ZAMBIA INFORMATION COMMUNICATION TECHNOLOGY (ICT) JOURNAL

Volume 1 (Issue 1) (2017) Pages 10-14

HOSUR: A Novel Measure for Evaluation of Image Segmentation Quality

Macmillan Simfukwe, Bo Peng and Tianrui Li
School of Information Science and Technology
Southwest Jiaotong University
Chengdu, China
Email: msimfukwe@mu.ac.zm, bpeng@swjtu.edu.cn, trli@swjtu.edu.cn
Douglas Kunda
School of Science, Engineering and Technology
Mulungushi University
Kabwe, Zambia
Email: dkunda@mu.ac.zm

***Abstract*—Image segmentation is one of the vital tasks in image processing. Nevertheless, no universally accepted quality measure for evaluating the performance of various segmentation algorithms or even different parameterizations of the same algorithm exists. In this paper, we propose a new segmentation evaluation measure, based on the fusion of HOG and SURF features. We call it the HOSUR. HOSUR exploits the local shape and corner information to evaluate the similarity between a given segmentation and its respective ground truth. It thus belongs to the category of supervised evaluation measures. Experimental results show accuracy of up to 85%.**

***Keywords*-image segmentation evaluation; HOG features; SURF features; data fusion;**

I. INTRODUCTION

Image segmentation is the partition of an image into homogenous and meaningful constituent parts called segments. It serves as a prerequisite stage for object detection and other subsequent operations in a computer vision system. Consequently, the quality of the image segmentation results has a direct impact on the performance of the entire computer vision system. The ability to assess the quality of segmentation results is essential for developing and improving segmentation algorithms. [1] summarize the significance of application-independent comparison schemes for image segmentations obtained from different methods/parameterizations as follows: (1) autonomous selection from among possible segmentations yielded by the same segmentation algorithm; (2) in order to place a new or existing segmentation algorithm on a solid experimental and scientific ground; and (3) in order to monitor segmentations results on the fly, so that segmentation performance can be guaranteed and consistency maintained. Image segmentation is a relatively ill-posed problem, thus rendering its evaluation difficult. Segmentation evaluation methods are divided into subjective and objective methods. Subjective methods entail that a human judge uses his intuition to assess the segmentation results quality as a measure of the performance of the segmentation algorithm that produced it. Subjective evaluation usually tends to be time consuming and may lead to inconsistent results, due to the variations in the visual capabilities of humans. Objective methods can be divided into analytical, empirical goodness and empirical discrepancy methods [2]. Analytical methods directly assess the actual segmentation algorithms from various perspectives such as the algorithms principle, complexity, efficiency and execution time. Unfortunately, these properties usually have no bearing on the quality of the segmentation results. Empirical goodness and empirical discrepancy methods directly assess quality of the segmentation results. The empirical goodness methods evaluate the segmentation results based on some goodness parameters which are relevant to the visual properties extracted from the original image and the segmented image. They do so without utilizing any prior knowledge and are therefore also referred to as unsupervised evaluation methods. Empirical discrepancy methods use a reference image called the ground truth or gold standard to evaluate the segmentation results, and are thus also referred to as supervised evaluation methods. During supervised evaluation, the segmentation is compared with the respective ground truth to assess the level of discrepancy between the two. In this paper, we propose a novel evaluation measure which we call the HOSUR. It is based on the fusion of Histogram of Oriented Gradients (HOG) and Speeded-Up Robust Features (SURF); i.e. it exploits the local shape and corner information extracted from the segmentation and its respective ground truth. This paper is organized as follows: in section 2 we review some commonly used supervised evaluation measures, in section 3 we present the HOSUR, in section 4 we present experimental results and in section 5 we conclude and state our future work.

II. EVALUATION OF IMAGE SEGMENTATION QUALITY

In the past two decades, a number of objective segmentation evaluation methods have been proposed [2] [3]. In

this section we present some of the most popular supervised segmentation evaluation measures.

The Boundary Displacement Error (BDE) [4]: The BDE is a boundary-based measure which evaluates segmentation quality by calculating the average displacement error of boundary pixels between a given segmentation S and its ground truth G. For a perfect segmentation, the BDE value is equal to zero.

Probability Rand Index (PRI [5]: The PRI takes a statistical perspective to segmentation evaluation. It counts the fraction of pairs of pixel labels that are consistent between the segmentation S and the ground truth G, taking the average across a set of ground truths so as to compensate for the scale variation of human perception. PRI takes values in the range [0, 1], where a score value of 1 indicates that the segmentation and the ground truth are identical.

