Your Gaze Betrays Your Age

Olivier Le Meur
University of Rennes 1 - IRISA, France

Antoine Coutrot
University College London, UK

Zhi Liu
Shanghai University, China

Pia Rämä
University of Paris Descartes, France

Adrien Le Roch
University of Rennes 1 - IRISA, France

Andrea Helo
University of Paris Descartes, France

***Abstract*—Visual attention networks are so pervasive in the human brain that eye movements carry a wealth of information that can be exploited for many purposes. In this paper, we present evidence that information derived from observers' gaze can be used to infer their age. This is the first study showing that simple features extracted from the ordered sequence of fixations and saccades allow us to predict the age of an observer. Eye movements of 101 participants split into 4 age groups (adults, 6-10 year-old, 4-6 year-old and 2 year-old) were recorded while exploring static images. The analysis of observers' gaze provides evidence of age-related differences in viewing patterns. Therefore, we extract from the scanpaths several features, including fixation durations and saccade amplitudes, and learn a direct mapping from those features to age using Gentle AdaBoost classifiers. Experimental results show that the proposed image-blind method succeeds in predicting the age of the observer up to 92% of the time. The use of predicted salience does not further improve the classification's accuracy.**

I. Introduction

Eyes are often compared to a window into the soul. Analyzing eye movements do provides a wealth of information about our personality, the cognitive state of our mind, our emotional state, to name a few [1], [2], [3]. Thanks to the advent of modern eye-trackers, capturing gaze with a high spatial and temporal resolution, a large amount of eye tracking data can be collected with a relative simplicity. Gaze is composed of a sequence fixations and saccades, called a visual scanpath. Fixations aim to bring objects of interest onto the fovea, where the visual acuity is maximum. Saccades are ballistic changes in eye position, allowing to jump from one position to another. Visual information extraction essentially takes place during the fixation period. The way we look within a visual scene, the way we jump from one location to another in order to inspect it accurately, the way we avoid looking at unpleasant stimulus can be diagnosis of the task at hand, our personality, our state of mind, or whether we suffer from a neurological disease or not [4], [5].

Amongst all the factors impacting gaze behavior, age is probably one of the most important. At birth, the visual system is limited but develops rapidly during the first years of life and continues to improve through adolescence [6], [7]. As we are getting older, the function of the visual system appears to deteriorate. Helo et al. [8] provided evidence of age-related differences in viewing patterns during natural scene perception. This development as well as aging influence saccade parameters. Saccade frequency, amplitude, and mean velocity are reduced and the velocity/amplitude distribution as well as the velocity profile become less skewed [9], [8]. An horizontal bias is also observed whatever the age, but is more pronounced when getting older [10]. Finally, the fixation durations decrease with age [8].

In this paper, we investigate whether we can infer the age of an observer only from his/her gaze data. Section II presents the eye tracking experiment, the influence of development on eye movements, the features that are extracted from the scanpaths and the classification algorithm. Results are presented in section III. Section IV concludes the paper.

II. Experiment and method

A. Experiment and influence of development

Experiment: We used the eye movement data from a total of 101 subjects participated in the experiments, including 23 adults and 78 children. These subjects were divided into 4 groups: 2 year-old group (18 participants, $M = 2.16$, $SD = 0.22$), 4-6 year-old group (22 participants, $M = 4.2$, $SD = 0.42$), 6-10 year-old group (38 participants, $M = 7.5$, $SD = 0.57$) and adults group (22 participants, $M = 28$, $SD = 4.48$). Participants were instructed to explore 30 color pictures taken from children's books for 10 seconds (see Fig. 1). They were instructed to perform a recognition test to determine whether an image segment presented at the center of the screen was part of the previous stimulus. Image resolution is 1024×768 pixels. The pictures were displayed on a CRT display at 1024×728 pixels viewed from a distance of 60 cm. Each trial consisted in a drift check followed by the presentation of a full-screen image for 10 seconds. Fixations shorter than 90 ms and fixations around blinks were discarded. For all the results reported in this paper, the first fixation has been removed (more details are available in [8]).

