Indian Journal of Science and Technology, *Vol 8(15), 70918, July 2015*

ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Segmentation Using Saliency–color Mapping Technique

B. Mohd Jabarullah[1*] and C. Nelson Kennedy Babu[2]

[1]Computer Applications, DTTE, New Delhi - 110068, India; Isma72@rediffmail.com
[2]Thamarabharani Engineering College, Tirunelveli – 627358, Tamil Nadu, India

Abstract

The accuracy of face image classification/recognition is absolutely based on the extraction of object of interest in the image. This can be achieved by identifying salient object and eliminating background and other unwanted details. Therefore, this paper proposes a method for object identification and segmentation, which is based on finding fused saliency maps by combining pixel-intensities, visual attracted location of the image and color in the CIEL*a*b* color space. This method uses quadrants and cluster center window of the image without transforming other domain like frequency domain, to obtain saliency maps, in order to minimize time, cost. The object of interest from the original image is segmented using spatial color mapping technique to map saliency image with actual input image using threshold. In order to get perfect segmentation, the input image is subjected to filter and enhanced. This method is effectively applied on data sources such as MSRA, IMM database, and random samplings. The obtained saliency map, subject to binarised, is compared with ground truth using statistical measures such as accuracy, precision and recall, and compared with other methods. As results of this comparison, the proposed method outperforms.

Keywords: Filter, Quadrants Images, Saliency Map, Segmentation, Spatial Color, Threshold

1. Introduction

The main objective of image processing is to obtain required information from the image without affecting important details of the image. In general, most of the images are corrupted by noises during the acquisition of the image which give an impact on the perfection of the required information. So, de-noising technique is applied on corrupted images to fulfill the requirement[1]. The enhanced images are now subjected to segmentation for further image analysis.

The segmentation is the process of partitioning the digital image into disjoint regions. The accuracy level of image classification is purely based on well defined segmented regions. False segmented region may lead to the entire system failure. Parameters which affect the segmentation process are illumination, weak object boundaries, inhomogeneous object region, high degree of spatial variability in scale, complicated background[2,3]. Therefore, many image segmentation techniques have been developed and most famous techniques are: region based, edge based, fuzzy logic based, partial differential equation based, artificial neural network based, threshold based image segmentation. The different types of image segmentation techniques are basically based on three properties such as color, intensity, orientation and texture of the image[4]. In this context, an automatic approach to soft color segmentation, which produces soft color segments with an appropriate amount of overlapping and transparency essential to synthesizing natural images for a wide range of image-based applications[5]. A very critical point of observation in the segmentation process is an image with complex background with uniform intensity level for both object of interest and background. In order to estimate background and removal, range and color are considered for the good segmentation on video

**Author for correspondence*

image. Selection of range gives solution to the problem if an image exists with shadow[6]. Background subtraction is performed by adopting Codebook background subtraction algorithm which fuses depth and color information to segment foreground regions, focused on video analytics. In this method, high performance and low cost depth sensor is used for depth estimation[7]. A paper on Bi-layer Segmentation of Live Video addresses the problem of accurately extracting a foreground layer from video in real time[8]. Apart from live video/video image segmentation, still image segmentation is also playing important role in digital image processing. In view of that, the object detection method is based on the Gestalt principle of symmetry[9]. Gestalt principle refers to theories of visual perception. These theories attempt to describe how people tend to organize visual elements into unified wholes when certain principles are applied. These principles are similarity, continuation, closure, proximity, symmetry, figure and ground. Salient object is detected using symmetry-saliency model, one of the Gestalt principles used to predict human eye fixation[10]. This method is compared with contrast-saliency model. Intensity, color and orientation maps are calculated and combined to form final saliency map. That is mainly based on contrast-saliency model. Winner-take-all is used to suppress less active location[11]. Saliency value is calculated using frequency domain processing[12]. A frequency tuned algorithm for computing saliency map to exploit all low and high frequency content using color and luminance[13]. Using luminance and color saliency maps are generated at different scales and normalized to get final saliency map and apply threshold on average saliency value to segment the salient object[14]. A novel method is proposed to compute saliency map based on local saliencies over random rectangular regions of interest[15] and similarly by combining band-pass filtering in frequency domain, image center and color[16] and inspired by these works, a modified proposed method is proposed by combining these two methods by using quadrant regions of image, spatial domain not in frequency domain, and pixel intensities, color and visual attention to yield higher saliency maps. Segmentation of saliency map is carried out by pixel wise color mapping technique.

