SCALABLE DISCREPANCY MEASURES FOR SEGMENTATION EVALUATION

C. ODET, B. BELAROUSSI, H. BENOIT-CATTIN

CREATIS, UMR CNRS #5515
INSA, Bât. B. Pascal, 69621 Villeurbanne, FRANCE

Abstract

In this paper, we propose a set of scalable discrepancy measures that may be applied for segmentation evaluation when a reference is known. The proposed measures take into account under and over detected points within an adjustable area. They give the intensity of the discrepancy and its relative position. Furthermore a scale parameter allows to adjust the accuracy of the measures.

1. INTRODUCTION

Segmentation is one of the most important step in low-level image analysis and computer vision. Indeed, the segmentation results affect all the following processing steps, such as feature extraction and quantization. A large number of segmentation techniques have been developed in the last decades and it is still increasing. Nevertheless, none of the existing segmentation algorithms can be generally applied and finding the appropriate one for a particular application is still quite a problem [1]. Furthermore, the segmentation result is impacted by the tuning of the chosen approach and by the image quality as well. In this context, having an appropriating tool to evaluate the segmentation results becomes necessary.

Many works deal with this problem. Zhang [2] classifies the segmentation assessment into the analytical methods and the empirical methods. The analytical methods consider the segmentation algorithms themselves. As an example, Canny [3] has given the main properties of an edge detector (good detection, good localization an unity) and has proposed an edge detector which optimizes a trade-off between good detection and good localization.

The segmentation approaches may optimize different criteria that can not be compared through the same analytical model. To overcome this problem, the empirical strategies deal with the segmentation results. The goodness empirical methods judge the quality of segmented images without any segmentation reference [4]. Three types of measure have been established [2] for region segmentation : goodness based on intra-region uniformity [5, 6], inter-region contrast [7, 8] and region shape [9]. For edge detection, human intuition based measures have been introduced [10] and recently, Meer [4] has proposed an edge detection assessment based on the bootstrap resampling technique.

The discrepancy empirical methods use the disparity between the segmented image and the reference image that can be given by a synthetic image or manually by a human expert. Several measures have been introduced. They are based on the number of miss-segmented pixels [11-13], the position of miss-segmented pixels [12, 14, 15], the number of objects in the image [15-17] and miscellaneous quantities [1, 15]. One of these measures is known as the Pratt Figure of Merit (FOM) [14]. An other discrepancy measure used in remote sensing [18] and pattern recognition [19], is the Hausdorff distance introduced by Huttenlocher [20]. The Hausdorff distance represents a measure of the spatial distance between two sets of points. Its major drawback is its sensitivity to noise. Derivatives versions of the Hausdorff distance have then been introduced like the partial directed Hausdorff distance [18, 20, 21]. Recently, Roman-Raldon [22] has introduced an hybrid measure of empirical goodness and empirical discrepancy. This measure is well adapted to the evaluation of low error results with respect to the theoretical edge.

In this paper, we propose a scalable discrepancy measure that can assess different error levels on binary segmentation results: Low error for edge detection evaluation as well as medium error in the context of image quality impact on the segmentation result.

2. A SCALABLE DISCREPANCY MEASURE

We propose a set of localization measures that can be used on a binary image (edges of a segmented region) under the knowledge of a binary reference image to evaluate the quality of the segmented edges. Such measures are called empirical discrepancy by Zhang [17].

2.1 Localization discrepancy measure

Two initial discrepancy measures are considered. They are defined by equations (1) and (2).

$$ODI = \frac{1}{No}.\sum_{i=1}^{No} d_o(i) \tag{1}$$

$$UDI = \frac{1}{Nu}.\sum_{i=1}^{Nu} d_u(i) \tag{2}$$

In equation (1), $d_o(i)$ is the distance between a point i of the segmented edge and the closest point of the reference one. In equation (2), $d_u(i)$ is the distance between a non detected point i and the closest point of the segmented edge. *No* and *Nu* correspond respectively to the number of over detected and under detected edge points.