Variation of Information (VOI) [6]: The VOI is based on information theory. This measure defines the discrepancy between a segmentation S and its ground truth G in terms of the information difference between them. For a perfect segmentation the VOI value is equal to zero.

Global Consistency Error (GCE) [7]: The GCE evaluates the degree of overlap between segments. Segmentations that are related in this fashion are deemed to be consistent because they could represent the image segmented at varying scales. For a perfect segmentation the GCE is equal to zero. The measures presented above that they all have different functional underlying principals and thus make different assumptions about segmentations. We can also notice that most measures perform their processing in the spatial domain. In our work, we seek to perform the processing in the transform domain so as to increase processing speed and also apply data fusion so that we can benefit from two distinct assumptions and approaches.

III. HOSUR

Data fusion is the combination of information from multiple sources to improve application performance. It has been applied in robotics and military fields with remarkable success. Some data fusion-inspired applications in image segmentation evaluation have been reported recently. [3] and [8] reported a co-evaluation framework for improving segmentation evaluation by combining various unsupervised evaluation methods. In [9], both supervised and unsupervised evaluation methods are combined for segmentation evaluation. [10] uses fusion of unsupervised evaluation measures to address the parameter selection problem for the graph cut segmentation algorithm. In this work we combine HOG and SURF features to develop the HOSUR. HOG features encode the local shape information from the regions within an image, while SURF features encode corner information. Details about HOG features can be found in [11], while details about SURF features can be found in [12]. Thus fusion of HOG and SURF features ensures that both local shape and corner information is utilized for segmentation evaluation. We have chosen to use features instead of raw pixel values so as to avoid the impact of illumination distortions that are inherent in original images. It has already been established that object shape is invariant to illumination, thus we can use HOG features to encode shape information. In essence, we would like to compare the shape(s) of ground truth objects with those of the segmentation. Further, we would like to assess whether the corners present in the ground truth occur in the corresponding positions in the respective segmentation. To encode corner information, we opt to use SURF features. Additionally, both HOG and SURF features have high discriminative power and for this reason they have been applied in object detection and classification problems [13] [14].

A. Histogram of Oriented Gradients Features

Each feature is denoted by $f(C, B, k)$, where C is the cell position, B is the parent block and k is the orientation bin number. Given an image $I(x, y)$, the gradients at the point (x, y) can be computed via convolution with a gradient operator as shown in equations 1 and 2.

$$G_x(x, y) = \begin{bmatrix} -1 & 0 & 1 \end{bmatrix} * I(x, y) \tag{1}$$

$$G_y(x, y) = \begin{bmatrix} -1 & 0 & 1 \end{bmatrix}^T * I(x, y) \tag{2}$$

The magnitude of the gradient at the point (x, y) is defined as

$$G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2} \tag{3}$$

The orientation of the edge at the point (x, y) is

$$\theta(x, y) = \arctan[\frac{G_y(x, y)}{G_x(x, y)}] \tag{4}$$

Dividing the orientation range $[-\frac{\pi}{2}, \frac{\pi}{2}]$ into K bins, the value of the k^{th} bin is given as

$$\varphi_k = \begin{cases} G(x, y) & \text{if } \theta(x, y) \in bin_k \\ 0 & \text{Otherwise} \end{cases} \tag{5}$$

The feature value is defined as

$$f(C, B, k) = \frac{\sum_{(x,y)\in C} \varphi_k(x, y) + \varepsilon}{\sum_{(x,y)\in B} G(x, y) + \varepsilon} \tag{6}$$

where ε is a small positive constant so as to avoid division by zero.

B. Speeded Up Robust Features

Suppose we have an image $I(x, y)$ and a given point in it $x = (x_i, y_i)$, the Hessian matrix $H(x, \sigma)$ at scale σ is defined as

$$H(x,\sigma) = \begin{bmatrix} L_{xx}(x,\sigma) & L_{xy}(x,\sigma) \\ L_{xy}(x,\sigma) & L_{yy}(x,\sigma) \end{bmatrix} \quad (7)$$

Where $L_{xx}(x, \sigma)$ is the convolution of image I at point x with the Gaussian second order derivative $\frac{\partial^2}{\partial x^2} g(\sigma)$. In similar fashion, we obtain $L_{xy}(x,)$ and $L_{yy}(x,)$. The second order derivatives can be approximated by 9×9 box filter templates with scale $s = \sigma = 1.2$. The convolutions of I and the box filter templates are denoted as D_{xx}, D_{xy} and D_{yy}, thus leading to

$$\frac{|L_{xy}(1.2)|_F |D_{xx}(1.2)|_F}{|L_{xx}(1.2)|_F |D_{xy}(1.2)|_F} \cong 0.9 \quad (8)$$

Where $|.|_F$ is the Frobenius norm.
The determinant of the Hessian matrix can then be approximated as

$$\det(H) = D_{xx} D_{yy} - (0.9 D_{xy})^2 \quad (9)$$

To compute the descriptor, we select an orientation square region which is centered at the point of interest and then divide the region into 4×4 sub blocks. Then we compute the Haar-wavelet responses in the vertical direction (dy) and horizontal direction (dx), with a filter size equal to $2s$ (s-scale) at sample points.

