Landscape and Urban Planning 148 (2016) 17–26

Contents lists available at ScienceDirect

Landscape and Urban Planning

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

Comparing saliency maps and eye-tracking focus maps: The potential use in visual impact assessment based on landscape photographs

Lien Dupont*, Kristien Ooms, Marc Antrop, Veerle Van Eetvelde

Geography Department, Ghent University, Krijgslaan 281, building S8, Ghent, B9000, Belgium

HIGHLIGHTS

- Saliency maps are reliable predictions of the human visual attention distribution.
- Non-salient objects provide an optimal visual integration into the landscape.
- Saliency maps can also be used to identify high visual impact designs (landmarks).

ARTICLE INFO

Article history:
Received 18 May 2015
Received in revised form 4 December 2015
Accepted 12 December 2015
Available online 6 January 2016

Keywords:
Correlation analysis
Viewing pattern
Visual conspicuity
Urbanisation
Visual landscape integration
Landscape design

ABSTRACT

In this study, we analyse how well saliency maps, which are theoretical predictions of the human viewing pattern, are correlated with human focus maps, obtained by tracking 42 observer's eyes while free-viewing landscape photographs ranging from rural to urban environments. The Pearson's correlation coefficient was calculated on the predicted and measured pixels' greyscale values. A relatively high correlation was obtained, indicating that the saliency maps can be used as reliable predictions of the human observation pattern and thus can predict which elements in a landscape will catch the attention. These findings are useful in visual impact assessment, a step in the planning process which is often not well elaborated or even skipped. Saliency maps could, for instance, be used to compare the conspicuity of different designs of a construction when simulated in photographs of the original landscape. As the visual impact of an object is reduced when its visual perception decreases, the least salient design will also have the lowest visual impact and will correspond to the best integration into the existing landscape. This method is easy and produces an objective measure of the degree of visual impact. However, as slight differences in correlation depending on the degree of urbanisation of the landscape were found, this methodology will not be equally reliable in all types of landscapes. Predictions of the viewing pattern in rural landscapes with a limited amount of buildings have been demonstrated to be most reliable. In more urbanised landscapes this reliability slightly decreases but nevertheless remains significant.

1. Introduction

When observing visual scenes, the resulting eye movements are not simply a set of random fixations. Instead, the fixations will exhibit a specific pattern (Humphrey & Underwood, 2009). The selection of locations to be fixated takes place according to a specific strategy, embedded in the human nervous system (Harel, Koch, & Perona, 2012). As it would be computationally too demanding to process the massive amount of incoming sensory information all the time, the nervous system constantly decides which parts of the available information will be selected for further, more detailed processing and which parts will be skipped. In addition, the selected parts are ranked by priority. The most important parts will be processed first, less important ones will follow later. This process is called 'selective attention'. As attention to an object is necessary for it to be perceived consciously (Harel et al., 2012), only a small part of the incoming information will thus reach visual awareness (Crick & Koch, 1998; Desimone & Duncan, 1995). This means that when observing images, attention will be allocated only to a limited part of the image. Two main aspects influence how the attention is distributed: the content of the scene (bottom-up, low-level process) and the

* Corresponding author. Tel.: +32 9 264 46 99; fax: +32 9 264 49 85.
E-mail addresses: Lien.Dupont@UGent.be (L. Dupont), Kristien.Ooms@ugent.be (K. Ooms), Marc.Antrop@UGent.be (M. Antrop), Veerle.Vaneetvelde@UGent.be (V. Van Eetvelde).

http://dx.doi.org/10.1016/j.landurbplan.2015.12.007
0169-2046/

cognitive characteristics of the observer (top-down, high-level process; Rajashekar, van der Linde, Bovik, & Cormack, 2008). While the fast bottom-up mechanism is always operating—although stronger in free-viewing situations—the top-down mechanism predominantly comes into effect when performing tasks (Borji, Sihite, & Itti, 2013; Land & Hayhoe, 2001; Navalpakkam & Itti, 2005; Parkhurst, Law, & Niebur, 2002; Rajashekar et al., 2008; Yarbus, 1967).