The remaining part of the paper is organized as follows: Section 2 presents Proposed Method for image segmentation, Section 3 presents Experimental Result and Analysis, Section 4 presents Conclusion.

2. Proposed Method

The input image I is a real world image with or without complex background. This image is probably affected by most frequent salt and pepper noise. Based on recent research on noises, the input images are subjected to Enhanced Switching Median Filter (ESMF) to remove noises. This noise free image is then converted into the CIEL*a*b* color space. This color space is perfectly produces various length of different colors in L* channel, a* channel and b* channel from lower to higher depth. This color space is best suited for saliency measurement to obtain salient object. The L* channel is sub imaged into four according to four quadrants and one cluster center window, of an L* channel. Now saliency value for each pixel in each sub image in L* is calculated as;

$$\mathrm{E_L} = \sum x f(x) \tag{1}$$

Where pdf: $f(\mathrm{x}) = \dfrac{\mathrm{nX}}{\mathrm{nS}}$

Then,

$$\mathrm{S_L} = \sum \left| \mathrm{I_{i,j}} - \mathrm{E_L}(\mathrm{x}) \right| \tag{2}$$

Similarly, for a* channel, apply the same procedure for L*

$$\mathrm{S_a} = \sum \left| \mathrm{I_{i,j}} - \mathrm{E_L}(\mathrm{x}) \right| \tag{3}$$

And for b* channel

$$\mathrm{S_b} = \sum \left| \mathrm{I_{i,j}} - \mathrm{E_b}(\mathrm{x}) \right| \tag{4}$$

Finally saliency map is obtained by applying Euclidean norm as

$$\mathrm{IS\ (I)} = \sqrt{(S_L^2 + S_a^2 + S_b^2)} \tag{5}$$

Pixel value of final saliency map is normalized in the interval [0,255] and subsequently subjected to ESM filtering.

In addition to intensity saliency map, Saliency maps of color and location are also derived, in[16]. The center of object of the image is more attracted to the human visual system. Therefore, object nearer to center and center object is more salient than the object is far away from the center. This can be defined by a Gaussian map as:

$$\mathrm{CS\ (I)} = \frac{1}{e^{\left(\frac{|x-c|^2}{\sigma^2}\right)}} \tag{6}$$

Where c is the center of image, x is location of center saliency object and σ is 114.

According to CIEL*a*b* color space, green-red color information is contained in a* channel and blue-yellow color information is contained in b* channel. Color space takes values from lower to higher i.e. greenish to reddish in a* channel similarly in b* channel, bluish to yellowish. So, based on both range of colors, highest positive values in both a* and b* channel refers red and yellow respectively. This is called warm color which is more attracted in the human visual system. Otherwise, less attracted color i.e. cold which has lower value.

In normalized cross correlation one subtracts mean, which means subtraction mitigates brightness variations and divides by standard deviation, which means the standard deviation migrates variations in the spread of data about the mean so that two images have similar mean and standard deviations. So that, channel a* is normalized in the range [0,1] using the following equation;

$$C_{an}(x) = \frac{C_a(x) - \mu\left(a^*\right)}{std\left(a^*\right)} \tag{7}$$

Similarly, the b* channel is also normalized in the range [0,1] using the following equation;

$$C_{bn}(x) = \frac{C_b(x) - \mu\left(b^*\right)}{std\left(b^*\right)} \tag{8}$$

For each channels, its color similarity is obtained through the following equation;

$$C_s = e^{\left(\frac{-ed}{\sigma^2}\right)} \tag{9}$$

Where ed is Euclidean distance on the mean color between channel a* and b* and σ is the standard deviation ie 0.25. If there is color similarity of a pixel in both channels, then eq. (9) assigns smaller values to these pixels. As these pixels have to gain higher saliency values in order to obtain salient object and to form saliency map. To achieve this, the color similarity is subtracted from higher level of the interval. It is defined as:

$$CLS\ (I) = 1 - C_s \tag{10}$$

According to visual inspection, lower pixel value in the channel is less attracted; magnify its value to suppress in the final saliency map.