Eye movements were sampled monoculary at 500 Hz using the EyeLink 1000 Remote eye tracker system (SR Research, Ontario, Canada) with a spatial resolution below $0.01°$ and a spatial accuracy above $0.5°$. Saccades were identified by deflections in eye position in excess of $0.1°$, with a minimum velocity of $30° s^{-1}$ and a minimum acceleration of $8000° s^{-2}$, maintained for at least 4 ms.

Influence of aging on eye movements: We first investigate the influence of development on fixation durations, saccade amplitudes and central bias. Results are presented in detailed in the supplementary materials [11], which shows that the

ISBN 978-0-9928626-7-1 © EURASIP 2017

Fig. 1. (a) and (d) examples of original stimulus; (b),(e) and (c),(f) represent fixation maps (red crosses indicate fixation) for 2 y.o. and adult group, respectively.

fixation duration significantly decreases from childhood to adulthood. The saccade amplitudes significantly increase between 2 year-old and 4-6 year-old conditions, between 4-6 year-old and 6-10 year-old conditions. There is no difference between 6-10 year-old and adult condition. The central bias, estimated as the mean of the Euclidean distance of fixation to the image center, significantly increases from 4-6 year-old to 6-10 year-old, and from 6-10 year-old to adult conditions. We also observe significant differences concerning the saccade duration and the saccade velocity. They significantly increase with age (see [11]). Hence, there is a significant difference between 2 year-old and 4-6 year-old conditions, between 4-6 year-old and 6-10 year-old conditions, between 6-10 year-old and adult conditions. Regarding the saccade velocity, there is not significant difference between 2 year-old and 4-6 year-old conditions. A significant difference is however observed between 4-6 year-old and 6-10 year-old conditions, between 6-10 year-old and adult conditions.

B. Extraction of gaze-based features

From a given scanpath, we extract its statistical properties describing the main characteristics of the ordered sequence of fixations and saccades. This description does not capture all the oculomotor information contained within a scanpath, but has to provide the best features for predicting observer's age. In the following, we define a fixation $\mathbf{f}$ as $\mathbf{f} = (x, y, t_s, t_e)$. The fixation location is given by (x, y). t_s and t_e represent the start and end time of the fixation, respectively. In a similar way, we define a saccade $\mathbf{s}$ as being $\mathbf{s} = (x_s, y_s, x_e, y_e, t_s, t_e)$, where (x_s, y_s) and (x_e, y_e) represent the start and the end locations, respectively.

From a given scanpath $sp_i = \{(\mathbf{f}_k, \mathbf{s}_k)\}_{k=\{1,\ldots,N\}}$ (N represents the number of fixations and saccades in the scanpath), we compute a set of properties related to the fixation duration, the saccade amplitude, the saccade velocity and the saccade duration. They are defined below:

- The duration of the k^{th} fixation $\mathbf{f}_d^k$ is the difference between the end time t_e^k and the start time t_s^k;
- The amplitude of the k^{th} saccade $\mathbf{s}_a^k$ is the Euclidean distance between the start point $\left(x_s^k, y_s^k\right)$ and the end point $\left(x_e^k, y_e^k\right)$;
- The duration of the k^{th} saccade $\mathbf{s}_d^k$ is the difference between the end time t_e^k and the start time t_s^k;
- The velocity of the k^{th} saccade $\mathbf{s}_v^k$ is given by dividing the saccade amplitude $\mathbf{s}_a^k$ by the saccade duration $\mathbf{s}_d^k$.

For each of these properties, we estimate 4 features, i.e. the median value, the mean value, the standard deviation and the first derivative (4×4 features). We also compute the average distance of the fixation points to the screen center (i.e. called d_c^k for the k^{th} fixation), the standard deviation as well as the mean of the absolute value of the first derivative (3 features). The latter feature allows us to capture aspects of gaze dynamics over the viewing duration.
We also compute the covariance matrix $C = \mathbf{A}\mathbf{A}^T$, where $\mathbf{A}$ is a 5-by-N matrix:

$$\mathbf{A} = \begin{bmatrix} \mathbf{f}_d^0 & \ldots & \mathbf{f}_d^{N-1} \\ \mathbf{s}_a^0 & \ldots & \mathbf{s}_a^{N-1} \\ \mathbf{s}_v^0 & \ldots & \mathbf{s}_v^{N-1} \\ \mathbf{s}_d^0 & \ldots & \mathbf{s}_d^{N-1} \\ d_c^0 & \ldots & d_c^{N-1} \end{bmatrix} \tag{1}$$

