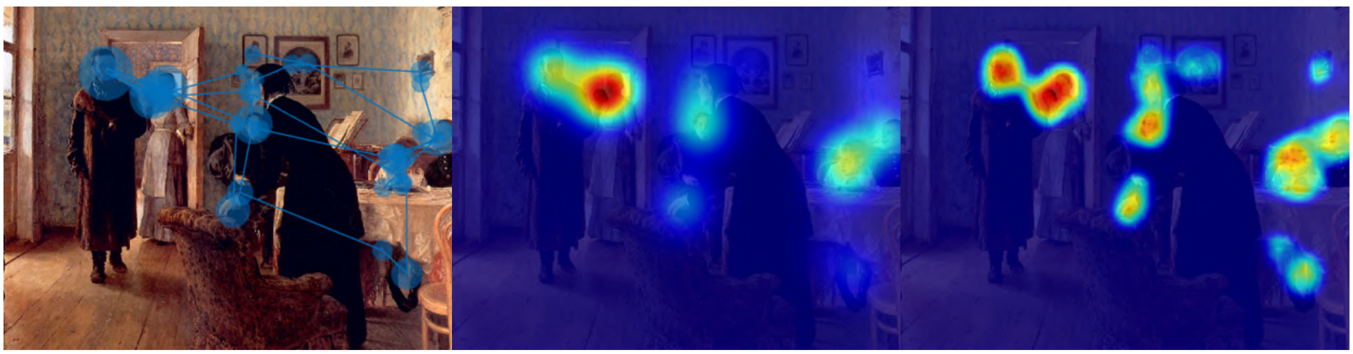
Exploiting the GBVS for Saliency aware Gaze Heatmaps

David Geisler
University of Tuebingen
Germany
david.geisler@uni-tuebingen.de

Daniel Weber
University of Tuebingen
Germany
daniel.weber@uni-tuebingen.de

Nora Castner
University of Tuebingen
Germany
nora.castner@uni-tuebingen.de

Enkelejda Kasneci
University of Tuebingen
Germany
enkelejda.kasneci@uni-tuebingen.de



(a) Fixation sequence on the painting *An Unexpected Visitor* from Ilya Repin. (b) Regular gaussian like fixation heatmap. (c) GBVS attention map with incorporated gaze signal.

Figure 1: (a) shows the sequential fixation signal, where the size of the circles encodes the fixation time. (b) shows the corresponding gaussian like fixation heatmap. (c) shows the output of the proposed approach, where the tracked fixations are incorporated into the GBVS attention map calculation.

ABSTRACT

Analyzing visual perception in scene images is dominated by two different approaches: 1.) Eye Tracking, which allows us to measure the visual focus directly by mapping a detected fixation to a scene image, and 2.) Saliency maps, which predict the perceivability of a scene region by assessing the emitted visual stimulus with respect to the retinal feature extraction. One of the best-known algorithms for calculating saliency maps is GBVS. In this work, we propose a novel visualization method by generating a joint fixation-saliency heatmap. By incorporating a tracked gaze signal into the GBVS, the proposed method equilibrates the fixation frequency and duration to the scene stimulus, and thus visualizes the rate of the extracted visual stimulus by the spectator.

CCS CONCEPTS

• **Human-centered computing → Heat maps**; *Visualization toolkits*; *Scientific visualization*; *Information visualization*; User models.

ETRA '20, June 2–5, 2020, Stuttgart, Germany

ACM ISBN 978-1-4503-XXXX-X/18/06. . . $15.00
https://doi.org/00.0000/0000000.0000000

KEYWORDS

Eye-Tracking; Visual Stimulus; Visual Perception; Scene Evaluation

ACM Reference Format:
David Geisler, Daniel Weber, Nora Castner, and Enkelejda Kasneci. 2020. Exploiting the GBVS for Saliency aware Gaze Heatmaps. In *ETRA '20: ACM Symposium on Eye Tracking Research & Applications, June 2–5, 2020, Stuttgart, Germany.* ACM, New York, NY, USA, 5 pages. https://doi.org/00.0000/0000000.0000000

