Visual Attention Saccadic Models Learn to Emulate Gaze Patterns From Childhood to Adulthood

Olivier Le Meur, Antoine Coutrot, Zhi Liu, Pia Rämä, Adrien Le Roch, Andrea Helo

Visual attention saccadic models learn to emulate gaze patterns from childhood to adulthood

Olivier Le Meur, Antoine Coutrot, Zhi Liu, Pia Rämä, Adrien Le Roch, Andrea Helo

***Abstract*—How people look at visual information reveals fundamental information about themselves, their interests and their state of mind. While previous visual attention models output static 2-dimensional saliency maps, saccadic models aim to predict not only where observers look at but also how they move their eyes to explore the scene. In this paper, we demonstrate that saccadic models are a flexible framework that can be tailored to emulate observer's viewing tendencies. More specifically, we use fixation data from 101 observers split into 5 age groups (adults, 8-10 y.o., 6-8 y.o., 4-6 y.o. and 2 y.o.) to train our saccadic model for different stages of the development of human visual system. We show that the joint distribution of saccade amplitude and orientation is a visual signature specific to each age group, and can be used to generate age-dependent scanpaths. Our age-dependent saccadic model does not only output human-like, age-specific visual scanpaths, but also significantly outperforms other state-of-the-art saliency models. We demonstrate that the computational modelling of visual attention, through the use of saccadic model, can be efficiently adapted to emulate the gaze behavior of a specific group of observers.**

***Index Terms*—saccadic model, scanpaths, saliency, development, age**

I. Introduction

OCULUS *animi index* is an old Latin proverb that could be translated as *the eyes reflect our thoughts*. Eye-movements, revealing how and where observers look within a scene, are mainly composed of fixations and saccades. Fixations aim to bring areas 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 sequence of fixations and saccades an observer performs to sample the visual environment is called a visual scanpath.

Thanks to the advent of modern eye-trackers, allowing us to capture gaze with a high spatial and temporal resolution, a large amount of eye tracking data can be collected with a relative simplicity. Given that the execution of eye movements is the result of a complex interaction between various cognitive processes, mining eye tracking data may provide many indications on our personality, on our mood, and more generally speaking, on the cognitive states of our mind. The way we explore our environment, the way we moves our eyes from one location to another in order to inspect it accurately, may reveal information about our cognitive state. For instance, Henderson et al. [1] inferred the task the participants are engaged in by analyzing eye-movements. Coutrot et al. used Hidden Markov Models to model scanpaths and use them to infer the task at hand or the presence of soundtrack [2]. Wang et al. [3] combined eye tracking with computational attention models in order to screen for mental diseases such as autism spectrum disorder (see also [4], [5]). Tavakoli et al. [6] investigated the influence of eye-movement-based features to determine the valence of images.

Predicting where we look within a scene is of particular relevance for many computer vision applications such as computer graphics [7], quality assessment [8], [9] and compression [10] to name a few.

There exist many computational models of overt visual attention [11]. Saliency models aim to predict the salience of a visual scene. They are based on low-level visual features including color, intensity, and orientation. They process these visual features at several scales using center-surround differences. This process filters out redundant information and outputs feature maps, one per channel. A final saliency map is obtained by combining these feature maps. In contrast with saliency models, saccadic models intend to predict the sequence of eye fixations, i.e. the fashion an observer deploys his/her gaze while viewing a stimulus on screen. Rather than computing an unique saliency map, saccadic models compute visual scanpaths from which scanpath-based saliency maps can be computed. As discussed later in this paper, saccadic models offer many advantages over saliency models. The most important one is the ability to tailor the saccadic model to a particular context, such as a particular type of scene, a particular population or to a particular task at hand [12].

Modelling the human visual attention is a complex task, because of the number of underlying biological mechanisms involved in the visual perception. One of the major difficulties is the high variability in eye-movements. This dispersion is due to many factors, which could be related for instance to the task at hand [13], the cultural heritage [14], the gender [15], [16] and observers' age [17]. The last factor, i.e. the age of observer, is the central concern of this paper.

In this paper, we designed an age-dependent saccadic model in order to reproduce the gaze behavior of a certain target age group. Recent studies give evidence that there exist age-related differences in viewing patterns while free-viewing scene perception [18], [19]. Fixation durations decrease and saccade amplitudes increase with age. These changes may be explained both by eye movement behavior and cognitive

M. Le Meur and M. Le Roch are with University of Rennes 1 IRISA, France. E-mail: olemeur@irisa.fr

Ms. Helo is with Laboratoire Psychologie de la Perception, University of Paris Descartes, and with Departamento de Fonoaudiologa, Universidad de Chile, Santiago, Chili.

Ms. Rama is with Laboratoire Psychologie de la Perception, University of Paris Descartes, CNRS (UMR 8242), France.

M. Liu is with School of Communication and Information Engineering, Shanghai Univ., China.

M. Coutrot is with University College London, UK.

processing. Like our skin, bones, and hair, the eyes undergo a metamorphosis as we grow older. Movement of the eye globe is accomplished by a system of extraocular muscles. Because of aging, the physical traits of these muscles evolve over time, leading to different responses. Another factor is related to cognition which also changes with age. Our perception of the world is indeed a constructive process which heavily relies on prior knowledge and past experiences. These top-down information influence the way we look within visual scene.

Being able to reproduce age-dependent visual deployment may have significant implications on computer vision applications, such as retargeting [20], video compression [10]. In addition, the proposed age-dependent saccadic model relies on a reliable scanpath signature shared within each age group. This signature is used to emulate the gaze behavior of a specific age group of observers. Understanding how this signature evolves with age could help to design new assisting applications for visually impaired people (e.g. people suffering from ARMD (Age-Related Macular Degenerescence)).

This paper is organized as follows. Section II presents saccadic models and focuses more specifically on the modelling framework proposed in [21], [22]. We will also stress how we can tailor this saccadic model for different age groups. Section III presents the eye tracking dataset that is used to determine the scanpath-based signature [18]. Section IV presents the age-dependent saccadic model and section V evaluates its performances. In Section VI, we discuss the results and draw some conclusions.

II. Saccadic model

A. Definition

Saccadic models aim to generate plausible visual scanpaths, i.e. the actual sequence of fixations and saccades an observer would do while viewing stimuli onscreen. By the term plausible, we mean that the predicted scanpaths should be as similar as possible to human scanpaths. They should exhibit similar characteristics, such as the same distributions of saccade amplitudes and saccade orientations. In summary, a saccadic model must predict how observer moves his gaze, but also where the observer looks.

Most existing saccadic models assume that gaze shifts follow a Markov process, meaning that the next gaze location depends only on the current one. In 2000, Brockmann and Geisel [23] generated a sequence of fixation points by considering a stochastic jump process, in which the transition probability density of shifting the gaze from one fixation to another depends on the product of a random salience field and the amplitude of the generated saccade. Boccignone and Ferraro [24] extended Brockmann's work, and modeled eye gaze shifts by using Lévy flights constrained by a bottom-up saliency map. Wang et al. [25] used the principle of information maximization to generate scanpaths on natural images. One interesting point is that they learned the distribution of saccade amplitudes from their own eye movements dataset in order to constrain the selection of the next fixation point. Liu et al. [26] went further by using a Hidden Markov Model (HMM) with a Bag-of-Visual-Words descriptor of image regions to account for semantic content. Tavakoli et al. [27] also

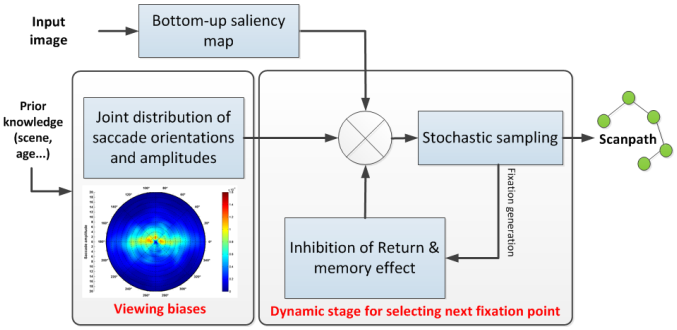


Fig. 2. Flow chart of the saccadic model proposed in [21], [22]. The model takes as input: the original image as well as prior information related to the type of the scene, the age of observers, etc. It outputs a set of visual scanpaths.

incorporated visual working memory and a Gaussian mixture to estimate the distribution of saccade amplitudes. Engbert et al. [28] proposed the *SceneWalk* model of scanpath generation based on two independent processing streams for excitatory and inhibitory pathways. Both are represented by topographic maps: the former represents the foveated saliency map whereas the latter is used for inhibitory tagging. These two maps for attention and inhibitory tagging are then combined. The next fixation point is selected thanks to a stochastic selection [28]. In [21], [22], Le Meur et al. proposed a model of scanpath generation by considering spatially-variant and context-dependent joint distribution of saccade amplitudes and orientations. The next subsection underlines its main components.

B. Le Meur's saccadic model

Predicted scanpaths result from the combination of three components, namely a bottom-up saliency map, viewing biases and memory mechanism, as illustrated by Fig. 2. In the following, we summarize the main operations involved in the method proposed in [21], [22].

