

SALIENT REGION DETECTION WITH OPPONENT COLOR BOOSTING

Guanqun Cao, Faouzi Alaya Cheikh

Gjøvik University College, Norway

{guanqun.cao, faouzi.cheikh}@hig.no

ABSTRACT

Numerous efforts have been made to detect salient regions in images. Mostly luminance-based saliency models are found in the literature, which ignore the important contribution of color in finding the local distinct image features. Methods about color saliency detection in the literature can only give indication of color salient points or derivatives. In this paper, we present a fast method for detecting the distinct color regions. This model is able to give inference of salient object in the accurate-to-contour level. A segmentation task is also performed based on the proposed color saliency model to show its strength in object outline definition. Finally an evaluation with both the eyetracking and the annotation is achieved for analyzing the model in an extensive dataset. The presented model outperforms the previous work in both tests.

Index Terms— visual saliency, feature detection, color imaging, opponent color, boosting

1. INTRODUCTION

Visual saliency is an important mechanism in our human visual system (HVS). It helps to reduce the overload visual information coming to the eye. Usually it is defined as the perceptual quality that makes an object, person, or pixel stand out relative to its neighbors and thus capture our attention. Visual attention results both from fast, pre-attentive, bottom-up visual saliency of the retinal input, as well as from slower, top-down memory and volition based processing that is task-dependent [1]. Region Of Interest (ROI) detection based on the attention mechanism has become an active research area recently. It is useful for object segmentation, image quality assessment, watermarking and so on. Itti et al. have proposed a biologically plausible saliency model based on the behaviour of neurons in the receptive fields of human visual cortex. However, due to the inadequate understanding of HVS, the biological model is not able to give very satisfying results and researchers' interest in computational models has been growing since. In this paper, we will constrain the computational saliency resulting from the low-level feature in the

The first author is pursuing his M.S. degree funded by the Erasmus Mundus CIMET program.

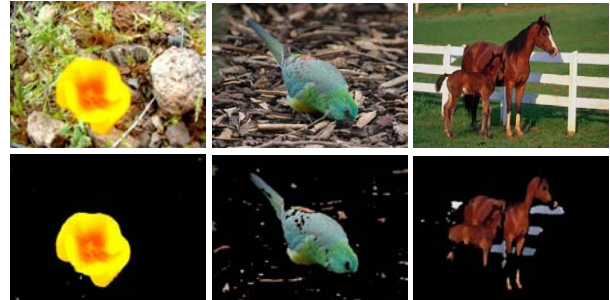


Fig. 1. The salient object extraction by the proposed model

visual field. That is to say, the higher levels of reasoning (e.g. knowledge and interest) will not be taken into consideration.

In general, all models adopt one or more low-level features such as intensity, color, or orientation, to reflect the visual distinctiveness. Almost all the saliency maps proposed based on these low-level features have low resolutions and ill-defined object boundaries. A recent model [2] outperforms the rest in that it is able to efficiently output full resolution saliency maps, establish well-defined boundaries of salient objects and disregard high frequencies arising from texture, noise and blocking artifacts. However, this model follows a problematic theory that the lightness of image regions is proportional to their saliency, which overlooks the importance of color in saliency.

Color is an very important factor that affects the saliency in images. The aim of color saliency models is to model how the HVS perceives the colors in the image as a function of its local spatial organization. Vazquez et al. [3] have illustrated that color contains more information than luminance in salient regions. They showed that if we represent the color image derivatives which have equal frequency (iso-salient derivatives) in a space, the distribution of RGB derivatives for the 40,000 images of the COREL dataset form an ellipsoid-like distribution. The distribution axis along the luminance direction is longer than the color direction, which indicates more color information is obtained in one unit of salient displacement. Therefore, in order to be consistent with neural mechanisms, the color feature should be emphasized when modeling the image saliency.

Color perception is not well represented in most recent saliency models through the RGB raw data. Color information must be exploited on the basis of chromatic channel opponencies in future color saliency modeling. Therefore in this paper, we boost the color saliency in the opponent color space to detect the region of interest.

