AUTOMATIC EXTRACTION OF SALIENT OBJECTS USING FEATURE MAPS

Ki Tae Park and Young Shik Moon

Department of Computer Science and Engineering, Hanyang University, Ansan, Korea

ABSTRACT

In this paper, we propose a technique for extracting salient objects in images using feature maps, regardless of the complexity of images and the position of objects. In order to extract salient objects, the proposed method uses feature maps with edge and color information. We also propose a reference map created by integrating feature maps, and a combination map representing the boundaries of meaningful objects that is created by integrating the reference map and feature maps. Candidate object regions including boundaries of objects from the combination map are extracted by convex hull algorithm. Finally, by applying a segmentation algorithm on the area of candidate regions, object regions and background regions are separated, and real object regions are extracted from the candidate object regions. Experimental results show that the proposed method extracts the salient objects efficiently, with 84.3% precision rate and 81.3% recall rate.

***Index Terms*—** Salient object extraction, Feature map, Reference map, Combination map, Segmentation

1. INTRODUCTION

In recent years, due to considerable efforts for the development of communication network and Internet services, there has been a rapid increase in the use of multimedia contents such as images, movies, and animations. Therefore, there has been a strong demand for an efficient system for multimedia information retrieval such as the content based image retrieval[1]. Early researches on image retrieval have suggested annotating images manually. However, text based image retrieval is impractical because the textural annotation is usually ambiguous and is expensive in terms of the time complexity[1,2]. An alternative approach to the manual annotation is content based image retrieval, in which images are indexed by their visual features(low-level features) such as color, texture, shape, and spatial layout. Generally, an image index is a set of features that are extracted from the whole image. However, even though humans want to retrieve images using high-level(semantic information) concepts such as tiger, bird, or airplane, it is difficult to represent a semantic information using only the low-level features[3,4]. Therefore, it is necessary to extract salient objects from an image for semantic image retrieval. A technique for extracting salient objects can also be utilized to many applications such as image retargeting and image compression as well as image retrieval.

2. PROPOSED METHOD

In this paper, we propose a technique for extracting salient objects in images using feature maps, regardless of the complexity of images and the position of objects. Fig. 1 shows the overall flowchart of the proposed system, which consists of five steps. In the first step, feature maps representing the boundary of objects in images are created by using edge and color information. In the second step, by extracting salient boundary regions from the feature maps, a reference map representing the boundaries where salient objects may exist is created. In the third step, by adding missed boundary regions to the reference map by comparing the reference map with the feature maps, a combination map which represents salient boundary regions is constructed. In the fourth step, candidate object regions containing objects and backgrounds are extracted by convex hull algorithm. In the final step, by applying a segmentation algorithm on the candidate object regions, object regions and background regions are separated, and the real object regions are extracted from the candidate object regions.

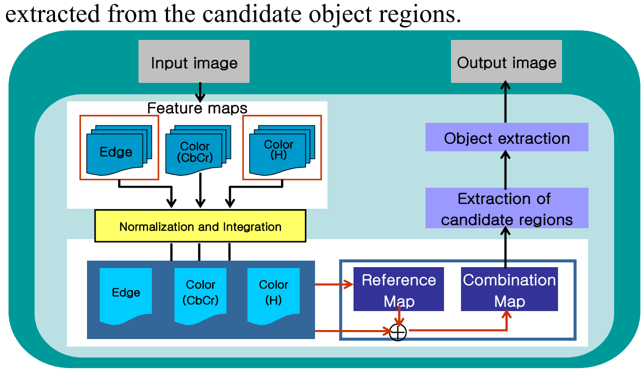

Fig. 1. Flowchart of the proposed system

2.1. Feature maps

The feature maps created by the proposed method are similar to the feature contrast maps proposed by Itti[5]. However, the contrast maps by Itti do not provide information about important regions and do not represent the boundary of object accurately. Therefore, we use edge information and color information to create the feature maps, instead of intensity, color, and orientation information. Fig. 2 shows the contrast maps by Itti's method and the feature maps by the proposed method. As shown in fig. 2, it is difficult to distinguish objects in the contrast maps, while the feature maps represent objects well.

