Cheng et al. EURASIP Journal on Image and Video Processing (2017) 2017:61
DOI 10.1186/s13640-017-0210-5

EURASIP Journal on Image and Video Processing

RESEARCH Open Access

Stereoscopic visual saliency prediction based on stereo contrast and stereo focus

Hao Cheng[1,2], Jian Zhang[2], Qiang Wu[2], Ping An[1*] and Zhi Liu[1]

Abstract

In this paper, we exploit two characteristics of stereoscopic vision: the pop-out effect and the comfort zone. We propose a visual saliency prediction model for stereoscopic images based on stereo contrast and stereo focus models. The stereo contrast model measures stereo saliency based on the color/depth contrast and the pop-out effect. The stereo focus model describes the degree of focus based on monocular focus and the comfort zone. After obtaining the values of the stereo contrast and stereo focus models in parallel, an enhancement based on clustering is performed on both values. We then apply a multi-scale fusion to form the respective maps of the two models. Last, we use a Bayesian integration scheme to integrate the two maps (the stereo contrast and stereo focus maps) into the stereo saliency map. Experimental results on two eye-tracking databases show that our proposed method outperforms the state-of-the-art saliency models.

Keywords: Stereoscopic image, Stereoscopic visual saliency prediction, Stereo contrast, Stereo focus

1 Introduction

Visual attention is a very important research topic in computer vision, as it is widely used in the field for many tasks, such as object detection [1] and video/image retrieval [2, 3]. Computational models of visual attention, which simulate the attention mechanism of humans, have been built by researchers in many fields, such as visual neuroscience, computer vision, and multimedia processing [4]. Visual attention enables the discovery of an object or region that efficiently represents a scene and, thus, harnesses complex vision problems, such as scene understanding.

The models of visual attention are usually divided into two categories: bottom-up and top-down [5]. The bottom-up model is a rapid data-driven task-independent process and is usually feed-forward. A prototypical example of a bottom-up model is the act of looking at a scene which has only one horizontal bar among several vertical bars, in which attention is immediately drawn to the horizontal bar [6]. Top-down model considers high-level cognitive features to quantify the visual saliency, such as human faces [7] and prior knowledge about the target [8]. Of these top-down features, prior knowledge about the target is difficult to model. Recently, a number of saliency models have incorporated both top-down and bottom-up feature detection in an effort to improve prediction accuracy [9]. Wei et al. [10] turned to background priors to guide the generic object level saliency detection. Goferman et al. [11] and Judd et al. [7] integrate high-level information, making their methods potentially suitable for specific tasks.

These models are mainly designed for 2D images. With the rapid development of 3D technology, many devices for stereoscopic capture have appeared. For example, the Panasonic 3D camera captures the stereoscopic images and video for 3D movies. The Kinect-1 device by Microsoft for the XBox captures both the color map and the depth map at the same time, which can generate the stereoscopic images (the depth map of the Kinect-1 may have holes that need to be smoothed [12], which may cause noise). These devices make up a number of applications for 3D images or videos, such as 3D rendering [13], 3D visual quality assessment [14], and 3D video detection [15]. These 3D applications increase the need for saliency modeling for 3D visual content.

* Correspondence: anping@t.shu.edu.cn
[1]The School of Communication and Information Engineering, Shanghai University, Shanghai, China
Full list of author information is available at the end of the article

Stereo saliency models can be classified into two categories according to the way they use the depth factor: stereo-vision models and depth-saliency models.

Stereo-vision models take into account the mechanisms of stereoscopic perception in the human visual system (HVS). This type of model considers the characteristics of depth factors and color information. Bruce and Tsotsos extended the 2D model, which uses a visual pyramid processing architecture [16], by adding neuronal units to model the stereo vision; however, they did not propose a computational model in that study. Based on our knowledge, designing the stereo-vision model is very difficult and we only find two models in [17], because the mechanisms of stereo vision still pose several research challenges, such as how to build then apply the model for the stereoscopic vision mechanism.

Depth-saliency models take depth saliency as a feature of saliency measurement, and methods of formulating and using depth saliency fall into two further categories. One category relies on a depth-saliency map (DSM) [17, 18]. The depth saliency is extracted from the depth map or disparity map (usually based on depth contrast or the depth pop-out effect) to create an additional depth-saliency map. The final result combines the 2D saliency maps (from 2D saliency models usually using color contrast, intensity, or image texture) and the depth-saliency maps (DSM). The other category builds the model directly. In other words, it builds the stereoscopic visual saliency prediction model by taking the mechanisms of stereoscopic perception in the HVS into account. It designs the model by fusing the depth and 2D features into the saliency measurement, based on the mechanisms of the HVS [19].

Kim et al. [15] designed a stereoscopic visual attention algorithm for 3D video based on multiple perceptual stimuli, which assumes that pixels closer to observers and at the front of the screen are more salient. Niu et al. [20] explored stereo saliency by analyzing the characteristics of stereo vision and proposed a depth saliency model for a depth map that would expand the 2D saliency model for stereo saliency analysis. However, the proposed model does not fully explore the relationship between the depth model and the 2D saliency model. Fan et al. [19] proposed a stereo saliency model based on region-level depth, color, and spatial information. Wang et al. [17] proposed a computational model that takes the depth factors as an additional visual dimension and provides a public database with a ground truth of eye-tracking data. Fang et al. [21] proposed a visual attention model for stereoscopic images based on the contrast between low-level features. However, they did not consider the characteristics of human stereo vision, such as the pop-out effect or 3D fatigue.

According to the above analysis, the key issue for a 3D visual saliency prediction model is how to adopt the depth factor and how to combine the depth factor with 2D information based on the mechanisms of HVS. In our earlier work [22], a novel saliency model for stereoscopic images was proposed. However, this model did not deeply exploit the HVS characteristics of the pop-out effect and comfort zone and only treated the depth information as a weight. In this paper, we deeply analyze two characteristics of the stereoscopic vision: pop-out effect and comfort zone. Based on these characteristics, we design two stereo-vision models for visual saliency prediction: one based on stereo contrast and the other based on stereo focus. We enhance these two models by clustering and then integrate them into the final stereoscopic saliency map.

The main contributions of this paper are as follows:

1. We propose a stereo contrast model for detecting stereo saliency. This model detects saliency based on color and depth contrast and the pop-out effect.
2. We propose a stereo focus model for detecting stereo saliency. This model detects the degree of focus via monocular focus and the comfort zone.
3. We propose an enhancement to increase the performance of the stereo contrast and stereo focus models.

The rest of the paper is organized as follows: In Section 2, we introduce the two mechanisms of stereo human vision for stereo saliency analysis. Section 3 proposes a new stereo visual saliency prediction method based on the stereo contrast and stereo focus models. Section 4 describes a quantitative comparison of the proposed model and state-of-the-art algorithms. Section 5 provides the research outcomes and future work.