Let $\mathcal{I} : \Omega \subset \mathcal{R}^2 \mapsto \mathcal{R}^m$ ($m = 3$ for RGB image) an input image and x_{t-1} a fixation point at time $t-1$. The next fixation point x_t is determined by sampling the 2D discrete conditional probability $p\left(x|x_{t-1}\right)$ which indicates, for each location of the definition domain Ω, the transition probability between the previous fixation and the current location x. The conditional probability $p\left(x|x_{t-1}\right)$ is composed of three terms as described in Eq. 1:

$$p\left(x|x_{t-1}\right) \propto p_{BU}(x) p_M(x,t|T) p_B(d(x,x_{t-1}),\phi(x,x_{t-1})) \quad (1)$$

where,

- $p_{BU}(x)$ represents the input 2D bottom-up saliency map. This saliency map is computed by a *traditional* saliency model, or by combining the results of several saliency models [29].
- $p_M(x,t|T)$ represents the memory state of the location x at time t, according to the T past fixations. This time-dependent term simulates the inhibition of return and indicates the probability to refixate a given location. As described in [21], $p_M(x,t|T)$ is composed of two operations: one for inhibiting the current attended location in order to favor the scene exploration. At the opposite, the

(a) (b) (c) (d) (e)

Fig. 1. (a) Original stimulus; (b) and (c) represent fixation maps (red crosses indicate fixation) for 2 year-old and adult group, respectively; (d) and (e) represent the actual saliency maps for 2 year-old and adults groups, respectively.

second term allows to recover the initial salience of the previous attended locations, favoring the re-fixation. An attended location requires T fixations before recovering the integrality of its salience.

- $p_B(d, \phi)$ represents the probability to observe a saccade of amplitude d and orientation ϕ. The saccade amplitude d, expressed in degree of visual angle, is the Euclidean distance between two consecutive fixation points x_t and x_{t-1}. The saccade orientation ϕ is the angle, expressed in degree, between these two consecutive fixation points. The joint probability of saccade amplitudes and orientations is learned from actual eye tracking data, by using kernel density estimation [30]. This representation implicitly encompasses gaze biases, which reflect the main tendencies of observers looking at well-defined stimuli. The joint probability is also content-dependent [22], indicating that our visual strategy depends on the stimulus displayed on screen (see also supplementary material[1]). By choosing the most relevant joint probability with respect to the displayed scene, the saccadic model can be fine-tuned for reproducing a specific visual behavior. This is one major difference between saccadic model and *traditional* saliency models. In Section III, we will see that the joint distribution of saccade amplitudes and orientations is a good candidate for representing the differences in visual deployment that exist between young children and adults.

When the three terms of the conditional probability $p\left(x|x_{t-1}\right)$ are known for all sites of the definition domain Ω, the next fixation point can be inferred. One obvious solution would be to consider the maximum a posteriori solution, also called the *Bayesian ideal searcher* in [31]. However, this solution is deterministic and fails to represent uncertainty about visual perception and perceptual interpretations [32]. Another way to model trial-to-trial variability, or in our context the dispersion between observers, is to assume a stochastic rule for choosing the next fixation point. In [21], a set of N_c samples is drawn from the conditional probability $p\left(x|x_{t-1}\right)$. The next fixation point is selected as being the sample having the highest bottom-up salience. This implementation is close to the one proposed in [28]. This form of stochastic selection is also known as Luce's choice rule [33]. It is important to underline that the number of samples drawn from the conditional probability controls the amount of dispersion between observers. A high number of samples (or candidates) would reduce the dispersion between observers. In the extreme case, where N_c tends to infinity, the inference of the next fixation point becomes deterministic and strongly similar to the *Bayesian ideal searcher*. At the opposite, when N_c is equal to 1, the amount of randomness is maximal providing the highest dispersion between observers.

This sampling strategy is obviously sub-optimal because the next fixation point is not necessarily the point having the highest probability to be attended. However, this strategy akin to probability matching [34] has been reported to be used by humans in a variety of cognitive tasks [35], [36].

In the next section, we show how this framework is able to capture and implement the specificities of gaze behaviour across the development of the visual system.

III. Eye movements from childhood to adulthood

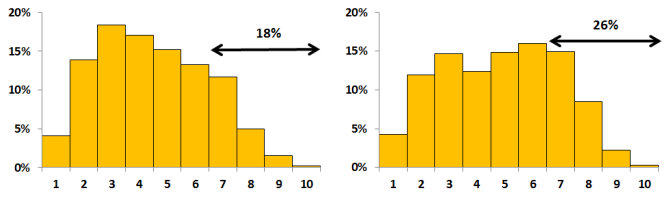
In this section, we analyzed an eye tracking data collected from observers of a wide range of ages. The main purpose was to investigate whether the joint distribution of saccade orientations and amplitudes learned from this raw eye tracking data is able to capture the gaze biases of different age groups. We already know that aging has an impact on the way we deploy our visual attention [37], [38]. If we succeed in quantitatively measuring the influence of development, the saccadic model described in the previous section could be tuned to replicate the gaze behavior of a specific age group.

A. Maturation of eye-movements

The visual system at birth is limited but develops rapidly during the first years of life and continues to improve through adolescence. Helo et al. [18] give evidence of age-related differences in viewing patterns during free-viewing natural scene perception. Fixation durations decrease with age while saccades turn out to be shorter when comparing children with adults. Materials and methods of this eye-tracking experiment are briefly summarized below.

1) Participants: A total of 101 subjects participated in the experiments, including 23 adults and 78 children. These subjects were divided into 5 groups: 2 year-old group, 4-6 year-old group, 6-8 year-old group, 8-10 year-old group and adults group. Participants were instructed to explore the images. The 4-10 year-old and the adults 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 (more details on experimental design is available in [18]).

[1]Available on Le Meur's webpage



(a) 2 year-old (b) Adults

Fig. 3. Distribution of visual fixations in function of the distance from the center and distributed into 10 crowns, numbered from 1 to 10 (1 is the centered crown). The y-axis represents a percentage of visual fixations.

2) Stimuli: Thirty color pictures taken from children books, as illustrated in Fig. 1 (a), are displayed for 10s. A drift correction is performed before each stimulus. The viewing distance is 60 cm. One degree of visual angle represents 28 pixels. For all the results reported in this paper, the first fixation has been removed.

3) Saliency map and center bias: Fig. 1 (b) to (e) illustrates fixation maps and saliency maps computed from eye tracking data of 2 year-old and adults groups. The saliency map is classically computed by convolving actual eye positions with a 2D Gaussian function which approximates the central part of the retina, i.e. the fovea [39]. The standard deviation is set to 28 pixels representing one degree of visual angle [40]. We observed that adults tend to explore much more the visual scene than 2 year-old children. In addition, the center bias is more important for the 2 year-old group than for the adult group. We quantify this trend by computing the ratio of fixations falling within centered crowns. For this purpose, a set of 10 concentric circles is used. The radius of each circle represents 10%, 20%,...,90%, 100% of the distance between the picture center and its top-left corner. The ratio of fixations falling within each crown (difference between two concentric successive circles) to the overall number of fixations is calculated. Fig. 3 plots these distributions for the 2 year-old and adults groups. The cumulative percentage of the last 4 crowns indicates that the center bias is more significant for young children than adults (26% of adults' fixations fall within these crowns, compared to only 18% for 2 year-old children). More results are presented in supplementary materials.

B. Joint distribution of saccade orientations and amplitudes

Following the method proposed in [21], we estimate the joint probability distribution of saccade amplitudes and orientations $p_B(d, \phi)$ for each age group. This nonparametric distribution is obtained by using a 2D Gaussian kernel density estimation. The two bandwidth parameters are chosen optimally based on the linear diffusion method proposed by [41]. The joint probability $p_B(d, \phi)$ is given by:

$$p_B(d, \phi) = \frac{1}{n} \sum_i K_h(d - d_i, \phi - \phi_i) \qquad (2)$$

where d_i and ϕ_i are the distance and the angle between each pair of successive fixations respectively. n is the total number of samples and K_h is the two-dimensional Gaussian kernel. Fig. 4 shows the joint probability distributions of saccade amplitudes and orientations (bottom row) in a polar plot representation. Radial position indicates saccadic amplitudes expressed in degree of visual angle. The top row of Fig. 4 shows the marginal probability distributions of saccade amplitudes.

A number of observations can be made: first, eye-movement patterns change with age. Saccade amplitudes are shorter in the 2 year-old group than in adults group. Saccade amplitudes increase with age. This first observation is consistent with the ones made in [18]. Regarding the saccade orientations, we observe a strong horizontal bias in the adult group which is also consistent with previous studies [42], [43]. This horizontal bias can be explained by several factors, such as biomechanical factors, physiological factors and the layout of our natural environment [44]. Regarding biomechanical factors, Van Renswoude et al. [44] stress the point that horizontal saccades require only the use of one pair of muscles whereas saccades in the other directions requires more than one pair of muscles [45]. This horizontal bias is less obvious for young children, even though it may exist [44]. Fig. 4 (bottom row) also shows that the distribution shape of the 2 year-old group (a) is much more isotropic than the adults' one (d), but with a marked tendency for making upward vertical saccades.