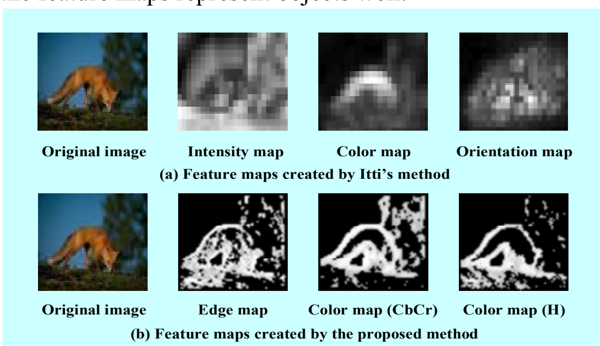

Fig. 2. Comparison of feature maps : Itti's method and the proposed method

2.1.1. Edge map

In order to create an edge map, an image is first resized to several smaller images such as 1/2, 1/3, and 1/4 of the original image. Then, by applying the Sobel edge operator to each resized image, edge images are generated. Finally, edge maps are created by equation (1), which calculates the mean of the differences between a center pixel and neighborhood pixels in edge images.

$$F_S(x,y) = \frac{1}{24}\left(\sum_{i=1}^{24}\left|I_S(x,y) - N_i\right|\right) \quad , \quad S = \{\frac{1}{2},\frac{1}{3},\frac{1}{4}\} \qquad (1)$$

where S is the scaling factor, and i denotes the index of neighborhood pixels. I_s and F_s denote the edge image and the corresponding feature map, respectively. By equation (1), three edge maps are created. Then, these edge maps are integrated to a representative edge map. In order to combine edge maps, each edge map is normalized into 1/4 size of the original image. Finally, by calculating the average of normalized edge maps, a representative edge map is created. The representative edge map is calculated by equation (2).

$$RFM_{Edge} = \frac{1}{3}\sum_{S=2}^{4} N\left(F_{1/S}^{Edge}\right) \qquad (2)$$

where $F_{1/S}^{Edge}$ denotes the resized map created by equation (1). $N(\bullet)$ represents a normalization of image into 1/4 size.

2.1.2. Color maps

Generally, a color map is known as a major feature to distinguish between salient objects and non-salient regions. In order to construct two color maps such as CbCr map and H map, we utilize Cb and Cr channels of YCbCr color model and the H channel of HSV color model. The process of creating color maps is similar to the process of creating the edge map. The CbCr map and the H map are constructed by equations (3) and (4). Fig. 3 shows three feature maps that are created from the original image.

$$RFM_{CbCr} = \frac{1}{2}\left(\sum_{S=2}^{4} N\left(F_{1/S}^{Cb}\right) + \sum_{S=2}^{4} N\left(F_{1/S}^{Cr}\right)\right) \qquad (3)$$

$$RFM_{H} = \frac{1}{3}\sum_{S=2}^{4} N\left(F_{1/S}^{Cb}\right) \qquad (4)$$

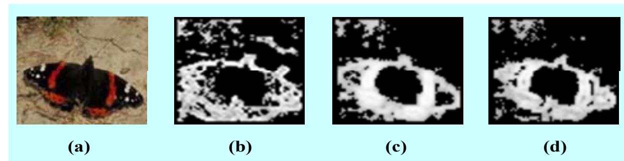

Fig. 3. Feature maps: (a) Original image, (b) Edge map, (c) CbCr map, (d) H map

2.2. Reference map

When an image contains cluttered backgrounds, it is not easy to identify object regions accurately. Therefore, we propose a reference map (RM) in order to reduce the influence of backgrounds. Since a reference map represents the boundary of objects excluding backgrounds in an image, a reference map can be utilized as a basis map for extracting salient regions. In order to create a reference map, we combine the edge map with two color maps, CbCr map and H map. The reference map is generated by equation (5) and the procedures for creating reference map is shown in Fig. 4.