The five row vectors are composed of fixation durations, saccade amplitudes, saccade velocities, saccade durations and distances from the screen's center, respectively. The covariance matrix $\mathbf{C}$ provides a compact and natural way for representing the correlation among features (15 features). Finally, we compute the histogram of saccade slopes. The slope is the angle between the saccade direction and the horizontal line. We use 36 bins (36 features), each representing 10 degrees.

Overall, we have 70 features ($4 \times 4 + 3 + 15 + 36$) extracted for 2686 scanpaths distributed as follows: 12.4% for 2 y.o., 23.01% for 4-6 y.o., 41.14% for 6-10 y.o. and 23.45% for adults.

C. Classification algorithm

To investigate a non-linear mapping of scanpath-based features to age, we use multi-class Gentle AdaBoost algorithm [12]. The strategy of adaptive boosting is to learn several weak classifiers that perform slightly better than chance. But as an ensemble, the combination of these weak classifiers usually provides a strong classifier with much better performance. We have implemented a classification tree as a weak learner. To get a relevant classification model, we followed a 10-fold cross validation approach consisting in partitioning the data set into training and testing sets. By default, the number of weak classifiers is set to 60. Results are presented with confusion matrices [13] as illustrated by Fig. 2. Green and red boxes correspond to correct and wrong classifications respectively. In each box, the number and the percentage representing either correct or incorrect classifications are given. The gray boxes provide the correct and wrong predictions per class. Finally, the overall correct and wrong predictions are given in the bottom-right blue box.

ISBN 978-0-9928626-7-1

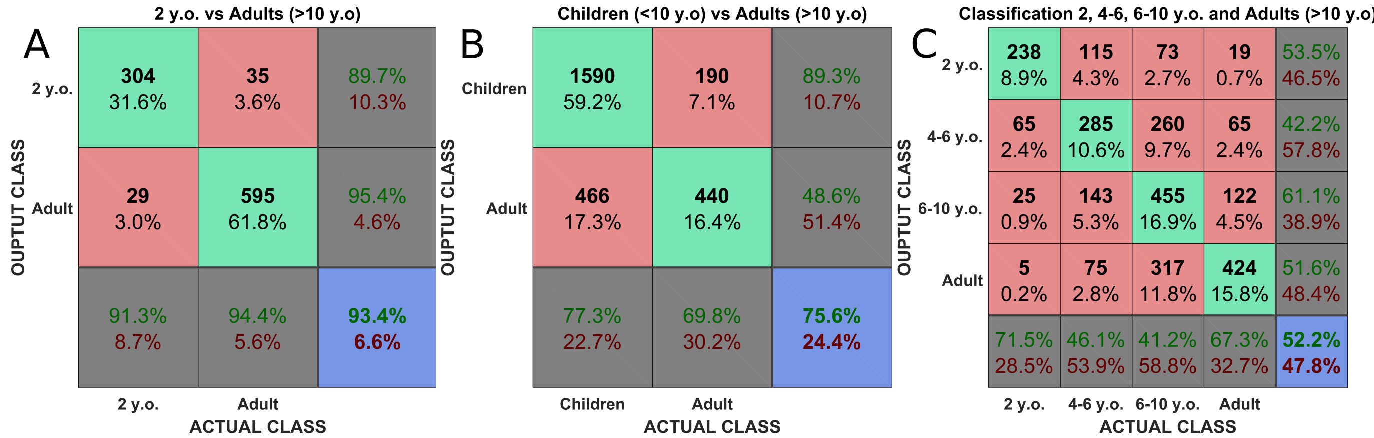

Fig. 2. Confusion matrix for the classification. (A) and (B) binary classifications between 2 y.o. and adult groups, and between children ($< 10y.o.$) and adult ($> 10y.o.$) groups. (C) is the four-class classification. Green and red boxes shows the number and percentage of correct classifications. The grey boxes provide the correct and wrong predictions per class. The overall correct and wrong predictions are given in the bottom-right blue boxes.