1 INTRODUCTION

Our eyes move around to perceive and understand the scene in order to compensate for our limited- but clearest- foveal vision. When viewing a scene, we frequently focus our attention, known as a fixation, before shifting to another area with a rapid eye movement known as a saccade. The selectivity of the focused scene locations is a highly optimized and developed process, mainly driven by two factors: 1.) visual scene features, extracted by the retina (bottom-up), and 2.) the interpretation of the extracted features regarding their semantic value by higher cognitive processes, and subsequent identification of the next fixation target (top-down) [Itti and Koch 2000; Itti et al. 1998; van Renswoude et al. 2019]. Modeling and understanding this reciprocal process is a long-term core topic in cognitive psychology and the computer vision community [Kotseruba and Tsotsos 2018].

While retinal feature extraction can be modeled using saliency maps such as GBVS [Harel et al. 2007], the selectivity of visual attention can be measured using eye tracking. Saliency maps are predominantly bio-physiologically inspired algorithms to predict potential fixation targets [Zhao and Koch 2013]. Therefore, regions emitting a strong, recognizable visual stimulus are identified and emphasized by replicating the retinal visual stimulus processing. Eye tracking, on the other hand, often tracks the pupil center and extrapolates the line of sight to a scene image. A subsequent typical visualization is to illustrate the extracted fixations as a heatmap overlay on the scene image. This is used to investigate the visual attention on the scene, and to identify areas of particular interest.

However, fixation heatmaps are subject to some limitations. For instance, slight shifts in the eye-tracking signal make it difficult to identify the scene parts that attracted the visual attention and which information of the scene was actually perceptible. In addition, depending on the implementation, long or frequent fixations on the same scene region may lead to a high density in the fixation heatmap. Hence, it is assumed that these regions are particularly relevant to the spectator since a comparatively high amount of visual information was extracted. However, frequent or long fixations may also be caused by difficult scene conditions such as low contrasts. Thus, the visual information may be harder to extract, and therefore requires longer or more frequent fixations to be perceived.

In this work, we propose to incorporate detected fixations from an eye tracking signal into the calculation of the GBVS attention map. The resulting heatmap equilibrates the measured visual attention to the retinal-perceivable stimulus, and thus visualizes the density of perceived information in the scene more accurately as pure fixation or saliency heatmaps.

Structure of the Paper: Section 2 gives a short introduction into state-of-the-art eye tracking visualization and saliency methods. Section 3 contains a comprehensive description of how the proposed approach takes place in the GBVS algorithm. Section 4 shows the exemplary application of the proposed visualization to different types of stimuli. The final sections 5 state the limitations of the presented approach and the final remarks.

2 RELATED WORK

The eye tracking community is a research powerhouse. Continuous improvements in tracking accuracy, precision, and availability over the last decades made eye tracking to one of the most eminent sensors in numerous research fields: Psychology, HCI, medicine, neuroscience, marketing, and many more. In particular, the success of recent years in the field of vision-based eye tracking has boosted the technology in terms of affordability, convenience, and usability for a broad community [Hosp et al. 2019; Santini et al. 2017b,a, 2019]. However, the ability to conduct comprehensive eye-tracking studies led to increasing demand for sophisticated methods for visualization and qualitative evaluation of the acquired data [Blascheck et al. 2015].

An initial exploratory step in eye tracking studies is often to examine the spatial location, duration, and frequency of fixations as a heatmap over the stimulus [Bylinskii et al. 2015]. This can be efficiently calculated over a large amount of data and gives a first impression of the distribution of visual attention on the stimulus [Dao et al. 2014; Duchowski et al. 2012]. But, in order to gain deeper insights into the data, an extensive repertoire of different visualization techniques is available, such as various saccade metrics [Burch et al. 2014; Kübler et al. 2016; Raschke et al. 2014], AOI hierarchies [Blascheck et al. 2016; Kurzhals et al. 2016b], or extensive interactive visualizations including stimulus and time domains [Kurzhals et al. 2016a, 2015; Kübler et al. 2015; Raschke et al. 2016], etc. A comprehensive overview can be found in the survey of [Blascheck et al. 2014] and [Blascheck et al. 2017].

