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

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.

© 2015 Elsevier B.V. All rights reserved.

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
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 similarity 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 (Notthdurft, 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 cannot 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...
at 50 mm in order to obtain equal visual angles (±31° × 21°). Third, a tripod was used to assure a constant shot height of 1.70 m. Fourth, the horizon was always placed at the same height in the photograph (2/3 of land, 1/3 of sky). Finally, all the photographs were taken in the same season to assure consistency about the condition of the foliage. The represented landscapes, 74 in total and ranging from rural to urban environments, are situated in Belgium and the north of France.

2.4. Eye-tracking apparatus

The eye-tracking experiment was performed in the Eye-tracking lab of the Department of Geography of the University of Ghent. A RED250 eye-tracking device, developed by SMI (Senso Motoric Instruments), was used to record the gaze pattern of the participants while observing the landscape images. This is possible as the eye-tracking technique consists of sending infrared light into the pupil of the observer (Duchowski, 2007). The reflected signal then provides information about the exact location of the point-of-regard on the screen (when calibrated) (Jacob & Karn, 2003; Poole & Ball, 2005). As such, all the stationary gaze positions (fixations) and interconnecting eye movements (saccades) are recorded (Poole & Ball, 2005). In this study, the threshold for determining when a position is stationary, and thus for defining a fixation, was set at 100 milliseconds in accordance with Inhoff and Radach (1998). Afterwards, this data can be ‘replayed’ and visualised on the observed image to gain insight into which areas in the image received attention. During the experiment, the participants were seated 60–80 cm (depending on the optimal calibration position) in front of a 22-inch colour monitor on which the photographs were displayed. Both eyes were tracked at a measurement rate of 120 Hz, which is equal to 120 measurements per second.

2.5. Procedure

The eye-tracking experiment consisted of free-viewing 74 landscape photographs, shown for 10 s each. The display order of the photographs was randomized to avoid the occurrence of effects originating from a fixed order. The participants were given no specific tasks but attentively observing the images. Free-viewing was chosen for two major reasons. First, we wanted to reproduce real life outdoor landscape observation conditions, which generally does not imply any tasks (Dupont et al., 2014). In addition, free-viewing most closely approximates natural viewing conditions (Parkhurst et al., 2002), which is what we aimed at. Second, the purpose of the study—comparing the viewing pattern of the participants with the prediction of the saliency map—requires free-viewing conditions. As saliency is based solely on bottom-up mechanisms of attention, the presence of top-down influences on the viewing pattern of the participants would make a proper comparison impossible. Although complete suppression of top-down influence cannot be achieved (Borji, Shiite, et al., 2013; Mannan, Kennard, & Husain, 2009), free-viewing reduces the task dependent top-down effects on eye movements to a minimum (Parkhurst et al., 2002).

Before starting the test, all the participants were given the same instruction text. After reading the instructions, a 9-dot calibration was performed to assure accurate measurements over the entire screen. After each image, the participants were asked to fixate a dot in the middle of the screen so that deviations from the initial calibration could be detected. When necessary a recalibration was performed. The drift correction dot also provided consistency on the starting point of the observation of each photograph.

2.6. Classification of photographs based on the degree of urbanisation

The ranking of the photographs, done by the participants, was obtained by using the Q-sort method like presented by Pitt and Zube (1979). This task was performed after the eye-tracking test in order to avoid biasing the observation pattern, which could occur when seeing the photographs for the second time. The participants were asked to sort the photographs depending on the amount of built area present in the landscape in order to obtain 5 classes of urbanisation. Therefore, the photographs were all presented on a desk. First, the participants had to pick out the 12 landscape photographs that were least characterised by built content. Second, the 12 photographs in which contained a maximum of built area were selected. This two-step procedure was repeated for the remaining 50 photographs but this time 16 photographs had to be selected each time. Finally, the last 18 photographs formed the last urbanisation class. This procedure provided 5 classes of urbanisation labelled as follows: Rural, Semi-rural, Mixed, Semi-urban, and Urban landscapes (Fig. 1). This classification was validated by objectively calculating the percentage of urbanised area in each photograph and comparing this to the score obtained from the Q-sorting. A correlation analysis confirmed a strong correlation between both variables (correlation coefficient of 0.959, $P<0.001$) (see Dupont, Ooms, et al. (2015) for further details). In consequence, the classification can be used for statistical comparisons.