2 Methodology

When watching a stereoscopic image, people experience different effects, such as the pop-out effect and deep-in effect [23]. When we watch a stereoscopic image/video, the pop-out effect occurs when an object looks like it is going to pop out of the screen and the deep-in effect occurs when an object looks like it is behind the screen. To obtain these two effects, we can control the parallax of objects, such as the negative or positive parallax as shown in Fig. 1. This finding is based on recent research on human stereo vision [24]. These effects cause viewers to feel immersed in the image, which is the most attractive aspect of stereoscopic images. Moreover, studies show that an object, which has the pop-out effect often, catches a viewer's attention [25]. This phenomenon provides a useful depth cue for stereo saliency analysis, since objects with a pop-out effect are usually more salient than objects that have a

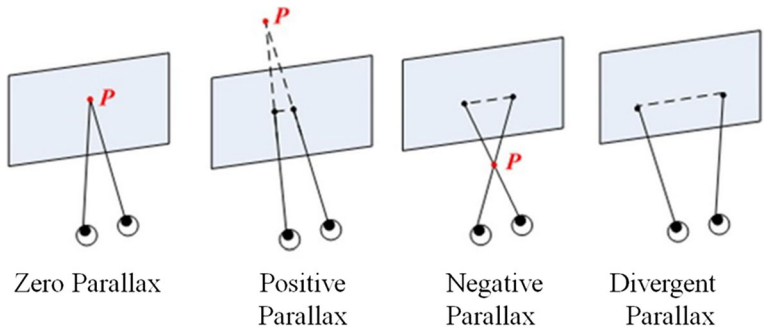


Fig. 1 Stereo perception based on the different parallax [24]

deep-in effect. We assume that the object with the pop-out effect tends to be more salient than the other objects. In addition, we use color/depth contrast for the stereo saliency analysis. Hence, we propose a stereo contrast model to simulate the pop-out effect by combining the color/depth contrast and pop-out value.

Another property of stereo vision is the viewing comfort zone based on the binocular information. Viewers may experience fatigue when they spend a long time watching stereoscopic images or video. The reason for this may be accommodation-vergence conflict or too much divergence [26, 27]. A good stereoscopic image needs to minimize 3D viewer fatigue. This conflict increases as the perceived depth of an object becomes further away from the screen, as shown in Fig. 2. The zone close to the screen plane is called the comfort zone. Photographers usually make sure the more important objects are in the comfort zone when they capture a stereoscopic image or video. This is another depth cue for saliency analysis: the object in the comfort zone tends to be more salient than other zones. Studies show that the object near the zero disparity plane is more salient than those which are away from the zero disparity plane, which can be described by the linear formulation [20]. When a person watches one salient object, this object should be in the focus region [9]. According to the above phenomenon, in the perspective of the comfort zone, this object should meet two conditions: one is that it is located in or near the comfort zone and the second is that it is in the focus region. Therefore, we use monocular focus and comfort zone to analyze stereo saliency. The monocular focus assumes that the salient object is usually located in the focus region. The comfort zone is treated as a weight to adjust the importance of the object located in the focus region. The proposed stereo focus model is based on the comfort zone and monocular focus.

In order to describe the two mechanisms of the human visual system: pop-out effect and comfort zone, we have chosen to develop our proposed model on a combination of the stereo contrast and stereo focus models of the stereo-vision model. The stereo saliency of an object can be determined by the values calculated from the stereo contrast and stereo focus models. However, in some cases, the values obtained by these two models can be substantially different. For example, if an object has negative parallax and is far from the comfort zone, or if the object has zero parallax, the two values are quite different. To obtain the benefits from two models and detect the saliency for different stereoscopic content, our stereo visual saliency prediction model considers both the stereo contrast model and the stereo focus model.

3 Proposed stereoscopic visual saliency prediction model

The proposed stereoscopic visual saliency prediction framework is shown in Fig. 3. To capture the structural

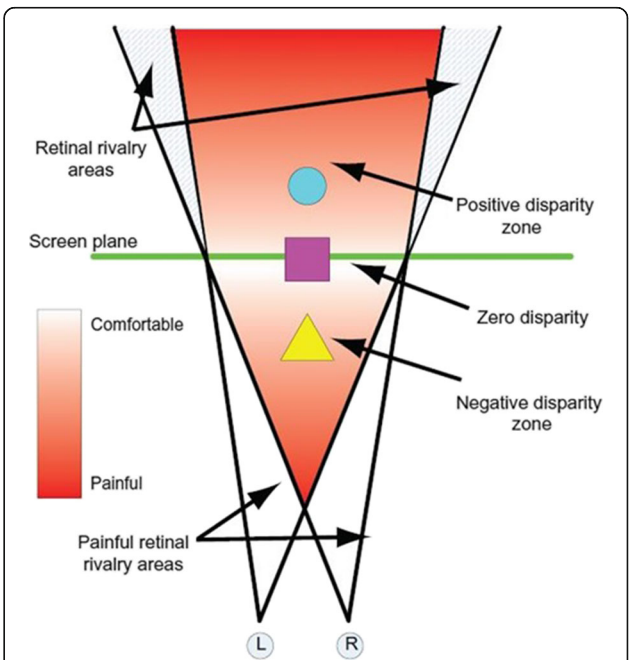


Fig. 2 Stereo comfort zone based on human stereo vision [20]

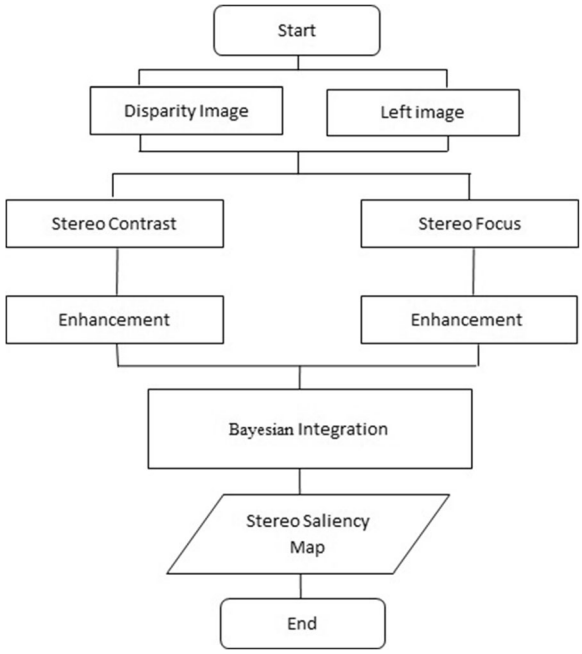


Fig. 3 The framework of the proposed stereo saliency model

information of the stereoscopic image, we first adopt a simple linear iterative clustering (SLIC) algorithm [28] for the segmentation. The SLIC algorithm can segment an input image (left image) into multiple uniform and compact superpixels. By controlling the number of superpixels in the SLIC algorithm, the image is segmented into multi-scale images. Then, we calculate the saliency values individually by applying the stereo contrast and stereo focus models for each superpixel based on the left image and disparity map. An enhancement is based on clustering and increases the performance of the two models according to the experiments. Multi-scale fusion is then used to form the pixel-level stereo contrast and stereo focus maps. Last, the two maps are integrated by Bayesian integration to form the final stereo saliency map.

3.1 Pre-processing

In this paper, we convert the stereoscopic images from the RGB color space to the hue-saturation-value (HSV) color space. Compared to the RGB color space, the HSV color space is more consistent with the characteristics of human vision attention, and using it leads to a saliency value with higher accuracy [27].

As mentioned previously, we conduct multi-scale visual saliency prediction. Based on the number of superpixels, the input image (left image) is segmented into a set of non-overlapping superpixels in the scale s using the SLIC algorithm. s represents the scale of the segmentation. We chose the SLIC algorithm as the segmentation method because it is a fast and highly efficient segmentation algorithm that is sensitive to the boundary of the object [29]. Each superpixel t is described by the mean color feature $\{H, S, V\}$, coordinates of the superpixels $\{x, y\}$, and the mean disparity value d, $x_t = \{H, S, V, x, y, d\}_t$. The entire image can be represented as $X = [x_1, x_2, \ldots, x_N]_s$.

3.2 Stereo contrast model

We propose the stereo contrast model based on the color/depth contrast and the pop-out effect to calculate the saliency value (using a disparity map to analyze the pop-out effect). According to the human vision system, human attention is sensitive to a contrast region that includes color contrast and depth contrast [25]. The colors of the salient region are distinctive and contrast with the other regions. The depth discontinuity region may attract the viewer's attention when view positions or angles are changed. Therefore, the distinctive region may attract the viewer's attention to color/depth information. According to [30, 31], humans pay more attention to those image regions that contrast strongly with their surroundings. Based on our observation, the distance between neighboring regions and the area of the region plays an important role in human visual attention. To simulate the above mechanism, we define the contrast value to measure the contrast of stereoscopic information.

