose a new eye tracking dataset in HD and 4K UHD of 37 high quality videos observed by more than 35 naive observers. The rest of this paper is organized as follows. Section 2 describes two related datasets on visual attention in UHD. Section 3 presents a new eye tracking

MM'Sys '16 Klagenfurt am Wörthersee, Austria

DOI: 10.475/123_4

setup adapted to UHD viewing conditions and used to create our dataset. Section 4 describes the proposed dataset. Section 5 gives dataset usage for future research works. Section 6 concludes the paper.
In the following, UHD exclusively refers to 4K resolution.

2. RELATED DATASETS

To our knowledge, only two recent datasets are used to study the effect of transition from HD to UHD on visual attention [9, 10, 7].

2.1 Ultra-eye

Ultra-eye is a publicly accessible dataset composed of 41 UHD and HD images [10]. HD images are downsampled from UHD with Lanczos filter. For each image, the dataset provides the list of fixation points and the fixation density maps. Eye movement data are recorded with the Smart Eye eye tracker using 20 naive subjects in two session (UHD then HD or HD then UHD). Images were presented in random order during 15 seconds in a test laboratory which fulfills the ITU recommendations. The viewing distance was respectively 1.6H and 3.2H, in UHD and in HD.
From the eye tracking data, the authors pointed out that viewing strategy and visual attention are significantly different in these two cases: UHD images can grab the focus of attention more than HD images. Moreover, several models of visual saliency were compared in HD and UHD scenarios, showing a reduction of model performance in UHD [9].
However, viewing behavior in video differs from static images, preventing the straightforward use of these observations for dynamic content.

2.2 UHD video saliency dataset of Shangai University

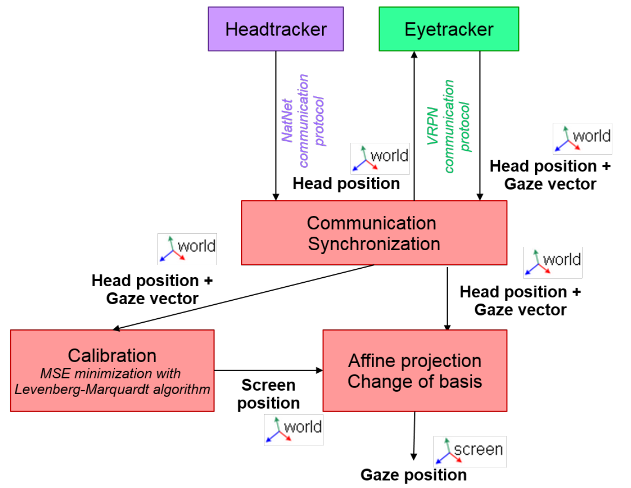
To our knowledge, the first and unique UHD video saliency dataset was published in [7]. These data come with a comparison of viewing behavior in UHD and HD scenarios. Eye movement data were recorded with the Tobii Eye-tracker X120 using 20 naive subjects in two sessions (UHD then HD). Fourteen videos of the SJTU 4K video sequences were used in native format (UHD) and downscaled to HD [11].
To analyze the gaze data, the new concept of aggregation maps (AGM) was introduced. It consists in the aggregation of all fixation points of one viewer for a video sequence in a unique map. With the AGM, an aggregation score (AGS) was computed which is an indicator of fixation concentration at the center of the screen. Thus, it was shown that viewer attention was more focused on the center of the screen in HD context.
However, the viewing distance in UHD and HD was constant, equal to 3H. It does not comply with ITU recommendations and stimulated visual angle is unchanged. Moreover, the fact that people always started with UHD scenario can skew the results because of memorization. Therefore, we propose to construct a new HD/UHD visual attention video dataset regarding ITU recommendations.

3. THE EYE HEAD TRACKER: A NEW EYE TRACKING SYSTEM FOR UHD

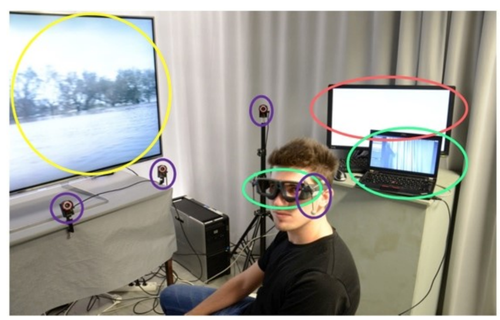
In this section we present a new eye tracking system adapted to large stimulated visual angle and used in our new dataset.