Saliency maps assess the stimulus by modeling the retinal signal processing to determine whether a scene area is particularly prominent in its immediate neighborhood, and therefore more likely to be perceived. The stimulus is evaluated by its intensity and its opponent color spaces: Driven by the neuronal circuit of the photoreceptors. Additionally, further feature spaces can be formed, such as edge orientation and difference formation of sequential images [Bian and Zhang 2008; Geisler et al. 2017; Harel et al. 2007; Hou and Zhang 2007; Itti et al. 1998; Zhang and Sclaroff 2013]. While these bottom-up approaches mainly reproduce the feature extraction of the retina perception, newer deep-learning-based approaches show great success in modeling the whole processes, from the retinal feature extraction up to the semantic interpretation of the visual cortex and higher cognitive levels. They include the recognition and evaluation of abstract forms regarding their object-relatedness and semantic relevance [Chen et al. 2017; Dubey et al. 2015; John et al. 2014; Kümmerer et al. 2014; Lee et al. 2016; Li and Yu 2015; Liu and Han 2016; Zhao et al. 2015; Zou and Komodakis 2015].

The proposed approach combines the worlds of fixation heatmaps and salience maps, as a novel visualization technique. The resulting heatmap provides insights into the extracted information rate of the scene and extends the existing visualization techniques towards a more stimulus-driven paradigm.

3 METHOD

Similar to most saliency map approaches, the GBVS algorithm is divided into 3 consecutive steps [Harel et al. 2007]:

(1) Extraction of a feature map M_t on a given image I_t.
(2) Calculation of an activation map A_t based on M_t.
(3) Normalization and combination of the activation map A_t.

Our approach amends step (2) by injecting the gaze signal g_t into the calculation of the activation map A_t. Steps 1 and 3 remain unchanged to the GBVS publication [Harel et al. 2007] and not further discussed here. The subscript t indicates the time domain since the gaze signal is given as a time series of consecutive fixation points. However, it also simplifies the handling with dynamic stimuli, such as videos. In the following, we assume that for each t exists a corresponding gaze signal g_t, as well as a stimulus I_t, respectively a feature map M_t.

The GBVS interprets the activation map as a state vector of a Markov model. The transition between two states is defined by a dissimilarity score over the feature map M_t. Thus, a random walk over the Markov model empowers those states that are dissimilar in the respective feature map. Analogous to the original GBVS, the dissimilarity between the two states i and j in the feature map M_t is defined as follow:

$$d_t(i,j) = \left|\log \frac{M_t(i)}{M_t(j)}\right|, \quad (1)$$

where $M_t(i)$ is the i-th value of the corresponding feature map M_t. The transition weight $w_t(i,j)$ between the two states i and j is defined as the product of their dissimilarity score $d_t(i,j)$ and a distance weight $F_w(i,j)$:

$$w_t(i,j) = d_t(i,j) \cdot F(i,j). \quad (2)$$

The distance weight adds a local sensitivity to the dissimilarity score. Thus, states that are dissimilar to their immediate neighborhood are

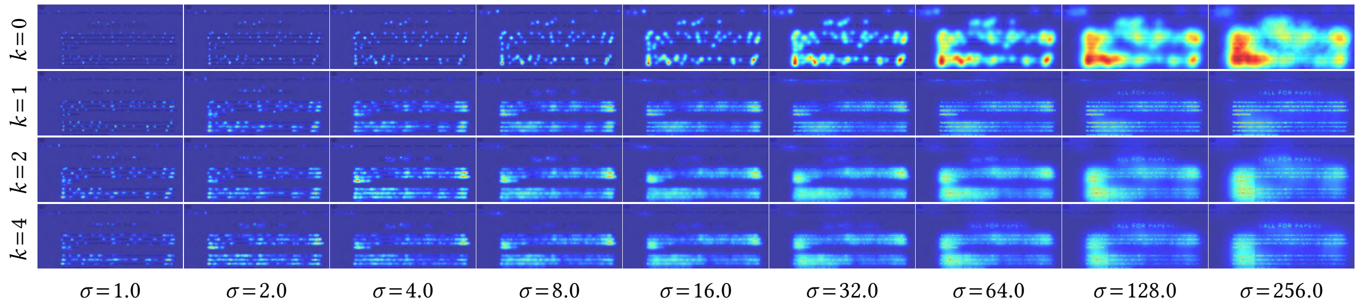


Figure 2: Influence of the parameters σ (horizontal) and k (vertical) on the adaptation of the fixations heatmap to the text stimulus. The input is the same as used in figure 3. For large k, the injected visual attention is increasingly distributed across the entire stimulus. The parameter σ should be chosen depending on the desired level of detail of the visualization. Using a text stimulus, it is usually reasonable to choose a high degree of detail (here $\sigma \leq 8$), in order to visualize the perception rate of single words or lines.