Once the sorting task was completed by all the participants, scores were assigned to each urbanisation class as depicted in Table 1.

Subsequently, the average score across participants was calculated for each photograph. Based on this value, the photographs were assigned to one of the five classes of urbanisation. This final classification, however, could not be effectuated unequivocally as a number of photographs seemed to be in the middle between two classes (scores close to 1.5, 2.5, etc.). Assigning these images to one of the two classes would be very arbitrary, which could bias the results. Consequently, all the photographs of which the score indicated doubt, were removed from the analysis, leaving 10 photographs in each urbanisation class, 50 in total.

<table>
<thead>
<tr>
<th>Urbanisation class</th>
<th>Rural</th>
<th>Semi-rural</th>
<th>Mixed</th>
<th>Semi-urban</th>
<th>Urban</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of photographs</td>
<td>12</td>
<td>16</td>
<td>18</td>
<td>16</td>
<td>12</td>
</tr>
<tr>
<td>Score</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>
the observed areas and obscure the unwatched areas. However, the presence of colours stemming from the original photograph in the focus map would not allow a proper comparison with the greyscale saliency map. Therefore, the original image was replaced by a white image. As a result, greyscale focus maps were obtained (3108 in total) with colour values ranging from 0 (black) to 255 (white) consistent with the saliency maps.

2.7.2. Comparison of focus maps with saliency maps

In order to be able to compare the focus maps of the participants with the theoretical saliency maps (for each photograph, 42 focus maps (one for each participant) was compared with the corresponding saliency map for that photograph (74 in total)) a number of operations were needed (Fig. 3). First, the focus and saliency maps (.jpg-images) were transformed into text-files (.txt) containing the values, which define the greyscale colour of each pixel. This was executed in ArcGIS 10.1 using the conversion command Raster to ASCII. The result is a 1050 × 1680 matrix of values for each focus and saliency map (their resolution differs from the resolution of the original photographs as the eye-tracking software automatically downscales all original and processed images to 1050 × 1680 images). Second, these matrices were rearranged into one column per image, working from left to right and starting with the first row of the matrix, then the second etc. During this operation, the average value per two adjacent pixels was calculated and stored in the final column. As a result, the column contained 882,000 records in total instead of the 1,764,000 (1050 × 1680) records when all pixels would have been included. This number, however, was too elevated to be handled properly and quickly by the SPSS software (see Section 3). In addition, this ‘downgrading’ is allowed since the accuracy of the eye-tracker is 0.5°, which corresponds to 54 pixels at a viewing distance of 60 cm. Consequently, averaging 2 pixels will not significantly affect the analysis. This is confirmed by an analysis of the distribution of the differences in value across the averaged pairs: for the saliency maps, 86.6% of the pairs had the same value and 99.6% had a difference of 1, while 83.1% of the pairs in the focus maps were equal and 99.7% had a difference of maximum 5. These differences are negligible considering that the values vary between 0 and 255. Finally, the datasets—one per photograph, 74 in total—necessary for the comparison were obtained by aggregating the columns of the focus map of each participant (42 columns) with the column of the corresponding saliency map (1 single column). This resulted in one table per photograph consisting of 43 columns (42 columns of focus values and 1 column of saliency values), each containing 882,000 records. As such each of the 3108 focus maps could be compared to the corresponding saliency map.

As the aim of the study is to analyse how close the viewing pattern of the participants is to the predicted saliency map, a comparison between both was performed based on the above-mentioned datasets. In particular, the correlation between each focus map (3108 in total) and its corresponding saliency map was determined by calculating the Pearson correlation coefficient, which is a statistic often used to compare human focus maps and saliency images (e.g., Borji, Tavakoli, Sihite, & Itti, 2013; Haass, Matzen, McNamara, & Czuchlewski, 2015; Overhani, Von Wartburg, Hügli, & Müri, 2003; Rajashekar, Cormack, & Bovik, 2004; Rajashekar et al., 2008). The overall correlation coefficient (based on 74 photographs) was calculated as well as the correlation coefficient of each of the five urbanisation classes (based on the 5 unequivocally classified photographs). Since raw correlation coefficients are not additive and thus average values cannot be computed, a Fisher’s Z transformation was performed onto the correlation coefficients (Sheskin, 2003). To check if significant differences in correlation occur between these classes a Kruskal–Wallis test in combination with a Dunn’s test was performed in SPSS. In addition, the

