Chapter 15
Estimating Visual Saliency for Omnidirectional HDR Images

Kenji Hara
Kyushu University, Japan

ABSTRACT

A unified decomposition-and-integration-based framework is presented herein for the visual saliency estimation of omnidirectional high dynamic range (HDR) images, which allows straightforward reuse of existing saliency estimation method for typical images with narrow field-of-view and low dynamic range (LDR). First, the proposed method decomposes a given omnidirectional HDR image into multiple partially overlapping LDR images with quasi-uniform spatial resolution and without polar singularities, both spatially and in intensity using a spherical overset grid and a tone-mapping-based synthesis of imaginary multiexposure images. For each decomposed image, a standard saliency estimation method is then applied for typical images. Finally, the saliency map of each decomposed image is optimally integrated from the coordinate system of the overset grid and LDR back to the representation of the coordinate system and HDR of the original image. The proposed method is applied to actual omnidirectional HDR images and its effectiveness is demonstrated.

INTRODUCTION

The computational determination of image-relevant areas that attract more visual attention of the human visual system is critical for accelerating and improving visual recognition. Thus far, a large number of visual attention models have been proposed in computer vision, artificial neural networks, biological science, and so on. In particular, the estimation of a visual saliency map, which is an evaluation of saliency of each image pixel to extract a region or an object of interest from a still image or a video image, has become useful tool in many applications such as object detection (Rutishauser et al., 2004), image retargeting (Avidan and Shamir, 2007), photograph ranking (Yeh et al., 2010), and image composition optimization (Liu, 2010); furthermore, research has been actively performed in this regard.

DOI: 10.4018/978-1-7998-3499-1.ch015

Based primarily on perceptual-psychological findings that scene contrast affects visual saliency (Einhauser, 2003; Parkhurst et al., 2002; Reinagel and Zador, 1999; Treisman and Gelade, 1980; Perazzi et al., 2012), existing saliency estimation methods (Achanta et al., 2009; Chang et al., 2011; Duan et al., 2011; Erden and Erden, 2013; Fang et al., 2016; Goferman et al., 2010; Hou and Zhang, 2007; Koch and Ulman, 1985; Tavakoli et al., 2011; Wang et al., 2011) are primarily bottom-up approaches that exploit various types of contrast measures with respect to image features such as intensity, color, local orientation, gradients, spatial frequencies, and other local descriptors. Moreover, a few hybrid bottom-up/top-down approaches that incorporate prior knowledge regarding tasks as a top-down cue exist. However, most of these saliency estimation methods are restricted to typical images with narrow field-of-view (FOV) and low dynamic range (LDR), as described below.

First, most of the abovementioned existing methods are tailored for narrow FOV images with uniform spatial resolution and without polar singularities. Consequently, a wide range of luminance information around 360-degree is dismissed as inappropriate for use. This is because omnidirectional images present two problems: nonuniformity in spatial resolution and singularity of spherical polar coordinates. Standard image processing calculations become impossible or unstable in the vicinity of poles or singular points of an omnidirectional image. Another limitation in conventional saliency estimation models is that they are likely difficult to be applied in high dynamic range (HDR) images that have more than the typical bit depth of 8 bits per pixel and channel, which can encompass a wide dynamic range of observed scenes such as outdoor scenes with bright sunlight and indoor scenes where artificial lights are much brighter than the remainder of the scene (Debevec and Malik, 1997; Reinhard, 2005; Spivak et al., 2009). This is because, as mentioned above, conventional bottom-up methods primarily use certain contrast features, and the feature measures are dependent upon the dynamic range of the input images. Meanwhile, it is difficult to collect accurate learning data by gaze measurement using 360-degree omnidirectional HDR images. This is because the viewing angle of the human eyes is only approximately 200°, and that HDR display devices are not popular yet owing to the high cost of the technologies. Hence, learning approaches such as deep learning are necessitated when targeting omnidirectional HDR images.

A new decomposition-and-integration–based method for saliency estimation of omnidirectional HDR images, which enables reuse of existing code of arbitrary saliency estimation method for typical images with narrow FOVs and LDRs is proposed herein. First, the proposed method decomposes a given omnidirectional HDR image into multiple partially overlapping LDR images with quasi-uniform spatial resolution and without polar singularities, both spatially and in intensity using a spherical overset grid called the Yin–Yang grid and a tone-mapping-based synthesis of multiple imaginary exposure-time images. For each decomposed image, an ordinary method for saliency estimation is then used. Finally, within the framework of convex optimization, the saliency map of each decomposed image is integrated from the coordinate system of the overset grid and LDR back to the representation of the coordinate system and HDR of the original image. Compared to the existing methods, the proposed method is excellent in precisely estimating the saliency maps of omnidirectional HDR images while responding to the spatial resolution unevenness, singularities at the poles, and brighter and/or darker regions, and is expected to be applied to automatic driving and robotics applications that require omnidirectional HDR image processing algorithms.

Related Content