How saccadic models help predict where we look during a visual task? Application to visual quality assessment

Olivier Le Meur, Antoine Coutrot

How saccadic models help predict where we look during a visual task? Application to visual quality assessment.

Olivier Le Meur; IRISA/University of Rennes 1; Rennes, France
Antoine Coutrot; University College London; United Kingdom

Abstract

In this paper, we present saccadic models which are an alternative way to predict where observers look at. Compared to saliency models, saccadic models generate plausible visual scanpaths from which saliency maps can be computed. In addition these models have the advantage of being adaptable to different viewing conditions, viewing tasks and types of visual scene. We demonstrate that saccadic models perform better than existing saliency models for predicting where an observer looks at in free-viewing condition and quality-task condition (i.e. when observers have to score the quality of an image). For that, the joint distributions of saccade amplitudes and orientations in both conditions (i.e. free-viewing and quality task) have been estimated from eye tracking data. Thanks to saccadic models, we hope we will be able to improve upon the performance of saliency-based quality metrics, and more generally the capacity to predict where we look within visual scenes when performing visual tasks.

Introduction

The human brain is amazingly complex and involves a sheer number of biological mechanisms. The eyes, often compared to a *window to the soul*, are an exogenous manifestation revealing how and where we pay attention to our surroundings. Eye movements are driven by bottom-up and top-down mechanisms to sample our visual field and to select the most important areas. The bottom-up guidance source is classically represented by a saliency map which indicates the most informative parts of our visual field. Top-down contributions are related to observers' goals [1] but also to their prior knowledge, motivations, mood and experience.

Since the very first computational models of visual attention, which date back to the 1980s [2, 3, 4], tremendous progress has been made in a number of directions. Various models, more or less biologically plausible and using different mathematical tools, have been proposed. A large majority of them is based on the seminal work of [5] who proposed a plausible computational architecture for predicting where we look at. From a set of feature maps processed in a massively parallel manner, a single topographic saliency map encoding the ability of an area to attract our gaze is computed. More recently, a special effort has been made to present several comprehensive benchmark studies [6, 7, 8]. To carry out these benchmarks, many hurdles have also been overcome. For instance, a comprehensive benchmark requires to collect large amount of eye-tracking data in various experimental conditions [9, 10, 11]. Benchmarks also need clear definition and design of similarity metrics in order to objectively assess the degree of similarity between the ground truth and the predicted data [12, 13].

While the picture is much clearer than 20 years ago, there are still a number of issues that need to be addressed. In a recent paper, Bruce et al. [14] dress a list of obstacles that remain in visual saliency modelling. They also discuss important points that need to be addressed, including spatial bias, context, scene composition and oculomotor constraints. We believe that saccadic models provide a new framework able to cope with many of these challenges.

In this paper, we investigate how to adapt saccadic model to a specific scene category or to a specific task observers have to perform. In the first section, we detail the main components of saccadic models and explain why they constitute a good alternative to classic saliency models. We will emphasize how they can be tuned to reproduce the visual behaviour of observers engaged in a task. In the second section, we will focus more specifically on the visual quality task assessment, and show how saccadic models outperform state-of-the-art saliency models.

Saccadic model

The common denominator between state-of-the-art computational models of visual attention is that they all output a 2D static saliency map. This representation, although interesting and convenient for many computer vision applications, is quite limited and cannot grasp the complexity of our visual system. Rather than computing a unique 2D saliency map from an incoming image, saccadic models aim to reproduce the ability of our visual system to make time-dependent saccadic eye movements.

Presentation

Saccadic models predict where an observer look at. In addition they have to provide plausible visual scanpaths, i.e. scanpaths exhibiting distribution of saccade amplitudes or saccade orientations similar to humans'.

Ellis and Smith [15] pioneered in the design of saccadic model by elaborating a general framework for generating visual scanpaths. They used a stochastic process where the position of a fixation depends on the previous fixation, according to a first-order Markov process. Bearing this assumption in mind, Hacisalihzade et al. [16] designed a computational model predicting the visual scanpaths. Later, in 1998, a new breakthrough was done by combining the framework of [15] with saliency models [4, 17]. From a static saliency map, a scanpath is generated by using winner-take-all (WTA) algorithm and inhibition-of-return (IoR) scheme. Brockmann and Geisel [18] used a Lévy flight to simulate the scanpaths. Boccignone and Ferraro [19] extended Brockmann's work, and modeled eye gaze shifts by using Lévy flights constrained by the saliency.