3.1 Description of EHT

Because of the larger stimulated visual angle in UHD, observers can need to move more their head and eye tracking systems may not be accurate enough at the edges of the screen. We developed a new setup to address this issue: the Eye Head Tracker (EHT). EHT is a combination of the mobile SMI eye tracking glasses and of the head tracker OptiTrack ARENA. We implemented an application which collects these two data in order to provide the gaze position in the screen plane as it is explained in Figure 2.



(a) EHT operating scheme.



(b) EHT setup in the viewing environment.

Figure 2: The Eye Head Tracker.

The EHT frequency is 30 Hz in binocular mode.

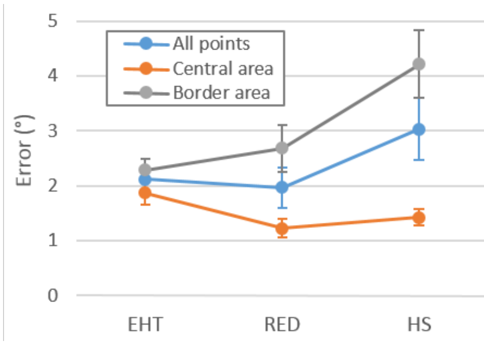
3.2 Evaluation of EHT

This setup was internally evaluated on 21 viewers along with two other systems: the remote SMI RED and the SMI Hi-Speed (HS) on a 65” UHD TV Panasonic TX-L65WT600E. The viewing distance was 1.5H, i.e. 120 cm. During the test, observers looked successively at 22 points displayed on the screen for two seconds.
Performance of eye trackers was mainly assessed through three metrics:

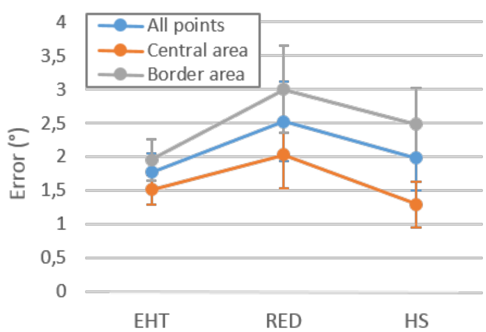
Accuracy: euclidean distance (in visual angle) between measured point and display point on the screen.

Robustness: euclidean distance (in visual angle) between measured point and centroid of all measured points for one display point on the screen.

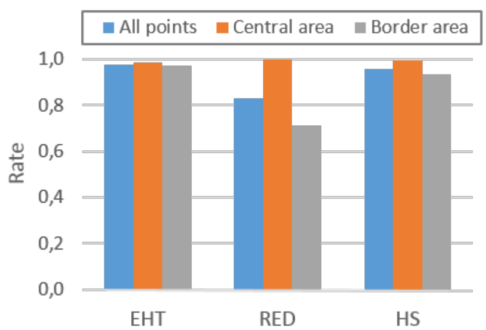
Recording rate: the ratio between real measured points and expected points (according to setup frequency).



(a) Accuracy



(b) Robustness



(c) Recording rate

Figure 3: Performance comparison of eye trackers. (Bars represent the standard errors.)

Figure 3 shows that EHT improves the number of recorded points mostly in border areas with a better robustness without loss of accuracy. To summarize, the advantages of the EHT are the non restriction of head movements, a large ocular field and a good accuracy at the edges.

4. DATASET DESCRIPTION

In this section, we describe the new HD and UHD video eye tracking dataset, freely available at http://ivc.univ-nantes.fr/en/databases/HD_UHD_Eyetracking_Videos/.

4.1 Video content

The dataset is composed of 37 native UHD high quality video sequences from seven content provider: SJTU Media Lab [11], Big Bug Bunny (Peach open movie project), Ultra Video Group, Elemental Technologies , Sveriges Television AB (SVT), Harmonic, Tears of steel (Mango open movie project).

In HD, the original sequences were downscaled with Lanczos-3 algorithm which was proven as the best filter both in terms of performance and perceptual quality [8].

The frame rate of the original sequences varies from 25 to 120 fps. They were uniformly played frame by frame with 25 fps in our test, causing some movements to appear a bit slower than in reality.

We did not use temporal downscaling methods because they often introduce more artifacts than the slowdown effect, particularly for non-integer ratios. Each source was cut to clips with a length of 8 to 12 seconds, producing a total of around 300 frames each.