The SURF feature vector is the sum of the wavelet responses over each sub region and is defined as

$$v = (\sum dx. \sum dy. \sum |dx|. \sum |dy|) \quad (10)$$

The HOSUR employs the ground truth in its evaluation. Therefore, it belongs to the class of supervised segmentation evaluation measures. The functional structure of the HOSUR is presented in figure 1.

The HOSUR takes a segmentation S and its respective ground truth G, as inputs and yields a final score as the quantification of the quality of S. Firstly, S and G are passed on to the feature extraction chamber, where the HOG and SURF features are extracted by the respective sub chambers. The extract HOG features sub chamber produces feature vectors $H = \{h_1, h_2, \cdots, h_m\}$ for S and $\acute{H} = \{\acute{h_1}, \acute{h_2}, \cdots, \acute{h_m}\}$ for G. The extract SURF features sub chamber produces feature vectors $S = \{s_1, s_2, \cdots, s_n\}$ for S and $\acute{S} = \{\acute{s_1}, \acute{s_2}, \cdots, \acute{s_n}\}$ for G. The HOG features from S and G are passed on to the HOG features similarity evaluator to compute the similarity between the feature vector components, as defined by equation 11.

$$Y_1(S,G) = \frac{1}{m}[\sum_{i=1}^{m}(h_i - \acute{h_i})^2] \quad (11)$$

Likewise, the SURF features from S and G are passed on to the SURF features similarity evaluator to compute the similarity between the feature vector components, as defined by equation 12.

$$Y_2(S,G) = \frac{1}{n}[\sum_{i=1}^{n}(s_i - \acute{s_i})^2] \quad (12)$$

The final score is computed as the mean value of Y_1 and Y_2 by the fusion chamber and is defined by equation 13.

$$Y(S,G) = \frac{Y_1 + Y_2}{2} \quad (13)$$

For a perfect segmentation the final score value is equal to zero. The HOSUR is thus bound at zero from below and has no upper bound. The smaller the score value, the better the quality of the segmentation.

IV. EXPERIMENTAL SET UP AND RESULTS

The HOSUR was compared to other supervised segmentation quality evaluation measures. We used images from the database presented in [15] for testing. This database contains 500 images and for each image, there is a pair of machine-generated segmentation results and multiple respective ground truths. Figure 2 shows some samples that we used.. The aim is to let an evaluation measure decided which one of the two, between segmentation 1 and segmentation 2, is better than the other.

However, it should be noted that the score values are dependent on the ground truth used for evaluation. For this reason we used multiple ground truths in our experiments and the final score is the mean value of the score values obtained for each respective ground truth. Given a segmentation S and a set of ground truths $G = \{G_1, G_2, \cdots, G_w\}$, the final score is defined by equation 14.

$$\overleftarrow{Y} = \frac{1}{w}\sum_{i=1}^{w} Y(S, G_i) \quad (14)$$

We conducted a comparison of our H2 with the BDE, VOI, GCE and PRI. For each image, there is a pair of segmentations and a ground truth. For each pair we ask 11 human evaluators to judge which of the two segmentations is better; and the segmentation that receives the majority votes is considered to be the better one. Then we compute the scores for the segmentations in each pair using the H2, BDE, VOI, GCE and PRI, and decide which segmentation in each pair is better, based on the scores. We then calculate the accuracy rates of the measures using equation 15.

$$AccuracyRate = \frac{D}{W} \times 100\% \quad (15)$$

where D is the number of times that the decision made based on a measures scores is the same as that made by the human evaluators for a particular pair of segmentations, and