More recently, Le Meur and Liu [20] extended Ellis and Smith's framework [15] to predict visual scanpaths of observers
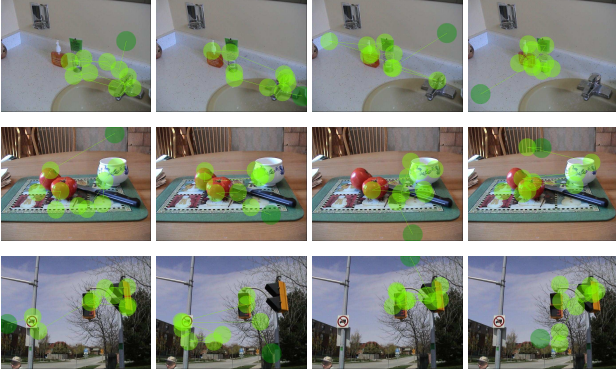
Figure 1. *Examples of simulated scanpaths. The dark green circle represents the first fixation point. Adapted from [20].*
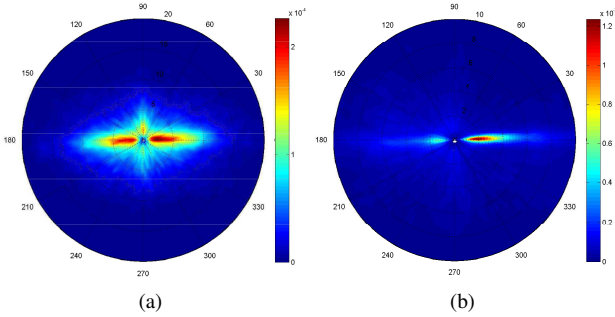

Figure 2. *Joint probability distribution of saccade amplitudes and orientations shown on a polar plot, (a) for natural scenes, (b) for webpages. Radial position indicates saccade amplitudes expressed in degree of visual angle. There is a marked tendency in fixation behaviour for making rather small horizontal saccades. For webpages, observers make horizontal small saccades in the rightward direction. Adapted from [20].*

while they freely watch a visual scene. Visual scanpaths are computed from both bottom-up saliency and viewing tendencies. Bottom-up saliency is represented by a saliency map whereas viewing tendencies (saccade amplitudes and saccade orientations) are inferred from public eye tracking datasets. Authors show that the predicted scanpaths exhibit similar trends to human eye movements in free-viewing condition. Predicted scanpaths are closer to human scanpaths than those generated by previous approaches. In addition, the saliency maps computed from the predicted scanpaths turn out to be more efficient than saliency maps computed by state-of-the-art saliency models. Figure 1 illustrates 4 predicted scanpaths, each composed of 10 fixations, on 3 different visual scenes.

In the following section, we further describe Le Meur and Liu's model and emphasize its ability to adapt to a specific scene category and tasks at hand.

Probabilistic framework for visual scanpath prediction

Let $I : \Omega \subset \mathcal{R}^2 \mapsto \mathcal{R}^m$ an input image ($m = 1$ for grayscale image and $m = 3$ for a color image) and $\mathbf{x}_t$ a fixation point at time t. To determine the next fixation point, we consider the 2D discrete conditional probability $p(\mathbf{x}|\mathbf{x}_{t-1}, \cdots, \mathbf{x}_{t-T})$ which indicates, for each location of the definition domain Ω, the transition prob-

ability between the past T fixations and the current location $\mathbf{x}$. This conditional probability $p(\mathbf{x}|\mathbf{x}_{t-1}, \cdots, \mathbf{x}_{t-T})$ depends on several key ingredients, that are listed below. However, we underline that this list is far from being exhaustive and may include other types of information. In addition, in order to limit the complexity of the method, the following components are supposed to be independent. The conditional probability $p(\mathbf{x}|\mathbf{x}_{t-1}, \cdots, \mathbf{x}_{t-T})$ is composed of:

- **A bottom-up saliency map:** this is a grayscale 2D map representing the saliency values for all locations $\mathbf{x} \in \Omega$. In [20] and in accordance with [21], this term is constant over time meaning that bottom-up influences do not vanish over time. However, to avoid any misunderstanding, if bottom-up influences are rather constant over time, this is not the case for top-down contributions which increase over time, leading to higher inter-observer variability;
- **A joint probability distribution of saccade amplitudes and orientations:** this joint probability represents the way our gaze is deployed within a scene. Figure 2 illustrates the polar plot of the joint probability distribution of saccade amplitudes and orientations computed from data collected on natural scenes (a) and on webpages (b) while observers freely look onscreen images. The saccade amplitude is the distance (expressed in degree of visual angle) between two fixation points $\mathbf{x}_t$ and $\mathbf{x}_{t-1}$. The orientation is the angle between these two points and is expressed in degree;
- **The memory state of the location x at time t:** The memory map indicates whether a location $\mathbf{x}$ can be re-fixated or not. This is a time-dependent term that simulates the inhibition of return and indicates the probability to re-fixate a given location. More details are given in [20].