emphasized while the impact of the dissimilarity score is attenuated with increasing distance. $F(i,j)$ is defined as an exponentially weighted square distance between the states i and j in their spatial dimension in the input image I_t:

$$F(i,j) = \exp\left(-\frac{(x(i)-x(j))^2+(y(i)-y(j))^2}{2\cdot\sigma}\right), \tag{3}$$

where $x(i)$ and $y(i)$ is the x- and y-coordinate of the i-th state in the respective input image I_t. The free parameter σ controls the shape of the exponential distance weight. The larger σ is chosen, the more weight is given to the dissimilarities of more remote states.

The final Markov transition matrix T_t is then assembled as follows:

$$T_t = \begin{pmatrix} 1 & w_t(0,1) & \dots & w_t(0,n) \\ w_t(1,0) & 1 & \ddots & w_t(1,n) \\ \vdots & \ddots & \ddots & \vdots \\ w_t(n,0) & w_t(n,1) & \dots & 1 \end{pmatrix}, \tag{4}$$

where n is the number of elements in the feature map M_t respective the input image I_t.

The activation map A_t is then calculated by k repeated multiplication with the transition matrix T_t:

$$A_t^{(k)} = T_t \cdot A_t^{(k-1)}. \tag{5}$$

Incorporate Gaze: Up to this step, the procedure follows the original GBVS algorithm. However, instead of initializing $A_t^{(0)}$ equally distributed, the gaze position is encoded as initial activation map:

$$A_t^{(0)} = q \cdot A_{t-1}^{(k)} + (1-q)\cdot(F(0,g_t),\dots,F(n,g_t)), \tag{6}$$

where $F(i,g_t)$ is the exponential weighted square distance between the recorded gaze position g_t and the spatial location of the i-th element in the activation map. In other words, the activation map is initialized by the measured visual activation from the eye tracking signal. Additionally, parameter $q \in [0,1]$ controls the influence of the previously calculated activation map A_{t-1} into the initialization of $A_t^{(0)}$. Thus, for $q > 0$, $A_t^{(0)}$ encodes the recently measured visual attention, but also the history of previous predicted attention areas. This smooths the resulting activation map $A_t^{(k)}$ in the temporal domain, and makes noise in the gaze signal less significant. However, it also poses the risk to generate a distorted activation map. For instance, on a dynamic stimulus: the previous predicted attentive area in frame I_{t-1} is located somewhere in frame I_t. Yet, $A_t^{(0)}$ provides values at this area and the Markov model will adapt it to the next salient region – which may not have ben actually focused on. Nevertheless, this effect only occurs if the content of the scene changes significantly, for instance on scene cuts in movies, or opening a new web page while browsing.

When generating static heat maps (such as Figure 1), it is common to ignore the temporal domain completely. In this case, q is set to zero. The overall heatmap $\mathbf{A}^{(k)}$ is then the weighted sum over $A_t^{(k)}$:

$$\mathbf{A}^{(k)} = \sum_t A_t^{(k)} \cdot b_t, \tag{7}$$

where the weighting b_t, for instance, can be chosen in relation to the fixation time.

Parameters: On regular gaussian like gaze heatmaps, σ models the area of visual attention (foveal perception) and/or the expected noise of the eye tracking signal, and thus controls the acuity of the resulting heatmap. In the proposed approach, σ controls the distribution of visual attention deduced from the fixation signal. But also how far the Markov model may adopt this distribution to the underlying stimulus in each iteration. The number of iterations is controlled by the parameter k. Whereby for $k=0$, $A_t^{(0)} = A_t^{(k)}$ corresponds to a regular gaussian like fixation heatmap of a single fixation point (respectively $\mathbf{A}^{(0)}$ overall fixation points). Figure 2 shows how the initial gaze heatmap $A_t^{(0)}$ is gradually distorted to the stimulus for each additional iteration over equation 5.

