Computers & Graphics xxx (2018) xxx–xxx

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Technical Section

Exploring visual attention and saliency modeling for task-based visual analysis

Patrik Polatsek[a], Manuela Waldner[b], Ivan Viola[b], Peter Kapec[a], Wanda Benesova[a,*]

[a] Faculty of Informatics and Information Technologies, Slovak University Of Technology in Bratislava, Slovak Republic
[b] Faculty of Informatics, Vienna University of Technology, Austria

ARTICLE INFO

Article history:
Received 3 October 2017
Revised 26 January 2018
Accepted 26 January 2018
Available online xxx

Keywords:
Information visualization
Eye-tracking experiment
Saliency
Visual attention
Low-level analytical tasks

ABSTRACT

Memory, visual attention and perception play a critical role in the design of visualizations. The way users observe a visualization is affected by salient stimuli in a scene as well as by domain knowledge, interest, and the task. While recent saliency models manage to predict the users' visual attention in visualizations during exploratory analysis, there is little evidence how much influence bottom-up saliency has on task-based visual analysis. Therefore, we performed an eye-tracking study with 47 users to determine the users' path of attention when solving three low-level analytical tasks using 30 different charts from the MASSVIS database [1]. We also compared our task-based eye tracking data to the data from the original memorability experiment by Borkin et al. [2]. We found that solving a task leads to more consistent viewing patterns compared to exploratory visual analysis. However, bottom-up saliency of a visualization has negligible influence on users' fixations and task efficiency when performing a low-level analytical task. Also, the efficiency of visual search for an extreme target data point is barely influenced by the target's bottom-up saliency. Therefore, we conclude that bottom-up saliency models tailored towards information visualization are not suitable for predicting visual attention when performing task-based visual analysis. We discuss potential reasons and suggest extensions to visual attention models to better account for task-based visual analysis.

1. Introduction

Visualization designers use a large variety of visual channels to effectively encode data. Due to the limited computational capacity of the brain, parsing and interpreting these visual channels cannot be performed rapidly. Instead, visual attention is serially directed to different regions in the visualization, and the information is gradually decoded. Visual attention is a set of cognitive processes that selects relevant information and filters out irrelevant information from the environment [3]. Attention is driven by both bottom-up and top-down factors. The aim of exogenous and very rapid *bottom-up* attention is to warn of impending danger. It is guided by low-level salient visual features which stand out from their neighborhood (the so-called "pop-out effect"), such as intensity or color contrasts, texture and motion. Visual channels used in information visualizations are perceived either with specialized receptors of the human visual system (e.g. red-green opponency [4], orientation or spatial frequency [5]) or by multiple receptors in the case of complex channels, such as shape. In contrast to this stimulus-driven attention, endogenous and much slower *top-down* attention is biased by cognitive factors. It involves prior knowledge, expectations, tasks and goals that enhance bottom-up attention. Real scene *perception*, referring to the organization and interpretation of sensory information, lies in the interaction between bottom-up and top-down processing of attention [6].

When users interpret visualizations, top-down factors of attention are incorporated in scene perception. Visual search is an important component of the process of interpreting visualizations. It is the process of finding a specific target object in a scene among non-targets. Visual attention thereby guides the user's gaze and the visual search, respectively. Understanding visual attention is therefore essential for selecting appropriate visual channels and designing effective visualizations.

Computational saliency models have been developed to predict users' visual attention (see Section 2). These models are quite accurate for simple stimuli and natural images [7–9]. While saliency models have also been used as a quality measure in the information visualization domain [10,11], it has been shown that these models' accuracy for predicting visual attention in visualizations

* Corresponding author.
E-mail addresses: patrik.polatsek@stuba.sk (P. Polatsek), waldner@cg.tuwien.ac.at (M. Waldner), viola@cg.tuwien.ac.at (I. Viola), peter.kapec@stuba.sk (P. Kapec), vanda_benesova@stuba.sk (W. Benesova).

https://doi.org/10.1016/j.cag.2018.01.010
0097-8493/

is significantly poorer than for natural images [12]. Matzen et al. [13] therefore recently introduced a saliency model tailored towards information visualization, and showed that fixations during exploratory visual analysis could be predicted more accurately.

While it has been shown that bottom-up factors captured by this new model have a strong influence on users' visual attention during exploratory visual analysis, it is still unknown how strong top-down guidance influences attention during task-based visual analysis. This work investigates human gaze behavior and saliency prediction when performing typical low-level analytical tasks with visualizations. To this end, we performed an eye-tracking study with charts from the MASSVIS database [1]. During this study, users solved three different low-level analytical tasks. We compared the data of this study with eye tracking data of the memorability experiment [2] with conditions closer to natural image viewing. We could show that fixations are more coherent between users, but correlate less with the visualization-tailored saliency map when performing a low-level analytical task. We discuss some potential extensions of saliency models to incorporate these added top-down factors.