Another element of consideration is the stochastic behaviour of the oculomotor system. It means that, given a set of parameters, an image and a history of recent fixation locations, the subsequent fixation cannot be entirely specified. Najemnik and Geisler [22, 23] proposed the Bayesian ideal searcher. It consists in determining the location $\mathbf{x}_t$ of the subsequent fixations by maximising the conditional probability $p(\mathbf{x}|\mathbf{x}_{t-1}, \cdots, \mathbf{x}_{t-T})$ given a set of known information. This model, unlike the human saccadic behavior, is deterministic. One solution to get a stochastic behavior is to pick up the next fixation point among N_c random locations and to select the candidate for which the saliency gain is the most important [20]. Note that the N_c random locations are generated according the conditional probability $p(\mathbf{x}|\mathbf{x}_{t-1}, \cdots, \mathbf{x}_{t-T})$. The parameter N_c controls the extent to which the model is stochastic: the model is deterministic for $N_c = 1$, whereas the variability between predicted scanpaths is maximal when $N_c \longrightarrow +\infty$.

One key advantage of saccadic model is its ability to adapt to various conditions. Figure 2 illustrates that eye movements pattern strongly depend on the scene content. On natural scene, the distribution is elongated on the horizontal axis. There are few vertical saccades and very few diagonal saccades. On webpages, we observe a strong tendency for making horizontal small saccades in the rightward direction. It simply indicates that the scanning of webpages is performed in a raster scan order [24]. Amongst the numerous factors that influence our visual attention, the task at hand is obviously one of the most important, as described by [1]. To illustrate this point, eye data collected by Ehinger et al. [25]
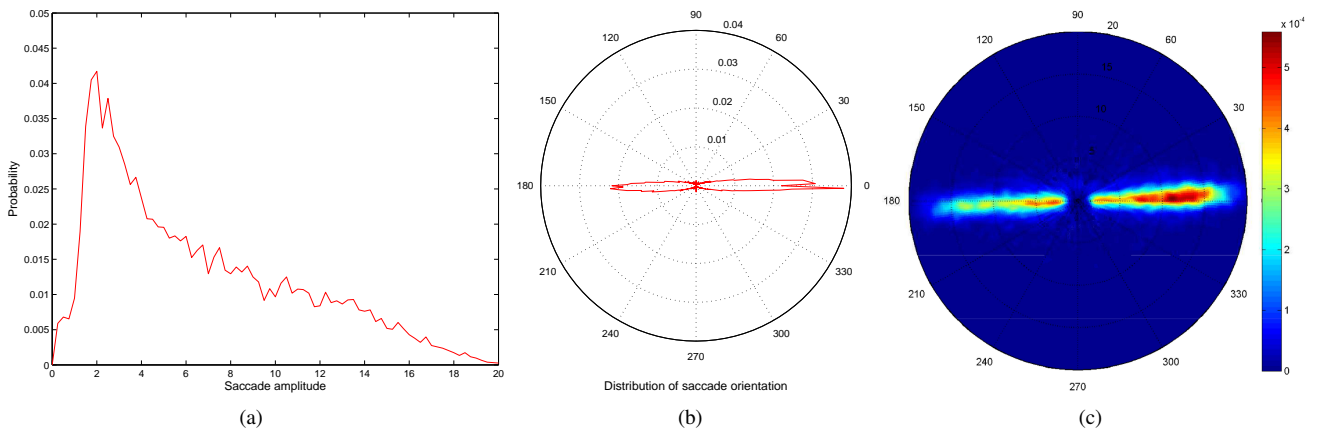
Figure 3. *(a) Distribution of saccade amplitudes; (b) Distribution of saccade orientations; (c) joint probability distribution of saccade amplitudes and orientations when observers look for people in an urban scene. Eye data collected by Ehinger et al. [25] are here used to demonstrate the influence of the task on the gaze deployment.*

are analysed. In this experiment, observers were instructed to decide as quickly as possible whether a person was present in urban images. Given these instructions, observers adapted their visual strategy to optimize the visual search. Figure 3 illustrates the distribution of saccade amplitudes (a), the distribution of saccade orientations (b), and the joint distribution of saccade amplitudes and orientations (c). This latter is clearly marked by the task at hand. Observers scan horizontally the scenes and make rather large horizontal saccades. Consequently, if we want to predict successfully human eye fixations in the context of this task, the saccadic model has to take into account the strategic mechanism deployed by observers for performing this task.

