Chapter 5
Modeling Visual Saliency in Images and Videos

Yiqun Hu
Nanyang Technological University, Singapore

Viswanath Gopalakrishnan
Nanyang Technological University, Singapore

Deepu Rajan
Nanyang Technological University, Singapore

ABSTRACT

Visual saliency, which distinguishes "interesting" visual content from others, plays an important role in multimedia and computer vision applications. This chapter starts with a brief overview of visual saliency as well as the literature of some popular models to detect salient regions. We describe two methods to model visual saliency – one in images and the other in videos. Specifically, we introduce a graph-based method to model salient region in images in a bottom-up manner. For videos, we introduce a factorization based method to model attention object in motion, which utilizes the top-down knowledge of cameraman for model saliency. Finally, future directions for visual saliency modeling and additional reading materials are highlighted to familiarize readers with the research on visual saliency modeling for multimedia applications.

INTRODUCTION

Visual saliency refers to the ability of any vision system to select a certain subset of visual information for further processing (Itti & Koch, 2001). This mechanism serves as the information processing bottleneck to allow only the "interesting" information related to current behaviors or tasks and ignores irrelevant information (Desimone & Duncan, 1995). Only the visual information in the "fovea" region is analyzed and processed while other information outside this field is suppressed (Eriksen & Yeh, 1985; Eriksen & St James, 1986). This ability is of evolutionary significance because it allows an organism to detect quickly possible prey, mates or predators in the visual world. Visual saliency is a complex concept and has diverse interpretations in psychology, neuroscience and vision research, leading to different research

DOI: 10.4018/978-1-4666-3994-2.ch005

methodologies as well as the evaluation criteria in these communities.

Visual saliency can be categorized based on taxonomies which are derived from different aspects of this mechanism. According to the target of saliency deployment, visual saliency has three forms:

1. **Feature-Based Saliency:** (Treisman, 1980) where saliency is attributed to different features;
2. **Space-Based Saliency:** (Wolfe, 1994; Tsotsos et al., 1995) where saliency is deployed at different locations; and
3. **Object-Based Saliency:** (Scholl, 2001; Grossberg & Raizada, 2000) where saliency is deployed on different objects/groups.

According to the control of saliency deployment, visual saliency can be driven in either bottom-up or top-down manner (Itti & Koch, 2001). In the bottom-up manner, visual saliency is purely driven by visual data itself. In top-down manner, high-level information like the goal and preferences of the observers can modulate and guide the deployment of saliency.

APPLICATIONS IN MULTIMEDIA

The two issues that limit the even more widespread use of multimedia content than in the present situation are their huge capacity and their high complexity. The use of visual saliency is a natural way to overcome these limitations by selecting relevant visual information and processing only the visual attention region. This mechanism can simultaneously improve the efficiency and robustness of various multimedia applications. In multimedia adaptation, images can be adapted (Chen et al., 2003) and browsed (Xie et al., 2006) or video sequences can be progressively transmitted for display (Hu et al., 2004) on small screen devices by preserving salient content. For Content-based Image Retrieval (CBIR) systems, detecting salient regions can improve the system performance by reducing the influence of cluttered background (Bamidele et al., 2004; Wang et al., 2004). Modeling visual saliency can also facilitate visual tracking due to the common issues that they address: salient content can be used to initialize, detect as well as recover tracking target (Brajovic & Kanade, 1998; Toyama & Hager, 1999; Yang et al.. 2007). Recently, media retargeting techniques (Shamir & Avidan, 2009; Wolf et al., 2007) have been reported that rely on visual saliency modeling to indicate important information that needs to be preserved. As digital cameras become ubiquitous, their technology aims to help the amateur photographer to capture pictures that are aesthetically much superior than before. Face detection is already available in many such digital cameras. Clearly, there is a role for visual attention that can improve the performance of this task, as also in others such as automatically focusing on a certain area of the scene that is visually salient and in automatic zooming into salient regions.

The remainder of this chapter is organized as follows. We first briefly review the related work about visual saliency modeling. From this review, two important issues about visual saliency modeling for multimedia and computer vision applications are defined. We introduce two methods in this chapter to target these two issues. First, we introduce a graph-based method for modeling spatial saliency in images. This method provides a possible solution to integrate both global and local information to detect visual saliency. Next, we formulate the problem of Attention-from-Motion (AFM) to detect Attention Object in Motion (AOM) from videos and propose a factorization based method to solve it. This method decodes the subjective attention of the cameraman from video content and uses this top-down knowledge to model spatial-temporal visual saliency. Finally, we highlight the directions and open issues for the future research in visual saliency modeling.