ODI estimates globally the discrepancy intensity between the over detected points and the reference edge while *UDI* corresponds to discrepancy intensity between under detected points and the segmented edge. However, one must know if the over (respectively under) detected points tends to be on one particular side of the reference (respectively segmented) edge. To get this information, we introduce the sign of the distances $d_o(i)$ and $d_u(i)$ to define the discrepancy relative position measures for the over detected points (named ODP) and the under detected points (named *UDP*) respectively given by equations (3) and (4). The sign of the distance is arbitrary chosen depending on the side where the over (respectively under) detected point is located, in respect to the reference (respectively segmented) edge.

$$ODP = \frac{1}{No}.\sum_{i=1}^{No} d_o(i) \times sign(d_o(i)) \qquad (3)$$

$$UDP = \frac{1}{Nu}.\sum_{i=1}^{Nu} d_u(i) \times sign(d_u(i)) \qquad (4)$$

Using these measures, one can analyze the differences between a segmented edge and a reference one having information concerning over detection (*ODI* and *ODP*) as well as under detection (*UDI* and *UDP*).

Such measures do not give only intensity discrepancy, like others discrepancy measures (Hausdorff distance [20] and its derivative, Pratt's Figure of Merit [14]...), but also the global relative position of the discrepancy which is an important point for quantitative analysis.

Nevertheless, these measures have been extended to get what we call the scalable discrepancy measures.

2.2 Scalable discrepancy measures

The way a human observer appreciates an edge image depends on the distance between the detected edge and the hypothetical true edge. However, this is true only for short distances. A detected edge pixel closed to the true edge is assumed to be linked to it, while pixels far away from the reference edge are no longer perceived. But since these pixels correspond to bad detected ones, they must be taken into account in an evaluation scheme. To address these remarks we introduce modifications of the previous set of measures to get the scalable discrepancy measures defined by equations (5-8)

$$ODI_n = \frac{1}{No}.\sum_{i=1}^{No}\left(\frac{d_o(i)}{d_{TH}}\right)^n \qquad (5)$$

$$ODP_n = \frac{1}{No}.\sum_{i=1}^{No}\left(\frac{d_o(i)}{d_{TH}}\right)^n * sign(d_o(i)) \qquad (6)$$

$$UDI_n = \frac{1}{Nu}.\sum_{i=1}^{Nu}\left(\frac{d_u(i)}{d_{TH}}\right)^n \qquad (7)$$

$$UDP_n = \frac{1}{Nu}.\sum_{i=1}^{Nu}\left(\frac{d_u(i)}{d_{TH}}\right)^n * sign(d_u(i)) \qquad (8)$$

First, we note that when an edge point is far away from the reference edge, the discrepancy measure should indicate the same value whatever the exact distance is. A saturation distance d_{TH} is then introduced in the measure. The value of this distance depends on the application and will fix what is to be considered far. Using an Euclidean distance, it means that inside a d_{TH} radius circle centered upon a reference edge pixel, the discrepancy distance is calculated, outside it is limited to d_{TH}. In this way, all bad edge points are effectively taken into account. To get normalized distance values (between *0* and *1*), each distance is divided by the threshold distance. So, the different discrepancy measures could be compared together.

Finally, when the distance is lower than d_{TH} (Figure 1), we introduce a scale parameter n that allows to weight differently edge points close to the reference and edge points close to the threshold value. Choosing small values of n ($0<n<1$) give more importance to the edge points very closed to the reference edge. This lead to a high accurate evaluation. At the opposite, large values of n consider points near the reference edge as correct and emphasize only the edge points near the threshold distance.

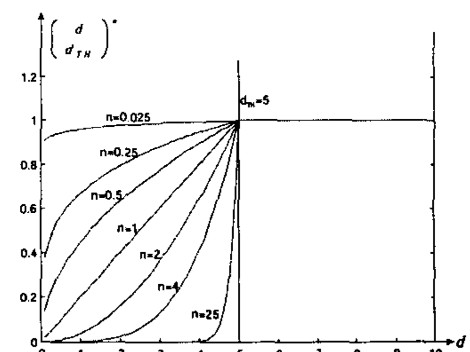

Figure 1: Evolution of a discrepancy measure function of the scale value n with a threshold distance d_{TH} set to 5.

Our scalable discrepancy measures (*Figure 2*) deal with over and under detection. It gives the global intensity of the discrepancy as well as its relative position. Furthermore, for a given application, the tuning of the two parameters d_{TH} and n allows to fix respectively which detected edge point has to be considered far (i.e. bad) and close (i.e. good).

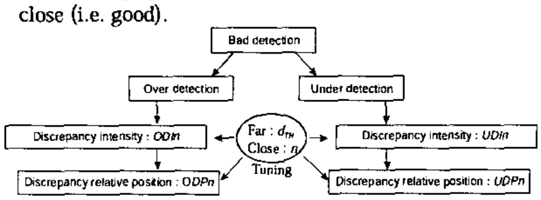

Figure 2: The scalable discrepancy measures tuning.

