A Model for Automatic Diagnostic of Road Signs Saliency

Ludovic Simon[(1)], Jean-Philippe Tarel[(2)], Roland Brémond[(2)]

[(1)]Researcher-Engineer
DREIF-CETE Ile-de-France, Dept. Mobility
12 rue Teisserenc de Bort, 78190, Trappes, France
lsimon@developpement-durable.gouv.fr

[(2)]Researcher
University Paris Est-INRETS-LCPC (LEPSiS)
58 Bd Lefebvre, 75015, Paris, France
tarel@lcpc.fr
bremond@lcpc.fr

Abstract

Road signs, the main communication media towards the drivers, play a significant role in road safety and traffic control through drivers' guidance, warning, and information. However, not all traffic signs are seen by all drivers, which sometimes lead to dangerous situations. In order to manage safer roads, the estimation of the legibility of the road environment is thus of importance for road engineers and authorities who aim at making and keeping traffic signs salient enough to attract attention regardless of the driver's workload.

Our long term objective is to build a system for the automatic estimation of road sign saliency along a road network, from images taken with a digital camera on-board a vehicle. This system will be interesting for accident analysis and prevention since it will enable a fine diagnostic of the road signs saliency, helping the road manager decide on which signs he must act and how (replacement or background modification). This should lead to improved asset management, road infrastructure maintenance and road safety.

What attracts driver's attention is related both to psychological factors (motivations, driving task, etc.) and to the photometrical and geometrical characteristics of the road scene (colours, background, etc.). The saliency (or conspicuity) of an object is the degree to which this object attracts visual attention for a given background. Road signs perception depends on the two main components of visual attention: objects pop-out and visual search. The first one is less relevant when the task is to search for a particular object, whereas one important part of the driving task is to look for road signs.

As most of current computational models of visual search saliency are limited to laboratory-situations, we propose a new model to compute visual search saliency in natural scenes. Relying on statistical learning algorithms, the proposed algorithm emulates the priors a driver learns on object appearance for any given class of road signs. The algorithm performs both the detection of the object of interest in the image and the estimation of its saliency. The proposed computational model of saliency was evaluated through psycho-visual experiments. This opens the possibility to design automatic diagnostic systems for road signs saliency.

1. Introduction

The points on which the driver focuses his gaze depend on the traffic situation, on the ongoing driver task and on the saliency of the objects relative to their background. Visual attention (Knudsen, 2007) can be thought of as a two components process: *object pop-out* and *visual search*. These two components are mixed during vehicle driving. The *object pop-out* is only due to the high saliency of the object which attracts attention, independently of the ongoing task. It is a bottom-up process which applies when, for instance, an observer is looking at a meaningless picture. The *visual search* which is a top-down process. It consists in a goal-driven process linked to voluntary attention, which depends on the driver's experience, the current task and motivation. Searching for a specific detail in a picture is an example of pure visual search.