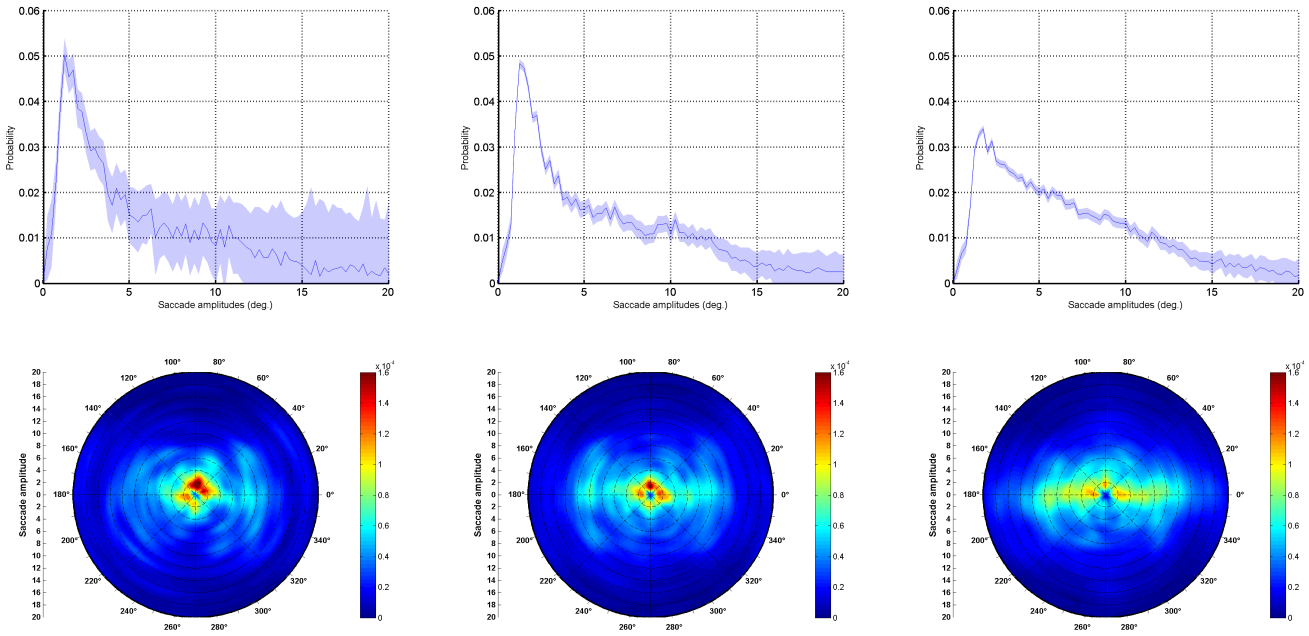
A two-sample two-dimensional Kolmogorov-Smirnov test [46] is performed to test whether the difference between the joint distributions illustrated in Fig. 4 is statistically significant. For two given distributions, we randomly draw 5000 samples and test whether both data sets are drawn from the same distribution. The tests show significant differences between 2 year-old and 4-6 year-old groups, and between 4-6 year-old and 6-8 year-old groups (all $p < .001$). There is no difference between 6-8 year-old and 8-10 year-old groups ($p = 0.2$). A significant difference is however observed between adults and 8-10 year-old groups ($p = 0.0049$). We reduced the within-group variance by increasing the sample size and merging the 6-8 and 8-10 year-old groups together. The resulting group is called the 6-10 year-old group.

In summary, these results suggest that the joint distribution of saccade amplitudes and orientations is able to grasp gaze behavior differences across age, as well as to reflect important features of development on the visual deployment.

C. To what extent do saliency models predict where infants and adults look?

There exist a number of saliency models working on different modalities, such as images [47]–[49], video sequences [50], [51], audio-visual video sequences [52] and compressed domain [51], [53] (see [11] for a taxonomy). In this study, ten saliency models are tested: BMS [54], AWS [48], AIM [55], ITTI [47], HOU [56], SUN [57], IMSIG [58], SIM [59], GBVS [60] and RARE2012 [61]. Performance is evaluated using the following metrics:

- The linear correlation coefficient (CC) is computed between two saliency maps. A value of 0 means that the two maps are uncorrelated.
- The similarity (SIM) is calculated based on the normalized probability distributions of the two maps [62]. The



(a) 2 year-old (b) 4-6 year-old (c) Adults

Fig. 4. Distribution of saccade amplitudes (top row) and polar plots of joint distribution of saccade amplitudes and orientations (bottom row) for different age groups: (a) 2 year-old group to (d) adult group. The light blue envelope on top-row curves represents the standard error of the mean, amplified by a factor of 2000. The 6-8 and 8-10 year-old distributions are not displayed for the sake of clarity. They are available in the supplementary materials.

similarity is the sum of the minimum values at each point in the distributions. SIM=1 means the distributions are identical whereas SIM=0 means the distributions are completely opposite.

- The Earth Mover's Distance (EMD) measures the distance between two probability distributions by how much transformation on one distribution would need to undergo to match another (EMD=0 for identical distributions).
- The metrics called AUC-Judd and AUC-Borji consist in considering the saliency map as a binary classifier to separate positive from negative samples at various thresholds (see [61], [63] for a review). A ROC analysis is then performed for computing the Area Under Curve: a score of 1 means that the classification is perfect, whereas a value of 0.5 is the chance level.
- The normalized scanpath score (NSS) measures the mean value of the normalized saliency map at fixation locations [64]. NSS=0 represents the chance level. A high positive value means that fixations fall within salient parts of the scene.

These metrics are complementary: the CC metric is used to compare two saliency maps, SIM and EMD compare two distributions whereas AUC-Judd, AUC-Borji and NSS compare a map with a set of fixations. Readers can refer to [40], [61], [63] for more details on these metrics.

The performances are given in Table I. The results were analyzed using a three-way mixed ANOVA design. Age groups (adults, 6-10 yo, 4-6 yo, or 2 yo) was the between-subjects variable; type of saliency model (GBVS or RARE2012[2]) and type of metric (CC, SIM, EMD, KL, AUC-Judd, AUC-Borji, or NSS) were the within-subjects variables. The three-way ANOVA yielded a significant main effect of age ($F(3, 95) = 17.55$, $p < .001$), model ($F(1, 95 = 4.87$, $p = 0.03$) and metric ($F(6, 90) = 784.84$, $p < .001$). The $metric \times age$ interaction is significant ($F(18, 276) = 8.10$, $p < .001$), as well as the $model \times metric$ interaction ($F(6, 90) = 59.91, p < .001$). The $model \times age$ interaction is not significant ($F(3, 95) = 2.53, p = 0.062$). Post-hoc Bonferroni comparisons show significant differences between all age groups ($p < .001$), except between adults and 6-10 yo, and between 4-6 yo and 2 yo ($p = 1$).

This analysis leads to several observations. First the influence of bottom-up factors such as saliency in eye movement behavior is significant for all age groups. This is specifically the case for GBVS model. Indeed GBVS model significantly outperforms RARE2012 model (paired t-test, $p << 0.01$). According to previous benchmarks of computational models of saliency [21], [65], this discrepancy in performance between these two models is unusual. This performance gap might be explained by two major differences between GBVS and RARE2012. The central bias, while intrinsically taken into account by GBVS, is not considered by RARE2012. As illustrated by Fig. 5 (a), we can observe black stripes all around GBVS saliency map, which may significantly improve the performance of the model [66]. Second, RARE2012 maps are much more focused than GBVS ones. The less focused

[2]We consider only these two models in the analysis because they are used as input for the saccadic model (see section V)

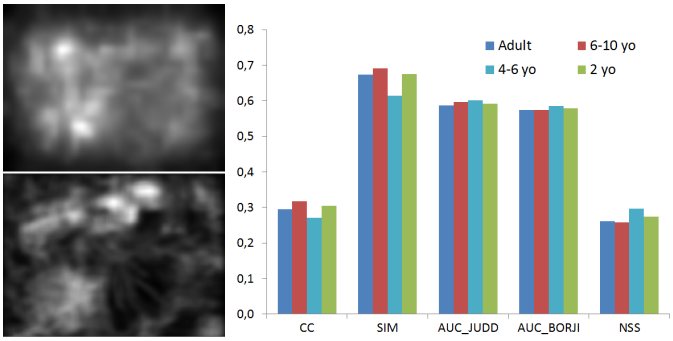
TABLE I
PERFORMANCE OF GBVS, RARE2012 MODELS AND SACCADIC MODEL (USING AN INPUT SALIENCY MAP COMPUTING FROM EITHER GBVS OR RARE2012). WE REPORT THE PERFORMANCES OF OUR SACCADIC MODEL FOR AN OPTIMAL N_c VALUE. THE BEST SCORES ARE IN BOLD. DETAILLED PERFORMANCES ARE GIVEN IN TABLE II.

Metrics	CC	SIM	EMD	AUC-Judd	AUC-Borji	NSS
Adults						
BMS	0.280	0.671	1.159	0.589	0.574	0.247
AWS	0.231	0.643	1.328	0.559	0.559	0.211
AIM	0.305	0.696	1.087	0.583	0.565	0.270
ITTI	0.353	0.702	0.962	0.615	0.589	0.311
HOU	0.239	0.644	1.299	0.567	0.560	0.217
SUN	0.129	0.649	1.260	0.539	0.534	0.117
IMSIG	0.301	0.688	1.170	0.601	0.580	0.264
SIM	0.139	0.659	1.209	0.535	0.536	0.130
GBVS	0.531	**0.731**	0.868	**0.644**	**0.634**	0.463
Our model	**0.636**	0.706	**0.759**	**0.644**	**0.639**	0.561
RARE2012	0.290	0.638	1.228	0.592	0.575	0.256
Our model	**0.566**	**0.701**	**0.787**	**0.640**	**0.630**	**0.492**
6-10 y.o.						
BMS	0.315	0.691	1.121	0.599	0.579	0.261
AWS	0.236	0.655	1.338	0.565	0.555	0.193
AIM	0.328	0.716	1.053	0.594	0.566	0.269
ITTI	0.357	0.719	0.947	0.619	0.583	0.289
HOU	0.236	0.652	1.321	0.567	0.556	0.193
SUN	0.149	0.666	1.252	0.544	0.538	0.124
IMSIG	0.317	0.692	1,131	0.596	0.575	0.259
SIM	0.148	0.675	1.198	0.539	0.534	0.126
GBVS	0.589	**0.761**	0.776	**0.661**	**0.640**	0.478
Our model	**0.686**	0.732	**0.744**	0.659	**0.640**	**0.562**
RARE2012	0.328	0.661	1.183	0.607	0.578	0.266
Our model	**0.617**	**0.717**	**0.792**	**0.648**	**0.630**	**0.505**
4-6 y.o.						
BMS	0.265	0.614	1.485	0.606	0.585	0.289
AWS	0.204	0.585	1.640	0.570	0.563	0.224
AIM	0.269	0.626	1.488	0.589	0.573	0.296
ITTI	0.304	0.634	1.330	0.627	0.597	0.334
HOU	0.200	0.581	1.626	0.574	0.566	0.221
SUN	0.121	0.586	1.660	0.541	0.542	0.136
IMSIG	0.271	0.614	1.478	0.602	0.586	0.296
SIM	0.116	0.592	1.624	0.532	0.533	0.124
GBVS	0.544	**0.691**	1.052	**0.690**	0.675	0.597
Our model	**0.673**	0.690	**0.806**	0.688	**0.681**	**0.745**
RARE2012	0.275	0.592	1.464	0.614	0.587	0.296
Our model	**0.602**	**0.683**	**0.900**	**0.678**	**0.672**	**0.661**
2 y.o.						
BMS	0.210	0.587	1.491	0.581	0.572	0.236
AWS	0.152	0.562	1.644	0.549	0.546	0.171
AIM	0.233	0.601	1.460	0.570	0.561	0.262
ITTI	0.266	0.607	1.332	0.602	0.582	0.301
HOU	0.166	0.562	1.674	0.549	0.548	0.193
SUN	0.078	0.563	1.617	0.525	0.526	0.083
IMSIG	0.231	0.591	1.472	0.582	0.572	0.261
SIM	0.062	0.564	1.594	0.513	0.515	0.065
GBVS	0.501	**0.662**	1.071	**0.674**	0.667	0.570
Our model	**0.579**	0.659	**0.906**	**0.674**	**0.671**	**0.666**
RARE2012	0.264	0.578	1.431	0.601	0.579	0.292
Our model	**0.517**	**0.639**	**1.014**	**0.662**	**0.653**	**0.585**

