Jean-Marie Beaulieu

BeaulieuJM.ca

[Val1988]
Quantitative Evaluation of Image Segmentation Techniques

Authors: Velarde Cesar, Jean-Marie **Beaulieu**

Conference: Canadian Conference on Electrical and Computer Engineering

Montreal, Canads

September 17-20, 1989, pp. 314-317

ISBN: 0-9694170-0-4

Abstract: The objective evaluation of picture segmentation techniques is an import ant and difficult topic. This paper presents an objective evaluation approach, based upon the comparison of the results of a given segmentation technique to either the ground truth, or to the results of another technique. Three performance criteria are defined : 1) the ability to extract the structure of the images, 2) the sensitivity to noise , and 3) the consistency between the results of two techniques. These performance criteria are measured for 4 segmentation techniques, over a set of artificial images. The results show that one of the techniques outperforms the others in retrieving the structure of the images. They also show that two techniques are less sensitive to noise than the others . Finally, we note that the four techniques produced essentially different picture partitions.

Quantitative Evaluation of Image Segmentation Techniques,
Velarde Cesar, Jean-Marie **Beaulieu**,
Canadian Conference on Electrical and Computer Engineering, Montreal, Canads,
September 17-20, 1989, pp. 314-317.
[Bibtex]

DOWNLOAD the author accepted version

QUANTITATIVE EVALUATION OF IMAGE SEGMENTATION TECHNIQUES

Cesar Velarde and Jean-Marie Beaulieu
Département d'Informatique, Université Laval,
Québec, Québec, Canada, G1K 7P4.

Abstract - The objective evaluation of picture segmentation techniques is an important and difficult topic. This paper presents an objective evaluation approach, based upon the comparison of the results of a given segmentation technique to either the ground truth, or to the results of another technique. Three performance criteria are defined: 1) the ability to extract the structure of the images, 2) the sensitivity to noise, and 3) the consistency between the results of two techniques. These performance criteria are measured for 4 segmentation techniques, over a set of artificial images. The results show that one of the techniques outperforms the others in retrieving the structure of the images. They also show that two techniques are less sensitive to noise than the others. Finally, we note that the four techniques produced essentially different picture partitions.

1. INTRODUCTION

A problem that has received much attention in image analysis is that of segmentation, i.e. the recognition of the set of regions or segments that compose an image or a picture. Let a vector (x,y) represent a point (i.e. a pixel) of the picture plane D, $(x,y) \in D$. Let f(x,y) represent the value of the pixel (x,y). The picture plane D can be thought of as being composed of c disjoint regions T_i. The set $T = \{ T_1, ..., T_c\}$ is called the ground truth or ideal segmentation of D. In an attempt to find T, an image segmentation technique decomposes D into n disjoint regions R_i, yielding a picture partition $P = \{ R_1, ..., R_n\}$.

Image segmentation is very important for many image analysis systems. In automatic detection and recognition of military targets, Goehrig and Ledford [3] consider that the maximum number of detected targets is determined at segmentation time. Another important problem is the choice of the best segmentation technique for a given application. In a recent comparative study of several segmentation techniques, Hartley et al [5] found substantial differences in their performances.

Despite the need for better segmentation techniques, and the importance of choosing the best one, only few objective evaluation approaches and studies have been proposed and reported [3], [5]. They essentially measure the similarity between T and P, for a particular set of images. The technique that consistently produces partitions that are very similar to the ground truth, is regarded as having a high value. The similarity between two partitions is measured by calculating the values of a set of previously defined parameters. The problem with these approaches is that the parameters and/or the rules used to calculate them, are usually task specific. They define the quality of a partition in high level terms. Therefore, they are difficult to apply or adapt to other segmentation tasks.

This work is supported by the National Science and Engineering Research Council of Canada

Picture segmentation is, however, a low level process. It employs general purpose image models and grouping criteria that are independent of the scene under analysis. Therefore, the techniques to evaluate them should rely upon the same kind of knowledge. They should be defined in general purpose and low level terms.