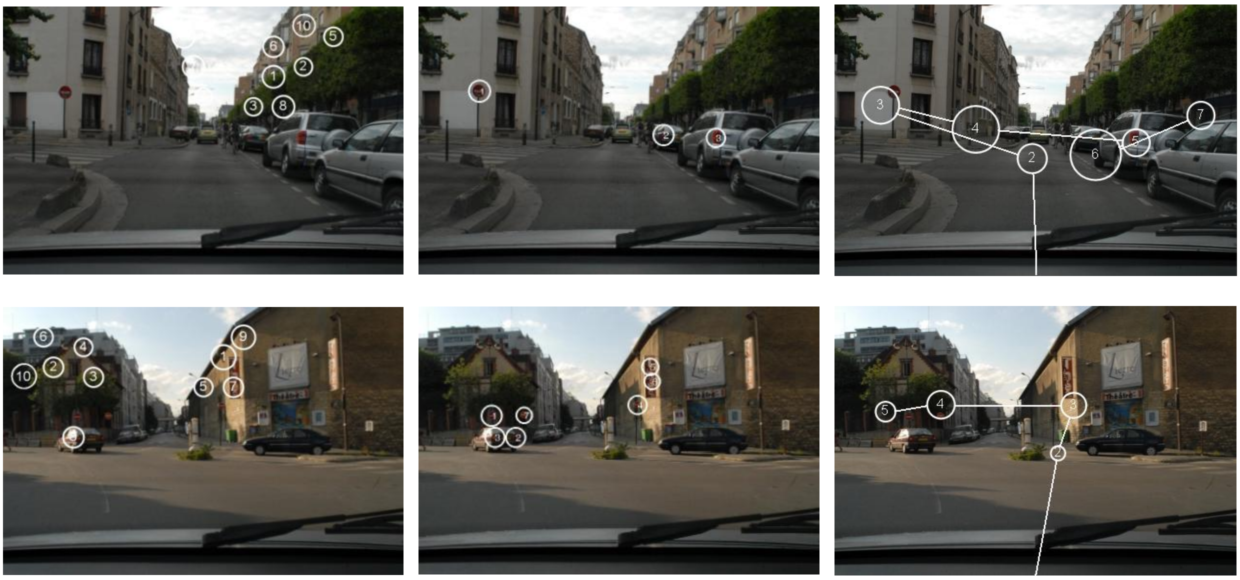

Figure 1: On the left, the focus of attention predicted by a bottom-up model (Itti, 1998)*. On the middle, focus points predicted as salient by our model for the search of no entry signs. On the right, scan-path of a subject searching for no entry signs. The focus points are labelled in decreasing order with a number inside each circle.*

Although several computational models of the saliency for object pop-out have been proposed in the last decade, due to the complexity of human behaviour, it is only very recently that a few computational models of the saliency for visual search have been proposed. The most popular computational saliency model was proposed in (Itti, 1998). This algorithm computes a bottom up saliency map based on a modelling of the low levels of the Human Visual System (HVS). This model was tested against oculometric data and it succeeds when the observer task is to memorize images, but as expected, it fails when the task is to search for an object, see (Underwood and others, 2006). Fig. 1 illustrates this limit on a road scene: on the left the focus points predicted by (Itti, 1998) are incorrectly spread on the whole image, whereas the observed scan-path is mostly along the line of horizon as displayed on right image.

To design an automatic system for the estimation of road sign saliency along a road network from images, a computational model of search saliency is thus requested. Until recently, there was no complete computational model, only theoretical models of the search saliency or computational models working only on simple laboratory situations and thus not on road environment. As a consequence, in (Simon and others, 2007), we proposed a new computational model of search saliency, based on learning the appearance of the sign(s) of interest in the images. As explained first in (Simon and others, 2007), our paradigm consists in linking the confidence in the detection with the search saliency, see section 2. Then, the approach was tested by psycho-visual experiments on no-entry sign, as described in section 3.

2. Saliency estimation based on sign appearance learning

The visual search task for a road sign is a detection problem in terms of pattern recognition. The search saliency of a given sign can be thus interpreted as a measure of the difficulty to detect it in an image.

2.1 Saliency estimation paradigm

To perform road sign detection in images, we used recently developed statistical learning algorithms. The learning is performed from a set of positive and negative examples of the signs to be detected, each example being represented as a feature vector. Positive feature vectors are samples of the appearance of the sign, while negative feature vectors are samples of the appearance of a possible background, see Fig. 2. This set of positive and negative vectors is usually called the learning database. From this database, the Support Vector Machine (SVM) algorithm (Schölkopf and Smola, 2002) is able to infer the frontier which splits the feature space into positive and negative parts. After this learning stage, the resulting classifier is able to set a positive or a negative label to any new feature vector, i.e to any window in a new image. This classifier can thus be used to perform sign detection in an image at several scales using sliding windows.

