Available online at www.sciencedirect.com

ScienceDirect

October 2017, 24(5): 68–76
www.sciencedirect.com/science/journal/10058885

The Journal of China Universities of Posts and Telecommunications

http://jcupt.bupt.edu.cn

# Extraction technique of region of interest from stereoscopic video

Lü Chaohui (✉), Pan Jiaying

School of Information Engineering, Communication University of China, Beijing 100024, China

## Abstract

A feature fusion approach is presented to extract the region of interest (ROI) from the stereoscopic video. Based on human vision system (HVS), the depth feature, the color feature and the motion feature are chosen as vision features. The algorithm is shown as follows. Firstly, color saliency is calculated on superpixel scale. Color space distribution of the superpixel and the color difference between the superpixel and background pixel are used to describe color saliency and color salient region is detected. Then, the classic visual background extractor (Vibe) algorithm is improved from the update interval and update region of background model. The update interval is adjusted according to the image content. The update region is determined through non-obvious movement region and background point detection. So the motion region of stereoscopic video is extracted using improved Vibe algorithm. The depth salient region is detected by selecting the region with the highest gray value. Finally, three regions are fused into final ROI. Experiment results show that the proposed method can extract ROI from stereoscopic video effectively. In order to further verify the proposed method, stereoscopic video coding application is also carried out on the joint model (JM) encoder with different bit allocation in ROI and the background region.

**Keywords** stereoscopic video, depth, saliency, ROI, Vibe

## 1 Introduction

An ROI is the region that people will focus on when watching images and videos. These regions are more attractive than other regions, and have higher saliency. There have developed a lot of methods for saliency detection and ROI extraction. Itti et al. [1] proposed a method that images are firstly filtered by Gaussian and downsampled to obtain the multi-scale image pyramids with the intensity, color and orientation features, and then exploited the central peripheral difference algorithm to get the multi-scale graphs of these features. Finally the fusion method is used to get saliency map. Achanta et al. [2] presented a frequency-tuned approach of computing saliency in images using low level features of color and luminance, which is fast and easy to implement. It also provides full resolution saliency maps. Xie et al. [3] proposed a computational saliency detection model which is implemented with a coarse to fine strategy under the Bayesian framework. First, saliency points are applied to get a coarse location of the saliency region. Then they compute a prior map for the Bayesian model to achieve the final saliency map based on the rough region. Cheng et al. [4] proposed a regional contrast based salient object detection algorithm, which simultaneously evaluates global contrast differences and spatial weighted coherence scores. Perazzi et al. [5] presented saliency filters, a method for saliency computation based on an image abstraction into structurally representative elements and contrast-based saliency measures, which can be consistently formulated as high-dimensional Gaussian filters. Their filter-based formulation allows for efficient computation and produces per-pixel saliency maps. Klein et al. [6] extracted ROIs with the knowledge of information theory, which measures the distribution difference of visual features between the center and its surrounding using Kullback-Leibler (KL) divergence. Imamoglu et al. [7] utilized the lower-level features produced by wavelet transform to extract ROIs and obtained better performance based on wavelet domain than frequency domain.

---

Received date: 07-06-2017

Corresponding author: Lü Chaohui, E-mail: llvch@hotmail.com

DOI: 10.1016/S1005-8885(17)60235-8

In the video saliency detection area, Zhai et al. [8] utilized scale invariant feature transform (SIFT) to establish the correspondence between the points of interest in the continuous video frame, and motion plane was detected. The color differences between the pixels were calculated to expand into a salient region of color, and the final salient region was acquired. The Vibe algorithm was proposed by Barnich et al. [9] to initialize the background model and update the background model through a randomly selected strategy. In order to solve the ghosting problem that occurred in the Vibe algorithm, Chen et al. [10] used the pixel value of the ghost region and compare it with the pixel value of the neighbor field to suppress the ghosting. Yin et al. [11] eliminated the false shadow part by detecting the hue, saturation, value (HSV) color space pixel values in the foreground region. Wang et al. [12] proposed a threshold based adaptive Vibe target detection algorithm. The improved algorithm changed the update rate of the background adaptively according to calculate the change of the centroid of the moving object.

According to the selection characteristics of HVS, the color feature, motion feature and depth feature of stereoscopic video are chosen as the visual and salient feature to be detected. A new feature fusion approach is proposed using the above three feature. Firstly, the extraction method of color region and moving region are introduced, and then depth salient region extraction is achieved. Finally, three feature regions are merged to obtain the final ROI. In order to further verify the validity of the final ROI of stereoscopic video, region based bit allocation scheme is performed on the JM encoder with different quantitative parameter (QP).

## 2 Extraction of the ROI

### 2.1 Extraction of color salient region

The existing image saliency detection algorithm can extract the ideal region in the image area when salient region is single, the salient region and the background region have significant color, brightness and texture difference. Such as Achanta et al. [13], Perazzi et al. [5], Cheng et al. [14] and so on. However, when the image composition or background is complex, such as the German heinrich hertz institute (HHI) book arrival stereo video sequence whose scene contains the object with different sizes and many kinds of color, the above algorithms are not valid for saliency detection. An improved method is proposed in this paper, whose block diagram is illustrated in Fig. 1.

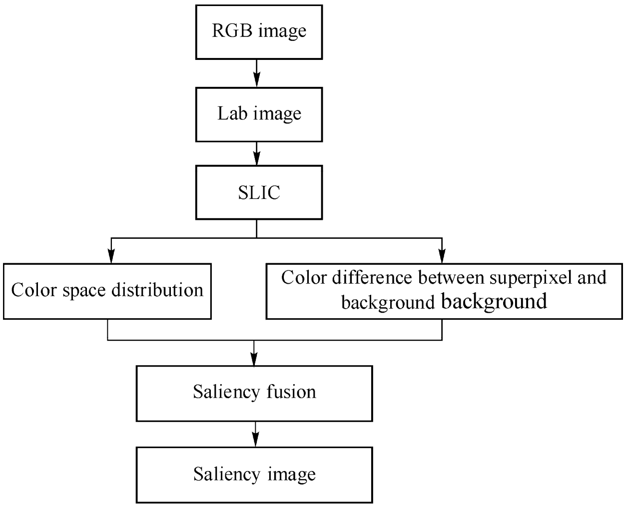

**Fig. 1** Block diagram of color saliency detection

Firstly, the input image is converted from the red green blue (RGB) color space to the Lab color space. Secondly, simple linear iterative clustering (SLIC) algorithm [13] is performed on the Lab color space to produce a compact superpixel that is generally uniform in size and maintains the color boundary. Then the saliency is calculated on the superpixel scale. The saliency of the image is described by the color space distribution of the superpixel and the color difference between the superpixel and the background pixel.

Color space distribution of the superpixels is calculated by the formula proposed in Ref. [5].

$$D_i = \sum_{j=1}^{N} \left\| P_j - \mu_i \right\|^2 \omega\left(c_i, c_j\right) \tag{1}$$

$$\mu_i = \sum_{j=1}^{N} \omega(c_i, c_j) P_j \tag{2}$$

where $D_i$ is the $i$th superpixel color space distribution. $P_j$ is the center position of the $j$th superpixel. $\mu_i$ represents the weighted average position of the $i$th superpixel. $\omega\left(c_i, c_j\right)$ is the color distance between $i$th superpixel and $j$th superpixel center, which is calculated in the Lab color space.

Eq. (1) represents the breadth of the spatial distribution of the color.

Color differences between superpixel and background pixel are calculated as follows:

1) For all superpixels in the image, each two pairs of superpixels in the RGB color space are calculated to get the color difference