3. RESULTS

Figure 3 shows synthetic test images (100x256 pixels) characterized by over-detected edge, under-detected edge and both.

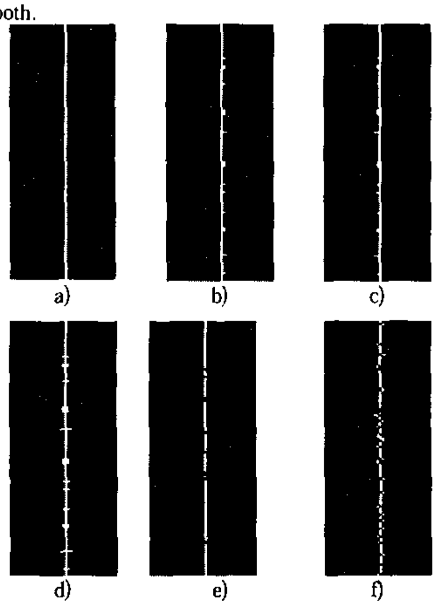

Figure 3: Synthetic test images.
a) Reference edge. b) Over detection on one side. c) Over detection on the other side. d) Over detection on both sides. e) Under detection. f) Under and over detection.

Table 1 gives the value of the scalable discrepancy measures for the test images. First, when the segmented edge is correct, all the measures are 0.

In the over-detection cases (images b-c-d) with no under detection, the under detection measures UDI_n and UDP_n are zero, which means that all the initial edge is found. The degree of over detection is indicated by ODI_n which is different from zero. ODP_n outlines the global local position of the over detection. An over detection on the right (image b) gives a negative value to ODP_n, and a positive one on the left (image c). ODP_n is zero when no side of over detection is advantaged (image d-f). Naturally, ODP_n is also zero for the under detection case (image e). When ODP_n and ODI_n are equal in absolute value, it means that all the over detected points are on the same side of the reference edge (image b-c).

When only under detection is found (image e), ODI_n and ODP_n are zero. UDI_n indicates the degree of under detection. and UDP_n the relative position of the under detected points. When UDP_n and UDI_n have the same absolute value (image e), it means that all the missed points are on the same side of the segmented edge.

Finally, in the case of image f, under and over detected points are underlined by ODI_n and UDI_n. Zero value of ODP_n indicates no specific localization in over detection. UDP_n compared to UDI_n indicates that the missed points are mostly on the left side of the segmented edge.

Image	a)	b)	c)	d)	e)	f)
ODI_n	0	0.052	0.052	0.052	0	0.048
ODP_n	0	-0.052	0.052	0	0	0
UDI_n	0	0	0	0	0.04	0.04
UDP_n	0	0	0	0	-0.04	-0.032

Table 1: Scalable discrepancy measures for n=1 and d_{TH}=5 obtained on the images given Figure 3.

As an example, Figure 4 gives for image f the evolution of the scalable discrepancy measures in function of the scale parameter n. It shows clearly that the accuracy evaluation degree grows when n parameter decreases.

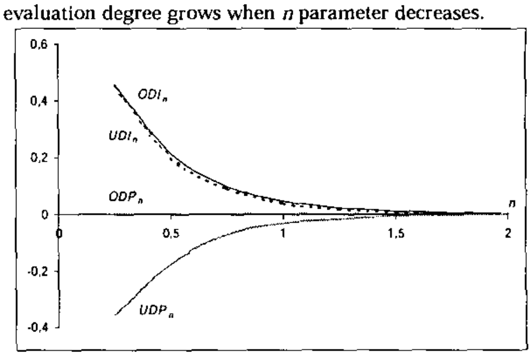

Figure 4: Localization measures evolution according to exponent distance for image of Figure 3-f

Table 2 contains other well known discrepancy measures which have been computed on our test images. The information given by these measures is quite simple but as a consequence relatively poor compared to the one given by the scalable discrepancy measures. The FOM measure does not consider under detection. Indeed image e returns *0* and images d and f return almost the same values. The Hausdorff distance is equivalent on the images b, c, d, f, meaning that for these images the maximum distance between the reference edge and the detected one is 2 pixels.

Discrepancy criterions	a)	b)	c)	d)	e)	f)
Pratt's FOM ($\alpha=1/9$)	0	0.14	0.14	0.24	0	0.23
Hausdorff Distance	0	4	4	4	1	4

Table 2: Known measures evaluated on the test images.

