NEW DISCREPANCY MEASURES FOR SEGMENTATION EVALUATION*

A. B. Goumeidane[1], M. Khamadja[1], B. Belaroussi[2], H. Benoit-Cattin[2], C. Odet[2]

1- Laboratoire de Traitement du Signal, Dépt d'électronique, Route d'Ain El Bey, 25000 Constantine, Algérie, tél/fax (213) 31 61 42 06, e-mail : m.khamadja@caramail.com

2- CREATIS, Bât. Blaise Pascal INSA, 20 Avenue A. Einstein 69621 Villeurbanne Cedex, France, Tél :(33) 472 43 85 62, Fax : (33) 472 43 85 26

Abstract

In this paper, we propose new evaluation measures for scene segmentation results, which are based on computing the difference between a region extracted from a segmentation map and the corresponding one on an ideal segmentation. The proposed measures take into account separately both under and over detected pixels. It also associates in its computation the compactness of the region under investigation.

1. INTRODUCTION

Image analysis usually refers to processing of images by computer with the goal to extract objects of interest. Image segmentation is one of the preliminary and most critical step in automatic analysis. Due to its importance, much efforts has been devoted to the segmentation process in the last decades, and a large number of segmentation techniques have been proposed in the literature [1][2]. But none of them can be considered as an universal technique or generally applicable one [3]. As a result, it transpires that the performance evaluation and comparison of segmentation results become indispensable. Nevertheless, it appears difficult to establish a measure capable of adequately evaluating segmentation quality, except for very constrained situations [4].

Some works dealing with segmentation assessment have been reported in the literature. In [5], a survey of major researches was presented. The author classified evaluation segmentation techniques into analytic methods, which assess algorithm performance by analysing their principles and properties, and empirical methods which judge the segmentation quality by applying the segmentation algorithm to test images, and by comparing the outcomes with expected segmentations, previously established. According to this survey, the empirical methods can be divided into two categories: Goodness methods and discrepancy methods.

In the first category, some desirable properties of segmented images, often based on human intuition are measured by some "goodness" parameters. The performance of the algorithms under investigation are assessed by the "goodness" measures. In the second category some reference images representing the expected segmentation must be available. The performance evaluation is achieved by comparing the results obtained by a segmentation algorithm with the reference result. It frees the evaluation from subjectivity and makes its results more accurate. Below, some of such methods are briefly reviewed.

In one of the earliest attempts, Yasnoff [6] has proposed a discrepancy measure based on the number of misclassified pixels and their position. Other discrepancy measures were developed such as the Pratt's FOM [7] and the distances proposed in [4][8-11] which measure the difference between two images, and can be used in a framework of segmentation evaluation. In [3], the author proposed a feature-based measure to judge the quality of segmented images. However, most of these works, are devoted to evaluation of techniques which produce boundaries maps [12-15] and there is still a lack of studies which deal with the evaluation of the resulting segmented regions.

In this paper, we propose discrepancy measures we called Distortion Rates that compute the difference between a reference binary shape and this shape obtained by a segmentation algorithm in order to assess the segmentation quality. The next section details the discrepancy measures. Results and interpretation are discussed in the last section.

* This work is supported by CMEP project number 02MDU559, French MAE and Algerian MESRS.

2. DISTORSION RATES

Let A and B be two binary shapes representing the same region placed in a common support, and let A be the region extracted from an ideal segmentation and B the corresponding one in the resulting segmented image. Ideally there would be a perfect matching between these two regions i.e. same shapes and complete superimposition without excess pixels of one region compared to the other one. However, this is a rare occurrence, and the differences between the two shapes must be computed in order to measure the two region's dissimilarity and assess the segmentation result.

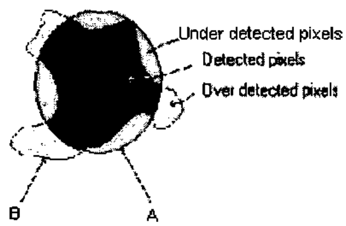

Figure1:The regions A and B superimposed.