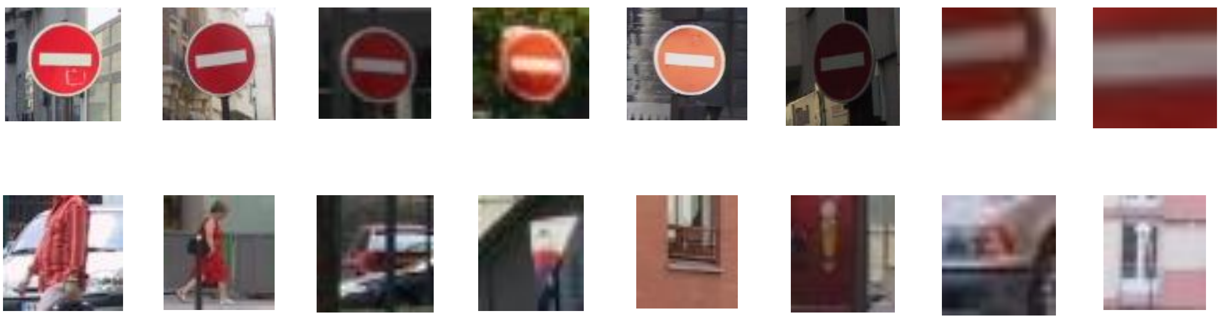

Figure 2: Positive (top) and negative (bottom) samples of the appearance of a no-entry sign.

The main advantage of the SVM is that the resulting classification function gives continuous values and not only binary values as with most classification algorithms. This classification value is related to the confidence in the pattern recognition: when it is higher than one, the recognition is positive, when it is lower than zero, the recognition is negative. The closer to zero the classification value is, the more hazardous the recognition decision. Our paradigm is thus to rely on a learning algorithm for modelling the appearance of the object of interest, and to define the search saliency of a given window image as a function of the classification value.

The input of the SVM algorithm, the learning database, is a set of feature vectors with positive or negative labels. In (Simon and others, 2008) are described the experiments which allowed us to select which features are adequate to represent the appearance of an image window. We found that a simple colour histogram is enough with the objective of “no-entry” sign detection for saliency estimation. The choice of the kernel is also of importance to achieve correct detection performance. As detailed in (Simon and others, 2008), the best choice appears to be the power kernel $k(x,x') = \|x - x'\|^{\alpha}$ which allows for implicit adaptation to the data density, unlike most classical kernels (α=1 in the following).

2.2 Saliency map

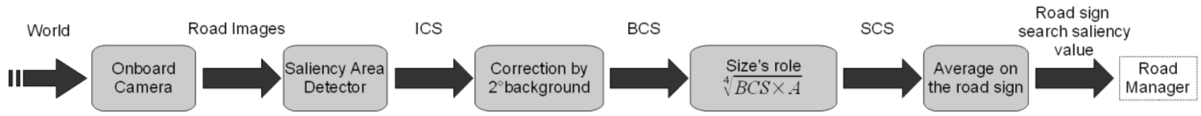

Figure 3: Scheme for the computation of the search saliency.

The SVM classifier is applied at different scales on sliding windows in a road image. Each scale results in a confidence map at a given scale of the positive classification values. The global confidence map for an image is built as the maximum of the confidence maps over the various scales. Finally, to take into account the saliency of the background around each detected road signs, the global confidence map is corrected by subtracting its local mean over the so-called background-window. The size of the background-window is of constant angular value set to 2 degrees, from our experiments in (Simon and others, 2009). The resulting map is defined as the Background related Computed Saliency (BCS) map. Notice that the BCS is independent of the size of the detected object, whereas it is known that the size plays an important role in the saliency. Thus, the positive connex components are extracted from the BCS map and for each component *i* the Search Computed Saliency (SCS) is computed as:

$$SCS(i) = \sqrt[4]{BCS(i)A(i)}$$

where B*CS(i)* is the average BCS over the connex component *i* and *A(i)* is its area. The search saliency of each road sign can be summarized by Fig. 3.

The field of applications of this computational model is not limited to the estimation of road signs saliency by inspection vehicle; it should also serve Advanced Driver Assistance Systems (ADAS), as explained in (Simon and others, 2009).