2.7. Data analysis

2.7.1. Creating the saliency maps and focus maps

For each photograph, a saliency map (Fig. 2) was created in Matlab using the GBVS (Graph-based Visual Saliency) algorithm as developed by Harel, Koch, and Perona (2006) and provided on http://www.klab.caltech.edu/~harel/share/gbvs.php (Harel, 2012). Out of other possible saliency algorithms the GBVS algorithm was chosen because it has been demonstrated to yield the highest correlation coefficients over datasets consisting of landscape scenes amongst other images (Borji, Tavakoli, Sihite, & Itti, 2013). As a result, it is assumed that for this kind of images, human predictions are more reliably predicted by the GBVS than by other algorithms (Harel et al., 2006).

The focus maps, based on the eye movements registered during the experiment, were created in BeGaze, the software package provided with the SMI eye-tracker (Fig. 2). This was achieved for each participant for each photograph (3108 focus maps in total). Generally, the focus map is projected onto the original image to highlight

Fig. 2. (a) Original landscape photograph, (b) Saliency map of the photograph, (c) Example of a focus map based on the fixations made when observing this photograph.
correlation between the Fisher’s Z values (correlation between focus maps and saliency maps) and the percentage of urbanised area (square root to obtain a normal distribution) was computed for the 74 images.

3. Results

The average Pearson correlation coefficient (after performing a Fisher transformation) over all photographs and all participants is 0.410 and was found to be significant (P<0.01), which indicates a medium positive correlation between the human focus maps and the theoretical saliency maps. The correlation coefficients calculated for the five urbanisation classes separately are all significant as well (P<0.01). Furthermore, significant differences in

correlation between the five classes were found (P<0.01) (see Fig. 4 and Table 2). In particular, the highest correlation coefficients are found for the semi-rural landscape photographs (0.530, P<0.01), it subsequently decreases for mixed landscapes (0.476, P<0.01) and semi-urban landscapes (0.391, P<0.01) to reach a minimum for the urban landscape photographs (0.327, P<0.01). Thus, when disregarding the rural landscapes, there seems to be a trend of decreasing correlation when the degree of urbanisation in the landscapes increases. This is also reflected in the correlation between the Fisher’s Z values and the square root of the percentage of urbanised area, which was found to be −0.470 (P<0.01) for all images and −0.557 (P<0.01) when the rural category images are excluded (as this category does not follow the linear relationship). This means that when the proportion of buildings in the

### Table 2

Results of the Kruskal–Wallis (ranks) and Dunn’s test per photograph type. N gives the number of observations. (A Fisher transformation was applied to the Pearson’s correlation coefficients.)

<table>
<thead>
<tr>
<th></th>
<th>Rural</th>
<th>Semi-rural</th>
<th>Mixed</th>
<th>Semi-urban</th>
<th>Urban</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson’s correlation coefficient 2100 N</td>
<td>1082</td>
<td>1440</td>
<td>1249</td>
<td>877</td>
<td>604</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>Mean rank per landscape type</td>
<td>Real mean values per landscape type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rural</td>
<td>Semi-rural</td>
<td>Mixed</td>
<td>Semi-urban</td>
<td>Urban</td>
<td></td>
</tr>
<tr>
<td>Pearson’s correlation coefficient 2100</td>
<td>0.438</td>
<td>0.530</td>
<td>0.476</td>
<td>0.391</td>
<td>0.327</td>
<td></td>
</tr>
</tbody>
</table>
4. Discussion

4.1. Validation of the methodology

Our analysis generates correlation coefficients, which are in the same order of magnitude as the ones reported by Borji et al. (2013) who provide an overview of the prediction performance of different saliency algorithms. While we found a mean correlation of 0.410, Borji et al. (2013) mention correlation coefficients varying between 0.280 and 0.450 for the GBVS. However, the images in this study did not solely consist of landscape photographs but of a mixture of landscape photographs, images of indoor environments and portrait images. Furthermore, the correlation coefficient as calculated by Borji et al. (2013) differs from the approach exhibited in this study. Borji et al. (2013) determine the correlation between the saccade frequency at each location and the corresponding saliency, while our correlation is calculated between the pixel information, consisting of the greyscale value, in both focus and saliency maps. As focus maps are calculated based on fixations, the greyscale value is related to the fixation density. Both correlation coefficients can thus be considered as similar, since fixations and saccades are inherently correlated. While these constraints impede a perfectly proper comparison, it nevertheless offers an indication of the validity of the method used in this paper.