We easily observe (Figure 1) that the error pixels namely the under detected and the over detected pixels have introduced distortions in B compared to A. Indeed, the under detected pixels have leaded to an internal distortion, while the over detected ones distorted B outside A. The measures we proposed deal separately with these two kinds of error and try to quantify these distortions relatively to the reference region, also taking into account it's compactness. Like it was proposed in [6], each error pixel is weighted according to its importance. We also use within these measures the importance of the pixels of the reference region according to their position in the region.

2.1 Internal distortion rate

The purpose of the internal distortion rate (IDR) is to bring an idea of the error amount induced by the under detected pixels. The computation of this rate associates both the number of the under detected pixels and their distance to the closest pixel of the reference background (outside of A). On the other hand, it takes into account the pixels of the reference region (A) and their distances to the background. The internal distortion rate is given by:

$$IDR = \frac{\sqrt{\sum_{i=1}^{K1} d_u^2(i)}}{\sqrt{\sum_{j=1}^{N} d_A^2(j)}} \times 100\,\% \qquad (1)$$

In equation (1), $d_u(i)$ is the distance between the under detected pixel i of B and the closest pixel of the reference background (external boundary of A), $d_A(j)$ is the distance between the pixel j of the region A and it's nearest pixel in the reference background, $K1$ is the number of under detected pixels and N is the number of pixels in A. *IDR* estimates the discrepancy between A and B, induced by the under detected pixels. The denominator's term combines both the reference region pixels and their distances to the background, which ensures the introduction of the compactness of A in the calculation of the measure. Intuitively, if A is elongated, it will be more sensitive to under detection than if it is compact (see section 3). Indeed, if we take two regions made up of the same number of pixels, a compact one and an elongated one, and if we remove the same part in the two regions, the distortion of the elongated one will be more perceptible than the one in the compact region. This fact strengthens our choice of including compactness of regions in our measure, which leads to a more accurate evaluation. Note that the IDR will be ranged between 0%, when all pixels are detected, and 100% when no pixel is detected.

2.2 External distortion rate

As mentioned above, the over detected pixels deform B outside A. A way to quantify this distortion is to compute the distances accumulation of these pixels from A. As it was done for the internal distortion rate, we include the compactness of A in this distortion rate.

The external distortion rate (EDR) is defined by:

$$EDR = \frac{\sqrt{\sum_{i=1}^{K2} d_o^2(i)}}{\sqrt{\sum_{j=1}^{N} d_A^2(j)}} \times 100\% \qquad (2)$$

In equation (2), $d_o(i)$ is the distance between the over detected pixel i of B and the closest pixel of the reference A, $d_A(j)$ is the distance between the pixel j of the region A and it's nearest pixel in the reference background, $K2$ is the number of the over detected pixels and N is the number of the pixels in A. *EDR* provides a way to evaluate the effects of the over detected pixels on the region B. Like with the *IDR*, the denominator guarantees that the compactness of A is considered, because each pixel of A is taken into account with its distance from the reference background. Note that *EDR* will always be equal or greater than 0%. 0% is achieved when there is no over detected pixel. However *EDR* can easily exceed 100% under certain conditions i.e. when the accumulation of over detected pixels distances is greater than the accumulation of the distances of the reference pixels.

3. RESULTS AND DISCUSSION

The figures 2, 3 and 4 show synthetic test images (256x256 pixels) that constitute our test image base.

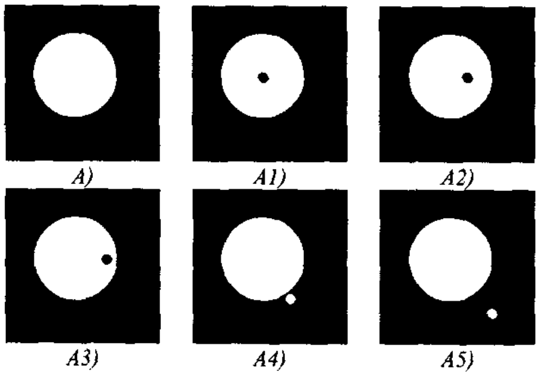

Figure 2: First group of synthetic test images.

