Evaluation and Limited Characterization of a Color Image Segmentation Method using Synthetic Images

Rodolfo Alvarado-Cervantes, Edgardo M. Felipe-Riveron and Rodolfo Alvarado-Martinez

Centro de Investigación en Computación, Instituto Politécnico Nacional, México
ateramex@gmail.com; edgardo@cic.ipn.mx; ralvarado@ateramex.com.mx

Abstract. We present an evaluation and a limited characterization of an own semi-automatic color image segmentation method using generated synthetic images with its associated ground truth. The synthetic images where designed to evaluate the efficiency of the resulting color information achieved from color segmentation algorithms. By the use of ROC curves and its analysis, we achieved the evaluation and some particular characteristics of our segmentation method, such as the level of sensibility related to the threshold selection and to the appropriate number of pixels to have by the color sample needed by the algorithm.

Keywords: Color segmentation evaluation; color segmentation characterization; synthetic color image generation; ROC curves evaluation.

1 Introduction

Image segmentation consists of partitioning an entire image into different regions, which are similar in some predefined manner. Segmentation is an important feature of human visual perception, which manifests itself spontaneously and naturally. All subsequent steps, such as objects recognition depend heavily on the quality of segmentation [1].

While the development of segmentation algorithms has attracted remarkable consideration, relatively fewer efforts have been spent on their evaluation [2], [3], [4] and [5]. Since none of the proposed automatic segmentation algorithms published are generally applicable to all types of images, and different algorithms are not equally suitable for particular applications, the performance evaluation of segmentation algorithms and its characterization are very important subjects in the study of segmentation [2].

For a long time the evaluation was limited to few real images acquired from particular application, that has as advantage that they are closer to reality although its intrinsic random nature makes them no suitable for analytical evaluation [2], [3], [6]. Many undetermined characteristics of those images prevent its use to compare different segmentation techniques because many phenomena are mixed and make it difficult to study its influence individually [2]. Another problem arrives from the lack

Advances in Computer Science and Engineering
Research in Computing Science 63, 2013 pp. 45–54
Paper Recived 14-09-2013 and Acepted 21-10-2013

of a Ground Truth (GT) that has to be obtained from 'experts' whose results always have intrinsic differences. This subjective and imprecise procedure is not appropriate for quantitative evaluation [3], [6].

So far segmentation evaluation methods can be divided in two groups: analytical and empirical. The analytical methods directly inspect and evaluate the segmentation algorithms themselves by analyzing their principles and properties. The empirical methods indirectly judge the efficiency of segmentation algorithms applying them to test images and measuring the quality of the results [2], [4], [5].

Several empirical methods have been proposed, the great majority can be classified in two types: goodness methods and discrepancy methods. In the first category some desirable properties of segmented images, often established according to human intuition, are measured by 'goodness parameters'. The performances of segmentation algorithms are judged by the values of goodness measures. In the second group the GT that presents the ideal or expected segmentation result must be first found. The actual segmentation results obtained by applying a segmentation algorithm are compared with the reference to count their differences. The performances of segmentation algorithms under investigation are then estimated according to discrepancy measures [2], [3], [4] and [5].

Receiver operating characteristics (ROC) curves are useful for organizing classifiers and visualizing their performance. ROC curves are commonly used in medical decision making, and in recent years have been used increasingly in machine learning and data mining research [10].

In this work an evaluation and characterization of a semi-automatic color image segmentation method using synthetic images generated with its associated ground truth is presented. The synthetic images where designed to evaluate the efficiency of achieved color information from a given segmentation algorithms. The system was applied to our semiautomatic color segmentation method already presented in international conferences [1], [7]. By the use and analysis of ROC curves we obtained some proper characteristics of the segmentation method under study such as its stability related to the threshold selection and to the selection of an appropriate number of pixels required by the color samples. This system may be useful for assessing the quality of use of the color information inside the segmentation algorithms in general.

2 Previous works

The first comprehensive survey on evaluation methods of image segmentation is presented in [2]. It brings a coherent classification of existing methods at the time. An up to date of 5 years of progresses in the subject is presented in [4] after the first survey. Another actualization is presented 5 years later [5], embracing together the principal methods of segmentation evaluation available until 2007.

In [3] a comprehensive survey on unsupervised methods of segmentation evaluation is presented. The authors propose a hierarchy of published methods at that time by summarizing the main branches and locating the group of unsupervised methods on it. They mention its advantages such as not requirement of GT to obtain quantitative results. They also propose the main lines of future research of this kind of methods.