4.2. Interpretation of the results

The relatively high, significant correlation coefficients—definitely in comparison to correlation coefficients found for other datasets or other saliency algorithms (see Borji, Tavakoli, et al., 2013 for an overview)—found between the human focus maps and the saliency maps indicate that the latter can be considered as fairly reliable predictions of the human viewing pattern in landscape photographs. The fact that the correlation decreases when the degree of urbanisation in the landscape increases implies that the viewing pattern appears to be less predictable when the amount of built content increases (e.g., Semi-urban and Urban landscapes). In less urbanised landscapes characterised by a restricted number of buildings (e.g., Semi-rural and Mixed landscapes) this predictability is higher. As the difference between the five landscape categories tested in this study is based solely on the criterion “amount of buildings”, it can be deduced that the predictability of the human viewing pattern seems to be influenced by the degree of urbanisation of a landscape. When buildings are sparse, they have been demonstrated to act as eye-catchers when observing landscape photographs (Dupont, Antrop, et al., 2015). This can explain the higher correlations in semi-rural and mixed landscapes because, as buildings visually often stand out of their surroundings (by colour, texture etc.), there is a high probability that they will be identified as highly salient by the saliency algorithm. When the proportion of buildings in a scene becomes too large, this effect fades out. Human observers then seem to ‘lose track’ and start looking around without clear targets to fixate upon (Dupont, Ooms, et al., 2015) (Fig. 5). In fact, this pattern emerges because photographs of more urbanised environments contain much more details and thus have a higher information content resulting in a less structured, scattered viewing pattern as people try to assimilate as much information as possible (Dupont, Ooms, et al., 2015) (Fig. 5). This viewing pattern could explain the lower correlations found in semi-urban and urban landscapes.

Fig. 5. Visualisations of the saliency maps (second row) and examples of one-observer focus maps (third row) for (a) Rural, (b) Semi-rural, and (c) Urban landscapes.
In a broader context, these findings—together with the results of Dupont, Antrop, et al., 2015—may point at a more general result, i.e., that buildings could be one of the determining factors guiding the observation pattern in landscape photographs. Several reasons could explain why buildings are so important in the visual exploration of the environment. First, the human eye has been demonstrated to select areas in an image containing a maximum of information (Reinagel & Zador, 1999). Salient regions with high contrast and thus high information content, like buildings in a ‘green’ landscape, will be fixated most (García, Hernández, & Ayuga, 2006). Second, the main function of selective attention is to direct our gaze towards elements of interest in our visual environment (Braun & Julesz, 1998; Hikosaka, Miyachi, & Shimojo, 1996). From an evolutionary point of view, these elements may be determined by the Prospect-refuge Theory formulated by Appleton (1975). In particular, this theory states that all creatures, including humans, unconsciously and instinctively perceive their environment in such a way that environmental information is obtained and stored in a form, which allows an easy and quick retrieval when needed to ensure survival. This form consists of classifying the landscape according to potential prospects and refuges. Prospects are defined as places, which offer an unimpeded opportunity to see, whereas sites providing the opportunity to hide from and protect against potential hazards are called refuges. The ability to see without being seen is important in determining one’s survival prospects (Appleton, 1975). Numerous examples for refuges can be mentioned but the most common concept of a refuge for modern man is a building (Appleton, 1975). Finally, because of their sharp vertical edges, buildings have high contrasts with their surroundings and could therefore catch more attention. This is related to the Gestalt principle of continuation, which states that people tend to continue shapes beyond their ending points (Koffka, 1935). When a landscape is ‘interrupted’ by a sharp edge, for instance of a highly contrasting building, the continuation is broken. This might result in an increase in the attention spent on this area.

The exceptionally low correlation found between the saliency and focus maps for the rural landscapes could be explained by the more monotonous character of these landscapes and thus by their low information content. This may cause boredom with observers, who may start looking around in order to find more interesting objects to fixate upon. The result is a less structured and more scattered and thus less predictable observation pattern (Dupont, Ooms, et al., 2015; Fig. 5).

