Indian Journal of Science and Technology, Vol 9(8), DOI: 10.17485/ijst/2016/v9i8/87907, February 2016
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Performance Evaluation of Image Segmentation using Objective Methods

D. Surya Prabha and J. Satheesh Kumar*

Department of Computer Applications, Bharathiar University, Coimbatore - 641046, Tamil Nadu, India;
spayrus@gmail.com, jsathee@rediffmail.com

Abstract

Background/Objectives: Image segmentation, a crucial and an essential step in image processing, determines the success of higher level of image processing. In this paper, a detailed study about different evaluation techniques based on subjective and objective methods have been discussed. **Methods/Statistical analysis:** An application specific characteristic of image segmentation paves a way for development of numerous algorithms. Traditionally subjective method of evaluation is used to determine the segmentation performance accuracy. As this evaluation method is quantitative and biased, a qualitative method of evaluation is demanded. This is done using the objective method of evaluation where discrepancy and goodness methods are used.Discrepancy method is used in widespread for predefined benchmark images where it has corresponding ground truth image for comparison. Goodness method is used for real time images where no ground truth image is available for comparison. These methods of objective evaluation are highly needed to validate the segmentation methods which are increasing rapidly in recent years. **Findings:** A detailed study of different evaluation methods are discussed and experimented over different segmentation methods. Boundary based methods like sobel, canny, susan, region based methods like region growing, thresholding and a hybrid method, combining boundary based and region based method are used for the purpose of experimentation.Experimental result shows that hybrid method performs better than other existing ones and also highlights the importance ofimage quality assessment method to identify a better segmentation technique for all type of images.

Keywords: Discrepancy Measures, Empirical Method, Goodness Measures, Image Segmentation, Objective Evaluation

1. Introduction

Image segmentation is an essential task in image analysis that plays an indispensable role in both the computer vision and image processing applications to have proper image understanding and accurate machine perceptions[1,2]. It is used to identify and segment required region of interest from the entire image scene. Due to its application specific characteristics and its importance in several applications, numerous image segmentation techniques are developed in the past few decades and yet more research works on segmentation are also being proposed[3]. Image segmentation algorithm broadly comes under two categories as boundary based and region based segmentation[4]. Boundary based segmentation method are based on pixels discontinuity property and region based segmentation methods are based on pixels similarity property[5, 6]. Hybridizing the concept of boundary based and region based method is competent to produce a better segmentation result. Surya Prabha and Satheesh Kumar proposed aHybrid combinationof edge detection based on color gradient and region growing for banana fruit segmentation[7]. Evaluation of these segmentation algorithms is a significant task which is required to prove the efficiency and effectiveness of algorithms. But there is less research on segmentation evaluation methods compared to numerous increasing segmentation algorithms. Generally, evaluation methods are classified into two broad category named as, subjective and objective method as in Figure 1.

*Author for correspondence

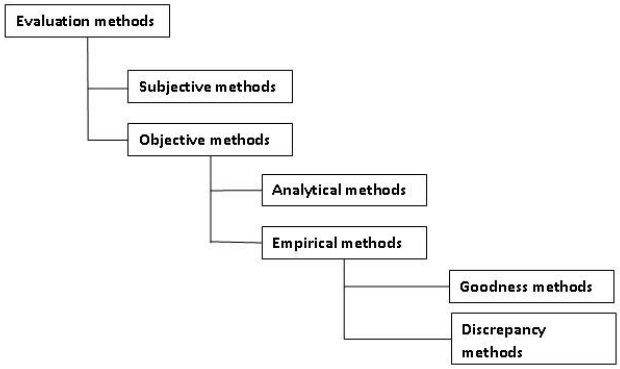

Figure 1. Different Evaluation Techniques

Subjective method is an evaluation based on human visual inspection which is biased, time consuming and expensive. Objective method does not involve human assumptions and assessments. It is further classified into two groups as analytical method and empirical method. Most of the segmentation methods are evaluated based on empirical method which is an indirect method of evaluation[8].Goodness evaluation method and discrepancy evaluation method are the two major classification of empirical method based on the use of reference or ground truth or gold standard image. Discrepancy evaluation method also known as supervised or relative evaluation methods evaluate the performance of segmentation algorithms by analyzing the similarity between segmentation algorithm applied output image and the ground truth image.