III. Results

A. Binary classification

In this first experiment, we evaluate the extent to which two age groups can be correctly identified. First, we consider a binary classifier composed of two classes: 2 year-old versus adult groups. Overall, 93.4% of the predictions are correct and 6.6% are wrong classifications (see Fig. 2 A). 35 adult scanpaths are incorrectly classified as 2 y.o., which corresponds to 3.6% of all 963 scanpaths in the data. 29 children scanpaths are also incorrectly classified, and this corresponds to 8.7% of all scanpaths. 94.4% (resp. 91.3%) of adult (resp. 2 y.o.) observers are correctly classified. Performances are clearly above chance (34% for children and 65% for adults). These promising results suggest that simple gaze-features are sufficient for getting a relevant 2 year-old versus adult classification. Table I presents the classifier performance when considering a subset of the 70 features. The best performances are observed when considering all features. However, we observe that a small number of simple gaze-based feature are sufficient to classify 2 y.o. children and adults with a good accuracy level. We also report in this table the Bookmaker Informedness [14] which specifies the extent to which the prediction is informed versus chance. The benefit of using Informedness is that it is influenced neither by class distribution nor population prevalence. Its value is between -1 and +1: +1 represents perfectly correct performance, while -1 indicates an incorrect response and 0 is the chance level.

A second test consists in considering children (i.e. participants younger than 10 y.o.) and adult (i.e. older than 10 y.o.) groups. Overall, 75.6% of the predictions are correct and 24.4% are wrong classifications (see Fig. 2 B). The classification performance is still significantly above chance for adults (23%) and slightly lower for children (76%). This result could be explained by the fact that the children group is composed of 2 y.o., 4-6 y.o. and 6-10 y.o. groups. Because of the quick maturation of the visual system, the difference between 6-10 y.o. and adult groups becomes small and subtle to determine. This explanation is supported by observations presented in section II-A; these observations indeed show a convergence of visual strategy when aging.

TABLE I
Accuracy and informedness of the binary classification (2 y.o. vs adult groups and children (< 10 y.o.) vs adult groups) in function of a subset of the 70 chosen features: S1=median, average, standard deviation and gradient of fixation durations; S2=similar to S1 but for saccade amplitudes; S3=similar to S1 but for saccade velocities; S4=similar to S1 but for saccade durations; S5=covariance matrix $\mathbf{C}$; S6=similar to S1 but for distance from center.

	All	S1	S2	S3	S4	S5	S6
2 y.o. vs adults							
Accuracy (%)	93.4	86.1	80.1	78.4	80.4	83.1	81.6
Informedness	0.84	0.68	0.55	0.52	0.56	0.62	0.59
$<$ 10 y.o. vs adults							
Accuracy (%)	75.9	64.1	62.2	64.8	66.3	68.5	65.5
Informedness	0.38	0.24	0.20	0.24	0.24	0.26	0.24

B. Multi-class classification

The multiclass problem is performed with the one-vs-all strategy. The four groups, namely 2 y.o., 4-6 y.o., 6-10 y.o. and adults, are considered. Overall, 52.2% of the predictions are correct (see Fig. 2 C). We observe that the classification performance varies a lot from one class to another. 71.5% of 2 y.o. scanpaths and 67.3% of adult scanpaths are correctly classified. The accuracy of the classification decreases to 46.1% and 41.2% for the 4-6 y.o. and 6-10 y.o. groups, respectively. These performances, although lower than those observed for 2 y.o. and adult groups, are above the chance-level; the chance-level of a given class is given by the proportion of scanpaths belonging to this class (see end of section II-A). We also observe that the incorrectly classified scanpaths are not evenly distributed over output classes. Most of them are classified into adjacent classes. For instance, for 2 y.o. class, the input scanpaths are classified as being either a 2 y.o. scanpaths in 71.5% of cases, a 4-6 y.o. scanpaths in 19% of cases, a 6-10 y.o. scanpaths in 7% of cases or an adult scanpath in 1% of