Implementation Details: The main limitation of GBVS is runtime and memory consumption. The transition matrix T_t grows in quadratic size with the input size n, and thus quickly exceeds the available memory (e.g. $> 9.4 \cdot 10^{10}$ elements on a VGA resolution). Additionally, the initialization of T_t requires a runtime complexity of $O(n^2)$. Both together, limit the GBVS to very low input resolutions, which leads to a loss of acuity. Thus, the standard parametrization of the GBVS toolbox limits the internal resolution to an edge length of 32px [Harel et al. 2006].

On closer examination, however, it is apparent that most values in T_t are extremely small and have no significant impact to the resulting activation map $A_t^{(k)}$. Thus, after applying a threshold l, the transition matrix T_t becomes predominantly sparse. Furthermore, assuming that $M_t \in [0,1]$, the elements of T_t, which potentially exceed the threshold l can be determined in relation to σ:

$$l < F(i,j), \tag{8}$$

and resolves to:

$$\sqrt{-2\cdot\sigma\cdot\log(l)} \geq \sqrt{(x(i)-x(j))^2+(y(i)-y(j))^2}, \tag{9}$$

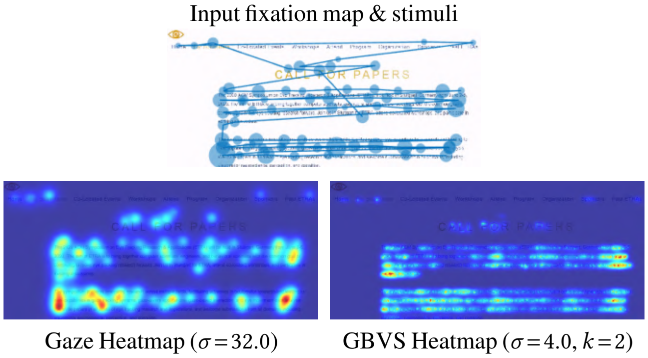


Figure 3: The top image shows the recorded fixation sequence on the ETRA 2020s *Call for Papers* website. The bottom left image shows a regular gaussian fixation heatmap ($\mathrm{A}^{(0)}$). The bottom right image shows the output of the proposed method ($\mathrm{A}^{(k=2)}$).

where the right term is the euclidean distance between the the i-th and j-th element in the feature map M_t. Thus, initializing T_t only requires the calculation of $2 \cdot \sqrt{-2 \cdot \sigma \cdot \log(l)}$ elements per row, since all other elements are not exceeding the threshold l. This reduces the actual runtime from $O(n^2)$ to $O(n)$. Similar considerations can be made for the initialization of $A_t^{(0)}$ (although this is not a bottleneck). However, due to the sparseness of T_t and $A_t^{(0)}$, solving equation (5) is much faster [Yuster and Zwick 2005].

4 EXPERIMENTAL DEMONSTRATION

Figures 1, 3, and 4 demonstrate the application of the proposed visualization on different stimulus types: the *An Unexpected Visitor* painting from Ilya Repin, the *Call for Papers* website from ETRA 2020 as text, and a short video snippet of *Big Buck Bunny* from the Peach open movie project [Roosendaal 2008]. The gaze signal was recorded by a Tobii Pro Spectrum at 1200Hz. The fixation locations and duration were extracted using the fixation filter I-VT provided by Tobii Pro Lab and default parametrization [Olsen 2012]. All stimuli were presented as full screen on the Monitor at 1920×1080 pixels.

On the text stimulus, it is recognizable how the Markov model depicts the measured visual attention to paragraphs, lines, down to single words and characters. Therefore, the acuity of the heatmap is increased, and consequently, interpretations about the perception rate to text passages are simplified. For instance, in the field of web design and advertising, the proposed model can help to analyze whether a certain area attracts the desired level of visual attention and whether the presented information was easily visual accessible to the spectator.