More generally, if we know the type of the scene, the task at hand, or any other high-level information, the behaviour of the saccadic model can be adapted in order to produce relevant visual scanpaths. In the following section, we investigate the strategic mechanisms used by observers while they have to score the quality of an image. Saccadic model of [20] is then adapted to predict the quality task gaze behaviour.

Tuning saccadic model for quality assessment task

The visual strategy deployed by observers for scoring the quality of an image is a complex process, involving both bottom-up and top-down factors. Although that the relative contributions of these factors are still an open question, many studies made the assumption that the performance of quality metric could be significantly improved by combining bottom-up saliency maps and distortion maps. The underlying idea is simple: an artefact appearing on a salient area would contribute more to the quality score than an artefact appearing on non salient areas. For most saliency-based metric [26, 27, 28, 29, 30], the use of saliency map consists in modifying the pooling strategy. The degree of saliency of a given pixel is used as a weight, giving more or less importance to the error occurring on this pixel location. In this section, we show that, instead of using a saliency model, it makes more sense to use a saccadic model fitted to the particular context of quality assessment.

To reach this objective, we use a subset of the eye tracking

dataset presented in [31]. This dataset is composed of 10 high-quality images that have been impaired by JPEG2000, with five quality levels, going from r_1 (i.e. smallest degradation) to r_5 (i.e. highest degradation). Eye movements have been recorded in free-viewing task for the ten original images and in quality task for the degraded images. The viewing duration was 8 seconds, and twenty observers were involved in the experiment. The standardized method DSIS (Double Stimulus Impairment Scale) was used to perform the subjective quality evaluation. More details on the experimental protocol are given in [31].

Gaze behaviour in free-viewing and quality task

The collected eye data are analysed for similarities of eye movements between the following configurations:

- FT: eye movements are recorded while observers freely look at unimpaired images (FT);
- Ref. QT: eye movements are recorded while observers look at the unimpaired images during the quality task (QT);
- Deg. QT: eye movements are recorded while observers look at the degraded images during the quality task (QT).

Figure 4 presents the plots of saccade amplitudes (a), saccade orientations (b) and the joint distribution of saccade amplitudes and orientations (c). We observe that observers performed smaller saccades in the condition Ref. QT compared to FT and Deg. QT conditions. The condition in which the saccades are the longest is Deg. QT. One could interpret these observations in the following way: an increase of the saccade amplitudes, as in the condition Deg. QT, could be referred to as an increase of visual exploration. This higher visual exploration could be used to look for local degradations. Unlike the condition Deg. QT, the visual exploration is less important in Ref. QT. Observers make small saccades probably to memorize small parts of the scenes. These observations are consistent with [31]; authors indeed noticed that the fixation durations were significantly longer in Ref. QT than in FT and Deg. QT. The condition Ref. QT could then be associated to a focal process (i.e. long fixation duration and short saccade), and the condition Deg. QT would be an ambient process (i.e. short fixation duration and long saccade) [32, 33].
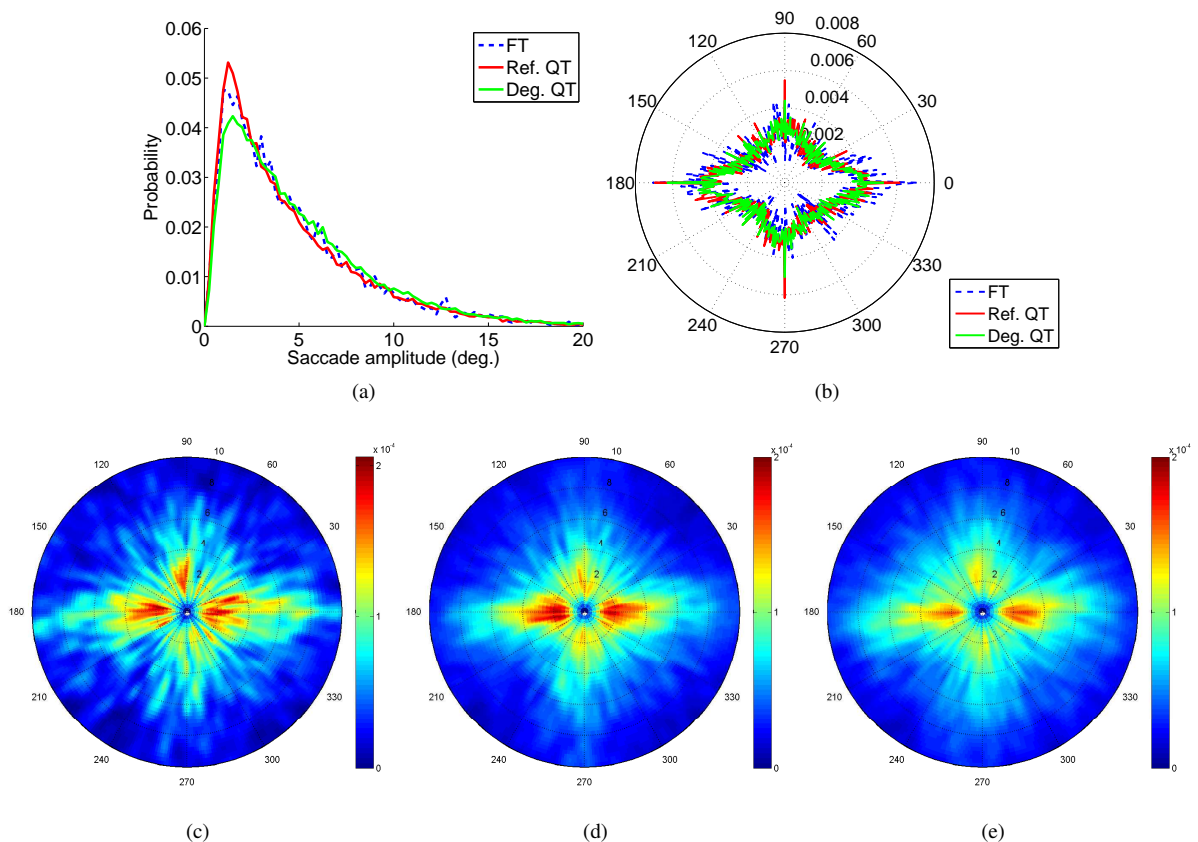
Figure 4. *Human saccadic behavior while observers freely view stimuli or perform a quality task. (a) Distribution of saccade amplitudes; (b) distribution of saccade orientations; (c), (d) and (e) represent the joint probability distribution of saccade amplitudes and orientations for reference images in free-viewing task (FT), for reference images in quality task (Ref. QT) and for the degraded images in quality task (Deg. QT), respectively.*