Ground truth image is usually generated by the human experts. But it is possible to generate the ground truth image only for synthetic images. So for real time images it is not possible to compare the segmentation output developed for specific application with a ground truth images. In these cases goodness method also known as stand alone or unsupervised evaluation method is used to evaluate the performance accuracy of segmentation algorithms. This paper reviews the different evaluation methods in detail and experiments the applicability of commonly used methodsfor different image segmentation methods.Boundary based methods like sobel, canny and susan, region based methods like region growing, thresholding and hybrid method by combining boundary and region based methods were evaluated in this paper by applying different evaluation techniques in both synthetic and real time images. The results from different evaluation techniques indicated that the hybrid method proposed by Surya Prabha and Satheesh Kumar reported a better performance when compared with other existing methods.

2. Subjective Evaluation Method

Subjective evaluation is a commonly used performance assessment method in literature. Human involvement is the basic requirement in this method as assessment is determined based on human visual inspection. Major challenge in this method is the varying result of human inspectors. Evaluation result obtained from this method is biased, expensive, time consuming and probability for accuracy is low in this method. Another major drawback is the requirement of large number of human inspectors. Parameter selection is also another problem faced in this method as it is biased and is based on favoritism.

3. Objective Evaluation Methods

This method provides a reliable comparison among the segmentation algorithms. It compares the performance of segmentation methods with golden standard based on properties of image like distance and similarity measure. It imitates certain characteristics from subjective method and it makes use of human expertise. Analytical objective and Empirical objective method are the two main classifications in objective evaluation[9].

3.1 Analytical Objective Evaluation

Analytical evaluation methods require prior knowledge to evaluate the segmentation algorithms by considering their nature, needs, and complications characteristics of the algorithms[10]. This method is complex and complicated to analyze and compare the algorithms performance as it is not reliable and consistent. Due to lack of proper theoretical knowledge of segmentation and the inability to extract all features from an image, this method is not preferred for evaluating the performance of segmentation algorithms.

3.2 Empirical Objective Evaluation

Generally empirical evaluation methods are the widely used evaluation technique to measure the performance of segmentation algorithms. In this technique, segmentation algorithm is applied on test images to evaluate the performance. This method is simple, faster, and reliable to produce accurate results. This method is competent to evaluate numerous set of segmented images automatically in a smaller period of time. Empirical method performs evaluation on the images either based on goodness measures or discrepancy measures[11].

3.2.1 Discrepancy Measures

Discrepancy evaluation method also termed as relative or supervised evaluation method, is used to evaluate the performance of segmentation methods based on the concept of using reference image or ground truth image. In order to define a ground truth image, human expertise is needed to have a hand drawn segmented result. This method is suitable in cases of images where images are predetermined and their golden standard images are generated with the knowledge of human expertise[12]. This method measures the relationship between output image of segmentation method and ground truth image. Discrepancy methods are broadly categorized into three groups based on similarity measure, distance measure and standard measure as in Figure 2. This method of evaluation is considered to produce an evaluation result with higher accuracy. One of the shortcomings of this method is generation of ground truth which is time-consuming, biased and tricky. Some of the commonly used discrepancy methods are discussed in detail in this section.

3.2.1.1 Receiver Operating Characteristics (ROC) Curve

ROC curve is a pixel based standard measure used to compare the ground truth image and output image of segmentation method based on the use of confusion matrix[13]. Factors involved in the confusion matrix generation are true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN) as in Figure 3. Sensitivity and 1-specificity are two measures required for plotting the ROC curve. Sensitivity or true positive rate or recall is the percentage of true positive pixels and its formula is,