In the particular case of landscapes, bottom-up processes will mainly drive the observation as people usually observe scenes freely and without a task in mind (Dupont, Antrop, & Van Eetvelde, 2014). Consequently, the distribution of fixations will be primarily guided by the content of the visual stimulus (e.g., landscape photographs). Of particular interest in this situation are saliency maps, which can be described as computationally generated focus maps, which encode for conspicuity or salience at each location in an image in a purely bottom-up fashion (Itti, Koch, & Niebur, 1998; Itti & Koch, 2000; Itti, 2005). Salience or saliency is defined as the distinct perceptual quality by which an item in the world stands out from its neighbours and therefore immediately catches the attention (Itti, 2007). A feature's salience is calculated based on its colour, orientation, and intensity information compared to its surround (Itti et al., 1998; Itti & Koch, 2000; Itti & Koch, 2001; Koch & Ullman, 1985; Peters, Iyer, Itti, & Koch, 2005). Objects which are in sharp contrast with or incongruent to their surroundings will thus 'pop out' in the saliency map and can be identified. This technique might be useful in landscape planning, architecture and design, and in particular in visual impact assessments of new projects—e.g., buildings, roads, bridges etc.—for estimating how well different scenarios are visually integrated in the surrounding landscape. As the visual impact of a new construction or modification is associated with its contrast with the background landscape, saliency maps obtained for different visualisations of the project can be used to objectively quantify these contrasts. As highly contrasting elements have been shown to capture people's attention (Itti, 2007), this measure can be used to assess the visual impact of a construction. However, before this method can be used and applied—which will not be done in this paper—empirical evidence of a substantial correlation between saliency maps of landscape scenes and focus maps, obtained from real observers who viewed the scenes, is required to demonstrate the validity of using saliency maps as predictions of the human viewing pattern in landscape photographs (which is the purpose of this study). This validity is very likely as eye movements have been demonstrated to be attracted to salient regions (Itti & Koch, 2000; Itti, 2005; Koch & Ullman, 1985). In fact, the similitude between saliency maps and human observation patterns has been confirmed in several studies (Harel et al., 2012; Humphrey & Underwood, 2009; Peters et al., 2005). However, for landscape photographs in particular this similarity has not yet been investigated thoroughly, while this analysis is an important first step in investigating the potential of saliency maps for objectively predicting a viewer's attention distribution in a landscape image and thus for identifying where and when objects are more likely to have a strong visual impact.

In this paper, we perform this analysis by investigating how well saliency maps approximate human focus maps when free-viewing landscape photographs by examining the correlation between both. As such, we check whether saliency maps can be used as reliable predictions of the viewing pattern in landscape visualisations and thus if they are usable for visual impact assessments. In addition, we examine if the result of this analysis is equal in different types of landscapes, ranging from rural settings to urban environments. This is of particular interest as the degree of urbanisation of a landscape has been demonstrated to have an effect on the observation pattern (Dupont, Antrop, & Van Eetvelde, 2015; Dupont, Ooms, Duchowski, Antrop, & Van Eetvelde, 2015).

2. Methods

2.1. Theoretical background of saliency

Saliency is solely based on the bottom-up attentional process (Itti et al., 1998), which is a fast and stimulus-driven mechanism (Parkhurst et al., 2002). In particular, for each pixel in the image the salience is calculated based on its colour, orientation, and intensity information compared to its surrounding (Itti & Koch, 2000; Itti & Koch, 2001; Koch & Ullman, 1985). As such, each pixel of the original image is ascribed a scalar value which indicates its salience (Itti, 2005; Peters et al., 2005). As the human eye tends to be attracted by salient objects in the visual environment (Itti, 2005), attention will first be attracted by the most salient region in the stimulus, i.e., the brightest area with the highest colour contrast and orientation change, then by the second most salient region etc. (Humphrey & Underwood, 2009). This guidance of the eye is completely driven by bottom-up mechanisms (Itti et al., 1998; Malcolm & Henderson, 2010). Shifting attention away from these regions will thus require voluntary top-down 'effort' (Itti & Koch, 2000; Itti & Koch, 2001) in order to surpass the bottom-up mechanisms of attention stemming from the characteristics of the visual stimulus (Nothdurft, 2005; Treisman & Gelade, 1980). This slower top-down process, determined by cognitive phenomena driven by the observer's expectations or intentions (Parkhurst et al., 2002), typically comes into play when performing tasks (Borji, Sihite, et al., 2013; Land & Hayhoe, 2001; Navalpakkam & Itti, 2005; Rajashekar et al., 2008; Yarbus, 1967), although the bottom-up guidance mechanism can never be completely ruled out (Parkhurst et al., 2002). As in free-viewing no tasks are involved, saliency maps have been especially successful in predicting fixations when free-viewing images (Foulsham & Underwood, 2008; Parkhurst et al., 2002; Peters et al., 2005;). For a mixture of images, a high correlation between saliency and human fixations has been confirmed in a number of recent studies (e.g., Borji, Sihite, et al., 2013; Humphrey & Underwood, 2009; Parkhurst et al., 2002; Peters et al., 2005).

2.2. Subjects

Forty-two subjects voluntarily participated in the eye-tracking experiment. They were given brief practical information about the test but no details were revealed with respect to the purpose of the study in order to avoid influencing their viewing pattern in advance. A mix of females (24) and males (18) aged between 22 and 65 was obtained. When applicable, the participants were asked to wear contact lenses instead of glasses if possible because otherwise the eye-tracker could erroneously lock onto the dark parts of the glasses instead of onto the pupil. For the same reason, mascara was prohibited. Before starting the test, the participants were asked about any aberrations of their eyes. The 42 selected subjects all had normal or corrected-to-normal vision.

2.3. Stimuli

As we are investigating how people observe landscapes, we use terrestrial landscape photographs in the eye-tracking test. This is allowed since numerous authors have confirmed the validity of using photographs as surrogates for real landscapes (e.g., Palmer & Hoffman, 2001; Zube, Simcox, & Law, 1987). In addition, performing the test in situ has many drawbacks of which the time consumption, the high cost and the difficulty in controlling the settings of the experiment are the most important.

The photographs were taken following a strict routine to allow an unbiased comparison between them. First, all photographs were taken with the same camera and have a resolution of 3888 × 2592 pixels. Second, the focal length of the objective was kept constant