GBVS maps might be an advantage in the context of this study. Indeed, the stimuli used in this experiments (see Fig. 1) are very dense, containing several areas of interest. In addition, except for the 2 year-old kids, all participants performed a recognition task which might favor scene exploration, and penalize too clustered saliency maps.

A second observation is related to the influence of salience in the four age groups. Regarding GBVS and RARE2012 models, the best match is obtained for the 6-10 year-old group when considering the CC, SIM and EMD metrics, for both saliency models. For the other three metrics, i.e. AUC-Judd, AUC-Borji and NSS, the best scores are obtained for the 4-6 year-old group.

Figure 5 (b) illustrates the average performance of the ten tested saliency models; each score is obtained by averaging the ten scores presented in Table I. We were expected to observe a notable trend towards lower performance with age. Indeed, some studies such as [18], [67] suggest that bottom-up factors decrease when aging while the role of top-down processes increases. Our results do not exhibit a clear and significant trend. However, this is not in agreement with [68] who recently came to the conclusion that saliency models are better for predicting adult saliency maps than infant saliency maps. This conclusion may also seem counterintuitive and does not agree with the aforementionned studies.



(a) (b)

Fig. 5. (a) GBVS (Top left) and RARE2012 (Bottom left) saliency maps for the original image in Fig. 1 (a). (b) Performance over 10 saliency models in function of age.

IV. AGE-DEPENDENT SACCADIC MODEL

In this section, we tailor Le Meur's saccadic model to the different age groups, namely, 2, 4-6, 6-10 year-old and adults. We perform three modifications to the purpose of our study. The first modification consists in using a joint probability density function $p_B(d, \phi)$ that has been learned from eye tracking data collected from different age groups, as presented in section III-B. This prior knowledge represents the viewing tendencies, expressed in this study in terms of saccade amplitudes and orientations, which are common across all observers of a given age. The use of such a prior is fundamental to constrain how we explore scenes and to generate saccade amplitudes and orientations that match those estimated from human eye behavior.

However, rather than using a unique joint distribution per age group, we use, as suggested in [22], a spatially-variant joint distribution. The image is then split into a non-overlapping 3×3 grid; for each cell in the grid, the joint distribution of saccade amplitudes and saccade orientations is estimated following the procedure detailed in section III-B (the polar plots of these distributions are given in supplementary materials). This spatially-variant prior is more appropriate for catching important viewing tendencies. One of the most important priors is the central bias. Indeed, as illustrated in [22] as well as in the supplementary materials, saccades located on the frame corners move the gaze towards the screen's center. This reflects our tendency to look near the screen's center, irrespective of the visual information at that location.

The second modification is related to the number of samples, N_c, which is drawn from the conditional probability $p\left(x|x_{t-1}\right)$. As presented in section II, this parameter can be used to tune the amount of randomness in the selection of the next fixation point. A low value results in a high dispersion between observers, and fosters the scene exploration [69]. A high value would reduce the dispersion. In our study, we evaluate the performance of the saccadic model for $N_c \in \{1, ..., 9\}$.

The third modification concerns the selection of the most appropriate candidate among the N_c candidates drawn from the conditional probability $p\left(x|x_{t-1}\right)$. In [21], the next fixation point is selected as being the candidate having the highest bottom-up salience. This selection rule is modified to take into account the probability $p_B(.,.)$, the bottom-up saliency $p_{BU}(.)$ and the distance d between the candidate and the previous fixation point. The next fixation point x^* is then selected as

$$x^* = \arg\max_{s \in \Theta} \frac{p_{BU}(s) \times p_B(d(s, x_{t-1}), \phi(s, x_{t-1}))}{d(s, x_{t-1})} \quad (3)$$

where, Θ is the set of N_c candidates, x_{t-1} is the previous fixation point and d is the Euclidean distance between the candidate s and the previous fixation point. This new rule allows us to favor the candidates that are close to the previous fixation point and featured by both a high probability to be attended and high bottom-up salience.

V. Performances

First, we evaluate the extent to which the predicted fixations fall within salient areas. Second, we test the plausibility of the generated scanpaths with respect to the actual scanpaths of the four age groups. Third, we evaluate the benefit to use dedicated age-dependent distributions of saccade amplitudes and orientations.

To perform these evaluations, we proceed as follows: for each image of the dataset and for each age group, we generate 20 scanpaths, each composed of 15 fixations. The first fixation is randomly chosen. The input saliency map, i.e. the term p_{BU} in equation 1, is computed using either the GBVS or RARE2012 model. These two models are chosen because of the good tradeoof between simplicity and performance [61]. From the generated scanpaths, a saliency map is computed by following the classical procedure, as described in section III-A (see also [39], [40]). These maps are called scanpath-based saliency maps.

A. Prediction of salient areas

Table I and II present the similarity degree between scanpath-based saliency map and the ground truth (i.e. either human saliency map or eye tracking data). Table I provides the performance of our saccadic model for an optimal value of N_c. We observe that our model significantly outperforms RARE2012 model, whatever age groups and metrics. Compared to GBVS model, our model performs better according to 4 metrics. A thorough statistical analysis is performed from the detailed scores given in Table II. The results were analyzed using 2 three-way mixed ANOVA designs.

The first one uses age groups (adults, 6-10 year-old, 4-6 year-old, or 2 year-old) as the between-subjects variable, type of saliency model (GBVS or GBVS-based saccadic model) and type of metric (CC, SIM, EMD, AUC-Judd, AUC-Borji, or NSS) as the within-subjects variables. For each metric and age group, we used the N_c value that led to the best result. The three-way ANOVA yielded a significant main effect of age ($F(3, 95) = 15.31$, $p < .001$), model ($F(1, 95) = 8.056$, $p = 0.006$) and metric ($F(5, 91) = 110.50$, $p < .001$). The $metric \times age$ interaction is significant ($F(15, 279) = 9.035$, $p < .001$), as well as the $model \times metric$ interaction ($F(5, 91) = 52.32$, $p < .001$). The $model \times age$ interaction is not significant ($F(3, 95) = 1.25$, $p = 0.29$). Post-hoc Bonferroni comparisons show significant differences between all age groups (all $p < .001$, except between 2 year-old and 6-10 year-old where $p = 0.035$), except between 4-6 year-old and 2 year-old ($p = 0.49$).

The second ANOVA analysis uses age groups (adults, 6-10 year-old, 4-6 year-old, or 2 year-old) as the between-subjects variable, type of saliency model (RARE or RARE-based saccadic model) and type of metric (CC, SIM, EMD, AUC-Judd, AUC-Borji, or NSS) as the within-subjects variables. For each metric and age group, we used the N_c value that led to the best result. The three-way ANOVA yielded a significant main effect of age ($F(3, 95) = 6.81$, $p < .001$), model ($F(1, 95) = 67.92$, $p < 0.001$) and metric ($F(5, 91) = 262.45$, $p < .001$). The $metric \times age$ interaction is significant ($F(15, 279) = 7.43$, $p < .001$), as well as the $model \times metric$ interaction ($F(5, 91) = 96.143$, $p < .001$). The $model \times age$ interaction is not significant ($F(3, 95) = 0.62$, $p = 0.60$). Post-hoc Bonferroni comparisons show significant differences between adults and 4-6 yo ($p < .001$), marginal differences between adults and 2 year-old ($p = 0.086$) and no difference between adults and 6-10 year-old ($p = 1$). There is a significant difference between 6-10 year-old and 4-6 year-old ($p = 0.01$) but not between 6-10 year-old and 2 year-old ($p = 0.57$). There is no significant difference between 2 year-old and 4-6 year-old ($p = 1$).

In summary, as shown in Table II, the proposed saccadic model performs better than GBVS and RARE2012 models. When the input saliency map of the saccadic model is the saliency map computed by RARE2012, the saccadic model significantly outperforms RARE2012 for all considered similarity metrics. These results are given on the right hand-side of Table II. We draw a similar conclusion for the CC, EMD and NSS metrics when the GBVS model is used to compute the input saliency map. Concerning the SIM, AUC-Judd and AUC-Borji metrics, the performances of the GBVS-based saccadic model are similar to GBVS model.

The proposed model performs well when N_c is in between 3 and 7, for all age groups. This shows a reasonable flexibility with the choice of the N_c parameter. As discussed in the next section, the parameter N_c appears to be much more important when it comes to generate plausible visual scanpaths.

B. Are visual scanpaths plausible?

Saccadic models predict salient areas as well as to generate scanpaths that present similar features as human scanpaths.

