Saliency Priority of Individual Bottom-Up Attributes in Designing Visual Attention Models

Jila Hosseinkhani, Carleton University, Ottawa, Canada
Chris Joslin, Carleton University, Ottawa, Canada

ABSTRACT

A key factor in designing saliency detection algorithms for videos is to understand how different visual cues affect the human perceptual and visual system. To this end, this article investigated the bottom-up features including color, texture, and motion in video sequences for a one-by-one scenario to provide a ranking system stating the most dominant circumstances for each feature. In this work, it is considered the individual features and various visual saliency attributes investigated under conditions in which the authors had no cognitive bias. Human cognition refers to a systematic pattern of perceptual and rational judgments and decision-making actions. First, this paper modeled the test data as 2D videos in a virtual environment to avoid any cognitive bias. Then, this paper performed an experiment using human subjects to determine which colors, textures, motion directions, and motion speeds attract human attention more. The proposed benchmark ranking system of salient visual attention stimuli was achieved using an eye tracking procedure.

KEYWORDS


INTRODUCTION

Human brain receives a massive amount of information when watching virtually any scene. The Human Visual System (HVS) is capable of processing this information rapidly and focusing on the salient regions of the scene. These selected regions which are more interesting to the subject are called salient areas. Human visual attention contains two types of processes: pre-attentive and attentive (Osberger, 1999).

Pre-attentive (subconscious) processing rapidly and automatically categorizes an image into regions in a spatially parallel manner to search for significant information across an image. However, the attentive (conscious) processing or focused attention incorporates the goals and desires of the viewer through the process of searching in a serial manner which is time consuming compared to pre-
attentive detection (Healey et al., 2012; Jing et al., 2017). Understanding the processing mechanism of HVS helps us to know how to properly prioritize and combine the visual stimuli as well as the low-level, mid-level, and high-level features in the design of attention models.

Physiological and psychological studies illustrated that the effective factors on visual attention and eye movements are categorized into bottom-up and top-down types (Gelasca et al., 2005). Bottom-up factors capture pre-attentive attention very quickly and have a strong impact on the human visual selection system. On the other hand, top-down factors capture the attention much slower and are influenced by bottom-up factors. Bottom-up and top-down factors are known as low-level and high-level features respectively. In the past two decades, researchers focused on designing visual attention models (VAMs) were inspired by the HVS to reduce the huge volume of data to more visually informative and important data. Saliency detection models or VAMs employ bottom-up and/or top-down factors to search for the salient part of data. Bottom-up based models use low-level attributes such as color, texture, size, contrast, brightness, position, motion, orientation, and shape of objects. Basically, these attributes are rapidly scanned and detected by the human visual system. However, top-down based models exploit high-level context-dependent attributes such as a face, human, animal, vehicle, text, etc. Both bottom-up and top-down factors can be exploited to design VAMs but because of the complexity and time limitation, few integrated approaches have been proposed that use both factors to detect the salient parts within a scene (Duncan et al., 2012; Jie et al., 2018).

To generate the saliency map, different feature maps are usually produced for bottom-up attributes first. Then, these maps are fused to produce the overall activation map indicating the most salient areas. It should be noted that the basic feature maps can be combined to generate top-down attributes as well.

The validation of the saliency maps is usually performed by comparing them with eye movement tracking datasets as the ground-truth data. Studies show that the human visual system is attracted to objects rather than locations (Banitalebi-Dehkordi et al., 2016). In fact, the pre-attentive part of the HVS firstly segments the scene into objects in a rapid scan (Correia et al., 2000). This segmentation is mostly performed based on the low-level attributes (Wang et al., 2018).

In this paper, we focus on the study of bottom-up attributes as visual stimuli such as color, texture, motion direction, object velocity, and object acceleration. Our goal is to investigate how they influence the HVS. The aim of this work is to achieve a ranking system to identify the hue range, texture pattern, motion direction, object velocity, and object acceleration that are most likely to be attractive for the HVS in terms of saliency.

We will provide a more detailed explanation about existing studies related to the impact of bottom-up attributes on the visual attention system in addition to our designed experiment in next sections. The rest of this paper is organized as follows: Background section reviews the related works and provides an overview of how the experiments have been performed to understand HVS response to the bottom-up attributes. Experimental methodology section describes the characteristics of the generated dataset for our experiment and its methodology. Section of result analysis contains the results of our proposed experiment for each individual attribute, and finally, the last section concludes the paper.

**BACKGROUND**

Visual attention includes the procedure of selecting the significant and interesting areas across visual information that humans receive in daily life. The selection procedure by the HVS is performed through eye movements. Researchers investigate how visual stimuli influence human eye movements in order to estimate the most salient regions in a scene and consequently design visual attention models to extract those regions.

In this section, we first provide derived conclusions and results of the existing works about bottom-up attributes. Then, a brief description of the relevant experimental studies will be presented.

In the literature, few experimental studies have been performed to investigate the impact of the bottom-up factors on the HVS and eye movements. This motivated us to design a comprehensive
Cognitive Processes of the Elderly Brain with MindGym Approach
www.igi-global.com/article/cognitive-processes-of-the-elderly-brain-with-mindgym-approach/157435?camid=4v1a

Application of Structural Properties of Seismic Data to Prediction of Hydrocarbon Distribution
www.igi-global.com/article/application-of-structural-properties-of-seismic-data-to-prediction-of-hydrocarbon-distribution/207744?camid=4v1a