This paper presents a low level evaluation approach, inspired from the one proposed by Rand [8] to evaluate clustering methods. This approach, like those discussed previously, assess the value of a given segmentation technique, by comparing its results against the ground truth, or against the results of another technique. However, this approach differs from other ones in that : 1) a task independent, low level measure of similarity between two partitions is used, 2) performance is measured not only as the ability to find T, but also in other aspects, and 3) artificial images are used.

Essential to this evaluation approach is the use of the Rand measure [8], which compares two partitions of a data set. This measure is general and involves only the results of a data partition. The Rand measure, used in cluster analysis, compares two partitions of a data set by examining if the relationship between two data points is the same in both partitions. Given two partitions X and Y of an image of N pixels, the similarity c between X and Y is :

$$c(X,Y) = \sum \delta_{i,j} \Big/ \binom{N}{2}$$

where

$$\delta_{i,j} = \begin{cases} 1 \text{ if pixels i and j belong to the same region in X and Y} \\ 1 \text{ if i and j belong to different regions in both X and Y} \\ 0 \text{ otherwise} \end{cases}$$

and where the summation is over all distinct pairs of pixels (i,j). Note that c(X,Y) varies between 0 and 1. If c(X,Y) = 1 then X = Y. Low values of c indicate a low similarity between X and Y.

The evaluation of picture segmentation techniques is a complex problem that involves many aspects (i.e. execution time, quality of the results, etc). In this paper, the evaluation is based upon three criteria: 1) the ability of the segmentation algorithm to extract the true structure of the image, 2) the sensitivity to noise, and 3) the consistency between the algorithm results.

Another distinctive aspect of the approach is the use of randomly generated images. The structure of these artificial images can be easily controlled and is simpler than the structure of real images. The effect of the segmentation algorithms on the segment shape become, then, more obvious. Moreover, artificial images simplify the experimentation by allowing the summation of results over a large set of test images.

The next section defines the three performance criteria. Section 3 describes the four evaluated segmentation techniques; Section 4 describes the methodology used and shows the results of the evaluations, and Section 5 summarizes and discusses the approach.

2. PERFORMANCE CRITERIA

In order to evaluate a picture segmentation technique, we need a set of test images I, $I = \{ I_i \}$. A partition P_i is obtained by applying a given segmentation technique to each I_i. The following performance criteria are employed.

a) Extraction of image structure.

The ground truth T_i of each image I_i, is the structure to be detected by a technique. The ability of a segmentation technique to extract the structure of the image, is represented by c_a, the mean of the Rand measure calculated between the partitions and the ground truth, $c(T_i,P_i)$.

b) Sensitivity to noise.

A characteristic, highly appreciated in image analysis, is the ability of a technique to cope with noise. Let I'_i be a perturbed image, generated by the addition of a random quantity to every pixel in a reference image I_i. This quantity is independently drawn from a distribution with a zero mean and a small variance. Let P'_i and P_i be two picture partitions, obtained by applying the same segmentation technique to I'_i and I_i, respectively. The sensitivity of the technique to noise is represented by c_b, the mean of the $c(P'_i,P_i)$ values. The ideal technique would produce $c_b = 1$. A low value of c_b shows a high sensitivity to noise.

c) Consistency of the results between techniques.

Picture segmentation techniques can be expensive in terms of computer execution time and storage requirements. An expensive technique could be replaced by another one, if the partitions produced by these two techniques are very similar. Let P'_i be the partition that can be obtained by applying a second segmentation technique to every image I_i. The consistency of the results of these two techniques is represented by c_c, the mean of the $c(P'_i,P_i)$ values.

3. SEGMENTATION TECHNIQUES

The following four segmentation techniques were employed for the experimentations:

HSWO) The Hierarchical Step-wise Optimization Algorithm [1].