### 4.3. Implications/possibilities/usefulness for visual impact assessment

According to García et al. (2006) three aspects need to be taken into account when building in relation to landscape: the landscape value, the location of the new project and the visual characteristics of the existing landscape (e.g., colours, textures, lines, etc.). We believe that our methodology (see further) can be particularly useful for evaluating the third aspect: the integration of a project into the landscape once a location has been selected. García et al. (2006) recommend a detailed study of the scene in which the construction is to be executed, including an analysis of the colours, textures, and lines of the main elements. As such, the design of the new development can be tuned to the existing landscape in order to attain an optimal visual integration. The main guidelines described by García-Moruno, Montero-Parejo, Hernández-Blanco, and López-Casares (2010) consist of avoiding sharp colour contrasts, introducing vegetation cover if necessary and be careful with vertical shapes as these catch more attention (Español, 1995), especially when exceeding the skyline. These aspects (colour, orientation, and brightness) are all taken into consideration in saliency images. Therefore, the method described below could be a promising tool for visual impact assessments.

In the ideal situation, a visual impact assessment is performed at the beginning of a project, before the final decision is made and before the actual works start on site. Computerized visualizations are often used to conceptualize the possible alternatives of a project (Lange, 1994; Pullar & Tidey, 2001). For new constructions, for example, different designs varying in form, scale, colour, materials and texture, can be evaluated (see e.g., VIA in the UK) in order to determine which design attains the optimal visual integration in the surrounding landscape (García, Hernández, & Ayuga, 2003; García et al., 2006; García-Moruno et al., 2010). Doing so prevents new constructions from visually violating the landscape, which occurs when the contrast between the new element and its surroundings is too large or when the new object simply defies the gist of the scene. However, this kind of assessment—and the consideration of visual aspects in the planning process in general—is rarely done (Lange, 1994; Schmid, 2001), and certainly not in an objective and quantifiable fashion (Hernández, García, & Ayuga, 2004; Möller, 2006). When performed, several computer-aided simulations of a project are generally produced and a photograph-based survey is conducted to choose the best option according to the opinion of experts, focus groups or—in the best case—the public (Palmer, 2015). However, this methodology is money- and time-consuming and it is often difficult to obtain a representative public opinion and reach consensus between public and experts. But above all, a clear, quantitative and objective methodology and/or guidelines (independent of the experts/planners) are missing (Lange, 1994; Minelli et al., 2014; Palmer, 2015; Uzzell & Jones, 2000). Consequently, in many countries, visual impact assessment caused by the design of the construction is not even compulsory, while for landscape quality control it is an indispensable step in the planning process (Lange, 1994; Schmid, 2001). Saliency maps could help to resolve this issue as they offer a number of advantages that could contribute to set up a standardized and transparent methodology for visual impact assessment.

The GBVS algorithm applied on landscape photographs allows the creation of saliency maps in an easy way. As demonstrated by the correlation coefficients, these maps are positively correlated with focus maps obtained from eye-tracking the viewing behaviour of a number of observers while free-viewing these photographs. Both identify features in the landscape scene which act as eye-catchers. Saliency maps can therefore be regarded as predictions of the human viewing pattern in landscape photographs. Objects which are indicated as salient in the saliency map, will have very high changes of catching the attention in practice, while non-salient elements will not. Since it has been demonstrated that when the visual perception of a construction is reduced, its visual impact is diminished too (Hernández et al., 2004), saliency maps have a potential to be used as a new objective tool for visual impact assessment. They could be used to evaluate the visual impact of different designs or scenarios of a new construction represented in a series of simulated photographs, showing the degree of integration in the existing landscape or the potential of creating new eye-catchers that will affect human perception and viewing behaviour. By comparing the saliency map of each simulation with the saliency map of the original landscape image, an objective measure can be obtained of how salient or eye-catching the different designs of the new object will be. In particular, the correlation between the saliency map of a design simulation and the saliency map of the original landscape photograph can be calculated using the method presented in this paper. When this procedure is repeated for all potential designs, a ranking can be drawn up indicating which designs approximate the original landscape most (highest correlation) and which deviate from the existing landscape (lowest correlation). As such, the visual impact of the different design
options can be compared. High correlations mean that there will not be large differences in saliency after inserting the new object and thus that the viewing pattern will not be affected. In this case, the new project will be well integrated into its surrounding landscape, will not catch the attention and thus will have a low visual impact. Low correlations reflect modifications in the saliency of the scenery after the new construction was inserted. As a result, the viewing pattern will change as well. Most probably, the new object will be more salient than the original landscape and will as a consequence catch the attention. As a result, the visual impact of the new development will be high, for example due to too sharp colour, texture or shape contrasts, which have been demonstrated to strongly influence the fixation pattern (Becker, Pashler, & Lubin, 2007; Underwood & Foulsham, 2006). Simulations generating low correlation coefficients will therefore not be visually well integrated into the existing landscape.