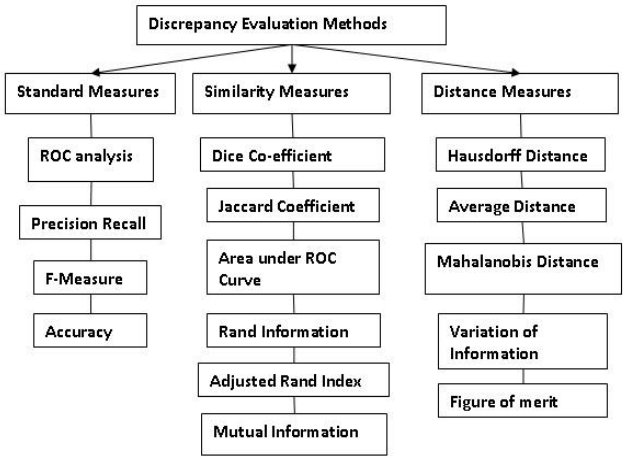

Figure 2. Different Discrepancy Evaluation Techniques

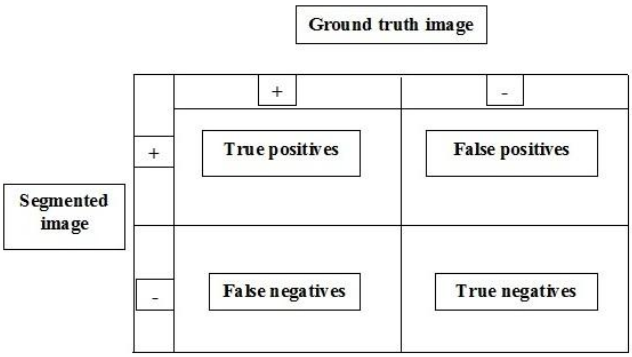

Figure 3. A confusion matrix

$$TPR = (True\ Positive)/(True\ Positive + False\ Negative) \quad (1)$$

1-Specificity or fallouts or False Positive rate or fallout is the percentage of false positive pixels and its formula is,

$$FPR = (False\ Positive)/((False\ Positive + True\ Negative)) \quad (2)$$

Higher percentage of sensitivity and 1-specificity assures that the segmentation method is of a good quality and has higher perfection.

3.2.1.2 Area under ROC curve (AUC)

It is a simple measurement metric used to measure the accuracy by reducing ROC curve result into a scalar value[14]. The value of this method is normalized between the range of 0 and 1. Higher value of AUC indicates a better performance of the segmentation. It is calculated using the formula,

$$AUC = \int_x^y f(a)\,da \quad (3)$$

where 'x' and 'y' are the minimum and maximum axis points in the curve with 'f(a)' a function partly above and below the curve. In simple words, AUC is the difference between the area above ROC curve and area below ROC curve.

3.2.1.3 Precision - Recall (PR) curves

Precision-recall curve is also a pixel based measure that uses confusion matrix to evaluate the algorithm's performance[15]. It is popularly used in information retrieval and pattern recognition. Precision or positive predictive value (PPV)) is a percentage of true positive pixels that are relevant and recall is a percentage of true positive pixels that are retrieved and is calculated based on the formula,

$$P = True\ positive/(True\ Positive + False\ Positive) \quad (4)$$

Precision gives information about the validity of segmentation result and recall gives information about the correctly identified edge pixels in an image. Higher value of precision and recall indicates a good performance by the segmentation method. Under segmentation is resulted in the segmentation method when the value of recall is low and over segmentation is resulted when value of precision is low.

3.2.1.4 F-Measure

F-measure is used to measure efficiency and success of segmentation based on the values of precision and recall. In order to have a single measure with higher effectiveness, a unimodal, F-measure is calculated by combining precision and recall. It is a harmonic mean that gives a precise result and is defined using the formula,

$$FMeasure = 2*(P*TPR)/(P+TPR) \quad (5)$$

3.2.1.5 Figure of Merit

It is based on the mean-square distance between all pixel pair points in segmented output image and ground truth image and assesses the similarity between them. This method is not only useful to assess the quality of edges but is also useful to assess the entire behavior of segmentation method. Its value ranges from 0 to 1 with higher value representing optimal segmentation result. It is calculated and assessed using,

$$FM = \left(\frac{1}{\max\{N_g, N_d\}}\sum_i^N \frac{1}{1+C*D_i^2}\right) \quad (6)$$

where'Ng'is the number of edge pixels in the ground truth image and 'N_d' is the number of edge pixels in the segmented output image. 'D' is the distance between detected edge pixel point and its accurate edge pixel point.

3.2.1.6 Dice Co-efficient

Dice co-efficient is a similarity measure mostly used in the medical image processing to evaluate the performance of segmentation algorithms which has a predefined ground truth information or data set[16]. It is calculated using the formula,

$$DC = \frac{2|M \cap N|}{|M|+|N|} \quad (7)$$

where 'M' is the non zero pixel element in ground truth image and 'N' is non zero pixel element is the segmented image.