In [6] a way to design synthetic images and a corresponding GT for evaluating segmentations algorithms is presented. They introduce a general framework and general design considerations. They also present a system for generating synthetic images in shades of gray taking into account theirs design considerations. The behavior of a segmentation method in gray images using thresholding is studied and some remarks are obtained.

In [8] is presented a database containing GT from segmentations produced manually by a group of people from a wide variety of color natural scenes. The authors define two related error measures that quantify the consistency between segmentations of different granularities. This measure permits to make comparisons between segmentations made by people and segmentations made by the computer of the same scene.

In [11] is presented a comparative study of 14 evaluation criteria of supervised image segmentation methods by generating edges. The study was done in two parts: (1) evaluation with synthetic images to characterize the general behavior of the algorithm is complemented by (2) an assessment over a selection of real color images. To get the GT of different peoples the authors mention that theirs propositions were merged, but do not give details of how they accomplished it.

3 Design and synthetic image generation

In [6] three important design considerations for synthetic images are presented: (1) Synthetic images should be appropriate for a quantitative study and should allow objective evaluations of its properties. (2) The synthetic images should reflect the main features of real images, i.e. corruption factors such as noise and blurring, variation of parameters such as size, shape, etc. (3). The system should allow the generation of images with progressive variations of each parameter. In this way the study of the influence of each individual parameter is possible.

In this work we try to extend the concepts and design considerations to RGB true color images presented in [6]. We created synthetic images with figure and background in color and selected the circle as the basic object to create our base image. The circle is appropriate because it is a base model of the objects of interest in various fields such as the detection of white blood cells in microscopic images, and others. It is also more difficult to detect than other objects made of straight lines by methods such as convolution with masks that highlight edges.

There are other techniques for generating synthetic images such as edge-oriented methods [11] or by the introduction of texture in the images [12]. In our case, the segmentation method is region-oriented without using the texture information; the synthetic images were designed with these considerations.

The basic colors selected for both figure and background were based on maintaining constant intensity to 0.5 and saturation to 0.3 and only varying the hue.

Hue was selected as the parameter because its change integrates the three RGB color channels together, making it more difficult to be processed by extending grayscale techniques to each color channel, thus forcing the segmentation algorithms in evaluation to use the color information holistically.

We varied the characteristics of the basic image (circle), such as size, shape, inclination, contrast, and color. A median filter was applied to the entire image (before the noise addition) to remove sharp edges in order to produce a blurring in some measure. Additive Gaussian noise was added to simulate white noise commonly present in real images [6]. The amount of Gaussian noise applied to images is based on the concept of signal to noise ratio (SNR), a high SNR means a recognizable signal with low noise respect to it; respectively a low SNR means high noise respect to the signal. We use as signal the contrast between figure and background: signal = abs (RGB (Figure)-RGB (Background))/3, extending its use in [6] to RGB images. A sample of the set of images can be seen in Fig. 1 where the vertical axis represent changes in size and horizontal axis represent changes in the amount of noise added for SNR = 32 (far left), 8, 4, 2 and 1 (far right).

About the images varying the color contrast, the common hue difference was decreased from 32 degrees to 2, where even to a human observer is hard to find any color difference (see Fig. 2 far right). Figure 2 shows a set of test images varying the color contrast adding a Gaussian noise in steps of SNR equal to two.

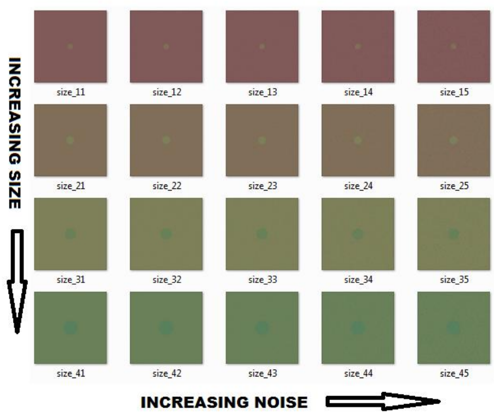

Fig. 1 Set of test images changing in size and signal to noise ratio.

Fig. 2 Set of test images changing the color contrast: 32, 24, 16, 8 and 4 degrees with additive Gaussian noise of SNR = 2

4 Experiments and results

We used the synthetic color images generated with its corresponding GT through the method described above. The color segmentation method requires samples from the operator having, typically from 1 to 9 pixels, which are representative of the desired color target. Samples taken were created trying to select pixels from a zone with a plain color as well as from the edges where the color presents a transition zone (Fig. 3 left).

