J. Vis. Commun. Image R. 49 (2017) 27–37

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Saliency detection for panoramic landscape images of outdoor scenes ☆

Byeong-Ju Han, Jae-Young Sim *

School of Electrical and Computer Engineering, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea

ARTICLE INFO

Article history:
Received 17 January 2017
Revised 9 June 2017
Accepted 2 August 2017
Available online 3 August 2017

Keywords:
Saliency detection
Panoramic image
Wide fields of view
Background estimation
Saliency refinement

ABSTRACT

Saliency detection has been researched for conventional images with standard aspect ratios, however, it is a challenging problem for panoramic images with wide fields of view. In this paper, we propose a saliency detection algorithm for panoramic landscape images of outdoor scenes. We observe that a typical panoramic image includes several homogeneous background regions yielding horizontally elongated distributions, as well as multiple foreground objects with arbitrary locations. We first estimate the background of panoramic images by selecting homogeneous superpixels using geodesic similarity and analyzing their spatial distributions. Then we iteratively refine an initial saliency map derived from background estimation by computing the feature contrast only within local surrounding area whose range and shape are changed adaptively. Experimental results demonstrate that the proposed algorithm detects multiple salient objects faithfully while suppressing the background successfully, and it yields a significantly better performance of panorama saliency detection compared with the recent state-of-the-art techniques.

1. Introduction

With the advent of various hand-held mobile devices equipped with cameras, we have run into the explosive growth of visual multimedia contents in our daily life. Accordingly, we have been interested in saliency detection techniques to find visually prominent regions from a large number of digital images efficiently. In particular, saliency detection has been also used to preserve contextually important data in various applications of image processing and computer vision, such as retargeting [1–3], segmentation [4–7], compression [8,9], and retrieval [10,11].

Most algorithms extract salient regions which exhibit highly distinct features compared to their surrounding regions, based on the concept of center-surround contrast. Moreover, additional prior knowledge for spatial layout of foreground objects and background can be also used: image boundaries are highly probable to belong to the background [12–17], while foreground salient objects are often located near the image center [18–23]. These assumptions have been successfully employed to improve the performance of saliency detection for *conventional* images with standard aspect ratios, where the width and height are not severely different from each other. Moreover, several datasets, such as MSRA [24], SED [25], DUT-OMRON [15] and ECSSD [26], have been introduced to provide conventional images with similar aspect ratios for the purpose of performance evaluation.

Recently, *panoramic* images, which yield wide fields of view, become popular multimedia contents and draw much attention in many practical applications. For example, virtual reality contents exhibit wide fields of view when used for wearable devices such as head-mounted display. Around view monitoring systems for autonomous vehicles use 360° panoramic images by combining multiple images captured at different viewing positions. These panoramic images can be directly obtained by using special devices, or can be generated by combining several conventional images with small aspect ratios using image stitching techniques [27]. Therefore, it takes more interest to develop efficient processing techniques for panoramic images. However, the assumptions used for detecting saliency of conventional images do not completely reflect the characteristics of panoramic images. Fig. 1 shows two typical panoramic landscape images of outdoor scenes, where the dog and the horses are perceived as salient. When we obtain saliency maps for these panoramic images using the existing state-of-the-art saliency detection techniques [13,15], the foreground objects as well as some background regions of the mountain and the grass are highlighted to be salient, as shown in the third and fourth rows in Fig. 1. In addition, we divide a panoramic image vertically into three sub-images with ordinary aspect ratios, and apply the saliency detection algorithm [13] on each of the three sub-images respectively. Then we stitch the three saliency maps together to generate a panoramic saliency map, as shown

---

☆ This paper has been recommended for acceptance by Zicheng Liu.

* Corresponding author.

E-mail addresses: bjhan@unist.ac.kr (B.-J. Han), jysim@unist.ac.kr (J.-Y. Sim).

http://dx.doi.org/10.1016/j.jvcir.2017.08.003

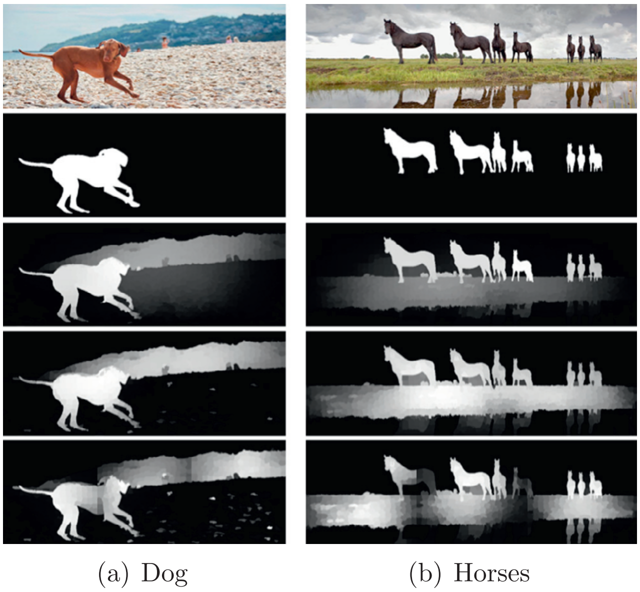

Fig. 1. Results of saliency detection for two panoramic images. From top to bottom: input panoramic images, the ground truth saliency maps, the resulting saliency maps obtained by using MR [15] and SO [13], respectively, and the saliency maps where we divide a panoramic image into three sub-images and detect saliency in each sub-image independently by using SO.