However, in this context, text reading is a relatively unambiguous challenge, since the text is very salient to its background. At the same time, the text is often the only element that attracts the visual attention of the reader. The strength of the proposed method of visual attention visualization is particularly evident in more complex stimuli as shown in figures 1 and 4. Considering the *An Unexpected Visitor* painting, the fixations are mainly on the faces in the scene, but also on some miscellaneous areas, such as hands, the paintings in the background, or feet. However, the regular fixation heatmap has a particularly pronounced fixation cluster on the face of the woman in the background. This can be attributed to the fact that this face is particularly difficult to perceive due to its low contrast. Yet, the long and frequent fixations in this area lead to a suppression of all other fixations, which can lead

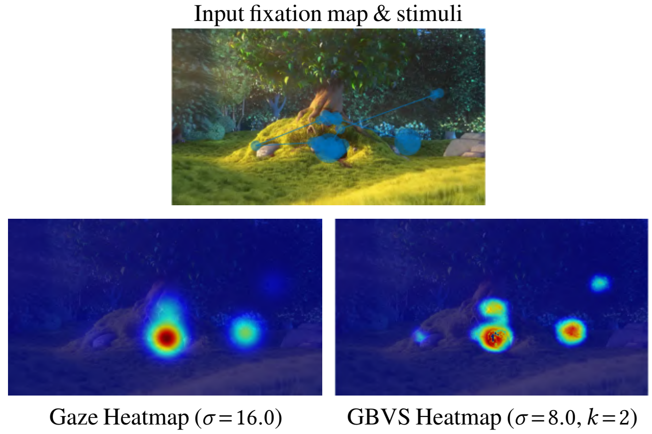


Figure 4: The top image shows the recorded fixation sequence on a short snippet of the video clip *Big Buck Bunny* from the Peach open movie project [Roosendaal 2008]. The bottom left image shows a regular gaussian fixation heatmap ($\mathrm{A}^{(0)}$). The bottom right image shows the output of the proposed method ($\mathrm{A}^{(k=2)}$).

to the interpretation that this area was of higher interest for the spectator. The GBVS generated fixation heatmap incorporates not only the fixation duration and frequency but also how accessible the stimulus in the region is to the observer. The result is a much more balanced fixation heatmap, where all the fixated heads are clearly pronounced.

5 FINAL REMARKS

The proposed method extends the well-known GBVS saliency algorithm by incorporating the measured visual attention. The resulting heatmap visualizes a predicted perception rate of scene areas for an individual or multiple spectators. However, as the most bottom-up saliency algorithm, GBVS uses exclusively intrinsic scene features to predict whether certain scene content is attractive for fixation. It turns out, this is very accurate for a free viewing scenario. Yet, various tasks may require the spectator to direct their visual attention to less saliency scene areas. The proposed algorithm might distort these fixation points to a close salient region and thus weigh the perception rate based on the wrong stimuli. This limitation can be compensated by using a small σ and high-resolution scene images but requires a high accuracy of the fixation point.

In practice, however, it has been shown that the proposed visualization generates more intuitive heatmaps than pure fixation heatmaps. Thus, the presented visualization provides an ingenious overview of the scene areas with a distinctive high rate of visual awareness.

REFERENCES

Peng Bian and Liming Zhang. 2008. Biological plausibility of spectral domain approach for spatiotemporal visual saliency. In *International conference on neural information processing*. Springer, 251–258.

Tanja Blascheck, Michael Burch, Michael Raschke, and Daniel Weiskopf. 2015. Challenges and perspectives in big eye-movement data visual analytics. In *2015 Big Data Visual Analytics (BDVA)*. IEEE, 1–8.

Tanja Blascheck, Kuno Kurzhals, Michael Raschke, Michael Burch, Daniel Weiskopf, and Thomas Ertl. 2014. State-of-the-Art of Visualization for Eye Tracking Data. In *EuroVis STAR*. http://dx.doi.org/10.2312/eurovisstar.20141173

Tanja Blascheck, Kuno Kurzhals, Michael Raschke, Michael Burch, Daniel Weiskopf, and Thomas Ertl. 2017. Visualization of eye tracking data: A taxonomy and survey. In *Computer Graphics Forum*, Vol. 36. Wiley Online Library, 260–284.