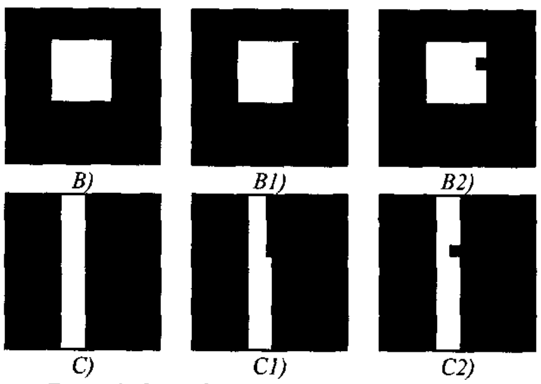

Figure 3: Second group of synthetic test images.

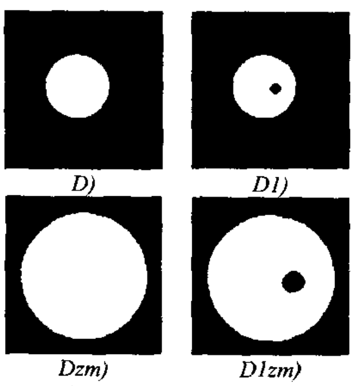

Figure 4: Third group of synthetic test images.

The first group of images (Figure 2) represents a set of compact shapes (circles). The first one (A) corresponds to the reference shape, while the other ones are the same shape with the same removed internal part but located in different parts of the circle (*A1, A2, A3)* and a little over detected part differently set (*A4* and *A5)*. The second group (Figure 3) contains two reference images with the same area (10000 pixels) but with a different compactness, a square (B) and an elongated rectangle (C). Images B1 and C1 represent the reference images with the same number of under detected pixels. A lower number of under detected pixels are presented on images B2 and C2. Finally in the last group (Figure 4), the two test images *D* and *Dzm* represent the same shape at different sizes with the same missing hole (D1 and D1Zm).

IDR and EDR are compared to two classical error measures, the Yasnoff distance and the Baddeley distance. Table 1 gives for the first image group (Figure 2) the distortion rates values (IDR and EDR), the Yasnoff distance (YasD) and the Baddeley distance (BdD). Table 2 and 3 gives the distortion rates values and the Yasnoff distance for the other images (B,C and D).

Image	*IDR (%)*	*EDR (%)*	*YasD (%)*	*BdD*
A1	26.41	0	1.47	0.24
A2	17.17	0	0.95	0.24
A3	7.29	0	0.4	0.24
A4	0	6.81	0.38	8.93
A5	0	25.21	1.4	20.38

Table1: Discrepancy measures computed on test images of Figure 2.

Image	*IDR (%)*	*EDR (%)*	*YasD (%)*
B1	10.29	0	0.33
C1	15.61	0	0.33
B2	10.22	0	0.33
C2	18.93	0	0.33

Table2: Discrepancy measures computed on test images of Figure 3.

Image	*IDR (%)*	*EDR (%)*	*YasD (%)*
D1	24,89	0	0.79
D1zm	24.92	0	3.14

Table3: Discrepancy measures computed on test images of Figure 4.

Table 1 shows that the new proposed distortion rates (IDR and EDR) allow us to know the type of error pixels responsible for the error by setting to 0 one of the rate. Nevertheless, the obtained values *IDR* and *EDR* evolve like the Yasnoff error, because the location of the under detected pixels has a direct importance over the obtained measure due to the modification of the compactness of the resulting shape. In the case of the presence of a hole caused by under detection, the Baddeley distance gives the same values regardless to the location of the hole. This is a major drawback of this distance, and we decide to give up the comparison with it for the rest of the images.

Table 2 shows that our measures are sensitive to the shape compactness, unlike the Yasnoff distance. Indeed, we can see for image *C* that, due to the elongated nature of the rectangle, IDR is more sensitive to error pixels than for the square in image B. Our measures have made a difference between these two cases whereas this has been unnoticed for the Yasnoff distance.