TABLE II

PERFORMANCE OF GBVS, RARE2012 MODELS AND SACCADIC MODEL. THE BEST SCORES ARE IN BOLD. A DAGGER, I.E. †, IS ADDED WHEN THERE IS A STATISTICALLY SIGNIFICANT DIFFERENCE (PAIRED T-TEST, $p < 0.05$) BETWEEN GBVS (RESP. RARE2012) AND GBVS-BASED SACCADIC MODEL (RESP. RARE2012-BASED SACCADIC MODEL). THE BULLET, I.E. $^\bullet$, INDICATES THE SCORES THAT ARE NOT STATISTICALLY SIGNIFICANT: THE PAIRED T-TEST IS PERFORMED IN THIS CASE BETWEEN THE HIGHEST SCORE (IN BOLD) AND OTHER SCORES OBTAINED BY VARYING N_c. ON THE LAST ROWS, *NSV dist.* MEANS NON SPATIALLY VARIANT JOINT DISTRIBUTION AND SVDIST2YO MEANS SPATIALLY VARIANT JOINT DISTRIBUTION OF 2 Y.O. GROUP.

Metrics	**CC**	**SIM**	**EMD**	**AUC-Judd**	**AUC-Borji**	**NSS**	**CC**	**SIM**	**EMD**	**AUC-Judd**	**AUC-Borji**	**NSS**
					Adults							
	GBVS model [60]						**RARE2012 model [61]**					
	0.531	**0.731**†	0.868	**0.644**	**0.634**	0.463	0.290	0.638	1.228	0.592	0.575	0.256
	GBVS-based saccadic model						**RARE2012-based saccadic model**					
Nc=1	0.471	0.706	**0.759**†	0.621	0.615	0.416	0.459	**0.701**†	**0.787**†	0.617	0.610	0.402
Nc=2	0.580	0.704	0.821$^\bullet$	0.641$^\bullet$	0.633	0.507	**0.566**†	0.698$^\bullet$	0.824$^\bullet$	**0.640**†	**0.630**†	**0.492**†
Nc=3	0.619$^\bullet$	0.686	1.029	**0.644**	**0.639**	0.541$^\bullet$	0.561$^\bullet$	0.671	1.094	0.639$^\bullet$	0.627$^\bullet$	**0.492**†
Nc=4	**0.636**†	0.676	1.099	0.643$^\bullet$	0.635$^\bullet$	**0.561**†	0.537$^\bullet$	0.644	1.250	0.630	0.619	0.467$^\bullet$
Nc=5	0.617$^\bullet$	0.651	1.244	0.634	0.629	0.540†	0.524	0.626	1.342	0.626	0.613	0.460$^\bullet$
Nc=6	0.615$^\bullet$	0.485	1.285	0.637	0.628	0.537	0.503	0.605	1.474	0.616	0.605	0.441$^\bullet$
Nc=7	0.620	0.638	1.368	0.635	0.629	0.543	0.500	0.599	1.442	0.619	0.605	0.441$^\bullet$
Nc=9	0.606	0.629	1.429	0.634	0.625	0.535	0.483	0.581	1.585	0.611	0.600	0.424
					6-10 y.o.							
	GBVS model [60]						**RARE2012 model [61]**					
	0.589	**0.761**	0.776	**0.661**	**0.640**	0.478	0.328	0.661	1.183	0.607	0.578	0.266
	GBVS-based saccadic model						**RARE2012-based saccadic model**					
Nc=1	0.479	0.716	**0.744**	0.620	0.608	0.390	0.448	0.716$^\bullet$	**0.792**†	0.621	0.602	0.367
Nc=2	0.649	0.732	0.740$^\bullet$	0.656	**0.640**	0.532	0.588	**0.717**†	0.816$^\bullet$	**0.648**†	0.629$^\bullet$	0.477
Nc=3	0.666	0.707	0.954	0.658	**0.640**	0.545	**0.617**†	0.698	0.982	**0.648**†	**0.630**†	**0.505**†
Nc=4	**0.686**†	0.698	1.011	0.659	**0.640**	**0.562**†	0.583	0.658	1.277	0.642$^\bullet$	0.621	0.475
Nc=5	0.655	0.675	1.162	0.655	0.635	0.535	0.569	0.639	1.330	0.639	0.616	0.463
Nc=6	0.667	0.661	1.258	0.646	0.631	0.545	0.560	0.625	1.355	0.637	0.613	0.455
Nc=7	0.667	0.655	1.283	0.647	0.631	0.546	0.547	0.608	1.494	0.630	0.607	0.448
Nc=9	0.660	0.647	1.361	0.646	0.629	0.544	0.525	0.586	1.655	0.627	0.600	0.429
					4-6 y.o.							
	GBVS model [60]						**RARE2012 model [61]**					
	0.544	**0.691**	1.052	**0.690**	0.675	0.597	0.275	0.592	1.464	0.614	0.587	0.296
	GBVS-based saccadic model						**RARE2012-based saccadic model**					
Nc=1	0.506	0.666	0.965	0.657	0.651	0.556	0.494	0.664	0.987$^\bullet$	0.652	0.647	0.541
Nc=2	0.630	0.690	**0.806**†	0.684	0.680$^\bullet$	0.692	**0.602**†	**0.683**†	**0.900**†	**0.678**†	**0.672**†	**0.661**†
Nc=3	0.660$^\bullet$	0.687	0.940	0.688	**0.681**	0.730$^\bullet$	0.592$^\bullet$	0.662	1.037	0.677$^\bullet$	0.667	0.651$^\bullet$
Nc=4	**0.673**†	0.669	1.020	0.683	0.675$^\bullet$	0.744$^\bullet$	0.569	0.632	1.216	0.671	0.655	0.624
Nc=5	**0.673**†	0.663	1.108	0.685	0.675$^\bullet$	**0.745**†	0.552	0.619	1.295	0.667	0.649	0.605
Nc=6	0.663$^\bullet$	0.651	1.131	0.680	0.669	0.737$^\bullet$	0.546	0.605	1.366	0.660	0.644	0.596
Nc=7	0.658$^\bullet$	0.645	1.221	0.677	0.670	0.730$^\bullet$	0.539	0.598	1.393	0.663	0.641	0.595
Nc=9	0.650	0.632	1.238	0.675	0.665	0.720	0.504	0.565	1.508	0.651	0.628	0.550
					2 y.o.							
	GBVS model [60]						**RARE2012 model [61]**					
	0.501	**0.662**	1.071	**0.674**	0.667	0.570	0.264	0.578	1.431	0.601	0.579	0.292
	GBVS-based saccadic model						**RARE2012-based saccadic model**					
Nc=1	0.385	0.624	1.157	0.628	0.622	0.445	0.413	0.627	1.136	0.628	0.629	0.468
Nc=2	0.556$^\bullet$	0.659	**0.906**†	0.670	0.670$^\bullet$	0.640	0.492$^\bullet$	**0.639**	**1.015**†	0.660$^\bullet$	**0.653**†	0.556$^\bullet$
Nc=3	0.577$^\bullet$	0.653	1.015$^\bullet$	**0.674**	**0.671**	0.661	**0.517**†	0.632	1.081$^\bullet$	**0.662**†	**0.653**†	**0.585**†
Nc=4	0.575$^\bullet$	0.634	1.129	0.667	0.663$^\bullet$	0.665$^\bullet$	0.475	0.599	1.304	0.653$^\bullet$	0.636	0.536
Nc=5	0.575$^\bullet$	0.635	1.130	0.668	0.663$^\bullet$	**0.666**†	0.476	0.599	1.305	0.653$^\bullet$	0.642	0.536
Nc=6	0.571$^\bullet$	0.628	1.166	0.670	0.662$^\bullet$	0.657$^\bullet$	0.468	0.582	1.419	0.647	0.631	0.532
Nc=7	**0.579**†	0.629	1.146	0.668	0.664$^\bullet$	0.665$^\bullet$	0.439	0.566	1.486	0.646	0.625	0.497
Nc=9	0.569$^\bullet$	0.616	1.321	0.658	0.655	0.661$^\bullet$	0.441	0.557	1.563	0.639	0.620	0.498
					Influence of joint distribution (Nc=4, Adults)							
	GBVS-based saccadic model						**RARE2012-based saccadic model**					
SVdist2yo	0.576	0.647	1.285	0.632	0.623	0.504	0.507	0.633	1.297	0.623	0.613	0.446
$p_B(d,\phi)=1$	0.497	0.667	1.106	0.631	0.621	0.430	0.362	0.630	1.323	0.603	0.589	0.318
NSV dist.	0.531	0.685	1.072	0.631	0.624	0.462	0.405	0.647	1.244	0.608	0.597	0.354

From the predicted scanpaths, we compute, for each age group and for both saliency models (i.e. GBVS and RARE2012), the 1D distribution of saccade amplitudes and the 2D joint distribution of saccade amplitudes and orientations. We evaluate the Kullback-Leibler (KL) divergence between these distributions and the distributions computed from eye tracking data. Fig. 6 plots the KL scores in function of the parameter N_c. We observe that the KL scores follow a U-shaped curve. The KL scores are higher for low and high values of N_c. A low value of N_c corresponds to high dispersion between observers whereas a high value reduces the randomness of the fixation point selection. The best KL scores are obtained for N_c in the range 4 to 6. More specifically, for each age group, we select the best N_c value in order to get the best compromise between salient area prediction and scanpath plausibility. For adults and 6-10 year-old groups, $N_c = 4$. For 4-6 and 2 year-old, $N_c = 5$.