Let $\mathrm{DC}(i, j)$ be the Euclidean distance between the vectorized superpixels i and j in HSV color space and $\mathrm{DD}(i, j)$ be the Euclidean distance between superpixels i and j in disparity. DC and DD are normalized to the range [0, 1]. We define the contrast measure $C(i, j)$ between superpixels i and j as:

$$C(i,j) = (1-a) \times \mathrm{DC}(i,j) + a \times \mathrm{DD}(i,j) \quad (1)$$

where a is a control weight to balance the color and disparity contrast. Although several approaches [17, 18, 32] combining depth-saliency maps with 2D visual features have been proposed, any specific and standardized approaches still lack the combination of saliency maps from depth with 2D visual features. The work in [17, 18] treats depth with the same importance as color. The work in [32] uses the adaptive weight for color and depth. In our experiments, we adopt a straightforward approach to merge color and depth contrast, treating depth contrast with the same importance as color contrast. We set $a = 0.5$ empirically.

Let $L(i, j)$ be the Euclidean distance between the position of superpixels i and j normalized to the range [0, 1]. According to the analysis above, we define the stereo contrast measure $S(i, j)$ between a pair of superpixels i and j based on color, disparity, and spatial information:

$$S(i,j) = \left(\frac{C(i,j)}{1 + c \times L(i,j)}\right) \times \omega_j \quad (2)$$

where ω_j is the number of pixels in superpixel j and c is a control value for spatial information (c = 3 in our implementation). As mentioned above, the saliency of a superpixel z can be defined by its stereo contrast measure as:

$$\mathrm{SC}_R(z) = \sum_{i \neq z, i \in R} S(z,i) \quad (3)$$

where R is the search range and $\mathrm{SC}_R(z)$ is the saliency value of superpixel z in the search range. Figure 4 shows the global and local search range. Then, we compute the global and local saliency maps.

When we compute the stereo contrast saliency value of the current superpixel, we do not compute all superpixels in the search range. We only choose the K most similar superpixels in the search range and use them to compute the stereo contrast saliency of the current superpixel. This is based on the experiments and [22], as using the k most similar superpixels to compute the stereo contrast can prevent the stereo contrast saliency value of an abnormal superpixel becoming too great. Therefore, in practice, to measure a superpixel's stereo contrast, we simply consider the K most similar superpixels. If the most similar superpixels are extremely different from the current superpixel, clearly all image superpixels are extremely different from it. In other words, to measure a superpixel's stereo contrast, there is no need to incorporate its stereo contrast value in all other superpixels in the search range. We simply consider K as the most similar superpixels. If most of the similar superpixels are extremely different from the current superpixel, clearly all image superpixels are extremely different from it. Therefore, we search for the K most similar superpixels k = {1, 2, ..., K}, $k \epsilon R$, where R is the search range. The local search is related to the search range R. (In practice, all distance is normalized to [0, 1] and we set $R = 0.3$ empirically.) Based on the observations of the experiments, we set K as 15 empirically. The local-global stereo contrast saliency of superpixel z is expressed as:

$$\mathrm{SC}'(z) = \sum_{k=1, k \in R}^{K} S(z,k) \quad (4)$$

According to the pop-out effect in Section 2, a region that has the pop-out effect may attract people's attention. Therefore, a pop-out effect describes the importance of the superpixel in stereoscopic saliency analysis. We treat the pop-out effect as a weight to enhance the stereo contrast saliency. Based on the work in [20] and our experiments, the superpixel of the pop-out effect can be represented by an exponential function of the disparity. We use d to represent the disparity, and d_z is the mean disparity for superpixel z which is normalized to [−1, +1]. Let o be the pop-out value for superpixel z. If $d_z < 0$, it means that the superpixel has a pop-out effect. The saliency of this superpixel should increase, and if $d_z > 0$, it means the superpixel has a deep-in effect and saliency should decrease. The pop-out value can be expressed as follows:

$$o_z = 2^{-d_z} \quad (5)$$

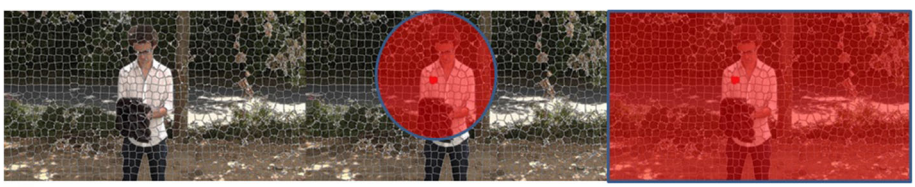
We use the local-global stereo contrast and the pop-out value to simulate the pop-out effect. Figure 5 is an example of a stereo contrast map. The stereo contrast $\mathrm{SC}(z)$ relies on the color/depth contrast, distance contrast, superpixel area, and pop-out value, which can be expressed as follows:

$$\mathrm{SC}(z) = \mathrm{SC}'(z) \times o_z \quad (6)$$

3.3 Stereo focus model

We propose a stereo focus model based on monocular focus and the comfort zone. According to the comfort zone as mentioned in Section 2, human visual attention can take the initiative to focus on the salient region by using monocular focus. Monocular focus can be detected by the focal blur [33], and we add the comfort zone to improve its accuracy.

For monocular focus, sharp edges of an object may be spatially blurred when projected on the image plane. The degree of the blur model [9] can measure the focus/defocus for the edges of the image by computing the differential-of-Gaussian (DOG) operation in a different



a SLIC **b** Local search **c** Global search

Fig. 4 a–c Global and local search range

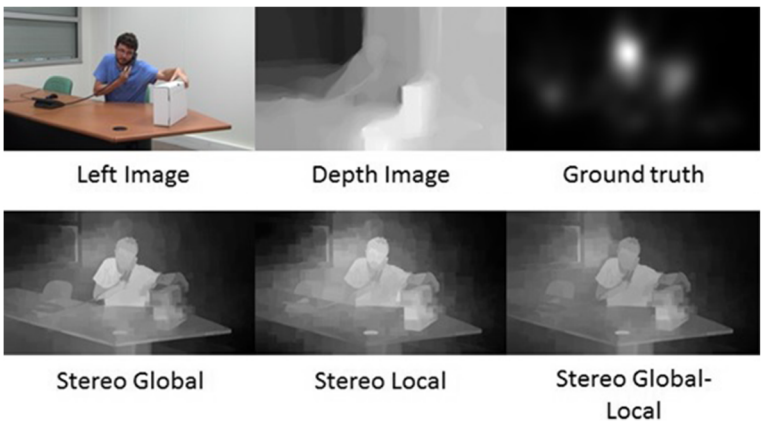


Fig. 5 An example of stereo contrast map

scale for the edge pixels. The monocular focus of the edge pixel p is $F_{2D}(p)$. This value is sensitive to the edge pixels and is easy to implement. However, it is a 2D focus measure and is only useful for the edge pixels of the image. For stereoscopic analysis, we expand this model to measure the edge of the stereoscopic focus by combining the monocular focus and the comfort zone. Then, we expand the stereoscopic focus model from edge to region.