2. PROPOSED SALIENCY MODEL

2.1. Color boosting saliency

Current color saliency models generally focus on the interest points or color edges, and most studies are based on a color saliency boosting function. Van de Weijer et al. [4] introduced the color saliency boosting function obtained from the iso-salient surface. Their work is based on the statistics of color image derivatives and uses the information theory to boost the color information content of an image. The main idea is inspired from the basics of information theory that specifies the rare events are more informative than normal events. This is based on the self-information factor defined by following formula,

$$I(v) = \log_2(1/p(v)) \quad (1)$$

where $p(v)$ is the probability of the descriptor v will occur in the descriptors of the image at hand.

The color saliency boosting function g is approximated from the distribution of image derivatives. According to the study by Van de Weijer et al., the shape of the distribution is quite similar to an ellipse which can be described by the covariance matrix M . However, the previous work limited the calculation of M based on the raw RGB data, which is insufficient to create a sphere-like iso-salient transformation. In this paper, the work is extended that the Gaussian derivatives of each opponent color in the image are computed to estimate M . We illustrate this equation as follows,

$$M = \begin{pmatrix} \overline{O_{1x}O_{1x}} & \overline{O_{1x}O_{2x}} & \overline{O_{1x}O_{3x}} \\ \overline{O_{1x}O_{2x}} & \overline{O_{2x}O_{2x}} & \overline{O_{2x}O_{3x}} \\ \overline{O_{1x}O_{3x}} & \overline{O_{2x}O_{3x}} & \overline{O_{3x}O_{3x}} \end{pmatrix} \quad (2)$$

where the elements in the matrix are computed as Equation 3. Note, i is the pixel coordinates in each Gaussian derivative component of the image I_x . The equation is shown below,

$$\overline{O_{1x}O_{1x}} = \sum_{i \in I_x} (O_{1x} - \overline{O_{1x}})(O_{1x} - \overline{O_{1x}}) \quad (3)$$

where $\overline{O_{1x}}$ is the mean of O_{1x} .

In order to be consistent with the color vision theory, the opponent color representation must be estimated through the LMS color space. Therefore, the chromatic color opponencies of each image are obtained in the following way. We firstly transform each image from the standard RGB space (sRGB) to the XYZ space, then the chromatic adaptation

matrix (M_{CAT02}) from the CIECAM02 model is adopted to quantify the tristimulus values XYZ into the LMS color space [5]. Finally, the opponent colors are computed from the LMS space by the transformation defined by Wandell in 1995 [6]. Due to the limits of this paper, we only illustrate the final transformation from the RGB color space to the opponent color space in Equation 4,

$$\begin{pmatrix} O_1 \\ O_2 \\ O_3 \end{pmatrix} = \begin{pmatrix} 0.3905 & 0.5499 & 0.0089 \\ -0.1764 & 0.4307 & -0.1164 \\ -0.1191 & -0.1739 & 0.8673 \end{pmatrix} \cdot \begin{pmatrix} R \\ G \\ B \end{pmatrix} \quad (4)$$

where O_1 is luminance component, O_2 is referred to as the red-green channel (G-R) and O_3 is referred to as the blue-yellow channel (B-Y).

After obtaining M from Equation 2, this matrix can be decomposed into an eigenvector matrix U , a diagonal matrix Σ , and an eigenvalue matrix V (Equation 5). Thereafter, we are able to transform the image into the iso-salient space using the color boosting function g , with which the ellipses are reshapes to spheres (Equation 6). The detailed formulae are shown below:

$$M = U\Sigma V \quad (5)$$

$$g(I_{opp}(x, y)) = U \cdot (\text{diag}(1/\text{diag}(\sqrt{\Sigma})) \cdot V^T) \cdot I_{opp} \quad (6)$$

where I_{opp} is the image in the opponent color representation and the diag function reduces a matrix to its diagonal elements. Note, this color saliency boosting function is comparatively sphere-like to the opponent representation of the original image. This is the extension to the previous work of boosting the RGB image, including [3, 4, 7]. After color saliency boosting, there is an increase in information context and a gain in photometric robustness.