Finally these three saliency maps are combined to form final saliency map. This is given as:

$$VMap = IS\ (I) \times CS(I) \times CLS(I) \tag{11}$$

The original image is pixel-wise mapped with final saliency map obtained by above method. If a pixel value in saliency map is less than or equal to T (Threshold value) then the corresponding pixel intensity value in the input image set to 0 otherwise no changes in the pixel intensity value of input image, as in equation (9).

$$I(x) = \left\{ I(x)\ \text{if}\left(\text{Intensity}(x)\right) > T\ @0\ \ \text{otherwise})\right\} \tag{12}$$

3. Experimental Result and Analysis

For experimental result and analysis of the proposed method, the experiment was conducted using three different data sources such as R. Achanta's Microsoft Research Asia Salient Object Database (MSRA SOD), IMM dataset[17] and Random Sampling (RS). Here, MSRA SOD database consists of 25000 actual images grouped into two (20000 images as one group and 5000 images as another group) and 1000 manually segmented ground truth images. Second one IMM dataset consists of 240 images of 40 different persons. Six images of each person are taken with different head angle (approx. +/- 30^o), different lighting conditions (spot / non spot) and different face expressions. In the third source of data, images are randomly collected from internet with complex background. Last two databases are manually segmented to create ground truths. The proposed method accepts any size of the image, as one of the parameter. Association between predicted saliency map and ground truth is almost similar.

These three data bases are tested by Contrast-Saliency color mapping method, Frequency Domain color mapping method and proposed method. Saliencies maps are generated by each of these methods, subjected to binarized, using adaptive threshold T i.e. twice the mean saliency of the image. Now, the similarity between binarized saliency image and ground truth is extracted using accuracy, precision, recall. In addition, it is necessary to calculate optimum run time of above three methods.

From the Figure 1 shows outcome of proposed method on MSRA SOD database, Figure 2 on IMM Database and Figure 3 on random samplings, it is observed that the segmented image is more accurate and almost similar to ground truth by visual inspection.

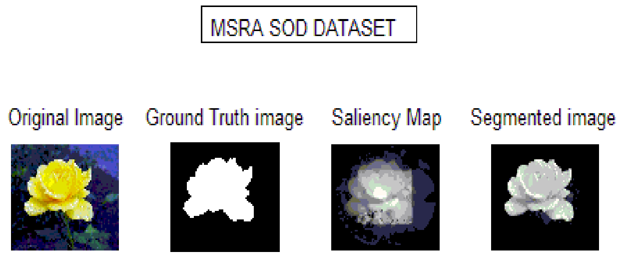

Figure 1. Outcome of proposed method on MSRA SOD database.

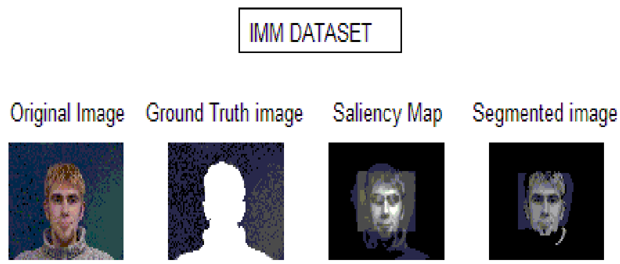

Figure 2. Outcome of proposed method on IMM database.

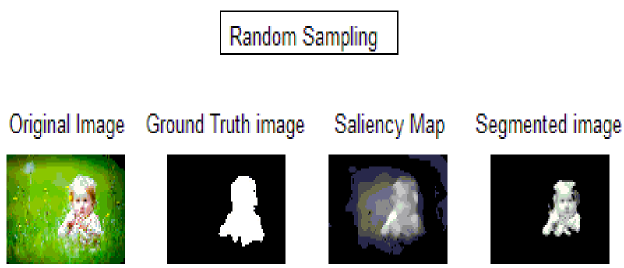

Figure 3. Outcome of proposed method on random sampling.

Let P is the positive instances, and N is the negative instances, then

$$\text{Accuracy}=\frac{(TP+TN)}{P+N} \quad (13)$$

Precision, PPV (Positive Predictive Value)

$$\text{PPV}=\frac{(TP)}{(TP+FP)} \quad (14)$$

Recall, TPR (True Positive Rate)

$$\text{TPR}=\frac{(TP)}{TP+FN} \quad (15)$$

Where TP- True Positive, TN – True Negative, FP – False Positive, and FN – False Negative. Further, to find optimum run time, system which was used to run all three methods developed by using MATLAB R2009b, on Intel core Duo CPU @2.20 GHz 1 GB RAM 32 bit operating system.

