An Information Theoretic Model of Saliency and Visual Search

Neil D.B. Bruce and John K. Tsotsos

Department of Computer Science and Engineering and
Centre for Vision Research
York University, Toronto, ON, Canada
{neil,tsotsos}@cse.yorku.ca
http://www.cse.yorku.ca/~neil

Abstract. In this paper, a proposal which quantifies visual saliency based on an information theoretic definition is evaluated with respect to visual psychophysics paradigms. Analysis reveals that the proposal explains a broad range of results from classic visual search tasks, including many for which only specialized models have had success. As a whole, the results provide strong behavioral support for a model of visual saliency based on information, supplementing earlier work revealing the efficacy of the approach in predicting primate fixation data.

Keywords: Attention, Visual Search, Saliency, Information Theory, Fixation, Entropy.

1 Introduction

Visual search is an important task in everyday functioning, but a consensus on the precise details of the system underlying visual search in primates has yet to be reached. Consideration of specific stimulus sets in a lab setting has allowed observation of some of the peculiarities of visual search in primates revealing surprising efficiency for some visual search tasks and surprising inefficiency for others. Despite the considerable interest and effort placed on the problem, and the growing body of data on visual search, explanation for various effects exists in many instances within only specialized models. One might view the ultimate aim of modeling in visual search to be a single model with the minimum set of requirements that captures all observed visual search behavior and additionally is based on some basic well defined principle. It is our view that our proposal Attention based on Information Maximization (AIM) satisfies the last of these requirements, and the intention of the remainder of the discussion is to address the extent to which the first of these requirements is satisfied. In the sections that follow, it is established that the model exhibits considerable agreement with a broad range of psychophysical observations lending credibility to the proposal that attentional selection is driven by information.

In [1] we described a first principles definition for visual saliency built on the premise that saliency may be equated to the amount of information carried

L. Paletta and E. Rome (Eds.): WAPCV 2007, LNAI 4840, pp. 171–183, 2007.

by a neuron or neuronal ensemble. It was demonstrated that such an approach reveals surprising efficacy in predicting human fixation patterns and additionally carries certain properties that make the proposal plausible from a biological perspective. An additional and perhaps more favorable test for a model that claims to represent the process underlying the determination of visual saliency in the primate brain, is the extent to which the model agrees with behavioral observations, and in particular, those behaviors that on first inspection may seem counterintuitive. It is with this in mind that we revisit the proposal that visual saliency is driven fundamentally by *information*, with consideration to a variety of classic psychophysics results. In this paper, we extend the results put forth in [1] to consideration of various classic psychophysics paradigms and examine the relation of qualitative behavioral trends to model behavior. It is shown that the model at hand exhibits broad compatibility with a wide range of effects observed in visual search psychophysics.

2 Saliency Based on Information Maximization

The following describes briefly the procedure for computing the information associated with a given neuron response or ensemble of neurons. For a more detailed description, including details pertaining to neural implementation, the reader should refer to [1]. Prior efforts at characterizing the information content of a spatial location in the visual field appeal to measures of the entropy of features locally. Some shortcomings of such a measure are highlighted in [1], but in short, local activity does not always equate to informative content (consider a blank space on an otherwise highly textured wallpaper). In the context of AIM, the information content of a neuron is given by $-log(p(x))$ where x is the firing rate of the neuron in question and $p(x)$ the observation likelihood associated with the firing rate x. The likelihood of the response a neuron elicits is predicted by the response of neurons in its support region. In the work presented here, we have assumed a support region consisting of the entire image for ease of computation, but it is likely that in a biological system the support region will have some locality with the contribution of neighbouring units to the estimate of $p(x)$ proportional to their proximity to the unit exhibiting the firing rate x. This discussion is made more concrete in considering a schematic of the model as shown in figure 1. A likelihood estimate based on a local window of image pixels appears to be an intractable problem requiring estimate of a probability density function on a high-dimensional space (e.g. 75 dimensions for a 5x5 RGB patch). The reason this estimate is possible is that the content of the image is not random but rather is highly structured. The visual system exploits this property by transforming local retinal responses into a space in which correlation between different types of cell responses is minimized [2,3]. We have simulated such a transformation by learning a basis for spatiochromatic 11x11 RGB patches based on the JADE ICA algorithm [4]. This is depicted in the top left of figure 1. This allows the projection of any local neighborhood into a space in which feature dimensions may be assumed mutually independent. The