Tanja Blascheck, Kuno Kurzhals, Michael Raschke, Stefan Strohmaier, Daniel Weiskopf, and Thomas Ertl. 2016. AOI Hierarchies for Visual Exploration of Fixation Sequences. In *Proceedings of the Symposium on Eye Tracking Research and Applications, ETRA '16*.

Michael Burch, Hansjörg Schmauder, Michael Raschke, and Daniel Weiskopf. 2014. Saccade Plots. In *Proceedings of Symposium on Eye Tracking Research and Applications*, ACM (Ed.). http://dx.doi.org/10.1145/2578153.2578205

Zoya Bylinskii, Michelle A Borkin, Nam Wook Kim, Hanspeter Pfister, and Aude Oliva. 2015. Eye fixation metrics for large scale evaluation and comparison of information visualizations. In *Workshop on Eye Tracking and Visualization*. Springer, 235–255.

Xiaowu Chen, Anlin Zheng, Jia Li, and Feng Lu. 2017. Look, perceive and segment: Finding the salient objects in images via two-stream fixation-semantic cnns. In *Proceedings of the IEEE International Conference on Computer Vision*. 1050–1058.

Thanh-Chung Dao, Roman Bednarik, and Hana Vrzakova. 2014. Heatmap rendering from large-scale distributed datasets using cloud computing. In *Proceedings of the Symposium on Eye Tracking Research and Applications*. 215–218.

Rachit Dubey, Joshua Peterson, Aditya Khosla, Ming-Hsuan Yang, and Bernard Ghanem. 2015. What makes an object memorable?. In *Proceedings of the ieee international conference on computer vision*. 1089–1097.

Andrew T Duchowski, Margaux M Price, Miriah Meyer, and Pilar Orero. 2012. Aggregate gaze visualization with real-time heatmaps. In *Proceedings of the Symposium on Eye Tracking Research and Applications*. 13–20.

David Geisler, Wolfgang Fuhl, Thiago Santini, and Enkelejda Kasneci. 2017. Saliency Sandbox-Bottom-up Saliency Framework.. In *VISIGRAPP (4: VISAPP)*. 657–664.

Jonathan Harel, C Koch, and P Perona. 2006. A saliency implementation in matlab. *URL: http://www. klab. caltech. edu/harel/share/gbvs. php* (2006).

Jonathan Harel, Christof Koch, and Pietro Perona. 2007. Graph-based visual saliency. In *Advances in neural information processing systems*. 545–552.

B. Hosp, S. Evazi, M. Maurer, W. Fuhl, and E. Kasneci. 2019. RemoteEye: An Open Source remote Eye Tracker. *Behavior Research Methods, BRM* (dec 2019).

Xiaodi Hou and Liqing Zhang. 2007. Saliency detection: A spectral residual approach. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. IEEE, 1–8.

Laurent Itti and Christof Koch. 2000. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision research* 40, 10-12 (2000), 1489–1506.

Laurent Itti, Christof Koch, and Ernst Niebur. 1998. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence* 20, 11 (1998), 1254–1259.

Vijay John, Keisuke Yoneda, B Qi, Zheng Liu, and Seiichi Mita. 2014. Traffic light recognition in varying illumination using deep learning and saliency map. In *Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on*. IEEE, 2286–2291.

Iuliia Kotseruba and John K Tsotsos. 2018. 40 years of cognitive architectures: core cognitive abilities and practical applications. *Artificial Intelligence Review* (2018), 1–78.

Thomas Kübler, Wolfgang Fuhl, Raphael Rosenberg, Wolfgang Rosenstiel, and Enkelejda Kasneci. 2016. Novel methods for analysis and visualization of saccade trajectories. In *European Conference on Computer Vision*. Springer, 783–797.

Matthias Kümmerer, Lucas Theis, and Matthias Bethge. 2014. Deep gaze i: Boosting saliency prediction with feature maps trained on imagenet. *arXiv preprint arXiv:1411.1045* (2014).

Kuno Kurzhals, Marcel Hlawatsch, Michael Burch, and Daniel Weiskopf. 2016a. Fixation-image charts. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*. 11–18.