Beaulieu and Goldberg [2] segment an image by a sequence of optimization processes. At each iteration a cost measure $C_{i,j}$ is calculated for each pair of adjacent regions (R_i,R_j). Then the pair (R_i,R_j) with the lowest $C_{i,j}$ is merged. This process continues until a predefined number of regions is obtained. A picture approximation model, suggests the utilization of the increase of the overall pixel variance around the region means, produced by the merging of R_i and R_j, as the cost measure $C_{i,j}$.

SM) The Split and Merge Algorithm [6]

Horowitz and Pavlidis present an algorithm based upon a pyramidal data structure. A region is considered as homogeneous if its gray level range is lower or equal to a given threshold. The data structure defines the way in which regions can be merged or split. A pyramid is a hierarchy of square regions where the regions of one level are split into four regular sub-parts to yield the regions of the next lower level. Therefore, the SM algorithm consists of 1) merging the homogeneous regions, if the resulting regions are also homogeneous, or 2) splitting the regions that are not homogeneous into their four sub-parts.

RG) The Region Growing Algorithm [4].

Gupta and Wintz follow a statistical approach for picture segmentation. Their technique attempts to produce homogeneous regions containing samples belonging to a same normal distribution. The picture is first divided into small cells (e.g. 2x2 pixels). The regions grow horizontally and vertically by sequentially examining the cells. A cell is merged with a neighbour region if they are similar. Otherwise, the cell forms the seed point of a new region. This process continues until every cell is examined and assigned to a region.

DPL) The Directed Pixel Linking Algorithm [7].

The algorithm of Narendra and Goldberg operates directly on a gradient picture, where a homogeneous region is a low gradient value area surrounded by a boundary with high gradient value. The algorithm first locates the local minimum points of the gradient picture. They are the root points around which regions will be formed. Each pixel is then linked to its neighbour with the lowest gradient value. The pixels pointing down to a same root point, define a region. A smoothing parameter is employed to prevent the formation of too small regions.

4. METHODOLOGY AND RESULTS

Artificial images are used to evaluate the techniques. Let I be a set of distinct images I_i, using the same picture plane D, which are generated such that: 1) there are N pixels in D, 2) all I_i have the same ground truth T, composed of c disjoint regions T_i, $N \geq c \geq 1$, 3) each T_i is associated to a given distribution whose mean and variance are known, 4) $T_i, T_j \in T$ cannot be associated to the same distribution if they are adjacent, and 5) the values of the pixels belonging to a region T_i, constitute a random sample from its associated distribution.

We now present the used methodology. For each performance criteria to evaluate, a set of 50 artificial images was created. These images are represented by arrays of 16x16 real values (or pixels). The values vary between zero and 64 (inclusively), and each value is obtained by sampling a normal distribution $N(\mu,\sigma)$.

Each segmentation technique was applied to each set of images. To make a fair evaluation, the SM, RG, and DPL techniques (section 3) were applied many times to the same image. At each time, their input parameters were slightly modified. Segmented images with more than 28 regions and less than 2 regions, were eliminated.

we now present the results:

a) Retrieval of structure.

To evaluate their ability to retrieve the structure, the techniques were applied to 50 images composed of four regions. The pixel values were drawn randomly from four normal distributions with different mean values for each region, $N_1(10,4)$, $N_2(45,4)$, $N_3(30,4)$, $N_4(35,4)$. The distribution N_i is associated to the region i. Each partition obtained was compared against the ground truth. The results are given in the graph of Fig. 1. We can observe from Fig. 1 that the evolution of c_a against the number of regions, is essentially the same for all the techniques: c_a increases with the number of segments. However, the HSWO technique clearly performs better than the others for partitions having between 3 and 6 regions. The maximum c_a value observed, $c_a = 0.94$, was obtained by the HSWO technique. This value is greater than the other 3 maximum c_a values of 0.87, 0.82, and 0.84, for the SM, RG, and DPL techniques, respectively. Moreover, this value $c_a = 0.94$ was obtained for partitions of 4 regions, the same number of regions as in the ground truth.

b) Sensitivity to noise.

