Signal Processing: *Image Communication* 39 (2015) 386–404

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

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

Perceived interest and overt visual attention in natural images

Ulrich Engelke [a,*], Patrick Le Callet [b]

[a] *Commonwealth Scientific and Industrial Research Organisation (CSIRO), 7005 Hobart, Australia*
[b] *LUNAM Université, Université de Nantes, IRCCyN UMR CNRS 6597 (Institut de Recherche en Communications et Cybernétique de Nantes), Polytech Nantes, France*

ARTICLE INFO

Available online 23 March 2015

Keywords:
Overt visual attention
Perceived interest
Region-of-interest
Eye gaze tracking
Image processing
Video processing

ABSTRACT

Region of interest (ROI) based image and video processing has attracted increased research efforts in recent years. The concept of perceptual ROI, however, is not always clearly defined leading to different interpretations between researchers related to bottom-up saliency (signal driven visual attention), top-down attention (subconscious, driven by higher cognitive factors, e.g. interest) or perceived interest. While all of these concepts are likely meaningful in the context of perceptual ROI based image and video processing, it is worth understanding how they are linked altogether. In this paper, the relationship between perceived interest and overt visual attention (which can cover both bottom-up and top-down attention) is studied. Towards this goal, a dedicated ROI selection experiment was performed and is analysed in detail, revealing deep insight into perceived interest in natural images. The outcomes are compared to an eye gaze tracking experiment representing overt visual attention in natural images. It is shown that there is indeed a strong relationship between perceived interest and overt visual attention for a wide range of natural scenes. We show that the relationship has a strong dependence on image content and presentation time during the eye gaze tracking experiment. Furthermore, eye gaze tracking data is revealed to have a high predictive value of primary ROI, particularly in case of the latter dominating over the remainder of the image. Both, the ROI and the eye gaze tracking databases are made publicly available to the research community.

1. Introduction

Digital image and video processing systems are typically designed as a trade-off between technical constraints and the quality of the output. To satisfy these constraints, technical parameters are chosen such as to optimise the system performance with respect to, for instance, the perceived visual quality or the intelligibility of the image or video. Similar to the technical limitations of these systems, our human visual system (HVS) also exhibits strong limitations regarding the computational complexity and the bandwidth with which we process visual information. The amount of visual information available at any instant in time, however, is vast and goes beyond the processing capabilities of the HVS. Therefore, several mechanisms are in place that facilitate to focus only on the most relevant information in any given context. First and foremost, these mechanisms consist of pre-processing in the early visual system and higher-order cognitive processes in the later stages of the HVS. The former is realised as non-uniform sampling in the retina of the human eye, which allows to process information with high accuracy only in the central point of focus, the fovea. The latter mechanism is referred to as visual attention [1] and facilitates that the focus of attention is shifted across the visual scene to the most salient or interesting locations. These mechanisms together achieve that in any given context the most relevant information is constantly favoured at the cost of less relevant information [2].

* Corresponding author. Tel.: +61 3 6237 5650.
E-mail address: ulrich.engelke@csiro.au (U. Engelke).

http://dx.doi.org/10.1016/j.image.2015.03.004
0923-5965/

The fact that not all visual information is equally relevant to the observer has been found to be an instrumental tool for further optimisation of image and video processing systems [3,4]. Typically, these systems first identify regions-of-interest (ROI) in the content and subsequently utilise these to improve system performance. In ROI-based image and video coding [5,6], for instance, the quality of perceptually relevant regions is prioritised over the quality of less relevant regions through non-uniform bit allocation, thus aiming to improve the overall perceived visual quality. In a communication context, ROI-based error resilience [7] takes into account the visual content in addition to the bit stream relevance in order to maximise protection of perceptually relevant information. In image retargeting [8,9], ROI information is used instead of or in addition to energy measures in the seam carving algorithm to identify the most relevant parts of the scene. In image retrieval [10,11], ROI information serves to improve database queries by taking into account the relevant and suppressing the irrelevant visual information. Finally, in quality assessment [12,13], ROI information is integrated into quality models to take into account the impact of distortions in relation to the relevance of the content. It has been found that ROI information is often beneficial in improving overall system performance for these and likely other image and video processing applications.