This categorization of visual fixations has been initially proposed by [33]. Focal fixations would be related to the processing of details and identification of object. Ambient fixations are likely to be involved in the processing of spatial arrangement, gist and scene's context. FT condition is in between Ref. QT and Deg. QT conditions. The polar plots illustrating the joint distribution of saccade amplitudes and orientations (Figure 4 (c) to (e)) confirm these observations. The joint distribution of the condition Deg. QT (Figure 4 (e)) is less focussed than the two others, suggesting higher visual exploration. Regarding the saccade orientations, the distributions as shown in Figure 4 (b) have similar shapes. We just observe more vertical saccades in the condition Ref. QT than in the two other conditions.

Performance

In this section, we evaluate the ability of saccadic model to predict where observers look at while free-viewing images or scoring their quality (i.e. quality task). We also compare the proposed saccadic model's performance to a purely bottom-up model of visual attention. We have chosen the model RARE2012, which is one of best state-of-the-art model [8].

Test procedure

For the free-viewing condition, the saccadic model takes in input: the RARE2012 saliency map computed on the original im-

age and the joint distribution of saccade amplitudes and orientations estimated from eye data collected in free-viewing (see figure 4 (c)).

For the quality task condition, the input saliency map is computed on the degraded image. Unlike the FT condition, the joint distribution of saccade amplitudes and orientations is estimated from eye data collected in quality task (see figure 4 (e)).