ISBN 978-0-9928626-7-1

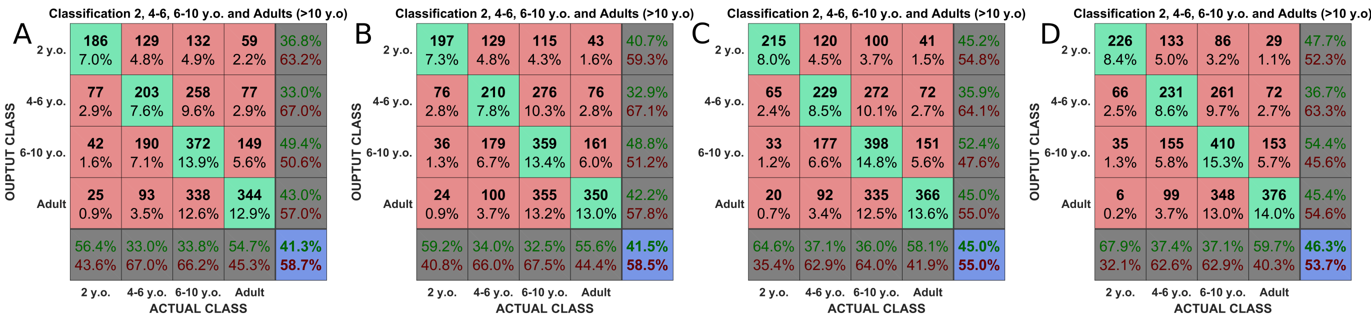

Fig. 3. Influence of the number of fixations on the 4-class classification. (A) first 6, (B) first 9, (C) first 12 and (D) first 15 fixations.

cases. For adults, the input scanpaths are classified as being either a 2 y.o. scanpaths in 3% of cases, a 4-6 y.o. scanpaths in 10% of cases, a 6-10 y.o. scanpaths in 19% of cases or an adult scanpath in 67.3% of cases. For the 4-6 y.o. and the 6-10 y.o. groups, the incorrectly classified scanpaths spread over the 2 y.o. (18%) and 6-10 y.o. (23%) classes, and over the 4-6 y.o. (23%) and adults (28%) classes, respectively. Increasing or decreasing the number of weak classifiers reduces the classification accuracy.

Results in Fig. 2 A and C show that the classification is more accurate with greater between-group age difference. As previously mentioned, this might be due to the developmental course of the principal oculomotor functions. Indeed, oculomotor functions develop rapidly during the first year of life and continue to develop during childhood with a relatively slow and gradual progression near to 10 years of age or even later [6]. It is therefore much more demanding to make the difference between 6-10 y.o. and adult scanpaths than between 2 y.o. and 6-10 y.o. scanpaths.

C. *Influence of the number of fixations on classification*

So far, all fixations, representing a viewing duration of 10 seconds, have been used for performing the classification. By using the same 70 features, the classification performance is assessed in function of the first 6, 9, 12 and 15 visual fixations. If we assume that the mean fixation duration is 250ms, these correspond to a viewing duration of 1.5s, 2.25s, 3s and 3.75s respectively. Results suggest that the performance of the four-class classifier steadily increases with the number of visual fixations taken into account (Fig. 3). From the first 6 fixations to the first 15 fixations, the overall accuracy increases from 41.3% to 46.3%. The more fixations we use, the better the classification is. By making a group-based analysis, we observe the best classification performances are for the 2 y.o. vs. adult groups. Even for 6 fixations, the accuracy is significantly above chance for both categories: 56.4% for 2 y.o. group (chance level is 12%) and 55.6% for adult group (chance level is 23%). Concerning the last two groups, namely 4-6 y.o. and 6-10 y.o., results are more contrasted. The accuracy for the 4-6 y.o. is good, i.e. 33% (chance level is 23%). Regarding the latter, the performance is poor, i.e. 33.8% (chance level is 41%). Classification errors mainly spread over the side groups, i.e. 4-6 y.o. and adult groups.