The last group of images (D) was chosen to underline the size influence on discrepancy measures. Indeed, a good evaluation measure should give the same value when the same region is evaluated regardless to the digital scale of the image. The results are presented in table 3. *IDR* gives nearly the same value for *D1* and *D1zm*, but the Yasnoff distance gives very different values for these two images. This shows that the new proposed measures are not sensitive to the sampling rate of the evaluated segmented region.

4. CONCLUSION

In this paper, we propose two distortion rates for evaluation of regions extracted from segmentation maps that treat separately the under detected and the over detected pixels. These discrepancy measures evaluate the difference between a detected shape and a reference one. Unlike other classical distances, they show a good behaviour in the presence of compactness issues and they are not sensitive to the image sampling rate. They can evaluate, with respect to homotetie and compactness the distortions that may come out following a segmentation.

5. REFERENCES

[1] R. M. Haralick and L.G Shapiro, "Image segmentation techniques", *Computer Vision, Graphics and Image Processing*. vol. 29, pp. 100-132, 1985.

[2] N. R. Pal and S. K. Pal, "A Review on image segmentation techniques", *Pattern Recognition*, vol. 26, pp. 1277-1294, 1993.

[3] Y. J. Zhang and J. J. Gerbrands, "Objective and quantitative segmentation evaluation and comparison", *Signal processing*, vol. 39, pp. 43-54, 1994.

[4] M. Beauchemin, K. P. B. Thomson and G. Edwards, "On the Hausdorff distance used for the evaluation of segmentation results", *Canadian Journal Of Remote Sensing*, vol. 24, n° 1, 1998.

[5] Y. J. Zhang, "A survey on evaluation methods for image segmentation", *Pattern Recognition*, vol. 29, n° 8, pp. 1335-1346, 1996.

[6] W. A. Yasnoff, J. K. Mui and J. Bacus, "Error measures for scene segmentation", *Pattern Recognition*, vol. 9, pp. 217-231, 1977.

[7] I. E. Abdou and K. Pratt, "Quantitative design and evaluation of enhancement/tresholding edge detectors", *Proceedings of the IEEE*, vol. 67, n° 5, pp. 753-763, 1979.

[8] A. J. Baddeley, "An error metric for binary images", *Robust Computer Vision*, pp. 59-78, 1992.

[9] D. L. Wilson, A. J. Baddeley and A. Owen, "A new metric for grey-scale image comparison", *International Journal of Computer Vision*, vol. 24, n° 1, pp. 5-18, 1997.

[10] D. Coquin and P.H Bolon, "Application of Baddeley's distance to dissimilarity measurement between grey scale images", *Pattern Recognition letters*, vol. 22, pp.1483-1502, 2002.

[11] J. Paumard, "Robust comparison of binary images", *Pattern Recognition Letters*, vol. 18, pp. 1057-1063, 1997.

[12] M. Heath, S. Sarkar, T. Sanocki and K. Bowyer, "Comparison of edge detectors: A methodology and initial study", *In Proc. of Computer Vision Pattern Recognition*, pp. 143-148, 1996.

[13] R. Roman-Roldan, J. F. Gomez-Lopera, C. Atae-Allah, J. Martinez-Aroza and P. L. Luque-Escamilla, "A Measure of quality for evaluating methods of segmentation and edge detection", *Pattern Recognition*, vol. 34, pp. 969-980, 2001.

[14] M. C. Shin, D. Goldgof and K. W. Bowyer, "Comparison of edge detector performance through use in an object recognition task", *In Proc. of Computer Vision and Pattern Recognition*, vol. 1, pp. 360-365, 1999.

[15] C. Odet, B. Belaroussi and H. Benoit-Cattin, "Scalable discrepancy measures for segmentation evaluation", *In Proc. of IEEE International Conference on Image Processing ICIP-02*, vol. 1, pp. 785-788, 2002.