This method is fast and easy and it allows a quantitative and scientific measure of the visual impact, which is widely demanded (Pullar & Tidey, 2001; Uzzell & Jones, 2000) as it facilitates the decision-making in choosing between different designs. Scenarios or designs having a high correlation with the existing landscape photograph, will have a low visual impact and will be well integrated into the scenery. It should be noticed that the method can also be used in the opposite case, e.g., when a design is intended to act as a landmark and thus needs a high level of conspicuity. In this case, the design with the lowest correlation corresponds to the highest visual impact.

4.4. Recommendations for applying the methodology in visual impact assessment

First, there is a large variety of saliency algorithms available, each with their own nuances and specifications (see Borji, Tavakoli, et al., 2013 for an extensive review and comparison). However, when applying the methodology in visual impact assessment, we strongly recommend using the GBVS algorithm for several reasons. The algorithm is freely available and easily accessible, which is not the case for other algorithms. But what is more important, our study demonstrates that in the specific case of landscape photographs, the correlation between the GBVS and human focus maps is relatively high and significant, which makes the GBVS a valid and suitable prediction of the human viewing pattern and thus suitable for use in visual impact assessment. While this might also be the case for other algorithms, this has not been tested yet and thus remains uncertain.

Second, top-down influences can never be completely excluded, even in free-viewing conditions (Parkhurst et al., 2002). For example, the observer’s interest, gender, mood, or cultural background may affect the eye movements (Borji, Shiite, et al., 2013). As a consequence, the viewing pattern of one observer will never be identical to the observation pattern of another observer. This also means that the focus map of one person can correspond more to the saliency map than the focus map of another person. Saliency maps can thus not be considered as predictions valid for all possible observers. However, as they make predictions of the viewing pattern purely based on bottom-up principles of attention guidance, which are unconsciously effective in each human being, saliency maps can be seen as useful predictions for most observers, at least in free-viewing conditions.

Third, the atmospheric and weather conditions under which a photograph has been taken, will affect the saliency. Pollution, fog, or rain can decrease overall contrast and thus decrease the saliency of all objects. Atmospheric attenuation increases with the distance to the objects viewed, which will become hazy, fuzzy and bluish, and details will fade out. In sunny weather conditions on the contrary, the contrast of colour and brightness will be enhanced, which increases the saliency and thus the visual attraction (García et al., 2006). In addition, the contrast will vary according to the relative position of the photographer and the sun. This is important when assessing the visual impact of objects from different viewpoints (Bishop, 2002). Photographs taken without direct sunlight have the most homogeneous contrast distribution and will therefore attain comparable saliency values. In order to be comparable, the different designs of a new construction need to have similar illumination conditions as the original landscape photograph.

Considering all these concerns, it is recommended to take the photographs under the most habitual observation conditions concerning weather and distance. Similarly, taking the photographs in one season is recommended. In usually cloudy regions, images should reflect this type of weather. The images should also be taken from points where people actually pass (e.g., roads, paths, vantage points, residential areas, etc.) and from where the new planned object will be seen (see also Palmer, 2015). As such, the distance can be determined as well (García-Moruno et al., 2010). Optimally, multiple views from where the project can be seen should be included in the assessment. Of course, the view from which the construction will be most seen can receive a larger weight in the final decision.

Finally, biasing the saliency of the planned object because of improper simulation techniques should be avoided (see Sheppard, 1989 for proper simulation methods). Inserting shiny elements in a shaded area in the scene will, for example, cause these objects to be incongruent. As a consequence of too sharp colour and brightness contrasts, erroneously high saliency values will be obtained indicating a high visual impact while in reality these elements might not at all catch the attention.