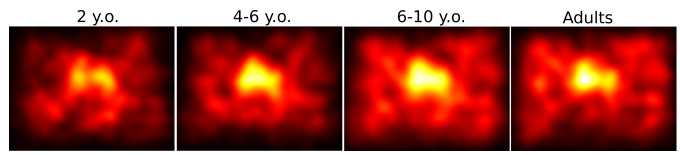

Fig. 4. Average human eye position map maps for the 4 age groups. These color maps are normalized with respect to adult map, for the sake of comparison.

D. *Saliency map features*

In the previous experiments, we have only considered features extracted from the scanpaths. Such features do not account for any saliency and spatial fixation location bias. In the following, we investigate whether the use of saliency maps may improve the classifier performance. The saliency map is a 2D topographic map representing the visual saliency of a corresponding visual scene [15]. We evaluate two kinds of saliency maps: one computed from eye-tracking data (see Fig. 4) and the other predicted from computational models of visual attention.

Human saliency map: The human eye position map is classically computed from human fixation map convolved by a 2D Gaussian kernel [16], [17]. The normalized saliency map represents the fixation density over the visual stimulus. The saliency map is reduced to 12×16 and is vectorized to make a feature vector having a dimension of 192. With this new set of features, composed of 70 gaze-based features and 192 saliency-based features, the classifier is again trained with the one-vs-all strategy on the four age groups. The overall accuracy of the classifier is 52.7%. The gain brought by saliency-based features is only 0.5%, compared to the previous accuracy (Fig. 2 C). Fig. 4 illustrates average saliency maps per age group. We observe significant differences, such as the saliency peak, the coverage [16] and the central bias.

Predicted saliency map: Rather than using human saliency map to perform the training, saliency maps are computed by using the RARE2012 computational model [18]. This model presents a good trade-off between performance and complexity. The overall accuracy of the classifier is 51.9% which is slightly below 52.2% (Fig. 2 C). It indicates that the predicted saliency maps do not bring added-value for predicting age. As expected, we observe that this saliency

ISBN 978-0-9928626-7-1

model can not reproduce the development influence on gaze behavior.

Rather than using the whole saliency map, we compute saliency at attended locations. The computation is performed over a patch of size $1^\circ \times 1^\circ$ centered on the fixation. First we standardize the input saliency map (zero mean and standard deviation equal to 1). We observe that fixated locations get less salient when aging. The average z-score salience at attended location is 0.145, 0.140, 0.087 and 0.059 for 2 y.o., 4-6 y.o., 6-10 y.o. and adult groups, respectively. Note that a significant difference is only observed between 4-6 y.o. and 6-10 y.o. groups (see [11]). By concatenating the 70 gaze-based features with the mean, median, standard deviation and first derivative of the predicted saliency values at fixated locations, the overall accuracy decreases to 51%. This is mainly due to the loss of accuracy in the 4-6 y.o. group (42.6% instead of 46.1%, see Fig. 2 C and [11]).

IV. Conclusion

In this paper, we demonstrate that the age of observers can be inferred from simple statistical properties of their eye movements. The performance of this classification can be further improved by using richer gaze descriptors (e.g. based on HMM [5]) or more sophisticated methods, such as deep learning. However training a deep network would require a much larger sample size. We believe that the rise of low-cost eye-tracker (e.g. webcam-based [19]) will soon provide large-scale eye-tracking datasets and bring the classification performance to the next level. Another issue raised by this study is the lack of computational saliency models for simulating the gaze behavior from childhood to adulthood. This echoes recent works on saccadic models, a range of saliency models able to capture the viewing patterns of a given population and use them to output the predicted scanpaths [20], [21], [22].

Acknowledgment

This work was supported by the National Natural Science Foundation of China under Grant No. 61471230.

References

[1] J. M. Henderson, S. V. Shinkareva, J. Wang, S. G. Luke, and J. Olejarczyk, "Predicting cognitive state from eye movements," *PLoS ONE*, vol. 8, no. 5, p. e64937, 05 2013.