Spatial perceptual information (SI) and temporal perceptual information (TI), as described in ITU-T P.910 recommendation [3], were computed for each sequence and are shown Figure 4. The spatial and temporal information as well as number of frames and native frame rate of each video sequence are available on the website of the dataset.

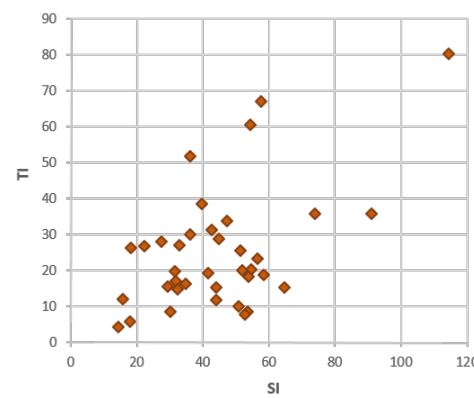


Figure 4: SI and TI of video sequences.

4.2 Eye tracking experiments

4.2.1 *Experimental setup*

The experiment was conducted in a test environment set as a standard subjective quality test condition according to ITU-R BT.500 [1]. The HD display used was a 46" Panasonic Full HD Vieta and the 4K display used was a 65" Panasonic TX-L65WT600E. The viewing distance was 1.5H, i.e. 120 cm, in UHD and 3H, i.e. 170 cm, as recommended in ITU-R BT.1769 [2].

We used the EHT eye tracker presented Section 2.

4.2.2 *Observers*

70 remunerated viewers participated in this subjective in two independent sessions for HD and UHD conditions. In HD, there were 17 males and 17 females, aged between 19 to 44 with an average age of 24.4 ($SD = 5.08$). In UHD, there were 18 males and 18 females, aged between 19 to 56 with an average age of 27.7 ($SD = 11.24$).

Correct visual acuity and color visions were assured prior to this experiment. The visual acuity tests were conducted with Monoyer chart for far vision and with Parinaud chart

(French equivalent of Jaeger chart) for near vision. All the viewers had either normal or corrected-to-normal visual acuity. The Ishihara plates were used for color vision test. All of the 70 viewers passed the pre-experiment vision check.

4.2.3 Procedure

UHD and HD were assessed in two different sessions with different observers to avoid any effect of memorization.
We adopted a free-looking approach in these experiments. Sequences were randomized for each observer. They were 2 seconds spaced out. The whole test lasted approximately 25 minutes.

4.3 Gaze data

For each video and each observer, the following gaze data are stored: eye identifier (0 for left eye and 1 for right eye) ; time (sec) ; eye position in X axis (px) ; eye position in Y axis (px). The origin (0,0) is in the upper left corner of the frame. If the eye was not tracked by the eye tracker, the X and Y positions are set as NaN.
The mean of successive left and right eye positions might be calculated to obtain binocular information.

4.4 Fixation points and saccades

A fixation is defined as the status of a region centered around a pixel position which was stared at for a predefined duration. A saccade corresponds to the eye movement from one fixation to another.
Most often, saliency maps are computed from fixation points rather than gaze points. Thus, we extracted fixation points and saccades from the gaze data following the method explained in [12].
More precisely, fixations are detected according four parameters:

- the fixation velocity maximum threshold, set as 30°/s;
- the maximum time between separate fixations, set as 75 ms ;
- the maximum visual angle between separate fixations, set as 0.5°;
- the minimum fixation duration, set as 100 ms.

For each source, we provide the following data about fixations: starting time of fixation (ms) ; end of fixation (ms) ; fixation position in X axis (px) ; fixation position in Y axis (px) ; number of gaze points in the fixation ; observer number.
We also provide saccade data between fixations as follows: starting time of saccade (ms) ; end of saccade (ms) ; position of start of saccade in X axis (px) ; position of start of saccade in Y axis (px) ; position of end of saccade in X axis (px) ; position of end of saccade in Y axis (px) ; saccade length (px) ; saccade orientation (°) ; observer number.