The evaluation was carried out in the following manner:

1. A database of synthetic color images with its correspondent GT was generated modifying gradually the basic image (in our case a circle) the following parameters: size, shape (changing the eccentricity towards increasingly elongated ellipses) with arbitrary angles, color contrast, varying the color difference between the object and background shrinking almost imperceptible (see Fig. 2 far right), noise varying from SNR of 32 to 1 applied to previous pictures. While generating the synthetic images, the system creates simultaneously its corresponding GT.

2. The semi-automatic color segmentation algorithm [1], [7] was applied to the synthetic images by taking samples from 1 to 9 pixels. Two examples of pixel samples are shown in Fig. 3. In some cases pixels were taken randomly in some other cases in line patterns from the center of the image (as in case of Fig. 3 right) where the operator cannot easily see the color to select because of the applied noise or by the low color contrast (as in Fig. 2 far right).

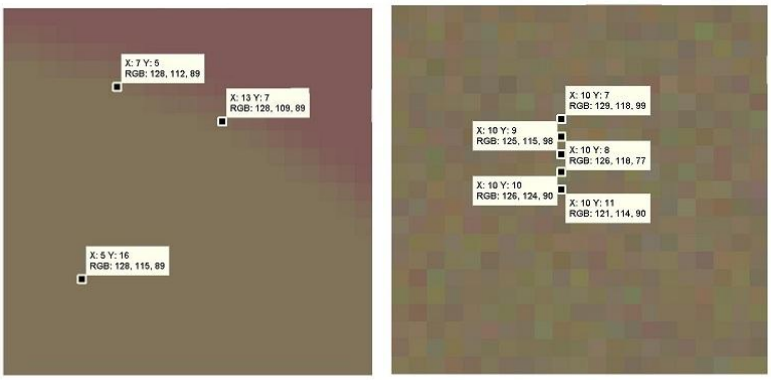

Fig. 3 Examples of pixel samples

3. Any test image was pre-processed; as post-processing a morphological closing filter with a disk of radius of 5 pixels was applied to all images to remove groups of connected pixels in number lower than 30 .

4. We displayed the results of the evaluation using ROC curves due to its ability of visualizing the successes and the failures and their interrelation [10]. They also have been successfully used to assess segmentation algorithms. We used ROC curves for True positives (TP) rate vs. False positives (FP) rate: TP rate = TP/P, FP rate = FP/N

using pixel classification either for object and background [10]. We also use graphs of TP rate, FP rate separated for convenience.

Figure 4 shows the ROC curves for varying size, shape and color contrast. As it can be appreciated, the majority of points (in red) are concentrated in the left corner, showing good performance of the segmentation algorithm in most cases with TP rate around 95% and FP rate lower than 0.1% in most cases. It was expected unvarying results from changes in size and shape because of the algorithm is pixel oriented. The color contrast ROC curve shows good performance of the algorithm even in cases where the color contrast was so low that to an observer is difficult to find any color difference (as in Fig 2 far right).

