Image and Vision Computing 32 (2014) 405–413

Contents lists available at ScienceDirect

Image and Vision Computing

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

A novel monochromatic cue for detecting regions of visual interest ☆

Chanho Jung [a,*], Wonjun Kim [b], Seungwoo Yoo [c], Changick Kim [d]

[a] IT Convergence Technology Research Laboratory, Electronics and Telecommunications Research Institute (ETRI), Yuseong-Gu, Daejeon 305-700, Republic of Korea
[b] Advanced Media Lab., Future IT Research Center, Samsung Advanced Institute of Technology (SAIT), Republic of Korea
[c] Qualcomm Research Korea, Seoul, Republic of Korea
[d] Department of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Yuseong-Gu, Daejeon 305-732, Republic of Korea

ARTICLE INFO

Article history:
Received 25 January 2013
Received in revised form 23 February 2014
Accepted 2 April 2014
Available online 13 April 2014

Keywords:
Regions of interest (ROIs)
Monochromatic cues
Visual attention
Taxonomy
Performance comparison of algorithms and systems

ABSTRACT

Finding regions of interest (ROIs) is a fundamentally important problem in the area of computer vision and image processing. Previous studies addressing this issue have mainly focused on investigating chromatic cues to characterize visually salient image regions, while less attention has been devoted to monochromatic cues. The purpose of this paper is the study of monochromatic cues, which have the potential to complement chromatic cues, for the detection of ROIs in an image. This paper first presents a taxonomy of existing ROI detection approaches using monochromatic cues, ranging from well-known algorithms to the most recently published techniques. We then propose a novel monochromatic cue for ROI detection. Finally, a comparative evaluation has been conducted on large scale challenging test sets of real-world natural scenes. Experimental results demonstrate that the use of our proposed monochromatic cue yields a more accurate identification of ROIs. This paper serves as a benchmark for future research on this particular topic and a steppingstone for developers and practitioners interested in adopting monochromatic cues to ROI detection systems and methodologies.

1. Introduction

Upon seeing images, humans have a natural tendency to pay more attention to some parts of the images. Over the last few decades, computational modeling of such a mechanism (i.e., selective visual attention) has attracted much research interest in computer vision because of its potential usefulness: many computer vision tasks (e.g., image compression [1,2], image quality assessment [3,4], image segmentation [5–7], image watermarking [8], image classification [9], image retrieval [10], surveillance [11], object detection and recognition [12], robot control, navigational assistance, etc.) can greatly benefit from the ability to identify regions of interest (ROIs), which attract the observer's attention, in an image. A large number of methods for detecting such visually conspicuous image regions have been reported in the literature so far, and they mainly belong to either of two broad categories: top-down and bottom-up. The top-down approaches are directed by prior expectations, i.e., they are task-specific or goal-driven. More specifically, with the proliferation of (pretrained) object recognition systems, they may attempt to estimate the likelihood of existence of interesting objects (e.g., face [13]) at each location and scale in an image. The vast majority of ROI detection methods focus on the behavior of bottom-up attention. The bottom-up approaches often use findings from psychophysics and physiology. Treisman and Gelade [14] suggest the feature integration theory (which is one of the most influential psychological models of bottom-up attention). They claim that a visual stimulus is composed of separable features (e.g., intensity, orientation, and color). Koch and Ullman [15] introduce the concept of saliency map, which quantifies visual attractiveness at each pixel in an image. Building upon their achievement, Itti and colleagues [16] suggest a biologically motivated computational model, and many variants of it are proposed (e.g., [17]).

The central aim of this paper is to study "monochromatic cues" in the modeling of bottom-up attention. Due to their importance in such a modeling [18,19], chromatic cues have been extensively studied in the literature [20–23], whereas investigating monochromatic cues has received less attention. Note here that the monochromatic cues are an important and beneficial complement to the chromatic cues, and this is the motivation of this study: for that reason, to detect salient regions, several authors have employed the monochromatic cues as well as the chromatic cues (e.g., [11,16,24,25]). We first present a taxonomy of existing methods to find ROIs from monochromatic cues, and then introduce a new validated monochromatic cue for the detection of ROIs in an image. We note that the proposed approach is targeted towards "natural scenes" (see examples in Figs. 1 and 7, 8, and 9). Our proposed method is inspired by the fact that human visual perception is highly adaptive and sensitive to structural information in images [26]: we show how such an attribute of human visual system (HVS) can be effectively extended to solve the problem of ROI detection and prove that, unlike existing biologically inspired methods, the proposed method has great ability of providing powerful contextual information

☆ This paper has been recommended for acceptance by Qiang Ji.
* Corresponding author. Tel.: +82 42 860 6571; fax: +82 42 860 6508.
E-mail addresses: peterjung@etri.re.kr (C. Jung), jazznova@kaist.ac.kr (W. Kim), yoos@qti.qualcomm.com (S. Yoo), changick@kaist.ac.kr (C. Kim).