5. DATA USAGE AND FUTURE WORKS

The main goal of this dataset is the comparison of visual attention and viewing behavior in HD and UHD. Different kind of analyses can be done: impact of viewing conditions and resolution on distribution of gaze points and fixations (Figures 9 and 10), comparison of saliency through fixation density maps (Figures 7 and 8), comparison of distribution of saccades (Figures 5 and 6), etc. Different indicators and metrics can be computed from these data as proposed in [5], in order to compare results in HD and UHD. Moreover, this dataset can be used to evaluate the performance of visual saliency models in HD and UHD, by comparing fixation density maps computed from acquired data with simulated saliency maps.
Futhermore, this dataset provides useful data for any researcher working on dynamic visual attention in videos (dynamic visual attention modelling, visual attention and quality of experience, saliency-based video compression, etc.). The main qualities of the dataset are the large number of sources and observers compared to previously published video saliency database, as well as the high quality of professional videos.

6. CONCLUSION

In this paper, we presented a new HD and UHD video eye tracking dataset on 37 high quality video sequences, respectively seen by 34 and 36 observers in HD and UHD. For each video sequence, gaze point, fixation and saccade data are provided.
The main objective of this dataset is the comparison of visual attention and viewing behavior in HD and UHD. Indeed, the emergence of UHD video format induces larger screens and involves a wider stimulated visual angle. Therefore, its effect on visual attention can be questioned since it can impact quality assessment, metrics but also the whole chain of video processing and creation.
Thanks to the variety of video sequences and the large number of observers, these data can be really useful for any study on visual attention in videos.

7. ACKNOWLEDGMENTS

This work is part of the UltraHD-4U project financed by the DGCIS through the European CATRENE program CAT111.

8. REFERENCES

[1] ITU-R BT.500-11 Methodology for the subjective assessment of the quality of television pictures. 2002.
[2] ITU-R BT.1769. Parameter values for an expanded hierarchy of LSDI image formats for production and international programme exchange. 2008.
[3] ITU-T Rec. P.910. Subjective video quality assessment methods for multimedia applications. 2008.
[4] ITU-R BT.2020. Parameter values for ultra-high definition television systems for production and international programme exchange. 2012.
[5] O. Le Meur and T. Baccino. Methods for comparing scanpaths and saliency maps: strengths and weaknesses. *Behavior Research Methods*, 45(1):251–266, 2013.
[6] O. Le Meur and Z. Liu. Saccadic model of eye movements for free-viewing condition. *Vision research*, 116:152–164, 2015.
[7] D. Li, G. Zhai, and X. Yang. Ultra high definition video saliency database. In *2014 IEEE Visual Communications and Image Processing Conference*. IEEE, 2014.
[8] J. Li, Y. Koudota, M. Barkowsky, H. Primon, and P. Le Callet. Comparing upscaling algorithms from HD to Ultra HD by evaluating preference of experience. In *2014 Sixth International Workshop on*

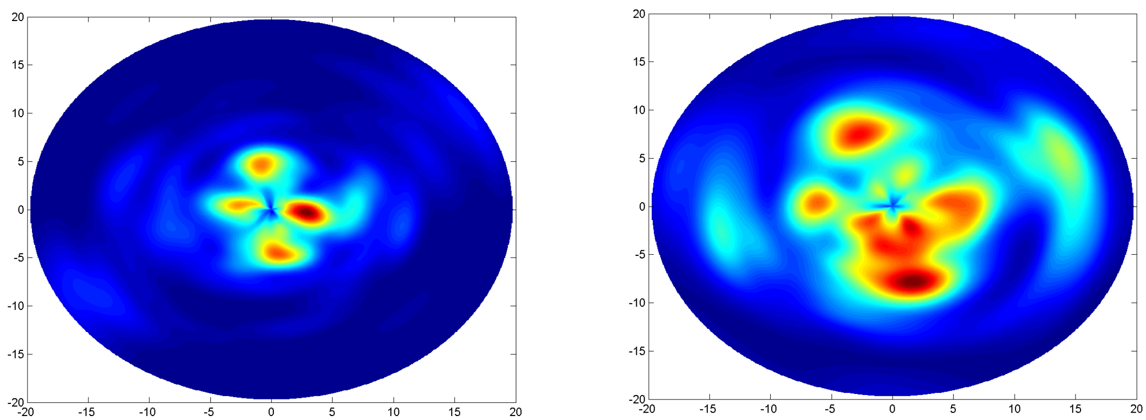
Quality of Multimedia Experience (QoMEX). IEEE, 2014.

[9] H. Nemoto, P. Hanhart, P. Korshunov, and T. Ebrahimi. Impact of Ultra High Definition on Visual Attention. In *Proceedings of the ACM International Conference on Multimedia - MM '14*. ACM Press, 2014.