To measure their sensitivity to noise, two sets of 50 random images were generated. The first set contains the reference images where each pixel value was drawn randomly from N(32,4). The second set, the perturbed images, results from adding a gaussian noise with zero mean and a standard deviation of 0.5 to the reference images. Both sets were segmented, and for each segmentation algorithm, the partitions obtained from the reference images were compared against the partitions of the perturbed images. The c_b values thus obtained are displayed in Fig. 2. The HSWO, RG, and DPL techniques, as seen from Fig. 2, exhibit a similar trend in the evolution of c_b against the number of regions. c_b decreases with the number of regions, until a minimum is reached. Then c_b increases with the number of regions. However, the SM technique does not follow this trend. For partitions having between 3 to 27 regions, c_b remains relatively constant. The high c_b value ($c_b = 0.85$) yielded by the SM technique, for partitions of 28 regions is probably a wildshot, because only 4 comparisons could be made. We conclude that the hierarchical techniques (HSWO and SM) are more sensitive to noise.

c) Consistency between algorithm results.

Finally, we assess how similar are the partitions produced by the tested algorithms. 50 random images, with pixel values drawn from N(32,4), were segmented. The partitions obtained by applying two different algorithms to the same image were compared. The results are shown in Fig. 3. In all cases we have a low Rand measure value, which confirms our intuitive expectation that the algorithms produce essentially different partitions.

5. CONCLUSIONS

The followed approach is appropriate for the objective evaluation of image segmentation techniques. The results of the experiments show that one technique, the HSWO algorithm, outperforms the others in finding the ground truth. However, it also turns out to be the most sensitive to noise. More work is required to correctly interpret these results. A more complete analysis will be given in [9].

In order to make this approach more flexible, the generated images should correspond more closely to real ones. The control of the parameters of the generating process would allow us to quantify, for example, the impact of the sizes or shapes of the regions, on the performances of several segmentation techniques.

REFERENCES

[1] Beaulieu, J. M. "Hierarchical picture segmentation by step-wise optimization," Ph. D. dissertation, Dep. Elec. Eng., Univ. Ottawa, Ottawa, Ont., Canada, 1985.

[2] Beaulieu, J-M., and Goldberg, M., "Hierarchy in Picture Segmentation: A Step-wise Optimization Approach," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 11, NO. 2, pp. 150-163, Feb. 1989.

[3] Goehrig, G. A., and Ledford, L.G., "Analysis of Image Segmentation Approaches with Emphasis on Performance Evaluation Criteria," SPIE, Vol. 252 Smart Sensors II, pp. 124-129, 1980.

[4] Gupta, J.N., and Wintz, P.A., "A Boundary Finding Algorithm and its Applications," IEEE Transactions on Circuits and Systems, pp. 351-362, April 1975.

[5] Hartley, R.L., Kitchen, L.J., Wang, C-Y., and Rosenfeld, A., "Segmentation of FLIR Images: A Comparative Study," IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-12, pp. 553-566, 1982.

[6] Horowitz, S.L., and Pavlidis, T., "Picture Segmentation by a Tree Traversal Algorithm," Journal of the Association for Computing Machinery, Vol. 23, pp. 368-388, April 1976.

[7] Narendra, P.M., and Goldberg, M., "Image Segmentation with Directed Trees," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-2, pp. 185-191, March 1980.

[8] Rand, W.M., "Objective Criteria for the Evaluation of Clustering Methods," Journal of the American Statistical Association, Vol. 66, pp. 846-850, December 1971.

[9] Velarde, C., "Quantitative evaluation of image segmentation techniques," M.Sc. thesis, Département d'Informatique, Univ. Laval, Quebec, QC., Canada, to be published.