Fig. 7 shows the distributions of saccade amplitudes (top row) and the joint distributions of saccade amplitudes and orientations for the age groups when considering the aforementioned values of N_c (bottom row)[3]. We observe that the distributions of saccade amplitudes computed with the proposed saccadic model have a similar shape when compared with actual distributions. We note, however, that the proposed model tends to generate larger saccades. The main peak of the predicted distributions is between 2 and 3 degrees of visual angle, whereas the main peak of actual distributions is about 2 degrees of visual angle. This discrepancy might be due to the computational modelling of the inhibition-of-return mechanism which does not entirely reflect the reality. A second explanation might be related to the computation of joint distributions, as well as how they are used. One of the strength of the proposed saccadic model is that we use spatially-variant joint distributions (see section V-C for more details). However, only 9 joint distributions are used to reproduce the gaze deployment, which might not be enough. Increasing this number would make sense but would require more fixation points in order to compute accurate and relevant distributions. Another concern pertains to the memory effect that is not taken into account. Indeed, there is a time dependency in saccade amplitudes. Small amplitude saccades tend to be followed by large amplitude saccades, which are followed by small ones [43], [70].

We also noticed that the key ingredient to produce plausible scanpath is not the input saliency map, as illustrated by Fig. 7. Although that GBVS and RARE2012 models generate saliency maps that have rather different saliency distributions, as illustrated by Fig. 5 (a), the saccadic model manages to produce plausible scanpaths in both cases.

The middle and bottom rows of Fig. 7 illustrate the joint distributions computed from GBVS-based scanpaths and RARE2012-based scanpaths, respectively. Compared to actual joint distributions shown in Fig. 4, we observe a similar evolution of the saccadic behavior. For the 2 year-old group, saccade amplitudes are rather small and isotropic. The horizontal bias as well as large saccades progressively appears with aging. The horizontal bias is very noticeable for adults groups.

[3]In supplementary material, more results are given, especially for low and high values of N_c.

C. Joint distribution influences

In this section, we discuss the influence of the joint distributions by comparing the performance of the proposed age-dependent saccadic model with those obtained by considering age-independent distribution, uniform joint distribution and spatially-invariant distribution. We perform these tests by considering the following setting: GBSV and RARE2012 model, adult groups and $N_c = 4$. We also emphasize that we do not need to consider different learning and training subsets for inferring the joint distribution of saccade amplitudes and orientations. Indeed, similarly to [21], [22], we observe a systematic tendency in visual deployment as soon as the population is homogeneous and watch similar stimuli.

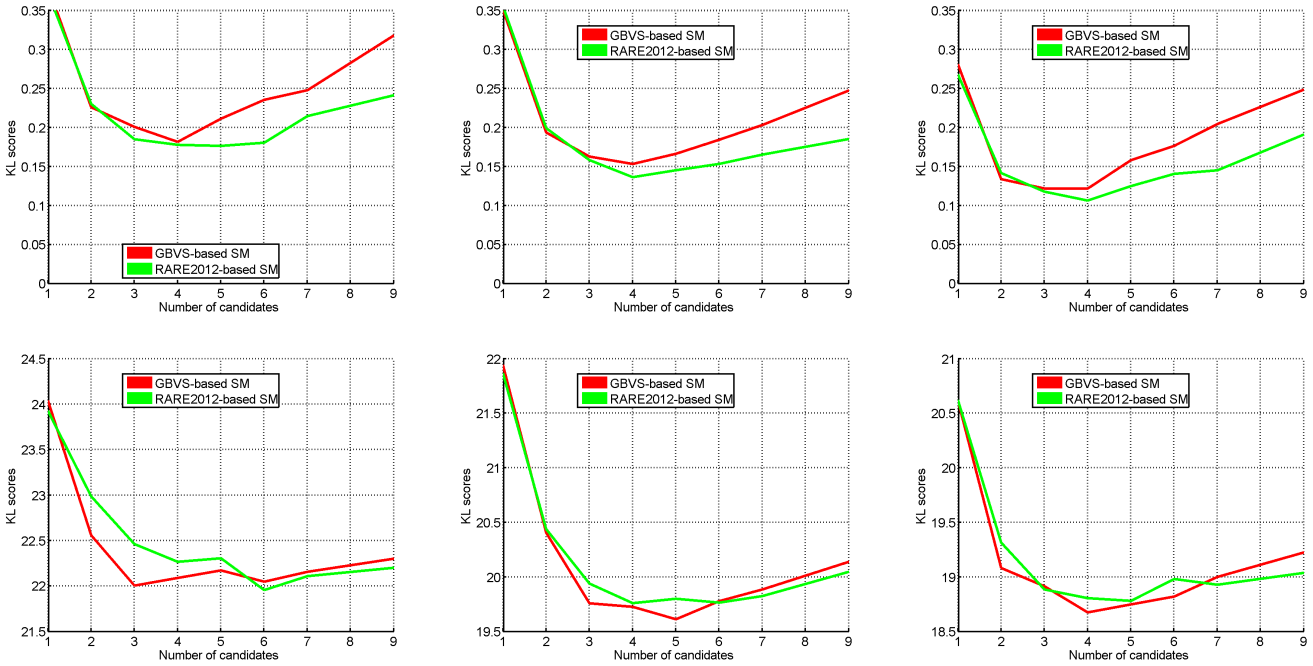
1) Age-dependent vs age-independent distribution: In this case, instead of using the spatially-variant joint distribution of adult group, we use the 2 y.o. spatially-variant joint distribution when computing adult scanpaths. We evaluate the performance of this modified model with the adult ground truth. Table II (bottom row called SVDist2yo) indicates that the ability to predict salient areas decreases when considering 2 y.o. distribution instead of adult one. In addition, when comparing the saccade amplitude distribution generated by this model (see Fig. 8 a)) with the best one (see top-right plot in Fig. 7), we observe that the predicted scanpaths are less plausible than those obtained with the model using adult distribution.

2) Uniform joint distribution: To further evaluate the influence of the joint distribution on the results, we set in equation 1, $p_B(d(x, x_{t-1}), \phi(x, x_{t-1})) = 1$, $\forall x \in \Omega$. In [71], Tatler and Vincent gave evidence that the viewing biases may be fundamental to predict where we look at. Results are presented in the bottom of Table II. As expected, the performances decrease, but they are still interesting. However, this solution does not allow us to generate plausible visual scanpaths as illustrated in Fig. 8 b).

3) Spatially-variant vs invariant joint distributions: As presented at the bottom of Table II, the use of spatially-variant joint distributions increases the performance of the saccadic model when compared to the saccadic model using a non spatially-variant joint distribution (see the acronym *NSV dist.*).

VI. Conclusion

In this paper, we show that saccadic models can be tailored for different age groups. Our saccadic model combines low-level salience, memory effects and viewing biases. Low-level salience is computed by state-of-the-art bottom-up saliency models. Memory effects represent the inhibition-of-return mechanism which performs the inhibition of an attended location in order to foster the scene exploration. The last component, i.e. viewing biases, provides fundamental information about how observers explore a visual scene. We show that these viewing biases evolve with the maturation of the visual system. We were able to capture differences in gaze behaviour between age groups with joint distributions of saccade amplitudes and orientations. This representation,



(a) 2 years-old (b) 4-6 years-old (c) Adults

Fig. 6. KL-divergence between the actual and the predicted distributions in function of the number of candidates N_c for the adults, 4-6 years-old and 2 years-old groups. Top row: the KL-divergence is computed between the actual distribution of saccade ampltiudes and the predicted one. Bottom-row: the KL-divergence is computed between the actual joint distribution of saccade amplitudes and saccade orientations and the predicted one.

which is learned from actual eye tracking data, turns out to be fairly different for 2 year-old, 4-6 year-old, 6-10 year-old and adult observers. By using this age-based visual signature, we showed that the proposed age-dependent saccadic model outperforms not only GBVS and RARE2012 saliency models but succeeds in generating scanpaths that match actual eye tracking data.

Obviously, the present saccadic model cannot fully account for the complex nature of overt visual attention. Although that the joint distribution of saccade amplitudes and orientations has a number of merits, it would be required to incorporate other known properties of gaze behavior, such as the fixation duration, the time dependencies between successive saccades and advanced scanpath statistics. These aspects will be tackled in future works.

This study may have a significant impact on some computer vision applications. For instance, it would allow us to tailor saliency-based image compression algorithms for observers of a specific age. Another example is related to image retargeting methods, which consist in reducing the image size while keeping the most visually important areas [20]. Most retargeting methods are based on importance maps that indicate the locations to preserve. Retargeting results could be improved by computing age-dependent importance map.

A side result of this work concerns the better understanding of the maturation of the visual system from childhood to adulthood, which could help to design new assistive applications for visually impaired people.

In supplementary material, a video sequence showing the maturation of eye movement behavior with respect to saccade amplitudes and orientations is provided. This video shows the influence of aging on saccade amplitudes and orientations, spanning from childhood to adulthood.

ACKNOWLEDGMENT

This work was supported in part 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, 2013.

[2] A. Coutrot, J. H. Hsiao, and A. B. Chan, "Scanpath modeling and classification with hidden markov models," *Behavior Research Methods*, pp. 1–18, 2017.

[3] S. Wang, M. Jiang, X. M. Duchesne, E. A. Laugeson, D. P. Kennedy, R. Adolphs, and Q. Zhao, "Atypical visual saliency in autism spectrum disorder quantified through model-based eye tracking," *Neuron*, vol. 88, no. 3, pp. 604–616, 2015.