According to our experiments, we use a comfort value to measure the comfort zone. The comfort value is a weight to indicate the object's importance by measuring the comfort zone. When multiple objects have zero or small disparity in the stereoscopic images and are located in the comfort zone, our observation is that their comfort values are similar. When they are far away from the zero disparity plane, their comfort values decrease sharply. Based on this observation, the comfort value complies with a Gaussian distribution. $v(p)$ denotes the comfort value of pixel p. This can be expressed as:

$$v(p) = \begin{cases} \exp\left(\dfrac{d_p^2}{-2\sigma_1^2}\right) & d_p \geq 0 \\ \alpha \times \exp\left(\dfrac{d_p^2}{-2\sigma_1^2}\right) + (1-\alpha) & d_p < 0 \end{cases} \tag{7}$$

where d_p represents the disparity of pixel p. σ_1 is the range of positive and negative disparity. α controls the weight of negative disparity. For negative disparity, we cannot directly follow the comfort zone model [20] to design our comfort value. The reason for this is that there is a conflict between the pop-out effect and comfort zone. If we directly use the comfort zone model [20] to measure saliency, in some cases, stereo contrast model and stereo focus model may give quite different results for an object with negative disparity, which will reduce the performance our proposed model. For example, if the pixel has a large negative disparity and is far from the comfort value, its pop-out value becomes big, and its comfort value is small. After the fusion of two models, the results may be not reliable. To reduce the errors caused by such conflicts, we increase the importance of the negative disparity in the comfort zone by using α to balance the comfort value of the negative disparity. There are two benefits in this modification. Firstly, this modification increases the importance of the pop-out effect for the object with the negative disparity. Secondly, it still keeps a high importance for the object in the comfort zone in stereoscopic saliency analysis. According to our experiments, our modification for the comfort zone works in most cases and improves the performance of the proposed model.

We set the comfort value as a weight, because the comfort value describes the importance of the stereo saliency analysis. We define the stereo focus value of the edge pixels p by combining the monocular focus value F_{2D} with the comfort value. This is expressed as:

$$F_{3D}(p) = F_{2D}(p) \times v(p) \tag{8}$$

It would be ideal to analyze the saliency for each object as a whole. However, it is difficult to segment an object accurately. Therefore, we compute the stereo saliency at the superpixel level instead. For each stereo focus value of the edge pixels, we filter it by using a Gaussian kernel of σ, equal to 1° of visual angle. This processing can effectively reduce noise, such as an isolated point. The stereo focus value of superpixel t relies on the stereo focus degree of all its pixels. Further, our observation is that a region with a sharper boundary usually stands out as being more salient. We set the boundary sharpness as a weight value, which can be represented by the stereo focus value of the boundary pixels. The stereo focus value SF(t) of superpixel t is formulated as:

$$\mathrm{SF}(t) = \frac{1}{m}\sum_{p\in B_t} F_{3\mathrm{D}}(p) \times \frac{1}{n}\sum_{q\in t} F_{3\mathrm{D}}(q) \tag{9}$$

B_t represents all the edge pixels in superpixel t, m is the number of edge pixels, and n is the number of all the pixels in superpixel t. The first term on the right-hand side of Eq. 9 is the average value of the stereo focus value for all the edge pixels. The second term is the average value of the stereo focus value for all the pixels in superpixel t. The stereo focus model is combined with the monocular focus and the comfort value. Figure 6 shows the example of the stereo focus map.

3.4 Enhancement

The stereo contrast model and stereo focus model are superpixel level. To make the salient region more distinctive and separated easily, we propose an enhancement based on clustering for the two models. In practice, we use the k-means algorithm to cluster N superpixels to K clusters via the value of superpixel t. For simplicity, we use SV to represent SC and SF (SV = SC = SF). To enlarge the difference between neighboring clusters, each value of superpixel t belonging to cluster k ($k = 1, 2, 3, \ldots, K$) is modified by considering its own value and the other superpixels in cluster k:

$$\mathrm{Sm}(t) = \delta \sum_{i=1,k_i\neq t}^{N_c} r_{tk_i}\mathrm{SV}_{k_i} + (1-\delta)\mathrm{SV}_t \tag{10}$$

where $\{k_1, k_2, \ldots, k_{\mathrm{Nc}}\}$ denotes the Nc superpixels in cluster k and t is one superpixel in cluster k. δ is the weight parameter. $Sm(t)$ is the value of superpixel t belonging to cluster k. r_{tk_i} is a weight value that relies on the value of superpixels t and k_i. The first term on the right-hand side of the equation is the weighted average of all the superpixels without superpixel t in cluster k, and the other is the weighted value of superpixel t. The weighted value is more sensitive to the spatial information of superpixel pairs:

$$r_{tk_i} = \frac{\exp\frac{\mathrm{SD}(k_i,t)}{-\sigma_2^2}}{\sum_{i=1,k_i\neq t}^{N_c}\exp\frac{\mathrm{SD}(k_i,t)}{-\sigma_2^2}} \tag{11}$$

$\mathrm{SD}(k_i, t)$ is the spatial distance between the superpixels k_i and t. σ_2 is a weight to control the range of the spatial information. After re-calculating the value of each superpixel, the values of the important superpixels in cluster k are enhanced. Figure 7 gives an example in which two maps computed by the stereo contrast and stereo focus models are processed by the enhancement.

Since the content of each superpixel may have more than one object or texture, a single scale segmentation scheme is not suitable for objects of different sizes. We conduct multi-scale segmentation based on controlling the number of superpixels in the SLIC algorithm. At each superpixel scale size layer, both the stereo contrast and stereo focus models are individually applied to calculate their respective saliency values. A multi-scale pixel-level fusion is introduced to fuse the results for each model. Through this fusion, the saliency value for each pixel is calculated based on multi-scale saliency and its texture information.

To deal with the values in the different scales, we adopt the method to fuse the multi-scale layered value [34]. This method considers the multi-scale value and its textural information, which uses the textural feature of the pixel and its corresponding superpixel as the weight value to average the multi-scale value. For each pixel, the saliency value relies on the saliency value of each scale and its corresponding weight. The weight considers the textural information that relies on the difference between the current pixel value and superpixel value.

3.5 Bayesian integration scheme

At this stage, two saliency maps have been built based on the stereo contrast and stereo focus models. The next step is to integrate them; however, as has been discussed [35], good individual saliency maps may become worse maps when they are combined by using weights. Therefore, we adopt a Bayesian model to integrate the two saliency maps [36]. For the Bayesian model, each pixel's saliency can be estimated by the posterior probability. The Bayesian integration approach is suitable for dealing with two saliency