The degree of similarity between predicted and human saliency maps is evaluated by the linear correlation coefficient [13]. To be more specific, in the context of free-viewing, human saliency maps are estimated from eye movements collected in free-viewing whereas, in the context of quality task, the human saliency map are computed from eye data coming from the quality task. Figure 6 presents the different saliency maps used in the test: (a) illustrates the original image (top) and JPEG2000 encoded images (*level* $= r_1$ and *level* $= r_5$, respectively); (b) are the corresponding saliency maps, where the bright areas correspond to the salient areas; (c) are the saliency maps computed by RARE2012 model; (d) and (e) are the saliency maps computed from the proposed saccadic models. We test three values of N_c: $N_c = 1$ means that the randomness is maximal. Increasing N_c reduces the randomness. This is reflected by the saliency maps that are more or less focussed. For a small value of N_c, the dispersion between predicted scanpaths is high (see figure 6 (d)). When N_c increases, the saliency maps are getting more focussed (see
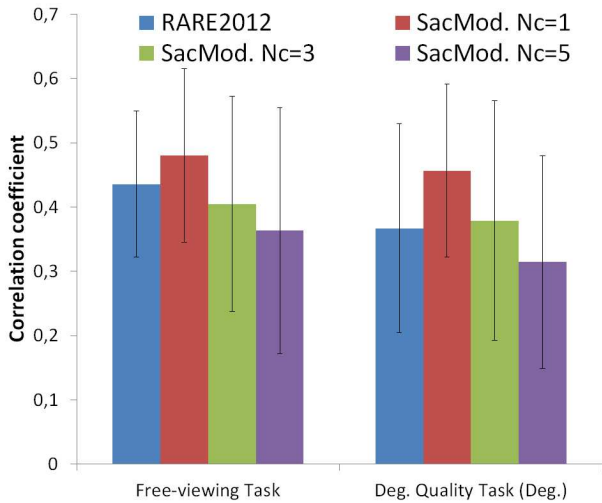
Figure 5. *Performance of RARE2012 and saccadic models (SacMod.) in free-viewing (right-hand side) and quality task conditions (left-hand side). Three values of N_c are tested. The error bars indicate the standard deviation.*

figure 6 (e)).

Results

Figure 5 presents the performance in term of linear correlation of RARE2012 model [8] and saccadic model. First, we observe that RARE2012 model performs better in the free-viewing task than in the quality task (when the degraded image is used). The linear correlation coefficient drops down from 0.43 to 0.36, respectively. There is, at least, one reason that might explain this drop in performance: RARE2012 model is not disturbed by JPEG2000 artefacts and provides similar saliency maps, whatever the level of impairment is (see figure 6 (c) middle and bottom images). However, the human saliency map is different because of the task at hand, i.e. the bottom-up contribution is likely less important during the quality task. The robustness of saliency models to visual degradations [34] is an advantage in free-viewing context. Unfortunately, this property could be seen as a disadvantage in the context of quality assessment.

Regarding the performance of the proposed saccadic model, the prediction is better than RARE2012 model, for $N_c = 1$ in both conditions. The gain is statistically significant in the condition Deg. QT (paired t-test $p < 0.05$). In the condition FT, the small number of original images (10) does not allow to get statistically significant gain. Results also show that the saccadic model performs similarly in both conditions, although that the predicted saliency maps are slightly different (see figure 6 (d)). Finally, the best performance are obtained when N_c is equal to 1. We remind that the amount of randomness is maximal when $N_c = 1$. When N_c increases, the saccadic model is less stochastic. In this case, the predicted saliency maps are more focussed as illustrated by Figure 6 (e).

Conclusion

In this paper, we demonstrate how we could tune saccadic model for better predicting where an observer look at during a visual task. The viewing bias of observers performing a visual task is inferred from eye data. Thanks to this understanding and description (e.g. joint distributions of saccade amplitudes and orientations, the variability between observers, etc), saccadic models can better reproduce how overt visual attention is allocated. For instance, we noticed that the gaze deployment differs substantially when observers free viewed images, looked for pedestrians and scored the quality of images.

The ability of saccadic model to reproduce the visual behaviour of an observer depends to a large extent on the quality of the viewing bias description. The use of the joint distribution of saccade amplitudes and orientations is a first step in this direction. However, considering saccades as independent events might be a strong limitation. Indeed, rather than assuming that saccades are independent events, it would probably be more accurate to consider that a saccade can be influenced by the preceding saccades. In addition, the fixation durations might play an important role as well. We will investigate these avenues in future works.

References

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

[2] J. J. Clark and N. J. Ferrier, "Modal control of an attentive vision system," in *ICCV*, pp. 514–523, IEEE, 1988.

[3] J. K. Tsotsos, S. M. Culhane, W. Y. Kei W., Y. Lai, N. Davis, and F. Nuflo, "Modeling visual attention via selective tuning," *Artificial intelligence* **78**(1), pp. 507–545, 1995.

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

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

[6] T. Judd, F. Durand, and A. Torralba, "A benchmark of computational models of saliency to predict human fixations," in *MIT Technical Report*, 2012.

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

[8] 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* **28**(6), pp. 642 – 658, 2013.

[9] A. Coutrot and N. Guyader, "An audiovisual attention model for natural conversation scenes," in *Image Processing (ICIP), 2014 IEEE International Conference on*, pp. 1100–1104, IEEE, 2014.

[10] A. Borji and L. Itti, "Cat2000: A large scale fixation dataset for boosting saliency research," *CVPR 2015 workshop on "Future of Datasets"*, 2015. arXiv preprint arXiv:1505.03581.

