A Biologically Inspired Saliency Priority Extraction Using Bayesian Framework

Jila Hosseinkhani, Carleton University, Ottawa, Canada

Chris Joslin, Carleton University, Ottawa, Canada

ABSTRACT

In this article, the authors used saliency detection for video streaming problem to be able to transmit regions of video frames in a ranked manner based on their importance. The authors designed an empirically-based study to investigate bottom-up features to achieve a ranking system stating the saliency priority. We introduced a gradual saliency detection model using a Bayesian framework for static scenes under conditions that we had no cognitive bias. To extract color saliency, we used a new feature contrast in Lab color space as well as a k-nearest neighbor search based on k-d tree search technique to assign a ranking system into different colors according to our empirical study. To find the salient textured regions we employed contrast-based Gabor energy features and then we added a new feature as intensity variance map. We merged different feature maps and classified saliency maps using a Naive Bayesian Network to prioritize the saliency across a frame. The main goal of this work is to create the ability to assign a saliency priority for the entirety of a video frame rather than simply extracting a salient area which is widely performed.

KEYWORDS

Attention Priority, Bayesian Networks, Eye-Tracking, Human Visual System, Network Traffic, Object Segmentation, Saliency Detection

INTRODUCTION

During the last few years, video streaming demand has gone through an incredible increase due to the immense expansion of multimedia communications. To tackle the resulting congestion of the enormous volume of data over the networks, more bandwidth, and reliable communications are required. This has led to a considerable amount of research within the field of video streaming, video compression, Quality of Service (QoS), and real-time traffic supporting. To overcome the video streaming problem from a network traffic perspective, we decided to utilize a semantic video analysis mechanism known as a saliency detection technique and we introduced gradual saliency concept. Regions of interest or salient areas in an image or video play an important role in the semantic analysis of visual data. Saliency detection is widely exploited in many applications such as content-based image/video retrieval, scene understanding, video surveillance, video summarization, event detection, and image/

video compression. The visual attention system includes the procedure of selecting the significant and interesting areas across visual information that humans receive in daily life.

We introduced the gradual saliency concept to distinguish different classes of saliency in a video frame. In this way, the most important and informative regions, i.e., salient regions are extracted to reduce the volume of video data in transmission. Our main goal is to provide a guidance for the encoder in order to decide what information need to be dropped and what information should form different video coding layers.

The Human Visual System (HVS) can process this information rapidly and focus on the distinct parts of a scene. Studies indicated that the effective factors on visual attention and eye movements are categorized into bottom-up and top-down categories (Healey et al., 2012; Duncan et al., 2012). Bottom-up factors capture unconscious 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 require previous knowledge about the scene. Saliency detection models or Visual Attention Models (VAMs) employ bottom-up and/or top-down factors to search for the salient part of visual data. Bottom-up based models use low-level features such as color, texture, size, contrast, brightness, position, motion, orientation, and shape of objects (Duncan et al., 2012). However, top-down based models exploit high-level context-dependent features 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 hybrid approaches have been proposed that use both factors. For a real-time video streaming application, it is essential to achieve a fast and simple saliency detection method with the aim of effectively controlling the network traffic. Therefore, we decided to avoid any cognitive bias in designing our model in order to speed up the procedure of saliency detection. Because considering cognitive bias requires machine learning based algorithms which are time consuming.

The origin of many attention models dates back to Treisman and Gelade's (1980) "Feature Integration Theory," where they stated which visual features are important to direct human attention over pop-out and conjunction search tasks. They found that human can detect and identify separable features in a parallel manner across a display (Treisman and Gelade, 1980). Wolfe et al. (1994) presented a model of human visual search behavior based on Treisman's work known as "guided search". The main idea of this model argues that attentional deployment of limited resources is guided by the output of the earlier parallel processes (Wolfe et al., 1994). Since saliency detection is very closely influenced by human visual perception, it is necessary to study the impact of visual stimuli on HVS. The lack of a comprehensive study in this area inspired us to design an experimental study in our previous works (Hosseinkhani and Joslin, 2018 and 2019) in order to precisely and comprehensively investigate the effect of bottom-up attributes on HVS in terms of saliency detection problem. Hosseinkhani and Joslin (2018) presented a ranking system for bottom-up features in both individual and combined scenarios to identify what hue range, texture pattern, motion direction, object velocity, and object acceleration are most likely to be attractive for the HVS. In the current work, we introduced the implemented Bayesian algorithm to extract the saliency in static scenes. The main application of this work is adaptive video compression and streaming to control the network traffic. Our goal is to obtain guidance for the video encoder about what parts of a video frame can form the base layer and what parts can be dropped in real-time video streaming application using an error resilient coding scheme.

The main contributions of this paper can be described as follows: 1) Introducing gradual saliency concept by providing the ability to assign prioritized saliency for the entire of an image/video frame rather than simply the salient regions. 2) Considering the intuitive factors using an eye-tracking based subjective experiment that leads to design biologically plausible feature maps. 3) Extracting salient areas with high quality and well-defined boundaries resulting in a full-resolution saliency map. In current literature, only the salient areas and objects are usually used to design the visual attention models (VAMs) and the saliency detection is considered as a binary classification problem.

Related Content

Metamorphic Testing of Image Classification and Consistency Analysis Using Clustering
Hemanth Gudaparthi, Prudhviraj Naiduand Nan Niu (2022). *International Journal of Multimedia Data Engineering and Management (pp. 1-20).*
www.irma-international.org/article/metamorphic-testing-of-image-classification-and-consistency-analysis-using-clustering/304390

On Combining Sequence Alignment and Feature-Quantization for Sub-Image Searching
Tomas Homola, Vlastislav Dohnaland Pavel Zezula (2012). *International Journal of Multimedia Data Engineering and Management (pp. 20-44).*
www.irma-international.org/article/combining-sequence-alignment-feature-quantization/72891

Using Semantics to Manage 3D Scenes in Web Platforms
Christophe Cruz, Christophe Nicolleand Marc Neveu (2005). *Encyclopedia of Multimedia Technology and Networking (pp. 1027-1032).*
www.irma-international.org/chapter/using-semantics-manage-scenes-web/17363

Capitalizing on Immersive Persistence as an Emergent Design Concept (A Position Paper)
Shalin Hai-Jew (2011). *Gaming and Simulations: Concepts, Methodologies, Tools and Applications (pp. 1901-1913).*
www.irma-international.org/chapter/capitalizing-immersive-persistence-emergent-design/49482

Gait Recognition Using Deep Learning
Chaoran Liuand Wei Qi Yan (2020). *Handbook of Research on Multimedia Cyber Security (pp. 214-226).*
www.irma-international.org/chapter/gait-recognition-using-deep-learning/253034