3.2.1.7 Jaccard Co-efficient

Jaccard Co-efficient is also similar to that of Dice co-efficient used to calculate the similarity between the two set of images and it also measures the variation or dissimilarity between two images[17]. It is calculated using the formula,

$$JC = \frac{|M \cap N|}{|M \cup N|} \quad (8)$$

and Jaccard distance is calculated using,

$$JD = \frac{|M \cup N| - |M \cup N|}{|M \cup N|} \quad (9)$$

where 'M' is the non zero pixel element in ground truth image and 'N' is the non zero pixel element is segmented image.

3.2.2 Goodness measures

Goodness method also known as unsupervised method or stand alone method of evaluation is used to evaluate the performance of segmentation based on the segmentation image characteristics. This method of evaluation does not require any predefined ground truth image and its prior knowledge for evaluation. This method is of very much useful for situations where it is not possible to collect the ground truth images. Properties of images like shape, region, color, texture, variance, uniformity and entropy are used as key factors to analyze the performance of segmentation methods. Based on these properties goodness measures are classified into different methods as in Figure 4. Some of the frequently used goodness measures are discussed in this section.

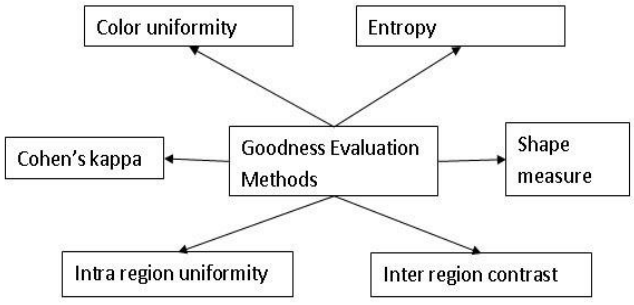

Figure 4. Different Goodness measure methods

3.2.2.1 Cohen's Kappa

It is a statistical method used to evaluate judgment based on the result of different persons by analyzing the level of agreement among those persons. This method is meaningful as it considers not only the observed agreements but also it considers the probability of agreements by chance. In image segmentation it is a pixel – by – pixel comparison. It considers and compares the pixels in segmented region of an image and the probability of pixels that can be found in segmented region of an image. Kappa value is calculated using the formula,

$$Kappa = \frac{O - E}{1 - E} \tag{10}$$

where 'O' is the observed pixels in segmented image and 'E' is the probability of having pixel by chance. The value of kappa is normalized to range from 0 to 1. Higher the value of kappa, better the performance of segmentation method.

3.2.2.2 Entropy

Entropy is another evaluation measure used to compute the randomness or information content in an image. It assists to calculate the uniformity measure in an image. This method of evaluation has derived its idea or concept from information theory and minimum description length principle where the data have discrete random distribution. It was calculated using,

$$E = -\sum_{i=1}^{a} e_i \log e_i \tag{11}$$

where 'e' represents the pixels frequency and 'i' represents the intensity value of pixel. Lower value of entropy assures less randomness in image information and vice versa for higher value of entropy which shows more randomness in image. Therefore for segmentation method with better performance entropy value will be lesser and for poor performance segmentation method entropy value will be higher.

3.2.2.3 Shape measure

It is used to evaluate the performance of segmentation method based on shape features[18]. It uses gradient value and neighborhood pixel values to determine the accuracy. It is calculated using the formula as follows,

$$M = \frac{1}{c}\{\sum_{(a,b)} s[I(a,b) - I_{J(a,b)}]G(a,b)s\left[I(a,b) - x\right]\} \tag{12}$$

where 'c' is a constant scalar value, 's' is an element wise step function, '$I(a, b)$' is the gray scale image, '$I_{J(a, b)}$' is the neighborhoods average value at each pixel locations of (a,b) for the image '$I(a, b)$' and '$G(a, b)$' is the gradient value for the image and 'x' is the threshold value.

3.2.2.4 Intra region uniformity

It is used to analyze the characters of segmented image based on its region uniformity[19, 20]. Inter region uniformity and Intra region uniformity is calculated for the segmented image in both their foreground image and background image. These values are analyzed by selecting an appropriate threshold value which exactly distinguishes the foreground and background in an image. Busyness is another feature used to evaluate the performance of segmentation method as it assumes that both background and foreground objects in an image are solid in shape with robust texture. Performance vector is also used for evaluation which provides all information related to region uniformity, region contrast and texture for evaluation.