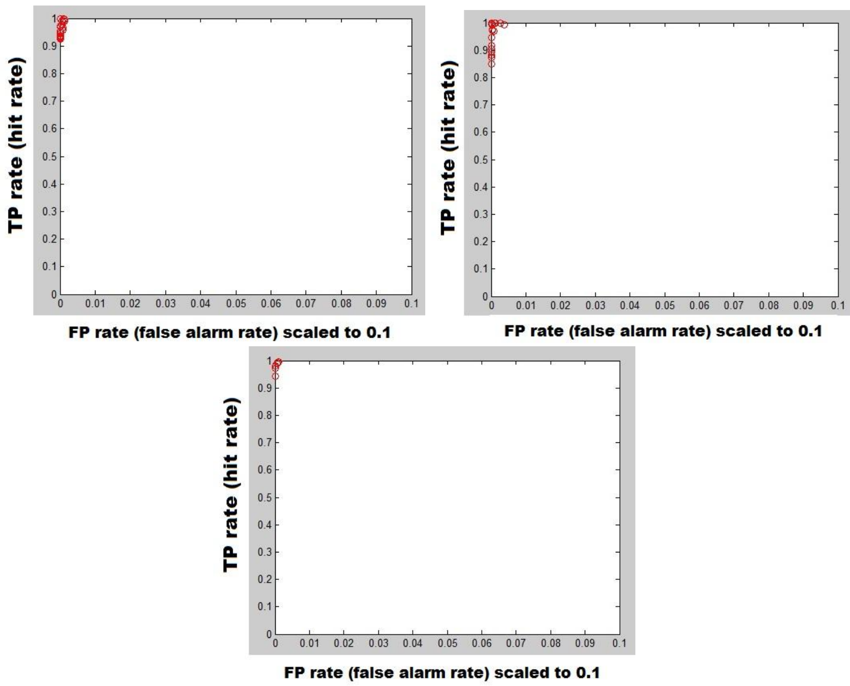

Fig. 4 ROC curves for size (left), shape (center) and color contrast (right)

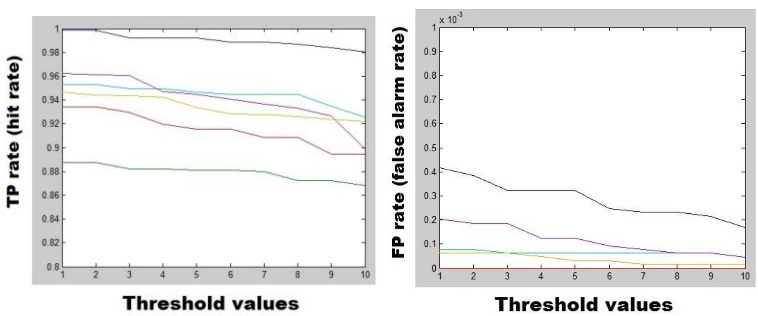

Fig. 5 Graphs for TP rates (left) and FP rates (right) for different thresholds

Figure 5 shows two graphs where it is plotted the behavior of the segmentation of six different images varying in size and shape modifying the threshold applied to the algorithm. The thresholds varied, using Otsu´s threshold [9] at a center value (selected by Otsu), 25 gray levels above and 25 levels below. Abscissa value 1 corresponds to Otsu at –25 gray levels, abscissa value 5 correspond to Otsu´s value, and abscissa value 10 to Otsu´s value at +25 gray levels above. As it can be observed from the graph the TP rate and FP rate values vary around 2% only, showing that the segmentation algorithm is scarcely sensitive to the thresholding value finally chosen.

Figure 6 (left) shows another graph where the TP rates (three graphs without marker) and FP rates (three graphs with marker) are plotted against the number of pixels per sample taken for the initial color sample. The blue line corresponds to the segmentation of different types of objects in images without noise, the red line corresponds to noise with SNR = 2 and the green line to noise with SNR = 1. The magenta line corresponds to FP rate with SNR = 2 that overlaps with FP rate of images with no noise, showing low error rates (near cero) in both cases. The cyan line corresponds to FP rate of noisy images of SNR = 1. It fluctuates showing the lower values near cero for 1, 3, 5 and 10 samples. From analysis of the graph we can point out that 5 pixels per sample is a good choice for most cases, having high TP rate and low PF rate despite noise. We can also observe that in cases without noise, only one pixel per sample is needed having TP rate around 90% and FP rate near cero. Figure 6 (right) shows the ROC curve corresponding to the noisiest cases were the points closer to the top-left corner correspond to the 3, 5 and 10 pixels per sample cases. Then, we can conclude that choosing samples of 5 pixels maximum the algorithm has good performance.

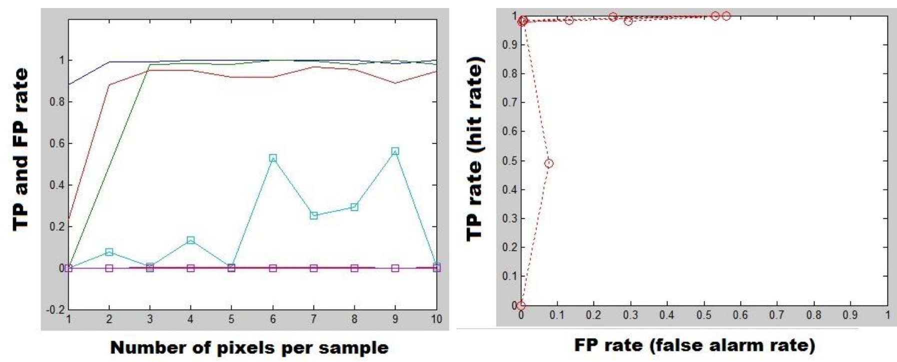

Fig. 6 Graph of TP rates for number of samples (left) and ROC curve for noise level SNR = 1 (right)

5 Conclusion

We presented a quantitative evaluation and characterization of a semi-automatic color image segmentation algorithm by generating synthetic images each with its corresponding ground truth image, with a circular object of interest gradually varying their size, shape, color contrast and amount of additive noise. We calculated the True Positive rate (TP rate) and False Positive rate (FP rate) for every image to obtain ROC curves of the results. This system may be useful for assessing the quality of use of the color information inside the segmentation algorithms in general.