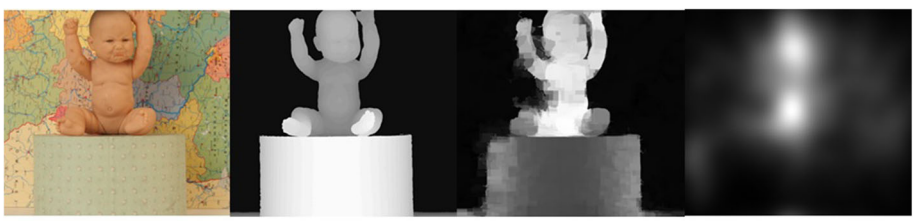


Fig. 6 The example of the stereo focus map

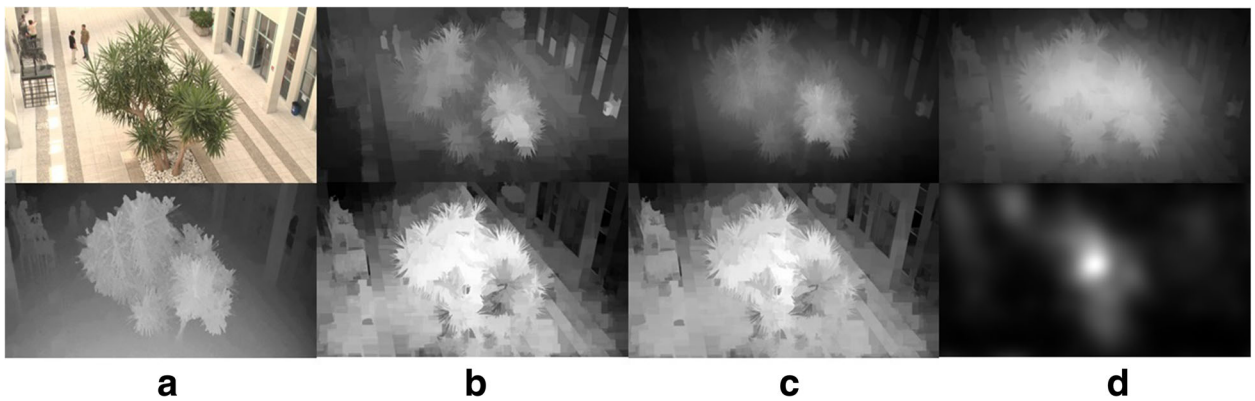


Fig. 7 An example of the proposed visual saliency prediction. **a** The original left image and depth map. **b** The maps computed by the stereo contrast and stereo focus models. **c** The maps after clustering. **d** Final saliency map and ground truth

maps. When we compute one saliency map, it treats the other saliency map as the prior while the current saliency map computes the likelihood. The specific steps are as follows: when we compute the saliency map $S_2{}'$ based on the Bayesian formula, using one saliency map S_1 computes the prior probability and using the other saliency S_2 computes the likelihood. After this, we use the saliency maps in the formula in the opposite way. In other words, S_2 then computes the prior and S_1 computes the likelihood. In this way, the saliency map $S_1{}'$ is computed. Finally, $S_1{}'$ and $S_2{}'$ are combined to obtain the final saliency map. Using this approach, it is possible to avoid reintroducing the noise in different saliency features, thereby obtaining a more accurate posterior probability. This model is very robust with regard to various types of images. After Bayesian integration, we use center bias to conduct post-processing to obtain the final stereo saliency map, because many datasets place the salient object or region in the center of the image [37]. Figure 7d is an example of the saliency map after Bayesian integration and center bias.

The complete visual saliency prediction algorithm can be summarized as:

Algorithm: Stereo visual saliency prediction based on stereo contrast and stereo focus
Input: Left image and disparity map
Output: Saliency Map
1. Multi-scale segmentation, superpixel number {600, 800, 1000, 1200}
2. For each scale $X=[x_1,x_2,\ldots,x_N]_s$
3. For each superpixel $t=[H, S, V, x, y, z]_t$
4. Stereo contrast: *SC(t),* in Eq.(6)
5. Enhancement for stereo contrast: *Sm_c(t)*, in Eq.(10)
6. Stereo focus: *SF(t),* in Eq.(9)
7. Enhancement for stereo contrast: *Sm_f(t),* in Eq.(10)
8. After multi-scale fusion, two pixel-level saliency maps are computed: S_i *(i=1, 2)*
9. Bayesian integration scheme: $S(S_1,S_2)$

4 Results and discussion

In this section, we evaluate the performance of our proposed model on two eye-tracking datasets [17, 18]. One supplies high-quality stereoscopic images and the other supplies low-quality stereoscopic images generated by Kinect-1. First, we present the quantitative metrics of evaluation for the proposed method in Section 4.1. To demonstrate the effect of the different component combinations of our algorithm, a performance comparison is given in Section 4.2. Last, we give a performance evaluation by comparing the proposed methods to state-of-the-art methods in Section 4.3.

4.1 Experimental setup

Our stereo saliency framework is based on the superpixel. In the experiment, we set the segmentation scale of superpixels in the SLIC algorithm. The number of superpixels was set as {600, 800, 1000, 1200}. The SLIC algorithm automatically adjusts the shape of each superpixel based on the segmentation scale and texture information of the image, which is sensitive to the boundary of the object. In stereo contrast, all distance is normalized to [0, 1] and we set $R = 0.3$ empirically. The main parameters of our proposed method are the number of clusters K and δ in Eq. (10). In the experiment, we varied K (K = 6, 8, 10, 12) and δ ($\delta = 0.4, 0.5, 0.6, 0.7$), and observed that the saliency results were insensitive to both parameters. We set the number of clusters $K = 10$ and $\delta = 0.5$. The parameters of σ_1 and σ_2 are given in Eqs. (7 and 11), we differed these values to [0.01, 3] and observed the saliency results. Then, we set $\sigma_1^2 = 0.8$ and $\sigma_2^2 = 0.6$. In Eq. (7), α is set to $\alpha = 0.5$, which is the same as in [22].

We used one of the databases from [17]. This database is consistent with the characteristics of the HVS and includes 18 high-quality stereoscopic images of various types (e.g., indoor scenes, outdoor scenes, and scenes containing various numbers of objects). Some images in the database were collected from the Middlebury 2005/2006 dataset [38], which has high-

accuracy depth maps, while others were produced from videos recorded using a Panasonic AG-3DA1 3D camera, which supplies high-quality left/right images. To avoid 3D fatigue resulting from conflict in the depth field (for example, one object is seen by the left eye but missed by the right eye), the degree of vergence in human vision was considered within the stereoscopic 3D viewing environment in this eye-tracking experiment. The disparity of the stereoscopic images used is within the comfortable viewing zone. The conflict in different depth fields will not be detected by observers during the eye-tracking experiments. The gaze points are recorded by the eye-tracker and processed by a Gaussian kernel to generate the fixation density maps, which are used as the ground-truth maps.

The other eye-tracking database was published in [18]. This database supplies low-quality stereoscopic images compared with [17] and has 600 stereoscopic images that include outdoor and indoor scenes. These stereoscopic images generated by Kinect-1 are diverse in terms of the number and size of objects and the degree of interaction or activity depicted. The stereoscopic images only have a resolution of 640×480 and may have some noise because the depth map by the Kinect-1 has some holes and needs to be smoothed. The stereoscopic image pair is produced by pre-processing, calibration, and post-processing. The eye-tracking data are captured in both 2D and 3D free-viewing experiments by the eye-tracker from 80 participants (ranging in age from 20 to 33 years old). Human fixation maps are constructed from the fixations of viewers to globally represent the spatial distribution of human fixations. Then, a Gaussian kernel is used to obtain the continuous fixation density maps as the ground-truth maps. This dataset supplies 2D and 3D fixation maps. To facilitate a comparison, we used 3D fixation maps as the stereoscopic 3D ground-truth maps.

To quantitatively evaluate the performance of the proposed model, we applied similar quantitative measuring methods to [17]. The performance of the proposed model was measured by comparing the saliency map with the ground-truth map supplied by the database. Because there are two images (left and right) for any stereoscopic image pair, we used the saliency map of the left image for comparison [17]. The area under the receiver operating characteristics curve (AUC) and the correlation coefficient (CC) were used to evaluate the quantitative performance of the proposed stereo visual saliency prediction model. Of these measures, the AUC is the area under the receiver operating characteristics (ROC) curve [39]. Using this score, human fixations were considered to be the positive set, and some points from the image were sampled to form the negative set. The saliency map S was then treated as a binary classifier to separate the positive samples from the negatives. By thresholding over the saliency map and plotting the true positive rate versus the false positive rate, an ROC curve was generated for each image. Then, the ROC curves were averaged over all images and the area underneath the final ROC curve was calculated as the AUC [40]. Perfect prediction corresponds to a score of 1 while a score of 0.5 indicates a level of chance. To compute the AUC, each eye fixation density map and saliency map were normalized to [0, 1]. In practice, we set different thresholds from [0.01, 1]. The LCC measures the strength of a linear relationship between the predicted saliency map and the ground-truth saliency map. When CC is close to $+1/-1$, there is almost a perfectly linear relationship between the two variables.