4. CONCLUSION

In this paper, we propose a set of scalable discrepancy measures. They evaluate, with a scalable accuracy, the result of a segmentation compared to a reference. Indeed, they take into account the under detected points, the over detected ones and their relative position. Furthermore, by the tuning of two parameters, they enable the study of different scale errors within an adjustable spatial area around the reference edge.

Such measures should be useful in many segmentation applications like segmentation algorithm adjustment, segmentation algorithm comparison or evaluation of segmentation algorithm to acquisition noise.

5. ACKNOWLEDGEMENT

This work is in the scope of the scientific topics of the PRC-GdR ISIS research group of the French National Center for Scientific Research CNRS.

6. REFERENCES

[1] Pal N.R. and Pal S.K., "A Review on Image Segmentation Techniques", *Pattern Recognition*, vol. 26, pp. 1277-1294, 1993.

[2] Zhang Y.J., "A Survey on Evaluation Methods for Image Segmentation", *Pattern Recognition*, vol. 29, n° 8, pp. 1335-1346, 1996.

[3] Canny J.F., "A Computational Approach to Edge Detection", *IEEE PAMI*, vol. 8, n° 6, pp. 679-698, 1986.

[4] Meer P., Matei B., and Cho K., "Input Guided Performnace Evaluation", *In Proc. Perf. Charac. Computer Vision*, pp. 115-124, 2000.

[5] Nazif A. and Levine M., "Low-level Image Segmentation: an Expert System", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, n° 5, pp. 555-577, 1984.

[6] Weska J.S., "A Survey of Threshold Selection Techniques", *CVGIP*, vol. 7, pp. 259-265, 1978.

[7] Levine M.D. and Nazif A., "Dynamic Measurement of Computer Generated Image Segmentation", *IEEE PAMI.*, vol. 7, n° 2, pp. 155-164, 1985.

[8] Otsu N., "A Threshold selection Method from Gray-level Histogram", *IEEE SMC*, vol. 9, pp. 62-66, 1979.

[9] Sahoo P.K., "A Survey of Thresholding Techniques", *CVGIP*, vol. 41, pp. 233-260, 1988.

[10] Heath M., Sarkar S., Sanocki T., and Bowyer K., "Comparison of Edge Detectors: A Methodology and Initial Study", *In Proc. of Computer Vision Pattern Recognition*, pp. 143-148, 1996.

[11] Shin M.C., Goldgof D., and Bowyer K. W., "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.

[12] Yasnoff W.A., Mui J. K., and Bacus J., "Error Measures For Scene Segmentation", *Pattern Recognition*, vol. 9, pp. 217-231, 1977.

[13] Lee S.U. and Chung S.Y., "A Comparative Performance Study of Several Global Thresholding Techniques for Segmentation", *CVGIP* vol. 52, pp. 171-190, 1990.

[14] Abdou I.E. and Pratt K., "Quantitative Design and Evaluation of Enhancement/Tresholding Edge Detectors", *Proceedings of the IEEE*, vol. 67, n° 5, pp. 753-763, 1979.

[15] Strasters K. and Gerbrands J.J., "Three-dimensional Image Segmentation Using a Split, Merge and Group Approach", *Pattern Recognition Letters*, vol. 12, pp. 307-325, 1991.

[16] Yasnoff W.A. and Bacus J.W., "Scene Segmentation Algorithm Development Using Error Measures", *AOCH*, vol. 6, pp. 45-58, 1984.

[17] Zhang Y.J. and Gerbrands J. J., "Objective and Quantitative Evaluation and Comparison", *Signal Processing*, vol. 39, pp. 43-54, 1994.

[18] Beauchemin M., Thomson K. P. B., and Edwards G., "On The Hausdorff Distance Used For The Evaluation Of Segmentation Results", *Canadian Journal Of Remote Sensing*, vol. 24, n° 1, 1998.

[19] Veltkamp R.C. and Hagedoorn M., "State of the Art in Shape Matching", Utrecht University, the Netherlands UU-CS-1999-27, 1999.

[20] Huttenlocher D.P., Klanderman G.A., and Rucklidge W.J., "Comparing Images Using The Hausdorff Distance", *IEEE PAMI*, vol. 15, pp. 850-863, 1993.

[21] Paumard J., "Robust Comparison of Binary Images", *Pattern Recognition Letters*, vol. 18, pp. 1057-1063, 1997.

[22] Roman-Roldan R., Gomez-Lopera J.F., Atae-Allah C., Martinez-Aroza J., and Luque-Escamilla P.L., "A Measure of Quality for Evaluating Methods of Segmentation and Edge Detection", *Pattern Recognition*, vol. 34, pp. 969-980, 2001.