4.5. Further research

The methodology presented in this paper must be considered as a first step in assessing visual impact and, in a broader context, as a contribution to the development of a method aimed at assessing acceptability of projects.

Our methodology needs to be validated by applying the saliency method on real simulations, differing in degree of visual impact. This visual impact can be determined by analysing observer’s viewing patterns to check which alternative is most eye-catching. As such, we can analyse if the most salient alternative indeed generates the lowest correlation and vice versa. However, different alternatives of a project are not always provided by the developer or the alternatives are not elaborated enough to create proper simulations of it. In this case, when comparing alternatives is not possible, it is difficult to determine if the proposed scenario will be visually integrated enough to be approved or not. A threshold for evaluating the correlation would resolve this issue and help policy makers decide on the approval of a project. However, for determining this threshold, more empirical research is needed. In particular, proposed simulations need to be developed and tested with eye-tracking and/or based on saliency maps in order to know which alternatives catch most attention (viewed first). In addition, people’s opinion about how well they think the project is visually integrated into the existing landscape must be probed (for example like proposed by Palmer, 2015) as it is important to know how well the objectively measured visual impact (eye-tracking or saliency maps) is related to human judgements of visual integration. The same steps can be repeated with photographs from executed projects which have already been built. Subsequently, the correlation with the existing landscape can be calculated for the saliency maps of the simulations and executed projects. When both, the correlation coefficient and the ratings/viewing pattern of the observers are compared for a large number of projects, it is possible to determine from which correlation threshold a project can be considered to be visually integrated into a landscape. By
including people's evaluation, this method could contribute to the more general concepts of landscape quality and acceptability.

Finally, at the moment, the GBVS code is only available for Matlab, which is an expensive mathematics software package not commonly accessible for most landscape architects. If our approach is to be used by landscape planners and architects it should be made more accessible. A solution would be to translate the Matlab-code into a Python-code, which can then be implemented in ArcGIS, a software program more often available to landscape professionals. The same Python-code can also be used in QuantumGIS, which is a free and open source geographic information system and for this reason even more accessible. This would largely improve the ability to use the approach presented in this paper.

5. Conclusions

The GBVS algorithm allows to produce saliency maps from landscape photographs in an easy fashion, at least when Matlab is available. These saliency maps have been demonstrated to be correlated with focus maps obtained from eye-tracking the viewing pattern of a number of observers while free-viewing landscape photographs. Thus, saliency maps can be considered as predictions of the human viewing behaviour, showing potential focus areas and identifying the features in a scene that will attract the attention. While the method still needs to be validated, we believe that saliency maps could be a promising tool for visual impact analysis in landscape architecture and design, urban planning and environmental impact assessment. Saliency maps from different simulations can be made and compared to the photograph of the original landscape. The correlation between the saliency maps of the simulation and the original photograph will then indicate the degree of integration in the existing landscape, offer a quantitative measure of the degree of visual impact of different features, and help to select the best scenario for a given purpose. For example, different simulations of one project could be examined and the least salient option will then represent the most optimal visual integration into the existing landscape.

Furthermore, the correlation between the saliency and focus maps seems to vary with the proportion of buildings visible in the photographs, suggesting a relation with landscape type and degree of urbanisation in particular. Our study points out that the prediction of the saliency maps increases when the amount of buildings in a landscape photograph decreases. The human viewing behaviour is thus best approximated in rural landscapes with a limited amount of built content. This means that the methodology for visual impact assessment presented in this paper will probably be more reliable when a new construction is to be executed in relatively rural landscapes. In more urbanised landscapes, this reliability will probably slightly drop as the correlation between the saliency maps and the human focus maps is a bit lower in this kind of landscapes. Nevertheless, all the landscape categories tested in this study generated relatively high correlations between the saliency maps and the focus maps, which we believe is sufficient to confirm their validity for visual impact assessment. While this methodology is relatively easy to execute and produces an objective measure of visual integration, more research is required in order know more about the feasibility and effectiveness when working with edited photographs. A trial-and-error study should be executed to answer a number of practical questions and address the potential teething troubles of the methodology.

References