In the same context of visual perfection, the outcome of the statistics performance evaluation, as shown in Table 1, the proposed method produces the high significant result of accuracy, precision and recall. As for as optimum run time is concern, the proposed method has little higher value than the rest, since the proposed method incorporated by three significances of image such as pixel wise intensities, visual attracted color and location, not like other two methods. Hence, from the observation, the proposed method is significantly better than the rest in terms of statistical measures.

Table 1. Performance evaluation

Sr. No	Method	Data Source	Optimum Run Time	Accuracy	Precision	Recall
1	Contrast-Saliency Color Mapping	MSRA	0.8673s	0.9236	0.9236	1.0000
		IMM	2.0092s	0.1269	0.1069	0.9800
		RS	0.3985s	0.8092	0.3219	0.5774
2	Frequency Domain Color Mapping	MSRA	1.0712s	0.9692	0.9087	1.0000
		IMM	2.1011s	0.0985	0.0985	1.0000
		RS	0.6112s	0.8142	0.1023	0.4168
3	Proposed Method	MSRA	1.0997s	0.9974	0.9927	1.0000
		IMM	2.1891s	0.2205	0.2205	1.0000
		RS	0.6248s	0.8522	0.3862	0.6669

4. Conclusion

The proposed method "Segmentation using Saliency-Color Mapping Technique" is incorporated by two applications such as background elimination and foreground segmentation from complex background real world images. This method has more success rate in detecting unknown objects in unknown scenes. This model is more close to human visual attention system. This proposed system compared with existing system and the result suggested that this method outperforms well in terms of statistics measures, cost, time and visual quality/perfection.

5. References

1. Jabarullah BM, Saxena S, Kennedy Babu CN. Survey on noise removal in digital images. IOSRJCE. 2012 Sep-Oct; 6(4):45–51.
2. Kherikhah E, Tabtabaie ZS. A hybrid face detection approach in color images with complex background. Indian Journal of Science and Technology. 2015 Jan; 8(1):49–60.
3. Zuva T, Olugbara OO, Ojo SO, Ngwira SM. Image segmentation, available techniques, developments and open issues. Canadian Journal on Image Processing and Computer Vision. 2011 Mar; 2(3):20–9.
4. Khan W. Image segmentation techniques: a survey. Journal of Image and Graphics. 2013 Dec; 1(4):166–70.
5. Tai Y-W, Tang C-K. Soft color segmentation and its alications. IEEE Transactions on Pattern Analysis and Maching Intelligence. 2007 Sep; 29(9):1520–37.
6. Gordon G, Darrell T, Harville M and Woodfill J. Background estimation and removal based on range and color. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 1999 Jun. p. 1–6.
7. Enrique J, Sanchez F, Diaz J, Ros E. Background subtraction based on color and depth using active sensors. Open Access Sensors. 2013; 13(7):8895–915.
8. Criminisi A, Cross G, Blake A, Kolmogorov V. Bilayer segmentation of live video. IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2006 Jun 17-22; 2006. p. 53–60.
9. Kootstra G, Kragic D. Fast and bottom-up object detection, segmentation, and evaluation using gestalt principles. Proceedings of the International Conference on Robotics and Automation (ICRA); 2011 May; IEEE; p. 3423–8.
10. Kootstra G, Bergstrom N, Kragic D. Using symmetry to select fixation points for segmentation. Proceedings of the International Conference on Pattern Recognition (ICPR); 2010 Aug; IEEE; p. 3894–7.
11. Itti L, Koch C, Neibur E. A model of saliency based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1998; 1254–9.
12. Hou X, Zhang L. Saliency detection: a special residual approach. IEEE Conference on Computer Vision and Pattern Recognition; 2007 Jun. p. 1–8.
13. Achnata R, Hemani S, Estrada F, Susstrunk S. Frequency-tuned salient region detection. IEEE International Conference on Computer Vision and Pattern Recognition; 2009 Jun; p. 1597–604.
14. Achanta R, Estrada F, Wills P, S Susstrunk. Salient region detection and segmentation. International Conference on Computer Vision Systems. 2008; 5008:66–75.
15. Vikram TN, Tscherep M, Wrede B. A saliency map based on sampling an image into random rectangular regions of interest. Pattern Recognition. 2012; 45:3114–24.
16. Zhang L, Gu Z, Li H. SDSP: A Novel Saliency detection method by combining simple priors. IEEE, ICIP; 2013. p. 171–5.
17. Stegmann MB. IMM face data set for research purposes. IMM Publications; IMM Technical report 2005-11.