[11] 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* **28**, pp. 802–817, May 2006.

[12] H. Jarodzka, K. Holmqvist, and K. Nystrom, "A vector-based, multidimensional scanpath similarity measure," in *ETRA*, pp. 211–218, ACM, 2010.

[13] O. Le Meur and T. Baccino, "Methods for comparing scanpaths and saliency maps: strengths and weaknesses," *Behavior Research Method* **45**(1), pp. 251–266, 2013.

[14] 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*, 2015.
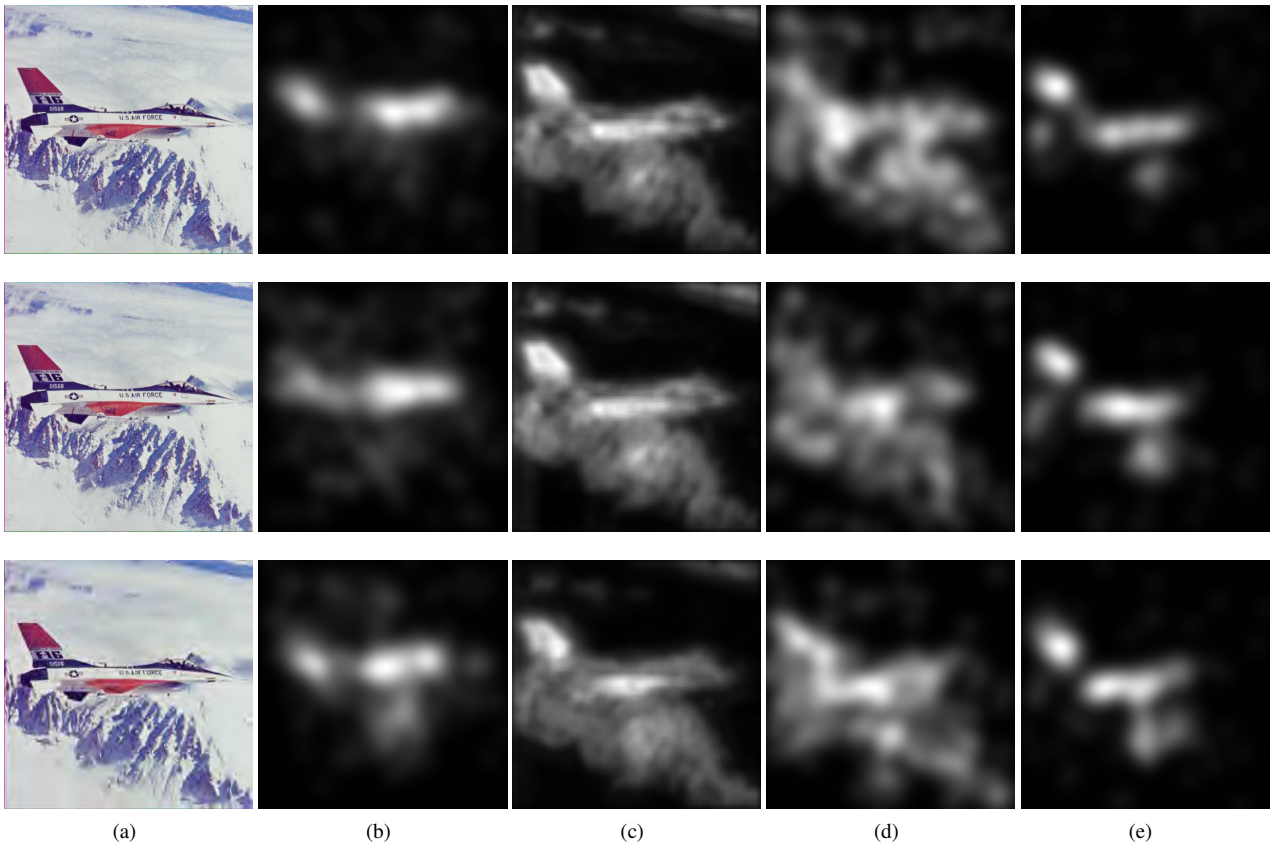
Figure 6. *(a) Original image (top to bottom: unimpaired, small JPEG2000 compression; high JPEG2000 compression) (b) Human saliency map (top: free-task; Middle,Top: quality-task); (c) RARE2012 saliency maps. (d) and (e) saliency map computed by the proposed saccadic model, for $N_c = 1$ and $N_c = 5$, respectively.*

[15] S. R. Ellis and J. D. Smith, *Patterns of statistical dependency in visual scanning*, ch. Eye Movements and Human Information Processing, pp. 221–238. Elsevier Science Publishers BV, (eds) Amsterdam, North Holland Press, 1985.