4. Experimental Results

Boundary based, region based and hybrid segmentation of boundary and region based methods performance are evaluated using the discrepancy and goodness measures.Sobel, canny and susan are the boundary based methods considered for comparing and for region based method, region growing and thresholding are considered. In hybrid method, a method combining color gradient and region growing algorithm proposed by Surya Prabha and Satheesh Kumar is used for comparison.The output images of these methods are shown in Figures 5 and 6. In measuring the performance of these segmentation methods using discrepancy method 20 benchmark images are taken from Berkeley segmentation dataset[20,21].To measure these segmentation methods performance for real-time images, goodness measures is used, 20 Banana images taken in real time are used for this purpose. Discrepancy methods like ROC curve, Area under ROC curve and goodness measure like entropy are used for experimentation as these are the frequently used methods for evaluation. ROC curve is mostly preferred for discrepancy methods due to its higher capability to produce accurate result. Comparably, Area under ROC curve derived from ROC curve has the ability to produce accurate result.

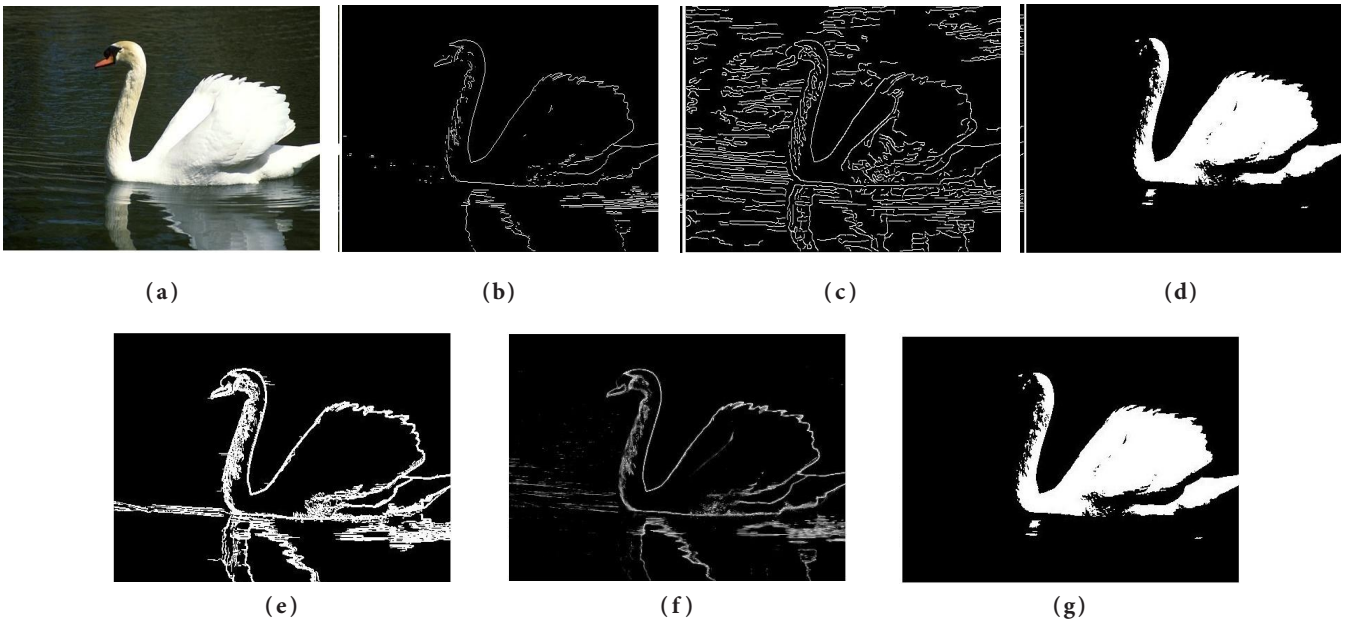

Figure 5. **(a)** Input benchmark image, Output images for **(b)** Sobel, **(c)** Canny, **(d)** Region growing, **(e)** Hybrid method, **(f)** Susan method and **(g)** Thresholding method