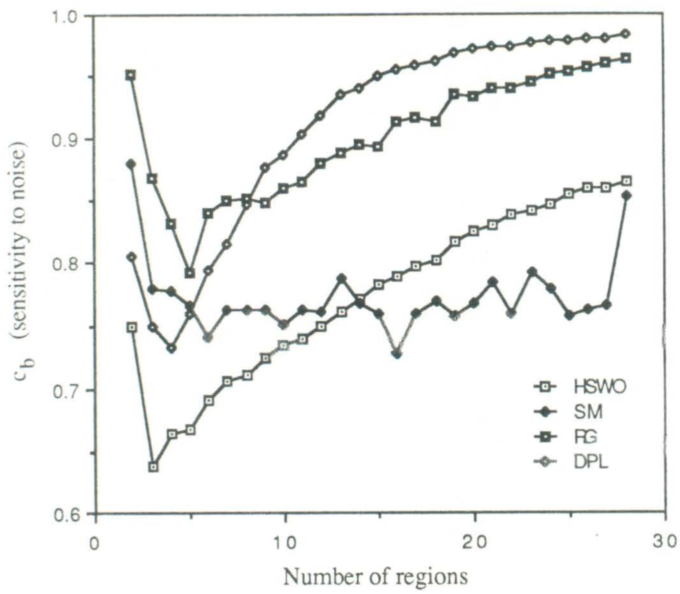

Fig. 2. Comparison of the sensitivity to noise of the 4 image segmentation techniques

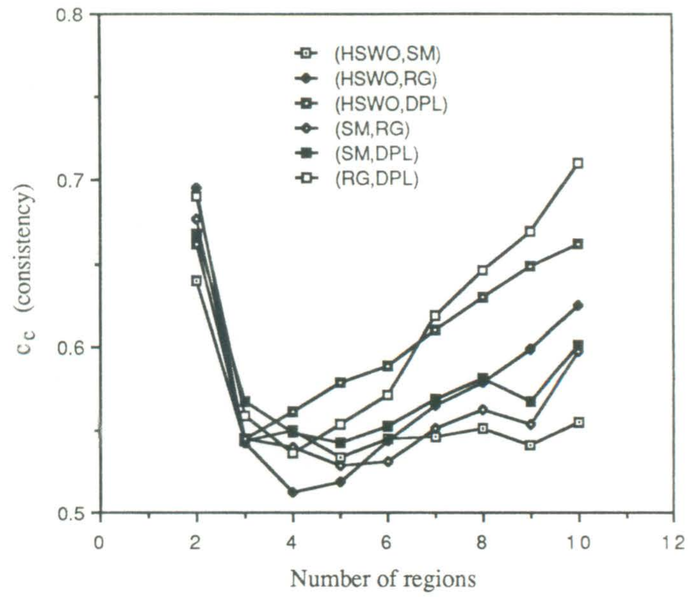

Fig. 3. Comparison of the similarity of the results produced by the 4 image segmentation techniques

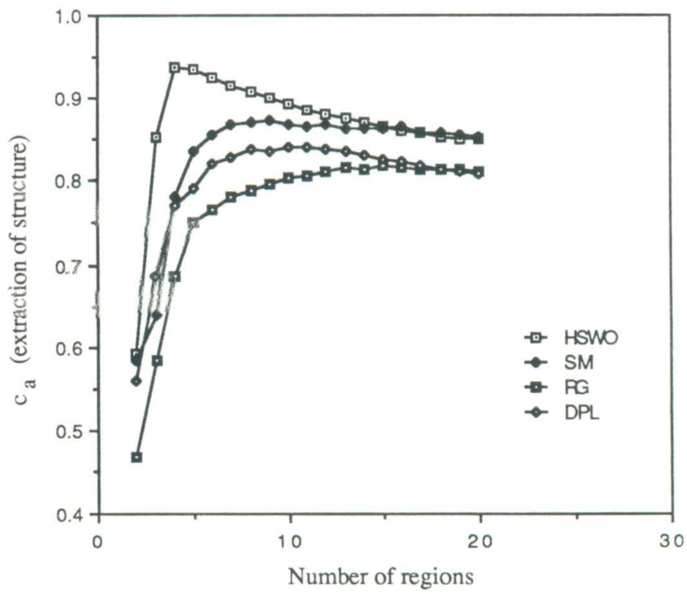

Fig. 1. Comparison of the ability of the 4 image segmentation techniques to extract the structure of the images