2. Visual attention models

In recent decades, various attention models have been proposed that differ in how they predict human visual attention. As pioneers, Itti et al. [4] defined a computational bottom-up saliency model using local center-surround differences of intensity, color and orientation features at multiple spatial scales. This approach of feature extraction has been adopted in many attention models. Harel et al. [14] also followed this approach. Their model computes saliency using graph-based dissimilarity measures. Hou et al. [15] introduced a model analyzing the frequency domain instead of the spatial domain to predict saliency. The bottom-up model presented by Zhang and Sclaroff [16] is based on the principle of figure-background segregation. The model identifies figures in a set of Boolean maps generated by thresholding feature maps. A work presented by Bruce and Tsotsos [17] defines saliency as the self-information of visual features of the image. Zhang et al. [18] proposed a Bayesian framework that incorporates top-down information dependent on the target's features with bottom-up saliency that is represented as the self-information derived from the statistics of natural images. Goferman et al. [19] proposed saliency detection based on patches with unique low-level features, visual organization rules according to which regions close to salient pixels are also salient and high-level factors, such as human faces. Vig et al. [20] and Cornia et al. [21] introduced saliency models that employ neural networks to predict fixations.

Visual saliency predicted by computational models can be applied in many areas of computer science including image processing [22–24], computer graphics [25], robotics [26,27], surveillance systems [28,29] and human-computer interaction [30,31]. Saliency models have been widely evaluated against different datasets that usually contain natural scenes and fixations from free viewing [7,32–35]. The benchmarks [36–38] show for some image databases a small difference between the state-of-the-art models and human inter-observer (IO) that outputs a fixation map integrated from other subjects than the one under test. The map serves as an upper bound for prediction accuracy. Generally, the prediction accuracy of the models is higher for simple images with few salient objects. However, predicting human fixations in complex images with multiple objects is a challenging task [39,40].

The models are commonly used to predict where the observer's attention is directed in natural images. However, they have also been used in visualization research to predict attention and to derive quality measures, respectively. For instance, Jänicke and Chen [11] describe a saliency-based quality metric for visualizations. It compares a saliency map using the cognitive model by Itti et al. [4] and importance of visualized data, generated automatically from data or manually by visualization designers. The metric is then computed as the difference of these maps. Attention in dynamic geovisualizations was studied by Gagarlandini and Fabrikant [10]. Saliency of dynamic visualizations was predicted by the spatiotemporal model by Itti et al. [41]. The highest saliency value was predicted in regions of the change. Saliency was also applied in volume visualizations to guide attention to selected regions [42]. Saliency was first determined for each voxel, and was then enhanced by center-surround operations between voxels inspired by the standard cognitive saliency model [4].

These works are based on the assumption that saliency models, originally developed for natural image viewing, are equally accurate for predicting the attention when interpreting visualizations. However, there are some notable differences between natural images and classic charts used in information visualization. Graphical marks, such as dots or lines, are usually abstract and simple compared to complex objects in natural images. Also, the background is mostly uniform and the visualizations contain a lot of textual information, such as labels and legends. Graphical marks and visual channels are chosen by a visualization designer according to design guidelines and visualization domain knowledge with the goal to expressively and effectively represent the underlying data. Thereby, visualization designers choose their visual channels to maximize the amount of information displayed [43]. Matzen et al. [13] also note that most saliency models tend to omit fine-grained visual features, like thin lines, which are highly relevant for information visualization.

Therefore, special variations of saliency models have been developed for information visualization. Lee et al. [44], for instance, introduced a saliency model for categorical map visualizations. They define point saliency as color difference of each point against its neighborhood. The class visibility quantifies the point saliency values that correspond to a given category. Most relevant for our work, Matzen et al. [13] proposed a novel saliency model that combines the model of Itti et al. with text saliency to predict saliency in data visualization with higher precision. The presented work evaluated saliency models on the MASSVIS database [1]. The results highlight the importance of text in visual attention since the model that relies only on text saliency outperforms classic saliency models in most evaluation metrics.

In our work, we will compute all above mentioned saliency models for the visualizations used in our experiment and compute the correlations to the obtained fixations from our eye tracking data.

3. Related work

To explore the applicability of saliency models beyond natural images, Borji et al. [7] compared the performance of four saliency models to eye tracking data obtained during free viewing of 20 different image categories, including abstract patterns and line drawings. In their study, saliency models predicted fixations surprisingly well for sketches. Matzen et al. [45] compared fixation maps of novices and professional analysts when analyzing synthetic aperture radar imagery to Itti et al.'s [4] saliency model. They showed that fixation maps of novices were more correlated with the saliency maps, compared to those of the professionals. They concluded that novices are much more likely to be directed by bottom-up salient features than experienced users.

Haass et al. [12] compared the performance of three different saliency models between the cat2000 dataset [7] and the MASSVIS dataset from Borkin et al.'s memorability experiment [2] using eight different comparison metrics. They found that saliency models performed worse for information visualizations than for the