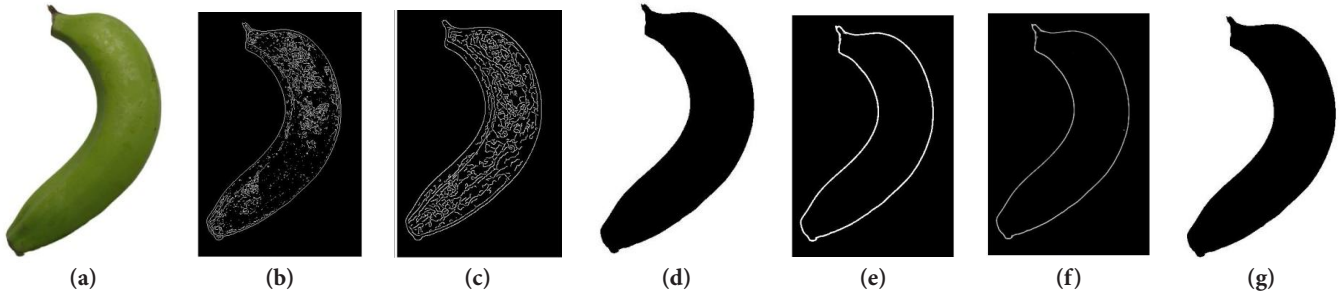

Figure 6. **(a)** Input real-time banana image, Output images for **(b)** Sobel, **(c)** Canny, **(d)** Region growing, **(e)** Hybrid method, **(f)** Susan method and **(g)** Thresholding method.

Comparative analysis output of ROC curve for benchmark images depicts that hybrid method of segmentation performs better than the existing boundary and region based segmentation for benchmark images as in Figure 7. Area under ROC curve also shows that the hybrid method shows a better performance as in Figure 8. Entropy is widely used goodness method to measure the performance of real time images. It exhibits the randomness value in image which helps to assess image information. For real time taken banana images, entropy is calculated and it indicates that the entropy value is less for hybrid method compared to other existing boundary based and region based methods as in Table 1. It depicts that hybrid method performs better than other methods.

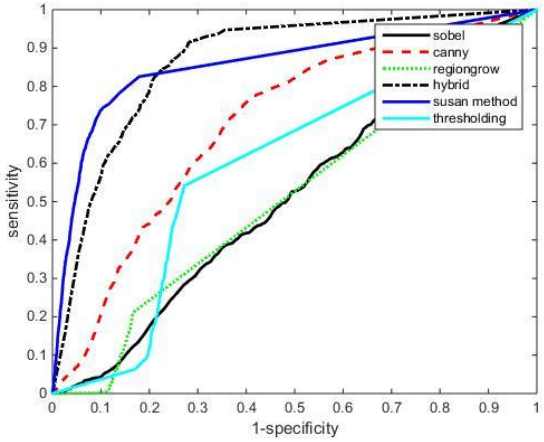

Figure 7. Output of Receiver Operating Characteristics (ROC) curve evaluation method

Figure 8. AUC Curve output of different segmentation methods

Table 1. Entropy value of different segmentation methods

Data Set	Sobel	Canny	Region Growing	Hybrid Method	Susan Method	Thresholding
1	0.1767	0.3251	0.8834	0.1474	0.2417	0.7948
2	0.1753	0.2614	0.9545	0.1715	0.2562	0.9014
3	0.1695	0.2644	0.8179	0.1559	0.2539	0.7711
4	0.1839	0.3004	0.9179	0.1608	0.3329	0.9150
5	0.1685	0.2964	0.8977	0.1540	0.2872	0.8902
6	0.1928	0.2431	0.9889	0.9865	0.3459	0.9865
7	0.1941	0.3058	0.8326	0.1509	0.2560	0.7377
8	0.1717	0.2755	0.8553	0.1297	0.2851	0.8495
9	0.2366	0.2658	0.9029	0.1575	0.5574	0.6555
10	0.1708	0.3062	0.9112	0.1443	0.3179	0.9056

5. Conclusion

Both subjective method and discrepancy based empirical method are evaluated with the support of human experts either directly or indirectly. Due to this factor, these methods are considered as biased and time consuming. Success and efficiency of discrepancy method is based on the accuracy of ground truth images as it compares the performance of segmentation methods with ground truth images. It is also noted that discrepancy methods are useful only for synthetic and pre defined image data set. For real time images, this method is not suitable and its evaluation accuracy is measured using goodness measures. Image segmentation methods like region based method, boundary based method and hybrid methods are applied in this paper to identify a method that performs better in both benchmark and real-time images. The widely used region based and boundary based methods like thresholding, region growing, sobel, canny, susan and hybrid method combining region and boundary methods are evaluated. Most commonly used evaluation methods such as ROC curve, AUC curve of discrepancy method and entropy of goodness method are experimented to analyze the performance of segmentation methods for both real time and benchmark images. The image quality assessment plays a vital role in assessing the performance of segmentation which is a key component in image analysis. The experiment result shows that performance of hybrid method is better than other existing segmentation techniques.