$$\begin{aligned} CRFM_{E\_CbCr} &= \frac{1}{2}\left(RFM_{Edge} \oplus RFM_{CbCr}\right) \\ CRFM_{E\_H} &= \frac{1}{2}\left(RFM_{Edge} \oplus RFM_{H}\right) \\ RM &= \frac{1}{2}\left(CRFM_{E\_CbCr} \oplus CRFM_{E\_H}\right) \end{aligned} \qquad (5)$$

where $\oplus$ denotes an operator combining two feature maps. *CRFM* is a map that is combined from two feature maps. By combining $CRFM_{E\text{-}CbCr}$ with $CRFM_{E\text{-}H}$, a reference map is finally created. As shown in Fig. 4, a reference map presents the boundaries of objects very well, excluding backgrounds efficiently.

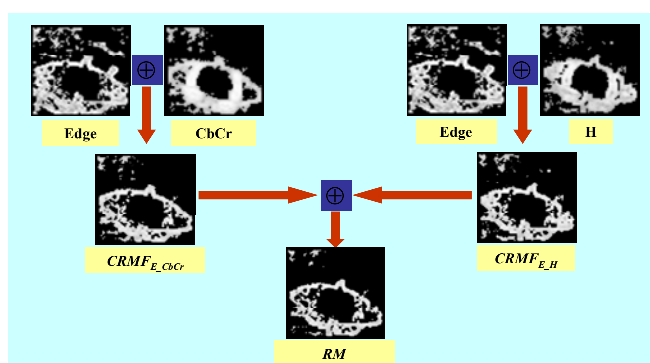

Fig. 4. Procedure for creating a reference map

2.3. Combination map

When a reference map is created by integrating feature maps, boundaries of objects may be over-eliminated. Therefore, in order to extract object regions accurately, it is necessary to recover the eliminated boundaries. A combination map(CM) is a map containing object regions, which is obtained by re-combining the reference map with three feature maps including edge map, CbCr map, and H map. The procedure for creating combination map consists of two steps, given in equations (6) and (7).

$$MRFM(x,y) = \begin{cases} RFM(x,y) & if\ RFM(x,y)\neq 0 \text{ and } RM(x,y)\neq 0, \\ \alpha RFM(x,y) & otherwise \end{cases} \quad (6)$$

$$CM = \frac{1}{3}\left(MRFM_{Edge} + MRFM_{CbCr} + MRFM_{H}\right) \quad (7)$$

where *MRFM* is the modified representative feature map based on the reference map, and α denotes a weighting factor between 0 and 1. In equation (6), if the pixels of *RFM* and *RM* at position (x,y) belong to the object boundary, *MRFM(x,y)* is set to the pixel value of *RFM*. Otherwise, *MRFM(x,y)* is set to the pixel value of *RFM* multiplied by the weighting factor. All the feature maps are modified by equation (6). Then, by combining the modified feature maps, a combination map is created. Fig. 5 shows combination maps and reference maps constructed from original images. Fig. 5 shows that the combination map is similar to the reference map if an object and background are clearly distinguishable. Otherwise, the combination map includes some object boundaries that have been excluded in the reference map.

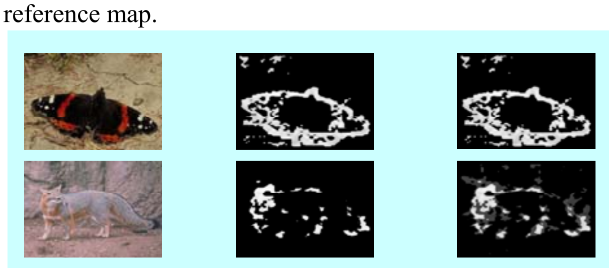

Fig. 5. Combination map and reference map created from example images: the left column represents original images, the middle column represents reference maps, and the right column shows combination maps.

2.4. Extraction of candidate object regions

After creating a combination map, candidate object regions are extracted. Since the combination map includes only object boundaries without interior regions, it is difficult to extract object regions directly from the combination map. Therefore, we extract candidate object regions containing object boundaries, before extracting real object regions. The procedure for extracting candidate object regions consists of two steps. First, noise regions that do not belong to the object regions are eliminated in the combination map. In the second step, candidate object regions are extracted by utilizing a convex hull algorithm[6,7]. Fig. 7 shows the result of extracting candidate object regions.