[2] H. Tavakoli, E. Rahtu, and J. Heikkika, "Stochastic bottom-up fixation prediction and saccade generation," *Image and Vision Computing*, vol. 31, pp. 686–693, 2013.

[3] A. Coutrot, N. Binetti, C. Harrison, I. Mareschal, and A. Johnston, "Face exploration dynamics differentiate men and women," *Journal Of Vision*, vol. 16, pp. 1–19, 2016.

[4] L. Itti, "New eye-tracking techniques may revolutionize mental health screening," *Neuron*, vol. 88, no. 3, pp. 442–444, 2015.

[5] A. Coutrot, J. Hsiao, and A. Chan, "Scanpath modeling and classification with hidden markov models," *Behavior Research Methods (in Press)*, 2017.

[6] B. Luna, K. Velanova, and C. F. Geier, "Development of eye-movement control," *Brain and cognition*, vol. 68, no. 3, pp. 293–308, 2008.

[7] E. Aring, M. A. Grönlund, A. Hellström, and J. Ygge, "Visual fixation development in children," *Graefe's Archive for Clinical and Experimental Ophthalmology*, vol. 245, no. 11, pp. 1659–1665, 2007.

[8] A. Helo, S. Pannasch, L. Sirri, and P. Rämä, "The maturation of eye movement behavior: Scene viewing characteristics in children and adults," *Vision research*, vol. 103, pp. 83–91, 2014.

[9] S. Dowiasch, S. Marx, W. Einhäuser, and F. Bremmer, "Effects of aging on eye movements in the real world," *Frontiers in human neuroscience*, vol. 9, p. 46, 2015.

[10] D. Van Renswoude, S. Johnson, M. Raijmakers, and I. Visser, "Do infants have the horizontal bias?" *Infant Behavior and Development*, vol. 44, pp. 38–48, 2016.

[11] O. Le Meur, A. Coutrot, Z. Liu, A. L. Roch, A. Helo, and P. Rämä, "Computational model for predicting visual fixations from childhood to adulthood," *arXiv preprint arXiv:1702.04657*, 2017.

[12] Y. Freund and R. E. Schapire, "A desicion-theoretic generalization of on-line learning and an application to boosting," in *European conference on computational learning theory*. Springer, 1995, pp. 23–37.

[13] S. V. Stehman, "Selecting and interpreting measures of thematic classification accuracy," *Remote sensing of Environment*, vol. 62, no. 1, pp. 77–89, 1997.

[14] D. M. Powers, "Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation," 2011.

[15] C. Koch and S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry," *Human Neurobiology*, vol. 4, pp. 219–227, 1985.

[16] D. S. Wooding, "Eye movements of large populations: Ii. deriving regions of interest, coverage, and similarity using fixation maps," *Behavior Research Methods, Instruments, & Computers*, vol. 34, no. 4, pp. 518–528, 2002.

[17] O. Le Meur and T. Baccino, "Methods for comparing scanpaths and saliency maps: strengths and weaknesses," *Behavior Research Method*, vol. 45, no. 1, pp. 251–266, 2013.

[18] N. Riche, M. Mancas, M. Duvinage, M. Mibulumukini, B. Gosselin, and T. Dutoit, "Rare2012: A multi-scale rarity-based saliency detection with its comparative statistical analysis," *Signal Processing: Image Communication*, vol. 28, no. 6, pp. 642 – 658, 2013.

[19] O. Ferhat and F. Vilarino, "Low cost eye tracking: The current panorama," *Computational Intelligence and Neuroscience*, pp. 1–14, 2016.

[20] O. Le Meur and Z. Liu, "Saccadic model of eye movements for free-viewing condition," *Vision research*, vol. 116, pp. 152–164, 2015.

[21] O. Le Meur and A. Coutrot, "Introducing context-dependent and spatially-variant viewing biases in saccadic models," *Vision Research*, vol. 121, pp. 72–84, 2016.

[22] O. Le Meur, A. Coutrot, A. L. Roch, A. Helo, P. Rämä, and Z. Liu, "Age-dependent sacadic models for predicting eye movements," in *ICIP*, 2017.