3. Experiments

3.1 Apparatus

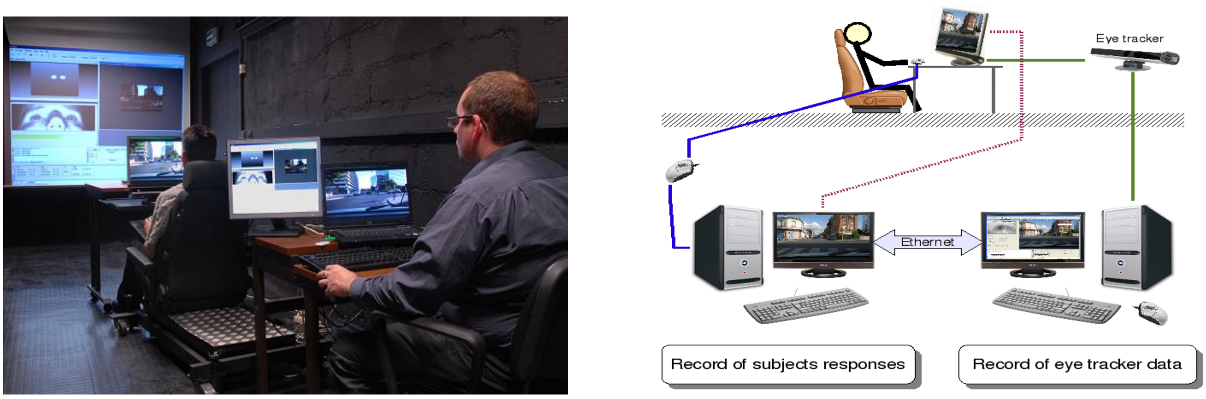

Figure 4: Psycho-vision laboratory, including an eye-tracker system used for psycho-visual experiments.

In order to test the proposed model, we conducted experiments in a display room which is photometrically controlled, as shown in Fig. 4. The room consists in subject's and advisor's screens. The display device is equipped with a remote eye-tracker (SMI) which tracks the subject’s gaze to record fixations positions and duration with an accuracy of 0.5 degree at a frequency of 50Hz. The screen is 19" wide with a viewing distance of 70 cm. Thus, the subjects saw the road scene with a visual angle of 20 degrees. For each subject, 40 road images were displayed, containing a total of 76 no-entry signs. These images were selected with various sign appearances and background, leading to a large variety of saliency levels for the observed signs.

3.2 Detection score and subjective saliency

Thirty subjects were asked to pretend they were drivers of the car from which the images were taken. The experiment consisted in two phases. In the first phase, the subjects were asked to count the no-entry signs, knowing that images would disappear after 5s. In the second phase, the subjects were asked to rate the saliency of each no-entry sign by giving a score between 0 and 10.

The analysis of the gaze fixations associated with the subjects answers (reported number of no-entry signs) allowed us to know whether the subjects detected each displayed no-entry sign. For a given sign, the mean percentage of detection over all subjects gives the Human Detection Rate (HDR).

In the second phase, due to the subjects variability in the use of the score scale, all scores were standardized enforcing for each subject the same Gaussian law with mean value 5. Subject Standardized Score saliency (SSS) (subject index *i*), related to subjective saliency (of sign *j*), is defined by:

$$SSS(i,j) = \frac{score_{i,j} - E_i(score_{i,j})}{\sqrt{E_i(score_{i,j} - E_i(score_{i,j}))^2}} + 5$$

3.3 Results

We first checked that the average detection rate HDR is related to the Subjective Standardized Score (SSS). As illustrated in the left of Fig. 5, the greater the SSS, the greater the HDR. This correlation between the subjective saliency score and the correct detection rate supports the proposed paradigm: the search saliency is an evaluation of the difficulty to detect an object.

Then, we studied the correlation between the Subjective Standardized Score (SSS) and the Search Computed Saliency (SCS). As shown in the right of Fig. 5, a linear relation can be observed between the SSS and the SCS. As a consequence, the proposed computational model of search saliency is correlated with the subjective saliency. The link between the HDR and SSS implies that the SCS is also correlated to the Human Detection Rate (HDR) which is a saliency indicator.