2.2. DoG model as the local feature detector

For extracting the local feature, we apply a similar technique to the one proposed by [2]. In brief, this saliency detection algorithm uses the Difference-of-Gaussian (DoG) filter for band pass filtering the image. The DoG filter is widely used in edge detection and becoming popular in feature extraction such as in the SIFT algorithm [8]. The DoG filter is given by:

$$\begin{aligned} DoG(x, y) &= \frac{1}{2\pi} \left[\frac{1}{\sigma_1^2} e^{-\frac{x^2+y^2}{2\sigma_1^2}} - \frac{1}{\sigma_2^2} e^{-\frac{x^2+y^2}{2\sigma_2^2}} \right] \\ &= G_1(x, y, \sigma_1) - G_2(x, y, \sigma_2). \end{aligned} \quad (7)$$

In order to ensure the salient regions will be fully covered and not just highlighted on edges or in the center of the regions, σ_1 is driven to infinity. This results in a notch in frequency at DC while retaining all other frequencies. To remove high frequency noise and textures, a small Gaussian kernel should be used. For small kernels, the binomial filter approximates the Gaussian one very well in the discrete case. In the experimental part we used the following filter $1/16[1, 4, 6, 4, 1]$ giving $\omega_{hc} = \pi/2.75$.

The method of finding the saliency map S for an image I of width W and height H pixels can thus be formulated as:

$$S(x, y) = \|I_\mu - I_{\omega_{hc}}(x, y)\|_2 \quad (8)$$

where I_μ is the mean image feature vector, $I_{\omega_{hc}}(x, y)$ is the corresponding image pixel vector value in the Gaussian blurred version (using a 5×5 separable binomial kernel) of the original image, and $\|\cdot\|_2$ is the L_2 norm. After the color boosting transformation, the image is separated into its three opponent channels and then smoothed by Gaussian blur. The Euclidean distance (the L_2 norm) is derived between the mean value of each channel and the blurred sub-image in its corresponding channel.

3. EXPERIMENTS AND ILLUSTRATIONS

3.1. Novel saliency detection

First, the original image is transformed into its opponent representation. The proposed color boosting matrix is constructed from the individual image and applied onto its three components, which as equation 6. Then, the image in the iso-salient color space is filtered by the Gaussian kernel. A mean value of each opponent channel from the blurred image is also computed. Finally, the Euclidean distances between the mean and the blurred image of the three channels form the final saliency map. We can illustrate the workflow of this process in Figure 2.

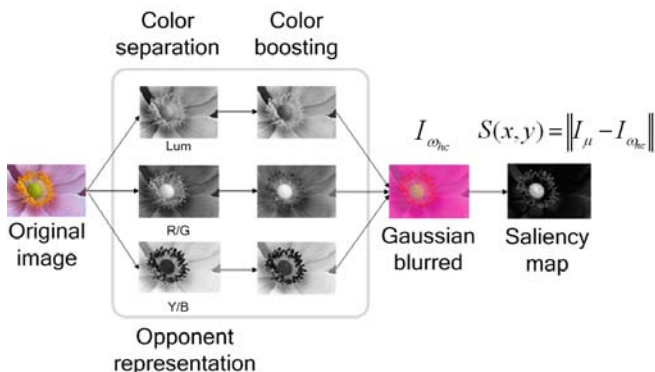


Fig. 2. Experimental framework

3.2. Comparison with state-of-the-art models

In [9], Hou and Zhang try to model the visual rarity from the frequency domain. For the first image in Figure 3, it is able to predict the flower as the attractive region. However, from the second image on, this model is greatly affected by the background of the image and could not give good prediction.

Achanta model [2] performs very well on the first two images. It is able to efficiently highlights the big salient region

in the image. However, from the third image, it shows its deficiency of accounting for too much lightness. The shadow region is always wrongly detected as a salient region of the object. For the fourth image, it shows that this model still suffers from the complex background especially in the presence of shadows. The fifth image also shows the highlight of the shadow of the horses.

With the color boosting, the novel saliency map is able to infer the most distinct color region/object from the whole visual scene. In the third image for example, the yellow flowers stands out of the image and instantly attract our attention. Our model is able to very accurately give the right estimation.

3.3. Segmentation based on ROI

The aim of this paper is to propose a novel saliency map based on color opponencies, and with the extension to the object segmentation level. The related previous work all made use of some sophisticated segmentation methods, such as K-means or mean shift clustering, to extract the salient regions. However, these methods are not able to emphasize the strength of saliency map on the object segmentation. Therefore, we adopt a simple method of binarizing the original image based on the saliency map. An adaptive threshold (T) is determined to extract the Region Of Interest (ROI) from the saliency map (Equation 9).

$$T = \frac{2}{W \times H} \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} S(x, y) \quad (9)$$

We give an illustration of the binarized masks from Achanta model [2] and the proposed model in Figure 4. And a further object extraction is illustrated in Figure 1 of the first page.



