Chapter 4
Computational Models of Visual Attention: A Survey

Rajarshi Pal
Institute for Development and Research in Banking Technology, India

ABSTRACT

Even the enormous processing capacity of the human brain is not enough to handle all the visual sensory information that falls upon the retina. Still human beings can efficiently respond to the external stimuli. Selective attention plays an important role here. It helps to select only the pertinent portions of the scene being viewed for further processing at the deeper brain. Computational modeling of this neuro-psychological phenomenon has the potential to enrich many computer vision tasks. Enormous amounts of research involving psychovisual experiments and computational models of attention have been and are being carried out all within the past few decades. This article compiles a good volume of these research efforts. It also discusses various aspects related to computational modeling of attention–such as, choice of features, evaluation of these models, and so forth.

INTRODUCTION

What is Visual Attention?

We, the human beings, are amazingly efficient in real-time interaction with the dynamic environment surrounding us. We, constantly, gather information about our surroundings through five senses. After analyzing or interpreting the information, the brain decides the course of action, but do we process all the incoming sensory information at the deeper levels of the brain? Let this question be put in a slightly different manner. All the time, our senses are actively sensing the outside world. All of that sensory information is routed towards the brain though nerves. But does all this information reach the deeper level of the brain where recognition and decision making takes place? For example, we cannot effectively listen to the important discussion on a television channel while we simultaneously carry out the conversation with a friend over the telephone. So, there is a limit to the number of things the brain can process simultaneously.

The human brain intelligently filters out the majority of incoming sensory information before it can reach the deeper levels of the brain. This phenomenon is known as attention. Attention helps

DOI: 10.4018/978-1-4666-4558-5.ch004

us to attend only a selective subset of sensory information. In the pretext of the subject matter of this article, we limit the discussion only to visual attention. Even, only a few selected portions of the visual stimuli sensed in the retina of the eye are able to draw our attention.

Why is a Computational Model Required?

Like humans, computer vision tasks also face the challenge of dealing with huge amount of input (Tsotsos, 1990). The attention mechanism of human vision has influenced computer vision researchers to restrict the computation in certain areas of input. As a result, modeling visual attention draws significant research effort within the past few decades. It gathers theories from psychology, neurobiology of human visual system, and other related topics. Psycho-visual experiments have provided some theoretical reasoning for saliency of a location or an object. Computer vision researchers try to fit various types of models on acquired salient data on the basis of these psycho-visual experiments.

Selective attention to relevant salient locations in a scene has various advantages. It reduces the computational burden by decreasing the amount of data to be processed. Moreover, suppression of irrelevant information ensures influence of only the relevant locations of the scene in the outcome of the system.

Applications of Visual Attention Models in Computer Vision

Computational modeling of visual attention enriches many application areas of image processing and computer vision. In these applications, scene contents get discriminative treatment based on their relative saliency. Some of these applications are listed here:

- Autonomous road following guided by artificial vision system is demonstrated in (Baluja & Pomerleau, 1997). This artificial vision system is capable of finding task-related salient locations in a scene.
- Artificial vision system having attention mechanism helps a human robot to easily interact with its surrounding (Stasse et al, 2000).
- Identified salient landmarks are used for robot localization and mapping as demonstrated in (Bur et al, 2006).
- An attention model is successfully applied to target detection tasks reported in (Gao et al, 2009; Itti & Koch, 1999b).
- Finding out salient regions is useful in image retrieval (Jing et al, 2002).
- Visual attention model has been applied to video shot matching in (Li & Lee, 2007). Shot matching is an essential step for video indexing and retrieval.
- Visual saliency model has been used in image segmentation (Ouerhani & Hugli, 2003a).
- Object tracking in dynamic scenes is also facilitated by attention model (Ouerhani & Hugli, 2003b).
- Computer graphics based applications needs rendering of salient locations in a scene with more details (Chalmers et al, 2003; Sundstedt et al, 2005). Some of these applications are graphics based story telling, playing a game, advertising a product, and so on.
- In image and video compression (Bradley & Stentiford, 2003; Itti, 2004), more bits are allocated to salient areas to retain the perceptual quality of the compressed version.
- Attention model is also used to display images for devices with various screen sizes to enhance viewing pleasure (Liu et al, 2003; Pal et al, 2012a). Methods of thumbnail