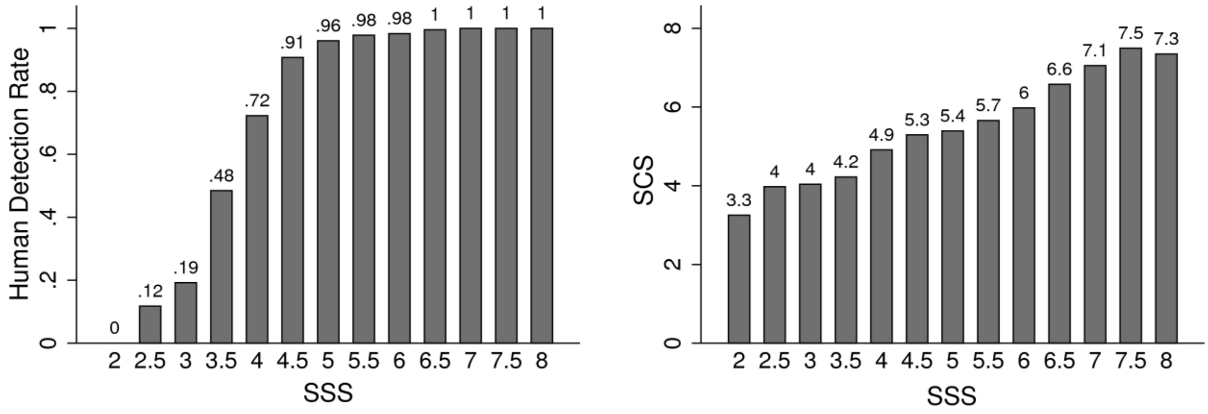

Figure 5: On the left, the correlation between the Subject Standardized Score (SSS) and the Human Detection Rate (HDR). On the right, the linear correlation between the SSS and the Search Computed Saliency *(SCS).*

The statistical analysis detailed in (Simon and others, 2009) showed that the proposed computational SCS model explains 56% of the variance between signs, and 39% overall. The same test using the 4th squared root of the road sign size instead of the proposed model SCS explains 46% of the variance between signs and 32% overall. Thus, the proposed computational saliency model improves the size-based mode by an increase of the explanation of the variance in an amount of 18%.

4. Conclusion

Most available computational models of the visual saliency are limited to objects pop-out, whereas the need for an automatic diagnostic system for road signs saliency implies to have a computational model of saliency during visual search task. We thus proposed a paradigm to define search saliency and a computational model to estimate the search saliency in an image within a search task using SVM. From our psycho-visual experiments, this computational model is correlated to subject's score of road sign saliency. In future work, we will test our computational model on other road signs, our goal being to develop a reliable automatic diagnostic system of road signs saliency.

5. References

1. Knudsen E.I. (2007), ‘Fundamental Components of Attention’, *Annual Review Neuroscience*, vol. 30, pp. 57-78.

2. Itti L., Koch C. & Niebur E. (1998), ‘A model of saliency-based visual attention for rapid scene analysis’, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, pp. 1254-1259.

3. Underwood G., Foulsham T., van Loon E., Humphreys L. And Bloyce j. (2006), ‘Eye movements during scene inspection: A test of the saliency map hypothesis’, *European Journal of cognitive Psychology*, vol. 18, pp. 321-342.

4. Simon L., Tarel J.-P. & Bremond R. (2007), 'A new paradigm for the computation of conspicuity of traffic signss in real images', *Proceedings of the 26th session of Commission Internationale de L'éclairage (CIE'07),* vol. 2, China, pp. 161-164.

5. Schölkopf B. & Smola A. (2002), '*learning with kernels', MIT Press, USA.*

6. Simon L., Tarel J.-P. & Bremond R. (2008), 'Towards the Estimation of Conspicuity with Visual Priors', *Proceedings of the International Conference on Computer Vision Theory and Applications (VISAPP'08),* Portugal, pp. 323-328.

7. Simon L., Tarel J.-P. & Bremond R. (2009), 'Alerting the Drivers about Road Signs with Poor Visual Saliency', *Proceedings of the IEEE Intelligent Vehicle Symposium (IV'2009),* China, pp. 48-53.