in the last row in Fig. 1. We see that the background regions of the mountain and the grass are still highlighted to be salient, and some foreground objects located at the boundaries of sub-images are not completely detected. The existing methods of MR and SO detect image saliency based on the assumption that the image boundaries usually belong to the background, which is suited for conventional images. On the contrary, in a typical panoramic image, the background is often composed of several homogeneous regions such as sky, mountain, and ground, which may rarely touch the image boundaries. Moreover, multiple foreground objects with different features and sizes are arbitrarily located in a panoramic image. Therefore, the existing methods often fail to detect salient objects and to suppress the background regions effectively in panoramic images.

In this work, we propose a saliency detection algorithm for panoramic landscape images of outdoor scenes. We observe several characteristics of typical panoramic images compared to conventional images. A panoramic image has a much larger width than height, and hence the background is distributed over horizontally elongated area. Moreover, the background is usually composed of several homogeneous regions such as sky, mountain, and ground. Also, a typical panoramic image includes multiple foreground objects with different features and sizes, which are arbitrarily located in an image. Due to these characteristics, it is difficult to devise a global approach to extract multiple salient regions directly from an input panoramic image. Therefore, we first estimate the background distribution in a panoramic image as global processing, and use the estimated background map as an initial saliency distribution which is then iteratively refined by performing locally adaptive saliency detection.

This work has the following contributions and benefits. As far as we know, this is the first work of saliency detection for panoramic images. We estimate the background in a panoramic image based on the observation that each homogeneous background region has similar pixel values and occupies a horizontally large connected area. Moreover, we perform saliency detection locally in a panoramic image, by applying the center-surround contrast only within a local surrounding area for each target region which changes its size and shape adaptively. We also think that this research may help to figure out the perception characteristics of human visual system for large-scale visual contents over a wide field of view.

The rest of this paper is organized as follows. Section 2 briefly reviews related work of saliency detection for conventional images. Sections 3 and 4 explain the proposed algorithms of background estimation and saliency map refinement for panoramic images, respectively. Section 5 presents the experimental results. Section 6 concludes this paper.

2. Related work

Saliency detection for conventional images can be implemented based on either top-down or bottom-up models. Top-down models [28–31] require high level interpretation usually provided by training sets in supervised learning. Contextual saliency was formulated according to the study of visual cognition: global scene context of an image is highly associated with a salient object [28]. The most distinct features are selected by information theoretic methods [29,30]. Salient objects are detected by joint learning of a dictionary for object features and conditional random field classifiers for object categorization [31].

In contrary, bottom-up models [12–23,32–34] do not require prior knowledge such as object categories, but obtain saliency maps by using low level features based on the center-surround contrast. They compute feature distinctness of a target region, e.g., pixel, patch or superpixel, compared to its surrounding regions locally or globally. For example, feature difference is computed across multiple scales, where a fine scale feature map represents the feature of each pixel while a coarse scale feature map describes the features of surrounding regions [32]. Also, to compute center-surround feature contrast, spatially neighboring pixels are assigned different weights [33], or random walk on a graph is used [34].

In addition, most bottom-up models can be combined with the assumptions of center prior or boundary prior. The *center prior* assumes that foreground salient objects are usually positioned near the image center, and thus assigns high saliency values to the image center [18–23]. A distance of pixels from the image center is combined with the other features to reduce the contribution of pixels far from the image center to compute object saliency [18,19]. To emphasize the region near the image center, an initial saliency map is multiplied by a Gaussian distribution centered at the image center [20] or at the center of the initial saliency map [21]. Multiple Gaussian distribution maps are also employed to weight the features of pixels adaptively according to the locations of salient objects in an image [22]. Convex-hull is used to estimate the center of a salient object when it is not strictly positioned at the image center [23]. However, this assumption puts a strict constraint on the location of foreground object in an image, and thus may not be applicable to various images.

The *boundary prior* assumes that image boundaries are highly probable to belong to the background, and is more flexible than the center prior [12–17]. Geodesic distance from a target region to image boundaries is computed, and the region with short geodesic distance is determined as background [12]. Also, more reliable saliency maps are obtained by computing the boundary connectivity using geodesic distance [13]. Moreover, they have an additional benefit to highlight the interior of salient objects as well, since the geodesic distances from image boundaries to the inside of foreground objects are always larger than that to the object boundaries. Multiple complementary background maps are estimated by measuring feature distinctness from that of the boundary superpixels [14]. Labeling the boundary nodes as non-salient, four background maps obtained by taking the four sides of an image respectively, are combined together to generate an initial saliency