4.2 Performance comparison with different combinations of components

Four main components were compared: stereo contrast, stereo focus, and enhancement and integration via the Bayesian scheme. The performance of different combinations of components is shown in Tables 1 and 2. SCM is the saliency map based on stereo contrast followed by multi-scale fusion. SFM is the saliency map based on stereo focus followed by multi-scale fusion. SCE is the saliency map based on stereo contrast followed by enhancement. SCE is the saliency map based on stereo contrast, followed by enhancement. OurWE is the proposed stereo saliency map without enhancement. Our model is the proposed stereo saliency map.

Table 1 indicates that SFM performs better than SCM on the database in [17] in AUC and CC. Table 2 shows that SFM performs better than SCM on the database in [18] with AUC and CC. The two models performed differently on each database, so using either one to form the saliency map would not result in good performance. Tables 1 and 2 show that the enhancement slightly improves the performance of the two models with AUC and CC. However, if we remove the enhancement from our proposed model, the performance of our model will be affected. In order to verify the improvement of the

Table 1 Comparison between different component orders in the database in [17]

Different combinations	AUC(→1)	CC(→1)
SCM	0.588	0.198
SFM	0.648	0.257
SCE	0.598	0.213
SFE	0.65	0.258
OurWE	0.864	0.557
Our model	0.881	0.656

Table 2 Comparison between different component orders in the database in [18]

Different combinations	AUC(→1)	CC(→1)
SCM	0.619	0.148
SFM	0.533	0.115
SCE	0.628	0.154
SFE	0.541	0.116
OurWE	0.849	0.37
Our model	0.861	0.419

enhancement, we conduct a significance test for our model and OurWe. For the dataset in [17], we use a paired-samples t test to compare the average performance of our model with the average performance of the OurWE model. For AUC, the improvement of the enhancement is not significant ($t(18) = 1.61$, $P(T \leq t) = 0.126$, $P < 0.05$). For CC, the improvement of the enhancement is significant ($t(18) = 3.09$, $P(T \leq t) = 0.0067$, $P < 0.05$). For the dataset in [18], we use an ANOVA to compare the average performance of our model with the average performance of the OurWE model. The improvement of the enhancement is significant in AUC ($F = 14.89$, P value = 0.00012, $P < 0.05$) and CC ($F = 114.948$, P value = 1.13E−25, $P < 0.05$). According to the results of the significant test, we can see there are three positive results and one negative result. We believe that the enhancement can increase the performance of our proposed model slightly.

From Tables 1 and 2, we can see that the contribution of stereo focus varies. In Table 1, stereo focus has a more important contribution than stereo contrast because the objects of the stereoscopic image from the database in [17] lie in different focus regions and stereo focus works more effectively. In Table 2, we can see that the contribution of stereo focus is less than stereo contrast because the content of the database in [18] is more sensitive to color/depth contrast. Thus, to deal with these different types of stereoscopic images, we designed our model based on both stereo focus and stereo contrast. Figure 8 shows examples of the proposed visual saliency prediction. We notice that the small cap is not detected as a salient region in the stereo focus model. The stereo focus is related to the monocular focus and comfort value. In this case, the zero disparity plane is at the big cap according to our comfort value. The monocular focus model detects the big cap as the focus region and the small cap is out of the focus region. Therefore, the salient region is the big cap region and the small cap is not the salient region in the monocular focus model. Even if we increase the weight of the comfort value (because the small cap is near the zero disparity plane and it pops out), it is not detected as the salient region according to the proposed stereo focus model. In stereo contrast model, the small cap is detected as the salient region because of the pop-out effect. Although the conflict between the stereo focus and stereo contrast still exists, our proposed model obtains the acceptable result that has the benefits from the stereo focus and stereo contrast models. This case shows that the stereo focus model may not work in the object with the negative disparity. For improving the performance of the proposed model, it is necessary to take the stereo contrast model into consideration.

4.3 Comparison of our proposed method with other methods

First, we compared the proposed model with other state-of-the-art methods [17]. We compared it with 2D saliency methods, mixed models, and stereoscopic 3D saliency models. The 2D saliency methods include IT [41], AIM [42], SR [43], and GBVS [44] (denoted as 2D model in Table 3). Mixed model means combining these 2D models with the depth saliency models proposed by [14] (denoted as 2D × depth (Chamaret)) and [17] (which have two models denoted as 2D + depth contrast and 2D + DSM). Model1, Model2, and Model3 were proposed by [17], which were computed by using the depth saliency model combining three 2D saliency models. We used a Bayesian integration [36] to process the 2D model and depth contrast saliency. For a fair comparison, we added center bias to process the results of the Bayesian integration. 2D + DSM considered the center-surrounded mechanisms. We then compared our proposed model with the stereoscopic 3D saliency model proposed by [45]. We should note that the stereo model in [45] has already taken the center bias into consideration. From Table 3, we can see that the performance is not improved significantly using the

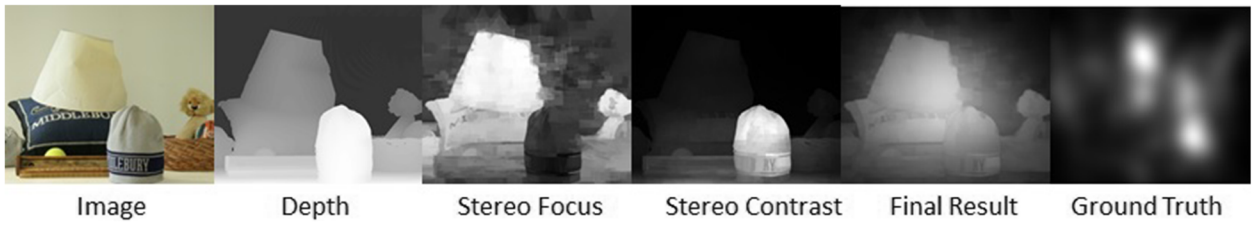


Fig. 8 An example of the proposed visual saliency prediction

Table 3 Comparison between the proposed framework with the others. DSM represents the depth saliency map in [17]

Model		AUC(→1)	CC(→1)
2D model	IT	0.538	0.137
	AIM	0.638	0.326
	SR	0.63	0.291
	GBVS	0.809	0.54
2D × depth (Chamaret)	IT × depth	0.54	0.137
	AIM × depth	0.636	0.299
	SR × depth	0.634	0.292
	GBVS × depth	0.771	0.515
2D + depth contrast	IT + depth contrast	0.596	0.211
	AIM + depth contrast	0.644	0.343
	SR + depth contrast	0.662	0.307
	GBVS + depth contrast	0.799	0.53
Bayesian integration	IT ⊕ depth contrast	0.668	0.254
	AIM ⊕ depth contrast	0.713	0.336
	SR ⊕ depth contrast	0.714	0.369
	GBVS ⊕ depth contrast	0.787	0.511
Center bias	CB(IT ⊕ depth contrast)	0.798	0.547
	CB(AIM ⊕ depth contrast)	0.830	0.61
	CB(SR ⊕ depth contrast)	0.844	0.629
	CB(GBVS ⊕ depth contrast)	0.856	0.632
2D + DSM	Model1	0.656	0.356
	Model2	0.675	0.424
	Model3	0.67	0.41
Stereo model [45]	CB (CNSP)	0.79	0.48
	CB (CNMC)	0.78	0.63
	CB (GNLNS)	0.77	0.65
Our model		0.881	0.656

depth information as a weighted value (2D × depth (Chamaret)) in AUC and CC. Directly using depth information as a weighted value for stereo saliency analysis does not achieve a good result because the method does not consider the actual characteristics of the depth information. By contrast, the performance of the 2D + DSM and 2D + depth contrast methods are better than the 2D × depth (Chamaret), precisely because both consider the characteristics of the depth information. Bayesian integration and center bias increase the performance compared with 2D + depth contrast methods. The performance of our proposed framework is the best of all the methods. Figure 9 gives the example of the proposed visual saliency prediction.