6. References

1. Prabha DS, Kumar JS. Three dimensional object detection and classification methods: A study. International Journal of Engineering Research and Science Technology. 2013; 2(2):33–42.
2. Kheirkhah E, Tabatabaie ZS. A hybrid face detection approach in color images with complex background. Indian Journal of Science and Technology. 2015; 8(1):49–60.
3. Prabha DS, Kumar JS. Assessment of banana fruit maturity by image processing technique. Journal of Food Science and Technology. 2015; 52(3):1316–27.
4. Gonzalez R, Woods R, Eddins S. Digital Image Processing Using MATLAB, Tata McGraw Hill Education Pvt. Ltd: New York, 2010.
5. Canny J. A computational approach to edge detection. IEEE Transactions on Pattern Analysis Machine Intelligence.1986; 8(6):679–98.
6. Marr D, Hildreth E. Theory of edge detection, Proceedings of the Royal Society of London. Series B, Biological Sciences.1980; 207(1167):187–17.
7. Surya Prabha D, Satheesh Kumar J. Hybrid segmentation of peel abnormalities in banana fruit. IJCA Proceedings on International Conference on Research Trends in Computer Technologies. 2013; 3. p. 38–42.
8. Pal NR, Pal SK. A review on image segmentation techniques, Pattern Recognition.1993; 26(9):1277–94.
9. Zhang YJ. A survey on evaluation methods for image segmentation, Pattern Recognition.1996; 29(8):1335–46.
10. Zhang YJ. Evaluation and comparison of different segmentation algorithms. Pattern Recognition Letters. 1997; 18(10):963–74.
11. Ji Q, Haralick RM. Efficient facet edge detection and quantitative performance evaluation. Pattern Recognition. 2002; 35(3):689–700.
12. Zhu SY, Plataniotis KN, Venetsanopoulos A. Comprehensive analysis of edge detection in color image processing. Optical Engineering.1999; 38(4):612–25.
13. Heath M, Sarkar S, Sanocki T, Bowyeryk K. Comparison of edge detectors – A methodology and initial study. Computer Vision and Image Understanding.1998; 69(1):38–54.
14. Yitzhaky Y, Peli E. A method for objective edge detection evaluation and detector parameter selection. IEEE

Transactions on Pattern Analysis and Machine Intelligence. 2003; 25(8):1027–33.
15. Avcbas I, Sankur B, Sayood K. Statistical evaluation of image quality measures. Journal of Electronic Imaging. 2002; 11(2):206–23.
16. Hore A, Ziou D. Image quality metrics: PSNR vs. SSIM. 20th International Conference on Pattern Recognition (ICPR), Istanbul. 2010. p. 2366–69.
17. Yao J, Han W, Summers RM. Computer aided evaluation of pleural effusion using chest CT images. IEEE International Symposium on Biomedical Imaging: From Nanoto Macro, Boston, MA, ISBI'09, Boston, MA. 2009; 241–44.
18. Cardoso JS, Corte-Real L. Toward a generic evaluation of image segmentation. IEEE Transactions on Image Processing. 2005; 14(11):1773–82.
19. Zhang H, Fritts JE, Goldman SA. Image segmentation evaluation: A survey of unsupervised methods. Computer Vision Image Understanding. 2008; 110(2):260–80.
20. Arbelaez P, Maire M, Fowlkes C, Malik J. Contour detection and hierarchical image segmentation. IEEE Transactions on Pattern Analysis Machine Intelligence. 2011; 33(5):898–16.
21. Martin DR, Fowlkes CC, Malik J. Learning to detect natural image boundaries using local brightness, color, and texture cues. IEEE Transactions on Pattern Analysis and Machine Intelligence.1986; 26(5):530–49.