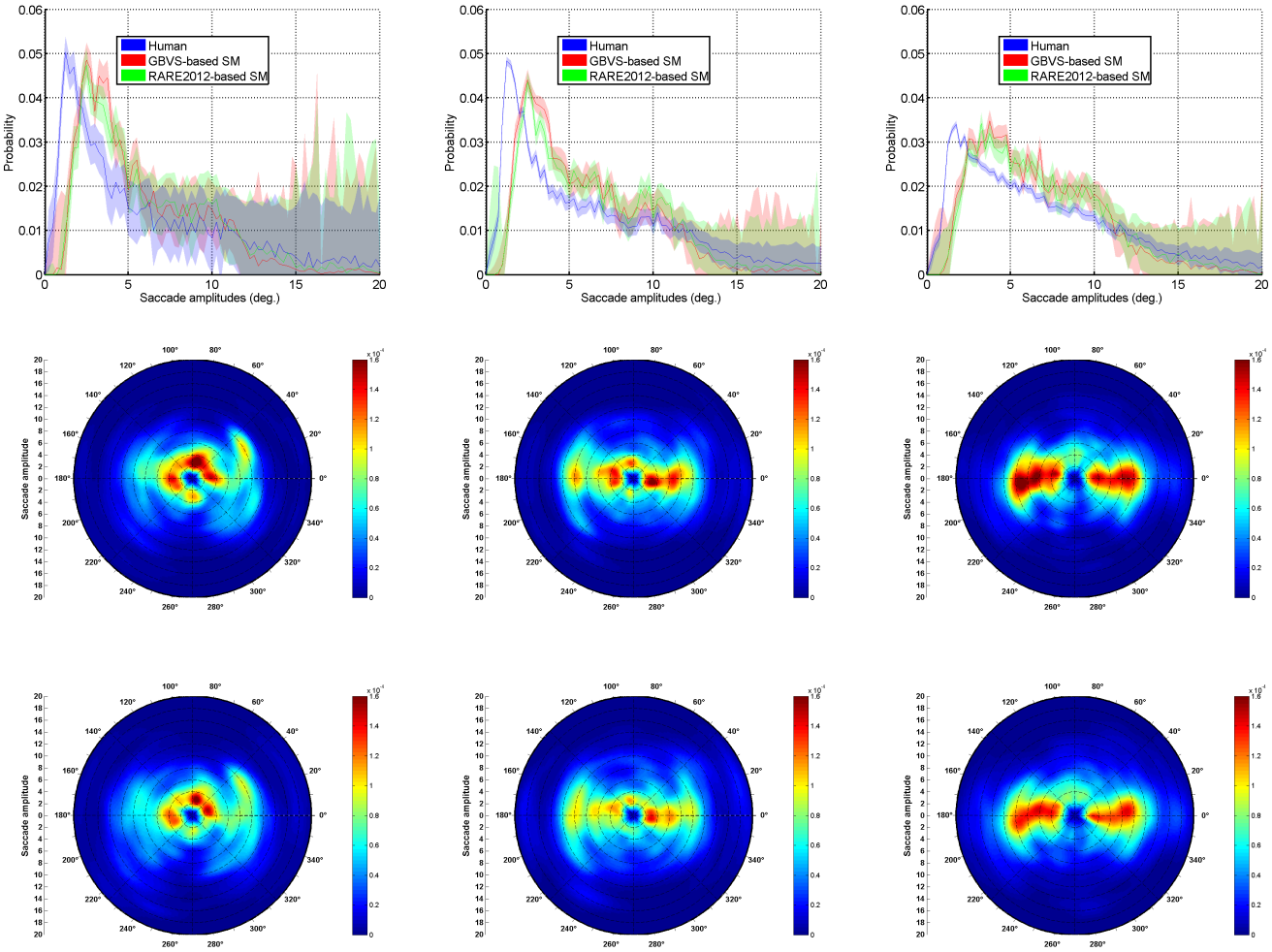
[4] P.-H. Tseng, I. G. Cameron, G. Pari, J. N. Reynolds, D. P. Munoz, and L. Itti, "High-throughput classification of clinical populations from natural viewing eye movements," *Journal of neurology*, vol. 260, no. 1, pp. 275–284, 2013.

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

[6] H. Tavakoli, , A. Atyabi, A. Rantanen, S. J. Laukka, S. Nefti-Meziani, J. Heikkilä *et al.*, "Predicting the valence of a scene from observers eye movements," *PloS one*, vol. 10, no. 9, p. e0138198, 2015.

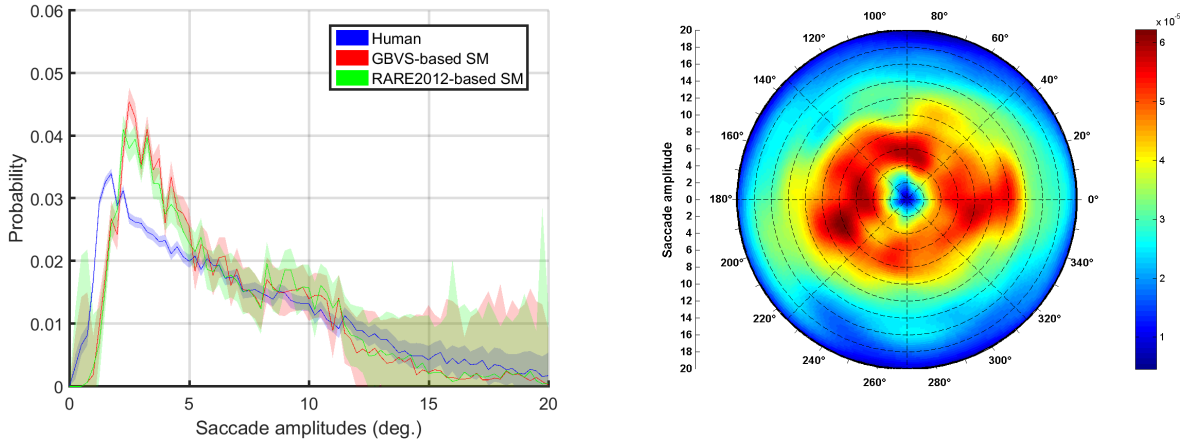
[7] P. Longhurst, K. Debattista, and A. Chalmers, "A gpu based saliency map for high-fidelity selective rendering," in *Proceedings of the 4th international conference on Computer graphics, virtual reality, visualisation and interaction in Africa*. ACM, 2006, pp. 21–29.

[8] A. Ninassi, O. Le Meur, P. Le Callet, and D. Barba, "Does where you gaze on an image affect your perception of quality? applying visual attention to image quality metric," in *2007 IEEE International Conference on Image Processing*, vol. 2. IEEE, 2007, pp. II–169.



(a) 2 years-old (b) 4-6 years-old (c) Adults

Fig. 7. Features of the predicted scanpaths. Top row: the actual and the predicted distributions of saccade amplitudes are plotted for $N_c = 5$, $N_c = 4$ and $N_c = 4$ corresponding to 2 year-old, 4-6 year-old and adults groups, respectively. Middle and bottom rows: joint distributions of saccade amplitudes and saccade orientations computed from GBVS-based saccadic model and RARE2012-based saccadic model, respectively.



(a) (b)

Fig. 8. (a) Actual and the predicted distributions of saccade amplitudes (see the difference with the top-right distribution in Fig. 7);(b) Joint distribution of saccade amplitudes and orientations when we do not consider viewing biases: $p_B(d(x, x_{t-1}), \phi(x, x_{t-1})) = 1$, $\forall x \in \Omega$. Note that the scale is not similar to previous used scales.

[9] H. Liu and I. Heynderickx, "Visual attention in objective image quality assessment: based on eye-tracking data," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 7, pp. 971–982, 2011.

[10] Z. Li, S. Qin, and L. Itti, "Visual attention guided bit allocation in video compression," *Image and Vision Computing*, vol. 29, no. 1, pp. 1–14, 2011.

[11] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 35, pp. 185–207, 2013.

[12] O. Le Meur and A. Coutrot, "How saccadic models help predict where we look during a visual task? application to visual quality assessment." *Electronic Imaging*, vol. 2016, no. 13, pp. 1–7, 2016.

[13] A. Yarbus, *Eye movements and vision*. Plenum Press: New York, 1967.

[14] R. Nisbett, *The geography of thought: how Asians and Westerners think differently... and why*. New York: Free Press, 2003.

[15] A. Miyahira, K. Morita, H. Yamaguchi, Y. Morita, and H. Maeda, "Gender differences and reproducibility in exploratory eye movements of normal subjects," *Psychiatry and clinical neurosciences*, vol. 54, no. 1, pp. 31–36, 2000.

[16] A. Coutrot, N. Binetti, C. Harrison, I. Mareschal, and A. Johnston, "Face exploration dynamics differentiate men and women," *Journal of Vision*, vol. 16, no. 14, pp. 16–16, 2016.

[17] 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, 2015.

[18] 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.

[19] H. L. Kirkorian and D. R. Anderson, "Anticipatory eye movements while watching continuous action across shots in video sequences: A developmental study," *Child Development*, 2016.

[20] M. Rubinstein, D. Gutierrez, O. Sorkine, and A. Shamir, "A comparative study of image retargeting," in *ACM transactions on graphics (TOG)*, vol. 29, no. 6. ACM, 2010, p. 160.

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

[22] 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.

[23] D. Brockmann and T. Geisel, "The ecology of gaze shifts," *Neurocomputing*, vol. 32, no. 1, pp. 643–650, 2000.

[24] G. Boccignone and M. Ferraro, "Modelling gaze shift as a constrained random walk," *Physica A: Statistical Mechanics and its Applications*, vol. 331, no. 1, pp. 207–218, 2004.

[25] W. Wang, C. Chen, Y. Wang, T. Jiang, F. Fang, and Y. Yao, "Simulating human saccadic scanpaths on natural images," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 441–448.

[26] H. Liu, D. Xu, Q. Huang, W. Li, M. Xu, and S. Lin, "Semantically-based human scanpath estimation with hmms," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 3232–3239.

[27] H. R. Tavakoli, E. Rahtu, and J. Heikkilä, "Stochastic bottom–up fixation prediction and saccade generation," *Image and Vision Computing*, vol. 31, no. 9, pp. 686–693, 2013.

[28] R. Engbert, H. A. Trukenbrod, S. Barthelmé, and F. A. Wichmann, "Spatial statistics and attentional dynamics in scene viewing," *Journal of vision*, vol. 15, no. 1, pp. 14–14, 2015.

[29] O. Le Meur and Z. Liu, "Saliency aggregation: Does unity make strength?" in *ACCV*, 2014.