This evaluation system was applied to evaluate the performance of our semi-automatic color segmentation algorithm of color images. In this algorithm color is the only discriminating feature. Behavior was assessed by varying size, shape, and color contrast, amount of additive noise, threshold and number of pixels taken in the sample. After an analysis of its corresponding ROC curves the algorithm shows good performance in most cases with TP rate around 95% and FP rate lower than 0.1%. The ROC curve for color contrast shows good performance of the algorithm even in cases where the color contrast is so low, that for a normal observer is difficult to find any color difference.

We made a characterization study of the algorithm behavior by varying the number of pixels per sample and the threshold value needed for final segmentation. We observed from analysis of the graphs that the segmentation algorithm barely sensitive to the threshold value. Even with changes of +/- 25 gray values above and below from Otsu´s chosen value, the obtained TP rate varied in 2% and FP rate in less than 0.1%. From a similar analysis of the graphs corresponding to the number of pixels samples we can observe that five pixel per sample give fine results of above 90% for the TP rate and less than 0.01% for the FP rate in average in most cases even with a high level of noise of SNR = 1 giving the best results in all cases of study.

Acknowledgements

The authors of this paper wish to thank the Computing Research Center (CIC), Mexico; Research and Postgraduate Secretary (SIP), Mexico; National Polytechnic Institute (IPN), Mexico; and CONACyT, Mexico, for their economic support to this work.

References

1. Alvarado-Cervantes R., Segmentación de patrones lineales topológicamente diferentes, mediante agrupamientos en el espacio de color HSI, M. Sc. Thesis, Center for Computing Research, National Polytechnic Institute, Mexico. (2006).
2. Zhang Y.J., A Survey on Evaluation Methods for Image Segmentation, Pattern Recognition, Vol. 29 No 8 pp. 1335-1346, Elsevier Ltd. (1996).
3. Zhang Y.J., A Summary of Recent Progresses for Segmentation Evaluation. In: Zhang Y.J., Advances in Image and Video Segmentation. USA IRM Press, pp. 423-439 (2007).
4. Zhang Y.J., Gerbrands J.J. On the Design of Test Images for Segmentation Evaluation. In: Signal Processing VI, Theories and Applications pp. 551-554, (1992).
5. Zhang Y.J., A Summary of Recent Progresses for Segmentation Evaluation. In: Zhang Y.J., Advances in Image and Video Segmentation. USA IRM Press, pp. 423-439 (2007).

6. Martin, D., Fowlkes, C., & Tal, D., et al., A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. Proceedings of the 8th International Conference on Computer Vision (2001), 2, 416-423.
7. Alvarado-Cervantes R., Felipe-Riveron E.M. and Sanchez-Fernandez L.P., Color Image Segmentation by means of a Similarity Function. In: I. Bloch and R.M. Cesar, Jr., Jacobo (Eds.), Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. LNCS Volume 6419 pp. 319-328. Springer, 15th Iberoamerican Congress, CIARP 2010, Sao Paulo, Brazil, 8-11 Nov (2010), ISBN: 978-3-642-16686-0.
8. Martin, D., Fowlkes, C., & Tal, D., et al., A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. Proceedings of the 8th International Conference on Computer Vision (2001), 2, 416-423.
9. Otsu N., A Threshold Selection Method from Gray-Level Histograms, IEEE Transactions on Systems, Man and Cybernetics, 9(1): pp. 62-66. (1979).
10. Fawcett T., An introduction to ROC analysis. Pattern Recognition Letters 27 (2006) pp. 861- 874.
11. Chabrier S., Laurent H., Rosenberg C., Zhang Y.J., Supervised evaluation of Synthetic and Real Contour Segmentation Results, 14th European Signal Processing Conference (EUSIPCO 2006), Fawcett Florence, Italy, September 4-8, (2006).
12. Haindl, M. - Mikes, S., Texture Segmentation Benchmark, 19th ICPR, Tampa 2008, IEEE Press, http://mosaic.utia.cas.cz.