[16] S. S. Hacisalihzade, L. W. Stark, and J. S. Allen, "Visual perception and sequences of eye movement fixations: A stochastic modeling approach," *Systems, Man and Cybernetics, IEEE Transactions on* **22**(3), pp. 474–481, 1992.

[17] L. Itti and C. Koch, "A saliency-based search mechanism for overt and covert shifts of visual attention," *Vision Research* **40**, pp. 1489–1506, May 2000.

[18] D. Brockmann and T. Geisel, "The ecology of gaze shifts," *Neurocomputing* **32**(1), pp. 643–650, 2000.

[19] G. Boccignone and M. Ferraro, "Modelling gaze shift as a constrained random walk," *Physica A: Statistical Mechanics and its Applications* **331**, pp. 207 – 218, 2004.

[20] O. Le Meur and Z. Liu, "Saccadic model of eye movements for free-viewing condition," *Vision research* **1**(1), pp. 1–13, 2015.

[21] B. Tatler, R. J. Baddeley, and I. Gilchrist, "Visual correlates of fixation selection: effects of scale and time," *Vision Research* **45**, pp. 643–659, 2005.

[22] J. Najemnik and W. Geisler, "Optimal eye movement strategies in visual search," *Nature* **434**, pp. 387–391, 2005.

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

[24] G. Buscher, E. Cutrell, and M. R. Morris, "What do you see when you're surfing? using eye tracking to predict salient regions of web pages," in *Proceedings of CHI 2009*, Association for Computing Machinery, Inc., April 2009.

[25] K. A. Ehinger, B. Hidalgo-Sotelo, A. Torralba, and A. Oliva, "Modelling search for people in 900 scenes: A combined source model of eye guidance," *Visual cognition* **17**(6-7), pp. 945–978, 2009.

[26] 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 *ICIP*, 2007.

[27] A. Ninassi, O. Le Meur, P. Le Callet, and D. Barba, "Considering temporal variations of spatial visual distortions in video quality assessment," *IEEE Journal of Selected Topics in Signal Processing, Special Issue On Visual Media Quality Assessment* **3**(2), pp. 253 – 265, 2009.

[28] 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* **21**(7), pp. 971–982, 2011.

[29] A. Guo, L. Zhao, D. ans Shaohui, X. Fan, and W. Gao, "Visual attention based image quality assessment," in *IEEE International Conference on Image Processing*, pp. 3297–3300, 2011.

[30] W. Y. L. Akamine and M. C. Q. Farias, "Incorporating visual attention models into image quality metrics," in *VPQM*, 2012.

[31] A. Ninassi, O. Le Meur, P. Le Callet, D. Barba, and A. Tirel, "Task impact on the visual attention in subjective image quality as-

sessment," in *Signal Processing Conference, 2006 14th European*, pp. 1–5, IEEE, 2006.

[32] B. Follet, O. Le Meur, and T. Baccino, "New insights into ambient and focal visual fixations using an automatic classification algorithm," *i-Perception* **2**(6), pp. 592–610, 2011.

[33] 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* **12**(3), pp. 473–494, 2005.

[34] O. Le Meur, "Robustness and repeatability of saliency models subjected to visual degradations," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*, pp. 3285–3288, IEEE, 2011.

Author Biography

Olivier Le Meur is associate Professor at the University of Rennes 1 (Ecole Supérieure d'Ingénieurs de Rennes). He is a member of the team SIROCCO, joint team between IRISA and INRIA. Olivier Le Meur got a Ph.D. from University of Nantes in 2005 and an HDR (Habilitation à Diriger des Recherches) of the University of Rennes 1 in 2014. Before joining the University in 2009, he was a project leader at the research center of Technicolor R&D. He is the author of more than 50 peer-reviewed publications in international conferences and journals in the field of image processing (IEEE TIP, IEEE PAMI, ICIP), computer vision (ECCV, ACCV) and applied perception (Vision Research). His expertise is in the field of image processing and cognitive sciences. In terms of research related to the proposal, Olivier Le Meur is strongly involved in the computational modelling of visual attention [11, 20] and models evaluation [13].

Antoine Coutrot holds an Engineering degree and a PhD in Cognitive Science from Grenoble-Alpes University (France). During his PhD, he studied the influence of sound on the visual exploration of dynamic scenes. Antoine is currently a Research Fellow at CoMPLEX, University College London (UK), where he develops statistical models and datamining approaches to understand how high and low level features guide our gaze.