Fig. 4. Illustration of segmentation based on ROI. The original image lies on the left, the mask generated by the Achanta model is in the middle, and the proposed model produces a mask on the right.

In order to evaluate the performance of the proposed model segmented in binarized masks, we conducted a twofold test from both the eyetracking data and the annotated masks. The approach of eyetracking is explicitly illustrated in Section 3.4. An accurate object-contour based ground truth database of 1000 masks drawn by Achanta et al. [2] from the widely used database [10] is also adopted for the comparison. Finally, the results and analysis are shown in Section 3.5.

3.4. Psycho-visual test on the real set

Many saliency models have been proposed in recent years; however, the ground truth database that they built on is problematic. [2, 3, 10] Liu et al. [10] created a large database consisting of around 5000 images, and 9 observers annotated what they considered to be salient regions with bounding boxes. Many research continued to carry out their research based on this database of showing the rough location of the salient region. Achanta [2] pointed out the inaccuracy of the bounding boxes and manually drew the salient regions with sharp edges. Valenti et al. [7] shortly discussed about the difference between the visual saliency and semantic saliency, but they limited their investigation with no further testing.

None of these computational models have been tested using an eye tracker. This is due to the fact of the complexity of processing the eyetracking data. However, no matter labeling with bounding box or manually drawing, the observers spend too much time focusing on one image and it could not mimic how the human really perceive the images. The number of observers is too little to represent the unbiased result. Also, unlike the shape saliency which these researchers focus on, the color saliency is influenced by its color distinctiveness and can hardly be expected by the simple annotation. The salient region may not be a fine textured structure as proposed by Achanta et al. Therefore, all the reasons above necessitates the performance of this psycho-visual experiment.

In order to evaluate the performance of the proposed saliency map in an extensive image dataset, 17 naive observers were invited to participate in a psycho-visual test with an iView XTM Hi-Speed eyetracking machine. The monocular mode with a data-capturing rate of 1250 Hz is used in our experiment. The participant is asked to be seated 50 cm in front of the monitor to have a downward gaze angle. The chin is stabilized onto the tray of the device.

The observers must pass the Ishihara test of normal color vision and have no pre-knowledge of the test images. The image database consists of 180 test images selected from Microsoft Research Asia [10] and some suitable pictures for the visual test. The participants were instructed to watch the images as they normal do. The experiment is divided into 3 sessions. Each session lasts for 5 minutes and 60 images will be shown in one session. Before each session, the observer will firstly be asked to look around the screen and the dominant eye will be tracked by the instructor. From the workstation PC, when the observers are well seated and the gaze can be captured through the lens, the slider of the pupil size is automatic adjusted.

After the experiment for each participant, an idf file will be stored in the workstation PC consisting of the observer's eye movement on the images. With a text converter software, the raw data is converted into a readable text file, which illustrates the points of eye fixation that each individual gives. Though the eyetracker capture the points of fixation, the indi-

vidual observes a nearby region around the fixation in reality. Therefore, it makes sense to process the attention map to obtain an area instead of displaying points.

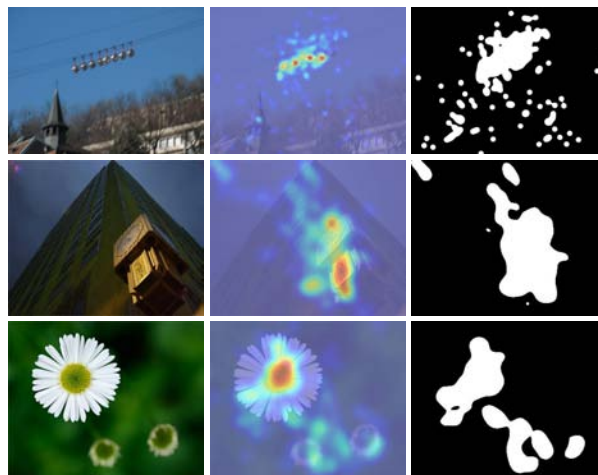


Fig. 5. The final attention maps and the binarized human saliency maps. The original images are located in the left column, the attention maps are presented in the middle, and the human saliency maps are on the right.