[30] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. London: Chapman & Hall, 1986.

[31] J. Najemnik and W. Geisler, "Simple summation rule for optimal fixation selection in visual search," *Vision Research*, vol. 42, pp. 1286–1294, 2009.

[32] S. J. Gershman, E. Vul, and J. B. Tenenbaum, "Multistability and perceptual inference," *Neural computation*, vol. 24, no. 1, pp. 1–24, 2012.

[33] R. D. Luce, *Individual choice behavior: A theoretical analysis*. Courier Corporation, 2005.

[34] D. R. Wozny, U. R. Beierholm, and L. Shams, "Probability matching as a computational strategy used in perception," *PLoS Comput Biol*, vol. 6, no. 8, p. e1000871, 2010.

[35] W. Gaissmaier and L. J. Schooler, "The smart potential behind probability matching," *Cognition*, vol. 109, no. 3, pp. 416–422, 2008.

[36] R. F. West and K. E. Stanovich, "Is probability matching smart? associations between probabilistic choices and cognitive ability," *Memory & Cognition*, vol. 31, no. 2, pp. 243–251, 2003.

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

[38] 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.

[39] D. S. Wooding, "Fixation maps: quantifying eye-movement traces," in *Proceedings of the 2002 symposium on Eye tracking research & applications*. ACM, 2002, pp. 31–36.

[40] 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.

[41] Z. Botev, J. Grotowski, and D. P. Kroese, "Kernel density estimation via diffusion," *The annals of Statistics*, vol. 38, no. 8, pp. 2916–2957, 2010.

[42] T. Foulsham, A. Kingstone, and G. Underwood, "Turning the world around: Patterns in saccade direction vary with picture orientation," *Vision Research*, vol. 48, pp. 1777–1790, 2008.

[43] B. Tatler and B. Vincent, "Systematic tendencies in scene viewing," *Journal of Eye Movement Research*, vol. 2, pp. 1–18, 2008.

[44] 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.

[45] P. Viviani, A. Berthoz, and D. Tracey, "The curvature of oblique saccades," *Vision research*, vol. 17, no. 5, pp. 661–664, 1977.

[46] J. Peacock, "Two-dimensional goodness-of-fit testing in astronomy," *Monthly Notices of the Royal Astronomical Society*, vol. 202, no. 3, pp. 615–627, 1983.

[47] L. Itti, C. Koch, and E. Niebur, "A model for saliency-based visual attention for rapid scene analysis," *IEEE Trans. on PAMI*, vol. 20, pp. 1254–1259, 1998.

[48] A. Garcia-Diaz, X. R. Fdez-Vidal, X. M. Pardo, and R. Dosil, "Saliency from hierarchical adaptation through decorrelation and variance normalization," *Image and Vision Computing*, vol. 30, no. 1, pp. 51–64, 2012.

[49] O. Le Meur, P. Le Callet, D. Barba, and D. Thoreau, "A coherent computational approach to model the bottom-up visual attention," *IEEE Trans. On PAMI*, vol. 28, no. 5, pp. 802–817, May 2006.

[50] O. Le Meur, P. Le Callet, and D. Barba, "Predicting visual fixations on video based on low-level visual features," *Vision research*, vol. 47, no. 19, pp. 2483–2498, 2007.

[51] Y. Fang, W. Lin, Z. Chen, C.-M. Tsai, and C.-W. Lin, "A video saliency detection model in compressed domain," *IEEE transactions on circuits and systems for video technology*, vol. 24, no. 1, pp. 27–38, 2014.

[52] A. Coutrot and N. Guyader, "How saliency, faces, and sound influence gaze in dynamic social scenes," *Journal of Vision*, vol. 14, no. 8, p. 5, 2014.

[53] Y. Fang, Z. Chen, W. Lin, and C.-W. Lin, "Saliency detection in the compressed domain for adaptive image retargeting," *IEEE Transactions on Image Processing*, vol. 21, no. 9, pp. 3888–3901, 2012.

[54] J. Zhang and S. Sclaroff, "Saliency detection: A boolean map approach," in *Proceedings of the IEEE international conference on computer vision*, 2013, pp. 153–160.

[55] N. Bruce and J. Tsotsos, "Saliency based on information maximization," *Advances in neural information processing systems*, vol. 18, p. 155, 2006.

[56] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. IEEE, 2007, pp. 1–8.

[57] 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*, vol. 8, no. 7, pp. 32–32, 2008.

[58] X. Hou, J. Harel, and C. Koch, "Image signature: Highlighting sparse salient regions," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 1, pp. 194–201, 2012.

[59] N. Murray, M. Vanrell, X. Otazu, and C. A. Parraga, "Saliency estimation using a non-parametric low-level vision model," in *Computer vision and pattern recognition (cvpr), 2011 IEEE conference on*. IEEE, 2011, pp. 433–440.

[60] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Proceedings of Neural Information Processing Systems (NIPS)*. MIT Press, 2006.

[61] 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.

[62] T. Judd, F. Durand, and A. Torralba, "A benchmark of computational models of saliency to predict human fixations," MIT, Tech. Rep. MIT-CSAIL-TR-2012, 2012.
[63] A. Borji, H. R. Tavakoli, D. N. Sihite, and L. Itti, "Analysis of scores, datasets, and models in visual saliency prediction," in *2013 IEEE International Conference on Computer Vision*. IEEE, 2013, pp. 921–928.
[64] R. J. Peters, A. Iyer, L. Itti, and C. Koch, "Components of bottom-up gaze allocation in natural images," *Vision Research*, vol. 45, no. 18, pp. 2397–2416, 2005.
[65] 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*, vol. 22, no. 1, pp. 55–69, 2012.
[66] N. D. Bruce, C. Wloka, N. Frosst, S. Rahman, and J. K. Tsotsos, "On computational modeling of visual saliency: Examining whats right, and whats left," *Vision research*, vol. 116, pp. 95–112, 2015.
[67] A. Açık, A. Sarwary, R. Schultze-Kraft, S. Onat, and P. König, "Developmental changes in natural viewing behavior: bottom-up and top-down differences between children, young adults and older adults," *Frontiers in psychology*, vol. 1, p. 207, 2010.
[68] A. Mahdi, M. Su, M. Schlesinger, and J. Qin, "A comparison study of saliency models for fixation prediction on infants and adults," *IEEE Transactions on Cognitive and Developmental Systems*, 2017.
[69] H. Martinez, M. Lungarella, and R. Pfeifer, "Stochastic extension to the attention-selection system for the icub," *University of Zurich, Tech. Rep*, 2008.
[70] P. Unema, S. Pannasch, M. Joos, and B. Velichkovsky, "Time course of information processing during scene perception: the relationship between saccade amplitude and fixation duration," *Visual Cognition*, vol. 12, no. 3, pp. 473–494, 2005.
[71] B. Tatler and B. T. Vincent, "The prominence of behavioural biases in eye guidance," *Visual Cognition, Special Issue: Eye Guidance in Natural Scenes*, vol. 17, no. 6-7, pp. 1029–1059, 2009.

Olivier Le Meur obtained his PhD degree from the University of Nantes in 2005. From 1999 to 2009, he has worked in the media and broadcasting industry. In 2003 he joined the research center of Thomson-Technicolor at Rennes where he supervised a research project concerning the modelling of the human visual attention. Since 2009 he has been an associate professor for image processing at the University of Rennes 1. In the IRISA/SIROCCO team his research interests are dealing with the understanding of the human visual attention. It includes computational modelling of the visual attention and saliency-based applications (video compression, objective assessment of video quality, retargeting).

Antoine Coutrot holds a B.S., and M.S. degrees in Engineering from Grenoble Institute of Technology, France, 2011 and a Ph.D. degree in Cognitive Science from Grenoble-Alpes University, France, 2014. He currently works as a Postdoctoral Researcher at University College London, United Kingdom. His research interests include eye-tracking, visual attention, multimodal signal processing, machine learning, and human behaviour modelling in general.

Zhi Liu (M'07-SM'15) received the B.E. and M.E. degrees from Tianjin University, Tianjin, China, and the Ph.D. degree from Institute of Image Processing and Pattern Recognition, Shanghai Jiaotong University, Shanghai, China, in 1999, 2002, and 2005, respectively. He is currently a Professor with the School of Communication and Information Engineering, Shanghai University, Shanghai, China. From Aug. 2012 to Aug. 2014, he was a Visiting Researcher with the SIROCCO Team, IRISA/INRIA-Rennes, France, with the support by EU FP7 Marie Curie Actions. He has published more than 130 refereed technical papers in international journals and conferences. His research interests include image/video processing, machine learning, computer vision and multimedia communication. He was a TPC member in VCIP 2016, ICME 2014, WIAMIS 2013, IWVP 2011, PCM 2010, ISPACS 2010, etc. He co-organized special sessions on visual attention, saliency models, and applications at WIAMIS 2013 and ICME 2014. He is an area editor of Signal Processing: Image Communication and served as a guest editor for the special issue on Recent Advances in Saliency Models, Applications and Evaluations in Signal Processing: Image Communication.

Pia Rämä holds a Ph.D. in Cognitive Neuroscience from University of Helsinki, Finland, 1998. He currently works as a CNRS scientist in Laboratoire Psychologie de la Perception, at University Paris Descartes, France. Her research interests include semantic development, lexical acquisition, and scene perception in young children.

Adrien Le Roch holds a M.S. in Engineering and a M.S. in Computer Science from the university of Rennes 1, France, 2016. He currently works as a software engineer in Scalian Eurogiciel in Rennes, France.

Andrea Helo holds a M.S. degree in Neuroscience from University of Chile, 2010 and a Ph.D. degree in Cognitive Neuroscience from Paris Descartes University, France, 2017. She currently works as an Assistant professor at University of Chile. Her research interests include eye-tracking, EEG, visual attention development, language acquisition and vision and language interaction during development.