The perceptual relevance of the visual content is typically determined using computational models that are trained and validated using eye gaze tracking data [14]. Amongst the most common models are the biologically plausible models [15,16] following the neural-based architecture by Koch and Ullman [17], which incorporate characteristics of the HVS known to contribute to visual attention. Recently, there has been also a strong trend towards content-based models [18–21], which often incorporate high-level semantic factors in addition to low-level features. Other models have been proposed based on statistical [22,23], information theoretic [24], or learning-based [25] approaches. One class of models that found particular interest is based on Bayesian methods [26–28]. In addition to saliency information, these models often incorporate prior information related to contextual effects and semantic information [27].

Despite their common goal of identifying the most relevant information in a visual scene, the type of relevance information that is predicted by the above models can be very different [3]. Some of the models focus on the prediction of saliency driven attention locations [15,16,24], whereas others aim at predicting ROI at an object level [18,19]. The former relates to visual locations that are standing out in relation to the remainder of the scene with respect to some low-level features. Saliency-driven bottom-up attention is usually fast and involuntarily controlled (exogenous attention). On the other hand, decisions on ROIs are strongly driven by top-down attention mechanisms that usually involve a voluntary control of the gaze shift (endogenous attention) and are strongly influenced by context and semantic information. The ground truth upon which the success of saliency and ROI prediction models is validated on is typically based on overt visual attention measured through eye gaze tracking experiments. The recorded gaze patterns, however, do not allow for a clear distinction between the various attentional mechanisms as they account for both bottom-up and top-down driven visual attention. They further do not allow for a direct insight into which objects or regions are perceived to be most interesting in the visual scene. Understanding the perceived interest in a scene, however, is instrumental for successful augmentation of image and video processing systems. We argue that dedicated experiments are needed to determine perceived interest in a visual scene.

In this paper, we are presenting a novel study to contribute to a better understanding of overt visual attention and perceived interest in natural scenes. The goals of this study are twofold. Firstly, we aim to quantify the relationship between overt visual attention and perceived interest for a wide range of natural images. This allows us to identify whether or not the locations that humans are looking at are also necessarily the most interesting ones. Secondly, we determine the success with which gaze patterns can predict ROI, which will allow for deeper insight into the legibility of using eye gaze tracking data for the creation of ROI maps and as a ground truth for the design of ROI prediction models. In order to address these issues, we conducted two dedicated experiments, an eye gaze tracking experiment and a ROI selection experiment to, respectively, measure overt visual attention and perceived interest of a number of observers when viewing natural scenes. The focus is here on the ROI selection experiment, as it describes a rather unconventional approach in comparison to the more commonly performed eye gaze tracking experiments. We therefore provide an extensive discussion and a detailed analysis of the ROI experiment to identify the perceived interest of human observers in natural image content. We then analyse and discuss the outcomes of the ROI experiments with regard to the results from the eye gaze tracking experiment. Furthermore, we classify the ROI maps into primary ROI, secondary ROI, and background, to determine which of these are best predicted by eye gaze tracking data. We are particularly interested in two factors, the image content and the presentation time during the eye gaze tracking experiment.

The remainder of the paper is structured as follows. In Section 2 we discuss in more detail some of the conceptual differences between overt visual attention and perceived interest and the related experiment methodologies; eye gaze tracking and ROI selections. Section 3 briefly summarises the eye gaze tracking experiment and Section 4 discusses and analyses in more detail the ROI selection experiment that we conducted. The relationship between the eye gaze tracking data and the ROI selections is then analysed in great detail in Section 5 and the success of using eye gaze tracking data for the prediction of ROI is evaluated in 6. Finally, the results are summarised and conclusions are drawn in Section 7. An overview of all images used and created in this work is presented in Appendix A.

2. Overt visual attention versus perceived interest

In this work we study the relationship between overt visual attention and perceived interest based on experimental data. While overt visual attention is commonly measured using eye gaze tracking, measuring perceived interest based on an ROI selection experiment is a rather novel and unconventional approach. In the following, we