Second, we used the published eye-tracking datasets in [18] with 600 3D images, including outdoor and indoor scenes, to evaluate performance. We used the 3D fixation maps as the ground-truth maps. Because we could not find the code of the DSM in [18], we could only compare our results with the best methods listed in their original paper. The comparative model is DSM, and the 2D saliency modes are IT [41], AIM [42], FT [46], GBVS [44], ICL [47], LSK [48], and LRR [49]. To compare the results of these models, we quantitatively evaluated their performance on the database of the proposed method, using AUC and CC [50]. The experimental results are shown in Table 4. Note that the AUC and CC values of the other existing models were taken from the original paper [18]. From this table, we see that the performance of our proposed model is the best of the 15 stereo visual saliency prediction models. Here, we notice that our proposed model does slightly better than the GBVS × DSM. The reason for this is that sometimes the pop-out effect and comfort zone will fail because the salient region may be located in the background or near the background. Therefore, although the results of our proposed model are

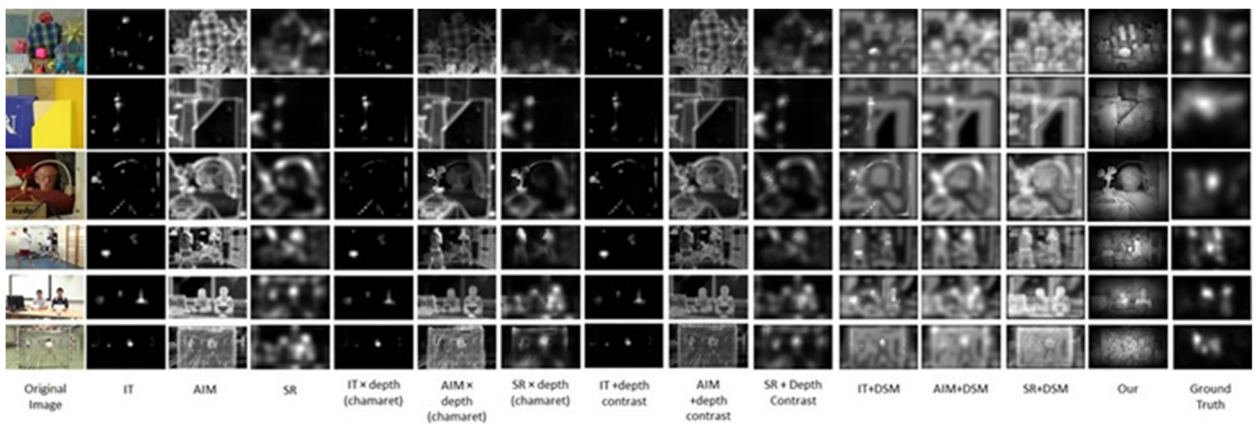


Fig. 9 Stereo comfort zone based on human stereo vision. DSM represents the depth saliency map in [17]

Table 4 Comparison between different 3D visual saliency prediction models

Component combination	AUC(→1)	CC(→1)
IT+DSM	0.849	0.375
IT×DSM	0.854	0.398
GBVS+DSM	0.851	0.39
GBVS×DSM	0.855	0.413
AIM+DSM	0.85	0.342
AIM×DSM	0.85	0.391
FT+DSM	0.797	0.315
FT×DSM	0.745	0.268
ICL+DSM	0.846	0.385
ICL×DSM	0.808	0.325
LSK+DSM	0.845	0.379
LSK×DSM	0.824	0.351
LRR+DSM	0.856	0.385
LRR×DSM	0.846	0.395
Our model	0.861	0.419

"+" means the combination by simple summation as in the study in [18]. "×" means the combination by point-wise multiplication [18]. DSM represents the depth saliency map in [18]

better than the other existing models, it is not much better than GBVS × DSM.

5 Conclusions

In this paper, we exploit two characteristics of stereoscopic vision and propose stereo visual saliency prediction based on stereo contrast and stereo focus. Stereo contrast is a product of color and depth contrast and the pop-out effect describes the contrast in objects. Stereo focus is based on the focus mechanism of human stereo vision, which describes the region of human focus. For each value of the two models, we individually enhanced the important region to make it more distinctive. The two values were individually converted into two saliency maps using multi-scale fusion. Lastly, both saliency maps were integrated using Bayesian integration. Experimental results show that our proposed model can process stereoscopic images from different stereoscopic capture devices to achieve the best performance on two eye-tracking databases compared to existing methods.

In the present study, even if the performance of the proposed model is good, our model still suffers from some limitations. The main one is that in some cases, the pop-out effect and comfort zone may fail in stereoscopic saliency analysis. For example, if the salient region is located near the background, the performance of our model will decrease. The reason for this is that this case is not suitable for our assumption that the salient region should be located in the comfort zone or have the pop-out effect. In the future, we will exploit more mechanisms of HVS for saliency analysis. We try to find out how to deal with the conflict between pop-out effect and comfort zone and how to improve the accuracy of the salient region if the pop-out effect and comfort zone are not working very well. Additionally, we will exploit more features (such as texture contrast, luminance contrast, the property of divergence, and different monocular focus approaches) to improve our proposed model in different color spaces.

Funding

This work was funded in part by the National Natural Science Foundation of China, under Grants U1301257, 61571285, and 61422111.

Authors' contributions

All authors have contributed equally to the text, while HC has implemented the algorithms and performed most of the tests. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Author details

[1]The School of Communication and Information Engineering, Shanghai University, Shanghai, China. [2]Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, Australia.

Received: 4 November 2016 Accepted: 22 August 2017
Published online: 02 September 2017

References

1. J. Lei, B. Wang, Y. Fang, W. Lin, P. Le Callet, N. Ling, C. Hou, A universal framework for salient object detection. IEEE Trans. Multimedia. **18**(9), 1783–1795 (2016)
2. Y. Fang, W. Lin, Z. Chen, C.-M. Tsai, C.-W. Lin, A video saliency detection model in compressed domain. IEEE Trans. Circuits Syst. Video Technol. **24**(1), 27–38 (2014)
3. Y. Fang, Z. Chen, W. Lin, C.-W. Lin, Saliency detection in the compressed domain for adaptive image retargeting. IEEE Trans. Image Process. **21**(9), 3888–3901 (2012)
4. A. Borji, L. Itti, State-of-the-art in visual attention modeling. IEEE Trans. Pattern Anal. Mach. Intell. **35**(1), 185–207 (2013)
5. A.L. Yarbus, B. Haigh, L.A. Rigss, *Eye movements and vision*, vol 2 (1967), pp. 5–10
6. A.M. Treisman, G. Gelade, A feature-integration theory of attention. Cogn. Psychol. **12**(1), 97–136 (1980)
7. T. Judd, K. Ehinger, F. Durand and A. Torralba, Learning to predict where humans look, 2009 IEEE 12th International Conference on Computer Vision, Kyoto, 2009, pp. 2106–2113.
8. S. Frintrop, E. Rome, H.I. Christensen, Computational visual attention systems and their cognitive foundations: a survey. ACM Trans. Appl. Percept. (TAP). **7**(1), 6 (2010)
9. P. Jiang, H. Ling, J. Yu and J. Peng, Salient Region Detection by UFO: Uniqueness, Focusness and Objectness, 2013 IEEE International Conference on Computer Vision (Sydney, 2013), pp. 1976–1983
10. Wei, Y., Wen, F., Sun, J.: Geodesic Saliency Using Background Priors. Google Patents. US Patent App. 14/890,884 (2013)
11. S. Goferman, L. Zelnik-Manor and A. Tal, Context-aware saliency detection, 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, 2010, pp. 2376–2383
12. M. Camplani, L. Salgado, Efficient spatio-temporal hole filling strategy for Kinect depth maps, Proc. SPIE 8290, (Three-Dimensional Image Processing (3DIP) and Applications II, Burlingame, 2012), pp. 82900–82900