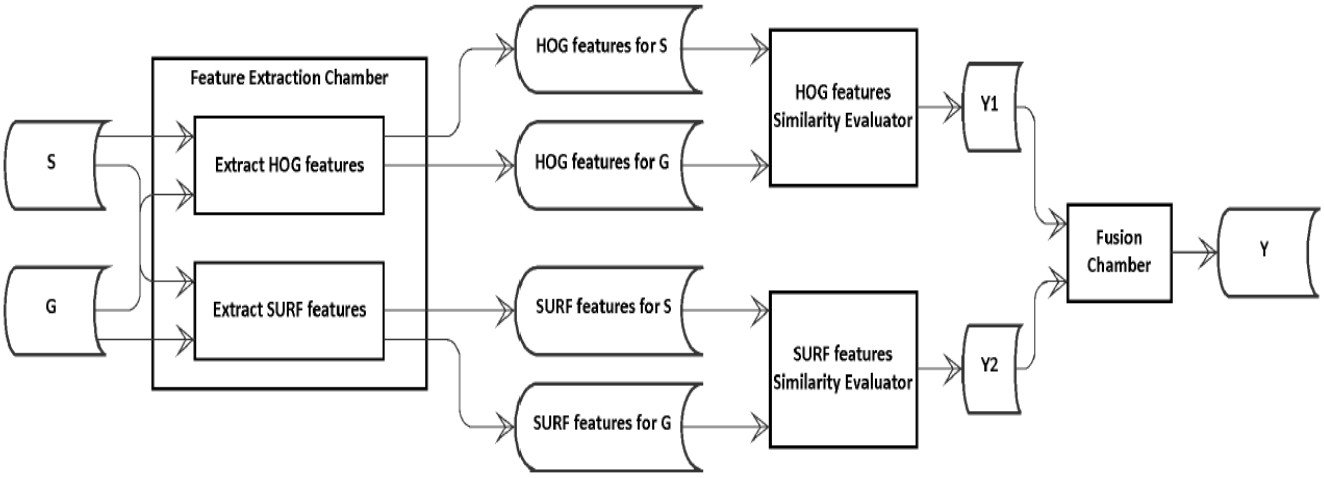

Figure 1. Functional Structure of HOSUR

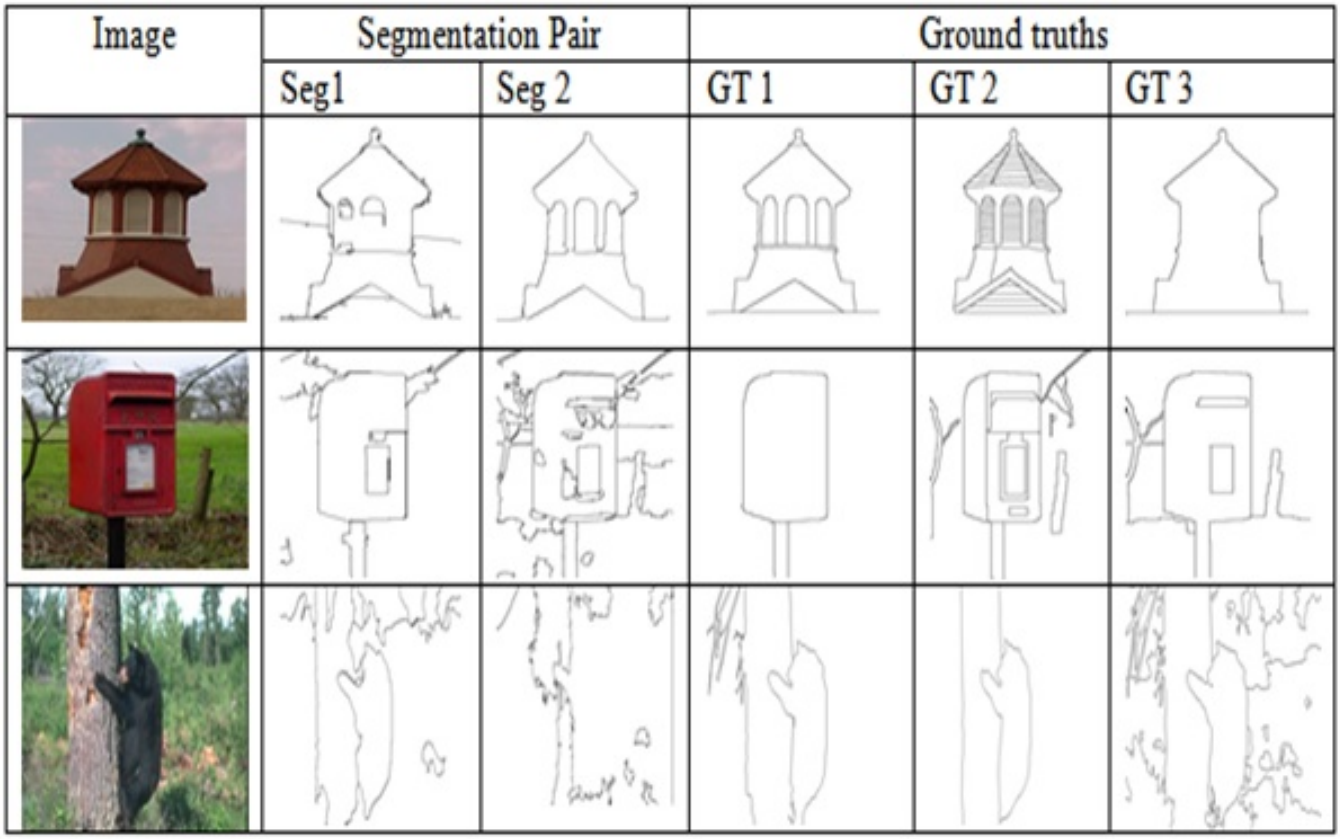

Figure 2. Image Samples [15]

Table I
HOSUR VS OTHER MEASURES

HOSUR	BDE	VOI	GCE	PRI
85%	80%	84%	82%	83%

W is the total number of segmentation pairs (500 in this case).

We used all the available ground truths in our database. Thus, the final score is the average of all the scores obtained across all available ground truths. Table 1 shows the accuracy rate of the HOSUR in comparison with other measures.

The HOSUR performs better than the other measures because: (1) it employs features instead of raw pixel information, thereby utilizing the knowledge about the shapes, edges and corners inherent in the segmentation and its respective ground truth; (2) it applies data fusion, thus allowing two feature extraction and representation schemes (HOG and SURF features) to complement each other, i.e. HOG features encoding shape information and the SURF features encoding corner and edge information.

V. CONCLUSION

In this paper, we have proposed the HOSUR. It is based on the fusion of HOG and SURF features, which allows for the exploitation of local shape and corner information for evaluation of segmentation quality. The performance of the HOSUR is better than that of the state of the art image segmentation evaluation measures on our test set consisting of 500 segmentation pairs and multiple ground truths. In future, we would like to incorporate more features in the HOSUR.

REFERENCES

[1] H. Zhang, J. E. Fritts, and S. Goldman, "Image segmentation evaluation: A survey of unsupervised methods," *Computer Vision and Image Understanding*, vol. 110, pp. 260–280, 2008.

[2] Y. J. Zhang, "A review of recent evaluation methods for image segmentation," in *International Symposium on Signal Processing and its Applications*.