The gaze map is obtain in such way. All the fixation points are summed up for all observers's gazes on each image. A frequency map is obtained from the aggregation of gaze points. Then, a gaussian kernel with the size of [150, 150] and the standard deviation of 10 is applied on the frequency map of the screen size (1024×768). The gauss maps are normalized using the maximum and minimum values and the normalized Gauss map is reshaped into three-dimensional color data using JET color map in Matlab having 64 color levels. Keeping in mind, during the experiment, the images are all displayed in the center of the monitor in their original size and the remaining surrounds are filled with the neutral color. Therefore, only the fixations within the size of each central image are remained.

Final attention maps are presented in Figure 5. The attention maps give an indication of the possible gaze regions from the fixation points. The more reddish a region shows that more individuals tend to gaze at. In order to have a more accurate gaze map, the same threshold T in Equation 9 is used to construct a binarized human saliency map. The last column of Figure 5 locates the final human saliency map for the further evaluation.

3.5. Results and analysis

After the computation of masks inferring the ROI from Achanta model and the proposed model, we computed the 2-D correlation coefficient between the binarized maps (human saliency maps or annotated maps) and both ROIs (Achanta

Models	AC	Cao
Gaze data	0.1933	0.2129
Annotated mask	0.5233	0.5276

Table 1. The correlation coefficients from both models. AC stands for the model by Achanta et al. [2]. Cao is the proposed model.

and the proposed model) respectively. The results of the mean correlation are shown in Table 1. Though both models do not have high correlation with the human salient regions, it is shown that the proposed model performs slightly better than the one with Achanta et al. The correlation between the manually annotated masks and the models are much higher and we can also see Achanta model performs no better than the presented model. It also indicates that, to investigate the real regions attracting the visual attention, the study should be a combination of color, lightness intensity, shape as well as the top-down cue.

4. CONCLUSION AND FUTURE WORK

In this paper, we proposed a new region-based saliency detection method with an emphasis on color. This approach is based on color boosting in an iso-salient color space and filtering by a DoG filter afterwards. According to Tremeau et al. [11], the future of visual attention models will follow the development of perceptual multi-scale saliency map based on a competitive process between all bottom-up cues (color, intensity, orientation, location, motion). Therefore, this color saliency model should be integrated with other features for modeling saliency in a bigger framework.

5. REFERENCES

- [1] E. Niebur and C. Koch, *The Attentive Brain*, chapter Computational architectures for attention, MIT Press, Cambridge MA, October 1995.
- [2] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, “Frequency-tuned salient region detection,” in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 20-25 2009, pp. 1597 – 1604.
- [3] E. Vazquez, T. Gevers, M. Lucassen, J. van de Weijer, and R. Baldrich, “Saliency of color image derivatives: a comparison between computational models and human perception,” *Journal of the Optical Society of America A*, vol. 27, no. 3, pp. 613–621, 2010.
- [4] J. van de Weijer, T. Gevers, and A.D. Bagdanov, “Boosting color saliency in image feature detection,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 1, pp. 150–156, Jan. 2006.
- [5] Mark D. Fairchild, *Color appearance model*, Wiley, second edition, 2005.
- [6] Brian A. Wandell, *Foundations of Vision*, Sinauer Associates, first edition, May 1995.
- [7] R. Valenti, N. Sebe, and T. Gevers, “Isocentric color saliency in images,” in *Image Processing (ICIP), 2009 16th IEEE International Conference on*, 7-10 2009, pp. 993–996.
- [8] David G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision*, vol. 60, pp. 91–110, 2004.
- [9] Xiaodi Hou and Liqing Zhang, “Saliency detection: A spectral residual approach,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition CVPR ’07*, June 17–22, 2007, pp. 1–8.
- [10] Tie Liu, Jian Sun, Nan-Ning Zheng, Xiaoou Tang, and Heung-Yeung Shum, “Learning to detect a salient object,” in *Computer Vision and Pattern Recognition, 2007. CVPR ’07. IEEE Conference on*, June 2007, pp. 1–8.
- [11] Alain Tremeau, Shoji Tominaga, and Konstantinos N. Plataniotis, “Color in image and video processing: Most recent trends and future research directions,” *EURASIP Journal on Image and Video Processing*, vol. 2008, no. 581371, pp. 26, 2008.