13. C. Chamaret, S. Godeffroy, P. Lopez, O. Le Meur, Adaptive 3D rendering based on region-of-interest, Proc. SPIE 7524, (Stereoscopic Displays and Applications XXI, San Jose, 2010) pp. 75240–75240
14. Q. Huynh-Thu, M. Barkowsky, P. Le Callet, The importance of visual attention in improving the 3D-TV viewing experience: overview and new perspectives. IEEE Trans. Broadcast. **57**(2), 421–431 (2011)
15. H. Kim, S. Lee and A. C. Bovik. Saliency Prediction on Stereoscopic Videos, in IEEE Transactions on Image Processing. **23**(4), 1476–1490 (2014)
16. N. D. B. Bruce and J. K. Tsotsos, An attentional framework for stereo vision, (The 2nd Canadian Conference on Computer and Robot Vision (CRV'05), Victoria, 2005), pp. 88-95
17. J. Wang, M.P. DaSilva, P. LeCallet, V. Ricordel, Computational model of stereoscopic 3D visual saliency. IEEE Trans. Image Process. **22**(6), 2151–2165 (2013)
18. C. Lang, T.V. Nguyen, H. Katti, K. Yadati, M. Kankanhalli, S. Yan, *Depth matters: influence of depth cues on visual saliency* (2012), pp. 101–115
19. X. Fan, Z. Liu and G. Sun, Salient region detection for stereoscopic images, 2014 19th International Conference on Digital Signal Processing, Hong Kong, 2014, pp. 454–458
20. Y. Niu, Y. Geng, X. Li and F. Liu, Leveraging stereopsis for saliency analysis, 2012 IEEE Conference on Computer Vision and Pattern Recognition (Providence, 2012), pp. 454–461
21. Y. Fang, J. Wang, M. Narwaria, P. Le Callet and W. Lin, Saliency detection for stereoscopic images, 2013 Visual Communications and Image Processing (VCIP) (Kuching, 2013), pp. 1–6
22. H. Cheng, J. Zhang, P. An and Z. Liu, A Novel Saliency Model for Stereoscopic Images, 2015 International Conference on Digital Image Computing: Techniques and Applications (DICTA) (Adelaide, 2015)
23. Beato, A.: Understanding comfortable stereography. Techical Report (2011)
24. Zhang, Z.Y., An, P., Zhang, Z.J., Shen, L.Q.: 2D/3D Video Processing and Stereo Display Technology (2010)
25. J. Hakkinen, T. Kawai, J. Takatalo, R. Mitsuya, G. Nyman, What do people look at when they watch stereoscopic movies?, Proc. SPIE 7524, Stereoscopic Displays and Applications XXI (San Jose, 2010), pp. 75240–75240
26. S. Yano, S. Ide, T. Mitsuhashi, H. Thwaites, A study of visual fatigue and visual comfort for 3D HDTV/HDTV images. Displays **23**(4), 191–201 (2002)
27. Mendiburu, B.: 3D Movie Making: Stereoscopic Digital Cinema from Script to Screen (2009)
28. R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Susstrunk, SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Trans. Pattern Anal. Mach. Intell. **34**(11), 2274–2282 (2012)
29. Y.-J. Lee, J.-B. Song, Autonomous salient feature detection through salient cues in an HSV color space for visual indoor simultaneous localization and mapping. Adv. Robot. **24**(11), 1595–1613 (2010)
30. W. Einhauser, P. Konig, Does luminance-contrast contribute to a saliency map for overt visual attention? Eur. J. Neurosci. **17**(5), 1089–1097 (2003)
31. M. M. Cheng, G. X. Zhang, N. J. Mitra, X. Huang and S. M. Hu, Global contrast based salient region detection, CVPR 2011 (Providence, 2011), pp. 409–416
32. Y. Fang, J. Wang, M. Narwaria, P. Le Callet, W. Lin, Saliency detection for stereoscopic images. IEEE Trans. Image Process. **23**(6), 2625–2636 (2014)
33. J.H. Elder, S.W. Zucker, Local scale control for edge detection and blur estimation. IEEE Trans. Pattern Anal. Mach. Intell. **20**(7), 699–716 (1998)
34. X. Li, H. Lu, L. Zhang, X. Ruan and M. H. Yang, Saliency Detection via Dense and Sparse Reconstruction, 2013 IEEE International Conference on Computer Vision (Sydney, 2013), pp. 2976–2983
35. V. Gopalakrishnan, Y. Hu, D. Rajan, Salient region detection by modeling distributions of color and orientation. IEEE Trans. Multimedia **11**(5), 892–905 (2009)
36. H. Lu, X. Li, L. Zhang, X. Ruan, M.-H. Yang, Dense and sparse reconstruction error based saliency descriptor. IEEE Trans. Image Process. **25**(4), 1592–1603 (2016)
37. A. Borji, M.-M. Cheng, H. Jiang, J. Li, Salient object detection: a benchmark. IEEE Trans. Image Process. **24**(12), 5706–5722 (2015)
38. D. Scharstein and C. Pal, Learning Conditional Random Fields for Stereo, 2007 IEEE Conference on Computer Vision and Pattern Recognition (Minneapolis, 2007), pp. 1–8
39. DAVID, M.G.: Signal detection theory and psychophysics. (1966)
40. N. Bruce, J. Tsotsos, Saliency based on information maximization. Adv. Neural Inf. Proces. Syst. **18**, 155 (2006)
41. L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. Pattern Anal. Mach. Intell. **20**(11), 1254–1259 (1998)
42. N. Bruce, J. Tsotsos, in *Advances in Neural Information Processing Systems*. Saliency based on information maximization (2005), pp. 155–162
43. X. Hou and L. Zhang, Saliency Detection: A Spectral Residual Approach, 2007 IEEE Conference on Computer Vision and Pattern Recognition (Minneapolis, 2007), pp. 1–8
44. J. Harel, C. Koch, P. Perona, in *Advances in Neural Information Processing Systems*. Graph-based visual saliency (2006), pp. 545–552
45. I. Iatsun, M.-C. Larabi, C. Fernandez-Maloigne, A visual attention model for stereoscopic 3D images using monocular cues. Signal Process. Image Commun. **38**, 70–83 (2015)
46. R. Achanta, S. Hemami, F. Estrada and S. Susstrunk, Frequency-tuned salient region detection, 2009 IEEE Conference on Computer Vision and Pattern Recognition (Miami, 2009), pp. 1597–1604
47. X. Hou, L. Zhang, in *Advances in Neural Information Processing Systems*. Dynamic visual attention: searching for coding length increments (2009), pp. 681–688
48. H.J. Seo, P. Milanfar, Static and space-time visual saliency detection by self-resemblance. J. Vis. **9**(12), 15 (2009)
49. C. Lang, G. Liu, J. Yu, S. Yan, Saliency detection by multitask sparsity pursuit. IEEE Trans. Image Process. **21**(3), 1327–1338 (2012)
50. N. Ouerhani, R. Von Wartburg, H. Hugli, R. Muri, Empirical validation of the saliency-based model of visual attention. Electron. Lett. Comput. Vis. Image Anal. **3**(1), 13–24 (2003)