[3] H. Zhang, J. E. Fritts, and S. A. Goldman, "A co-evaluation framework for improving segmentation evaluation," in *Proc. SPIE 5809, Signal Processing, Sensor Fusion and Target Recogntion XIV*.

[4] J. freixenet, X. Munoz, D. Raba, J. Marti, and X. Cufi, "Yet another survey on image segmentation: Region and boundary information integration," in *European Conference on Computer Vision*.

[5] C. Pantofaru and M. Hebert, "A comparison of image segmentation algorithms," Carnegie Mellon University, Tech. Rep.

[6] M. Meilai, "Comparing clusterings - an axiomatic view," in *22nd International Confereing on Machine Learning*.

[7] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its applications to evaluating segmentation algorithms and measuring ecological statistics," in *8th International Conference on Computer Vision*.

[8] H. Zhang, S. Choletti, and S. A. Goldman, "Meta-evaluation of image segmentation using machine learning," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.

[9] J. Lin, B. Peng, T. Li, and Q. Chen, "A learning-based framework for supervised and unsupervised image segmentation evaluation," *Internation Journal of Image and Graphics*.

[10] B. Peng and O. Veksler, "Parameter selection for graph cut image segmentation," in *BMVC*.

[11] C. Tomasi, "Histograms of oriented gradients," Duke University, Tech. Rep.

[12] B. Herbert, A. Ess, T. Tuytelaars, and L. V. Gool, "Speeded up robust features," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[13] N. Dalal and B. Tiggs, "Histograms of oriented gradients for human detection," in *IEEE Conference on Computer Vision and Pattern Recognition*.

[14] J. He, Y. Xie, X. Luan, X. Niu, and X. Zhang, "A tv logo detection and recognition method based on surf features and bag-of-words model," in *2nd IEEE International Conference on Computer and Communications*.

[15] M. Simfukwe, B. Peng, and T. Li, "A data fusion-based framework for image segmentation evaluation," in *International Conference on Intelligent Computing (ICIC2016)*, Lazhou, China, Aug.