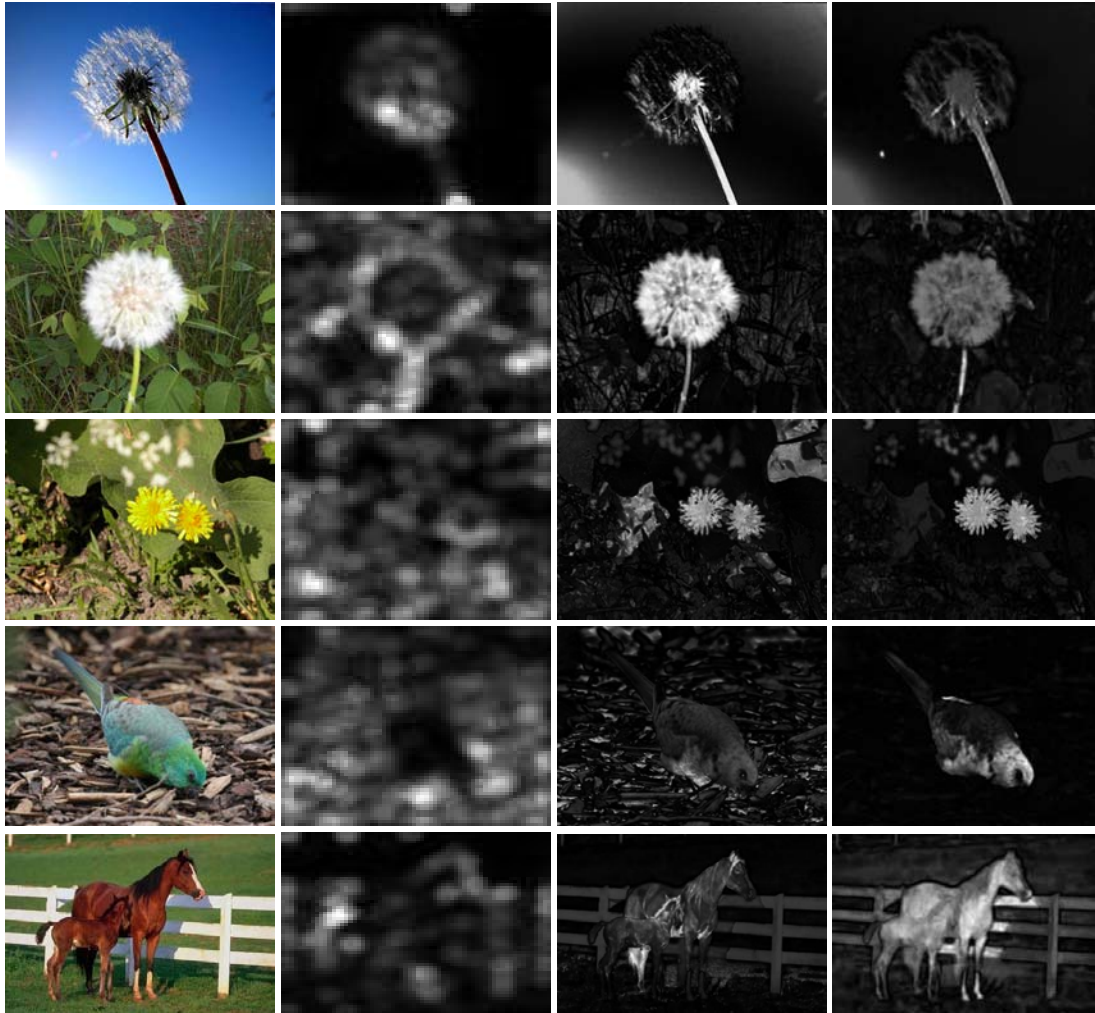


Fig. 3. From the left to the right, the first column are the original images, the second column are the saliency maps from Hou's method [9], the third column indicate the saliency detection by Achanta et al. [2], and the last column are our new saliency maps.