Kuno Kurzhals, Marcel Hlawatsch, Florian Heimerl, Michael Burch, Thomas Ertl, and Daniel Weiskopf. 2015. Gaze stripes: Image-based visualization of eye tracking data. *IEEE transactions on visualization and computer graphics* 22, 1 (2015), 1005–1014.

Kuno Kurzhals, Marcel Hlawatsch, Christof Seeger, and Daniel Weiskopf. 2016b. Visual analytics for mobile eye tracking. *IEEE transactions on visualization and computer graphics* 23, 1 (2016), 301–310.

T. C. Kübler, K. Sippel, W. Fuhl, G. Schievelbein, J. Aufreiter, R. Rosenberg, W. Rosenstiel, and E. Kasneci. 2015. *Analysis of eye movements with Eyetrace*. Vol. 574. Biomedical Engineering Systems and Technologies. Communications in Computer and Information Science (CCIS). Springer International Publishing. 458–471 pages.

Gayoung Lee, Yu-Wing Tai, and Junmo Kim. 2016. Deep saliency with encoded low level distance map and high level features. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 660–668.

Guanbin Li and Yizhou Yu. 2015. Visual saliency based on multiscale deep features. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 5455–5463.

Nian Liu and Junwei Han. 2016. Dhsnet: Deep hierarchical saliency network for salient object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 678–686.

Anneli Olsen. 2012. The Tobii I-VT fixation filter. *Tobii Technology* (2012), 1–21.

Michael Raschke, Dominik Herr, Tanja Blascheck, Michael Burch, Michael Schrauf, Sven Willmann, and Thomas Ertl. 2014. A Visual Approach for Scan Path Comparison. In *Proceedings of the Symposium on Eye Tracking Research and Applications, ETRA '14*. ACM. http://dx.doi.org/10.1145/2578153.2578173

Michael Raschke, Bernhard Schmitz, Michael Wörner, Thomas Ertl, and Frederik Diederichs. 2016. Application Design for an Eye Tracking Analysis based on Visual Analytics. In *Proceedings of the Symposium on Eye Tracking Research and Applications, ETRA '16*. Rezensiertes Poster.

Ton Roosendaal. 2008. Big Buck Bunny. In *ACM SIGGRAPH ASIA 2008 Computer Animation Festival (SIGGRAPH Asia '08)*. Association for Computing Machinery, New York, NY, USA, 62. https://doi.org/10.1145/1504271.1504321

T. Santini, W. Fuhl, D. Geisler, and E. Kasneci. 2017b. EyeRecToo: Open-Source Software for Real-Time Pervasive Head-Mounted Eye-Tracking. In *12th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2017)*.

T. Santini, W. Fuhl, and E. Kasneci. 2017a. CalibMe: Fast and Unsupervised Eye Tracker Calibration for Gaze-Based Pervasive Human-Computer Interaction. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*.

T. Santini, D. Niehorster, and E. Kasneci. 2019. Get a Grip: Slippage-Robust and Glint-Free Gaze Estimation for Real-Time Pervasive Head-Mounted Eye Tracking. In *Proceedings of the 2019 ACM Symposium on Eye Tracking Research & Applications (ETRA)*.

Daan R van Renswoude, Ingmar Visser, Maartje EJ Raijmakers, Tawny Tsang, and Scott P Johnson. 2019. Real-world scene perception in infants: What factors guide attention allocation? *Infancy* 24, 5 (2019), 693–717.

Raphael Yuster and Uri Zwick. 2005. Fast sparse matrix multiplication. *ACM Transactions On Algorithms (TALG)* 1, 1 (2005), 2–13.

Jianming Zhang and Stan Sclaroff. 2013. Saliency detection: A boolean map approach. In *Proceedings of the IEEE international conference on computer vision*. 153–160.

Qi Zhao and Christof Koch. 2013. Learning saliency-based visual attention: A review. *Signal Processing* 93, 6 (2013), 1401–1407.

Rui Zhao, Wanli Ouyang, Hongsheng Li, and Xiaogang Wang. 2015. Saliency detection by multi-context deep learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1265–1274.

Wenbin Zou and Nikos Komodakis. 2015. Harf: Hierarchy-associated rich features for salient object detection. In *Proceedings of the IEEE international conference on computer vision*. 406–414.