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

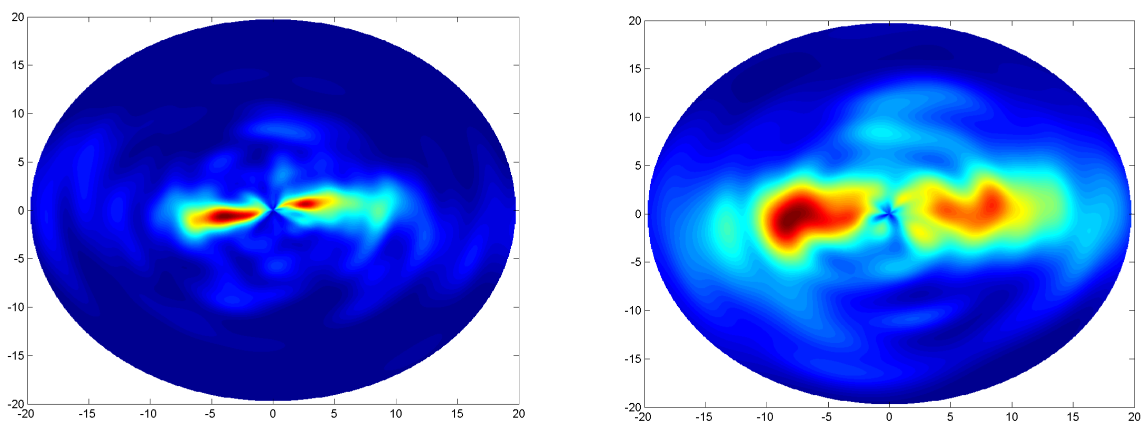
[11] L. Song, X. Tang, W. Zhang, X. Yang, and P. Xia. The SJTU 4K video sequence dataset. In *2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX)*. IEEE, 2013.

[12] Tobii Technology. User Manual - Tobii Studio. 2014.



(a) HD (b) UHD

Figure 5: Polar distribution of saccades between 0 et 20° length in the whole video sequence Beauty. (Distributions are calculated following the method presented in [6].)



(a) HD (b) UHD

Figure 6: Polar distribution of saccades between 0 et 20° length in the whole video sequence Bosphorus.

(a) HD

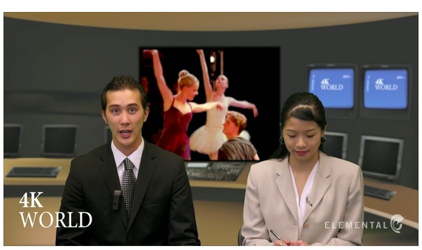
(b) UHD

Figure 9: Gaze points (red) and fixations (blue) for all observers (Big Bug Bunny, sequence 1, frame 40).

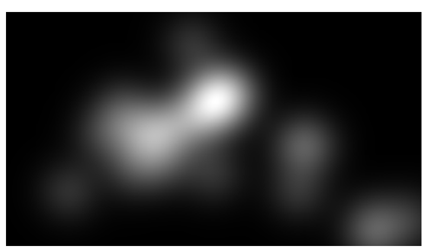
(a) HD

(b) UHD

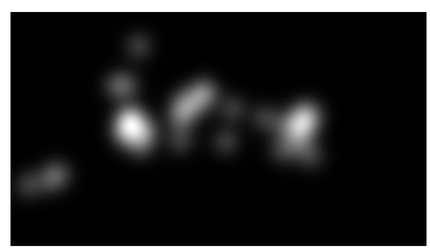
Figure 10: Gaze points (red) and fixations (blue) for all observers (Big Bug Bunny, sequence 2, frame 100).



(a) Original frame.



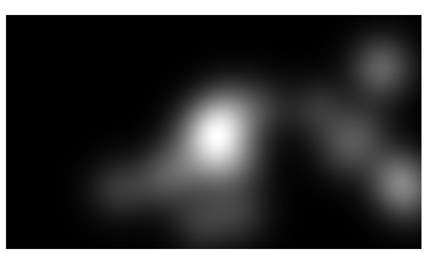
(b) Fixation density map in HD.



(c) Fixation density map in UHD.

Figure 7: Example of fixation density maps in HD and UHD. Video sequence News_ProRes, frame 50.

(a) Original frame.



(b) Fixation density map in HD.

(c) Fixation density map in UHD.

Figure 8: Example of fixation density maps in HD and UHD. Video sequence Traffic_and_Buildings, frame 150.