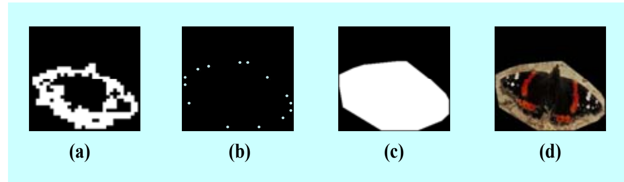

Fig. 7. Result of candidate object regions extracted by convex hull algorithm: (a) combination map after noise regions eliminated, (b) convex hull points, (c) candidate object region in combination map, and (d) candidate object region in the real image.

2.4. Extraction of salient objects

In Fig. 7(d), it is noted that the candidate object region contains both the object region and the background region. Therefore, to identify the object from the candidate object region, we utilize an image segmentation algorithm proposed by Wang[8]. After segmenting the candidate object region into sub regions, the regions adjacent to the boundary of convex hull are eliminated as background. Since the candidate object region contains the minimal backgrounds, our method is efficient to extract the salient object. Fig. 8 shows the result of extracting the salient object.

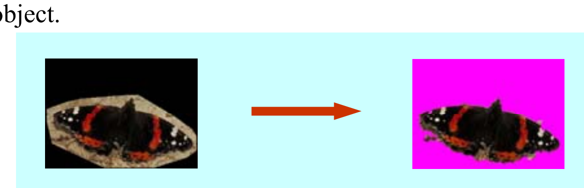

Fig. 8. Extraction of salient object

3. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed method, we selected 160 images from 10 categories of the Corel Photo Gallery as our test images. Two kinds of experiments have been carried out to evaluate the performance. The first one is to evaluate the extraction of candidate object regions. The second one is to evaluate how accurately salient objects are extracted. In the first experiment, the result of the proposed method is compared with other two methods including Itti's method[5] and Hu's method[6]. Even though it is very difficult to compare the performance quantitatively, Fig. 9 shows that the proposed method extracts candidate object regions more properly than the other two methods. While the existing methods do not include the whole object regions in the candidate object regions, the proposed method contains the object regions properly. The reason is the existing methods do not consider the edge information of objects. In the second experiment, we evaluate how accurately salient objects are extracted. In order to evaluate the performance, the ground truth representing salient object

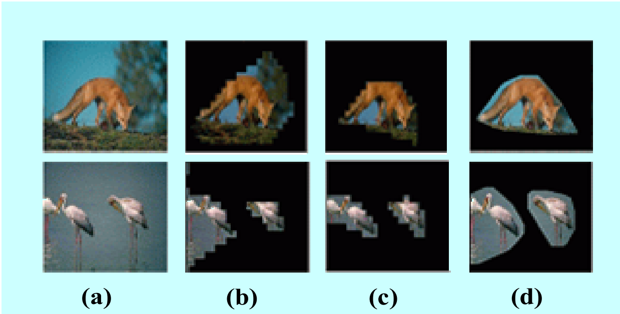

Fig. 9. Performance comparison of extracting candidate object regions: (a) original image, (b) result by Itti's method, (c) result by Hu's method, and (d) result by the proposed method

regions is obtained by a human operator. Then, the object regions generated by the proposed method are compared with those of the ground truth. The precision and recall are used as the performance measures. The precision is a measure of how accurately salient objects are extracted comparing with the ground truth. The recall is a measure of how well the salient object regions extracted by our method represent the salient object regions. Both measures are expressed in equations (8) and (9).

$$precision = \frac{N(S_M \cap S_A)}{N(S_M)} \times 100 \quad (8)$$

$$recall = \frac{N(S_M \cap S_A)}{N(S_A)} \times 100 \quad (9)$$

where S_M is a set of pixels in the ground trut, S_A is a set of pixels in the extracted object regions, and $N(S_M \cap S_A)$ denotes the number of identical pixels between the ground truth and the extracted object regions. $N(S_M)$ is the number of pixels in the ground truth, and $N(S_A)$ is the number of pixels in the extracted object regions. Fig. 10 shows the results of extracting salient objects from various images. The precision and recall are shown in Table 1. On the average, the proposed method achieves 84.3% precision rate and 81.3% recall rate.