http://dx.doi.org/10.1016/j.imavis.2014.04.001
0262-8856/

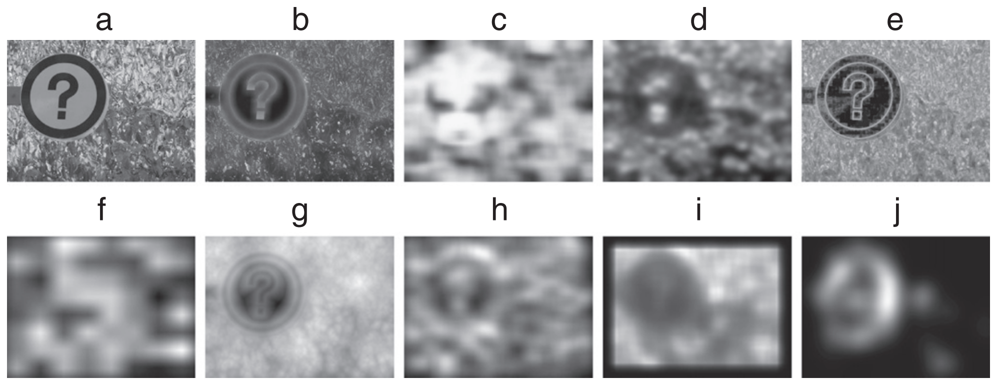

Fig. 1. Comparing some selected ROI detection techniques. (a) Original image (8 bits/pixel). (b) Ma and Zhang [28]. (c) Seo and Milanfar [29]. (d) Kim et al. [11]. (e) Mancas et al. [31]. (f) Gopalakrishnan et al. [25]. (g) Rosin [32]. (h) Guo et al. [2,37]. (i) Wang et al. [40]. (j) Ours. Throughout this paper, the larger the pixel intensity is, the more likely the pixel attracts the observer's interest. Note that, unlike existing methods, our proposed method gives robust results even in cluttered natural scenes (see also Figs. 7, 8, and 9).

regarding ROIs. We finally conducted an empirical comparative evaluation of the algorithms.

The rest of this paper is organized as follows. In Section 2, a taxonomy of previous techniques to detect ROIs from monochromatic cues is provided. Section 3 describes the details of our proposed method. Experimental results are reported in Section 4, followed by concluding remarks in Section 5.

2. Detecting regions of interest from monochromatic cues: A taxonomy of previous work

In this section, we present a taxonomy of existing bottom-up methods to find ROIs from monochromatic cues. One important advantage of the bottom-up approaches is that they do not need any priors. Based on the nature of approach and features used, these methods may be broadly divided into two categories: 1) biologically inspired methods and 2) purely computational methods. ROI detection methodologies in each category may be further classified, i.e., spatial-domain and frequency-domain approaches or local and global approaches. Table 1 summarizes the taxonomy of 16 different ROI identification techniques using monochromatic cues.

The biologically inspired methods attempt to simulate mechanisms of preattentive vision: specifically, they focus on the neurobiological hypothesis that human vision may preferentially respond to “high contrast stimuli” [27]. These approaches first extract low-level visual features such as intensity, orientation, and texture, then identify ROIs with local or global contrast analysis (e.g., [16]). Ma and Zhang [28] compute the local spatial contrast of image intensity at each location. The authors argue that the locations with high feature contrast also often have rich information. Liu et al. [24] propose to calculate such a visual feature in a multi-scale manner. Seo and Milanfar [29] suggest a bottom-up method motivated by the center-surround contrast mechanism [16,30] of preattentive biological vision. Rather than taking the classical image features, they propose the use of nonparametric kernel density, which is designed to capture local data structure, as a feature. Kim et al. [11] also exploit the center-surround paradigm based on an ordinal signature of feature distribution. The rationale behind this approach is the fact that the ordinal measure of edge orientation histogram indicates

Table 1
Taxonomy of 16 different approaches to detect ROIs from monochromatic cues.

Publication	Model	Domain	Features	Analysis	Approach
Itti[2] [16]	BIO	S	Intensity and orientation	L	Center-surround hypothesis
Ma [28]	BIO	S	Intensity	L	Feature contrast Fuzzy growing
Liu[2] [24]	BIO	S	Intensity	L	Multi-scale contrast
Seo[1] [29]	BIO	S	Nonparametric kernel density	L	Center-surround hypothesis
Kim[1,2] [11]	BIO	S	Ordinal signature of edge orientation	L	Center-surround hypothesis
Mancas [31]	BIO	S	Local mean Local variance	G	Information theory (self-information)
Gopalakrishnan[2] [25]	BIO	S	Orientation	L-G	Feature entropy contrast
Rosin [32]	COM	S	Edge density	G	Threshold decomposition Distance transform Thresholding
Deng [33]	COM	S	Edge density	G	Closed curve detection
Cohen [34]	COM	S	Contour closure Contour convexity Contour size	G	Contour fragment grouping
Caron [35]	COM	S	Power law distribution Inverse power law distribution	G	Clustering
Hou [36]	COM	F	Spectral residual	G	Fourier transform
Guo[1] [2,37]	COM	F	Phase spectrum	G	Fourier transform
Bian[1] [38]	COM	F	Flattened amplitude spectrum	G	Fourier transform
Xu [39]	COM	S-F	Pseudo-Wigner–Ville distribution	L	Information theory (Renyi entropy)
Wang [40]	COM	S	Sparse coding basis functions	L	Site entropy rate

(BIO = BIOlogically inspired model, COM = purely COMputational model, S = Spatial, F = Frequency, L = Local, and G = Global).
[1] In this paper, we deal with only still images.
[2] Note that, for fairness, only (isolable) monochromatic cues are used in the implementations of [11,16,24,25] in this paper.