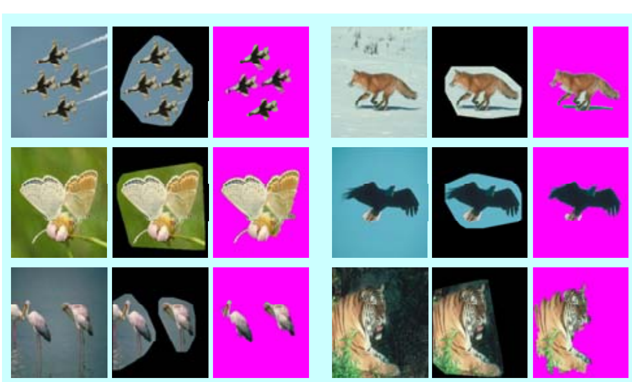

Fig. 10. Results of extracting salient objects

Table 1. Performance of the proposed method

Class	precisio n	recall	Class	precision	recall
Dogs	73.7 %	87.9 %	Airplanes	86.2 %	96.1 %
Foxes	80.1 %	80.8 %	Eagles	87.5 %	87.1 %
Hawks	87.2 %	81.2 %	Butterflies	91.1 %	86.3 %
Horses	88.9 %	67.8 %	Tigers	80.0 %	74.1 %
Antelopes	80.3 %	76.2 %	Waterfowl	87.6 %	75.7 %
			Average	84.4 %	81.3 %

4. CONCLUSIONS

In this paper, we proposed a technique for extracting salient objects in images using feature maps, regardless of the complexity of images and the position of objects. Experimental results have shown that the proposed method extracts the salient regions and salient objects efficiently, with 84.3% precision rate and 81.3% recall rate. As a future work, an objective measure that quantitatively evaluates the performance of object extraction needs to be developed.

5. ACKNOWLEDGMENTS

This work was supported by the Korea Science and Engineering Foundation.

6. REFERENCES

[1] A. W. M Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content Based Image Retrieval at The End of The Early Years," *IEEE Trans. on PA MI*, Vol. 22, pp. 1349-1380, 2000.

[2] H. K. Lee and Y. S. Ho, "A Region Based Image Retrieval System Using Salient Point Extraction and Image Segmentation," *Int. Conf. on PCM2002, LNCS 2532*, pp. 209-216, 2002.

[3] S. Kim, S. Park, M. Kim, "Central Object Extraction for Object Based Image Retrieval," *Int. Conf. On CIVR2003, LNCS 2728*, pp. 39-49, 2003.

[4] S. Y. Kwak, B. C. Ko, and H. Byun, "Automatic Salient Object Extraction Using The Contrast Map and Salient Points," *Int. Conf. On PCM2004, LNCS 3332*, pp. 138-145, 2004.

[5] L. Itti, C. Koch, and E. Niebur, "A Model of Saliency-based Visual Attention for Rapid Scene Analysis," *IEEE Trans. on PA MI*, Vol. 20, pp. 1254-1259, 1998.

[6] Y. Hu, X. Xie, W. Y Ma, L. T Chia, and D. Rajan, "Salient Region Detection Using Weighted Feature Maps Based on The Human Visual Attention Model," *Int. Conf. On PCM2004, LNCS 3332*, pp. 993-1000, 2004.

[7] A. M. Andrew, "Another Efficient Algorithm for Convex Hulls in Two Dimensions," *Information Processing Letters*, pp. 216-219, 1979.

[8] D. Wang, "Unsupervised Video Segmentation Based on Watersheds and Temporal Tracking," *IEEE Trans. on Circuits and System for Video Technology*, Vol. 2 pp. 539-546, 1998.