



HAL
open science

Overview on evaluation criteria in image segmentation procedure based upon watershed transform

Arij Haj Khalifa, Ramzi Mahmoudi, H el ene Laurent, Mohamed Hedi Bedoui

► To cite this version:

Arij Haj Khalifa, Ramzi Mahmoudi, H el ene Laurent, Mohamed Hedi Bedoui. Overview on evaluation criteria in image segmentation procedure based upon watershed transform. 7th Medical Applications of Informatics and New Approach Workshop, Nov 2014, Monastir, Tunisia. hal-03112524

HAL Id: hal-03112524

<https://hal.science/hal-03112524v1>

Submitted on 16 Jan 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destin ee au d ep ot et  a la diffusion de documents scientifiques de niveau recherche, publi es ou non,  emanant des  tablissements d'enseignement et de recherche fran ais ou  trangers, des laboratoires publics ou priv es.

Overview on evaluation criteria in image segmentation procedure based upon watershed transform

A. HADJ KHALIFA¹, R. MAHMOUDI^{1,2}, H. LAURENT³, and M. H. BEDOUI¹

¹Laboratoire Technologie Imagerie Médicale - Faculté de Médecine de Monastir, 5019 Monastir - TUNISIE

²Laboratoire A3SI, Groupe ESIEE - Cité Descartes, BP 99 - 93162 Noisy-le-Grand Cedex- FRANCE

³Laboratoire PRISME, ENSIB - 88 Boulevard Lahitolle 18000 Bourges- FRANCE

Many image segmentation algorithms have been developed. Given the multitude of results after segmentation, it is difficult to find the best result. To solve this problem, evaluation methods have been proposed. In this paper, we present an overview of various evaluation criteria and a comparison between them.

Index Terms - image, segmentation, evaluation.

I. INTRODUCTION

The main objective of an evaluation criterion is a comparison of different results after segmentation to find the best result. Several classifications of evaluation methods have been proposed, but in general they can be classified into two categories: supervised and unsupervised evaluation. The first is based on what is called a "ground truth". It makes a comparison between the results obtained and the proposed reference. The second category is based on the segmented image without a priori information.

Section 2 provides a detailed state of the art of existing evaluation criteria in the literature with their advantages and disadvantages.

Section 3 is a summary of the best criteria for each category based on comparative studies already completed.

II. OVERVIEW OF THE EVALUATION CRITERIA

A. Supervised evaluation

The principle of the supervised criteria is to judge the image quality of the result of an image processing algorithm by exploiting a priori knowledge. This knowledge is generally a reference to an ideal result called ground truth (GT). This reference may be a synthetic image or path made by experts.

1) Evaluation with synthetic ground truth

The synthetic ground truth is a synthetic image created by a program. There are two classes of evaluation as a segmentation contours or regions.

a) Evaluation of region segmentation

Evaluation of segmentation algorithms is based on calculations of distances between bad segmented pixels and

regions to which they belong or on the pairs of regions with a maximum recovery.

Criteria	Publication date
Yasnoff	1977
Vinet	1991
Hamming	1995
Karla-falah	1995
Martin	2001

- Yasnoff [1]

Yasnoff et al. proposed measures of misclassification resulting classes of segmentation and classes in a reference segmentation.

The errors are calculated from a confusion matrix.

$$YAS1(k) = 100 \times \frac{(\sum_{i=1}^n M_{ik}) - M_{kk}}{\sum_{i=1}^n M_{ik}}$$

$$YAS2(k) = 100 \times \frac{(\sum_{i=1}^n M_{ik}) - M_{kk}}{(\sum_{j=1}^n \sum_{i=1}^n M_{ij}) \sum_{i=1}^n M_{ik}}$$

These errors relate to the misclassified pixels and proportional to the distance between a poorly segmented pixel and the region to which it should belong.

$$YAS3(I_R, I_{R_{ref}}) = \frac{100}{\text{card}(I_R)} \times \sqrt{\sum_{a \in I_R, a \notin R_a} \min_{b \in R_a} d(a, b)}$$

- Vinet [2]

Consists of measuring the dissimilarity calculating the maximum overlap between the two regions segmentation results and give the percentage of mismatched pixels between these two segmentations.

$$VIN(I_L, I_{L_{ref}}) = card(I) - \sum_{C'} card(L_i \cap L_j^{ref})$$

- Hamming [3]

Is to match two regions R^1 and R^2 from two segmentation results with a maximum coverage.

$$HAM(R^1, R^2) = 1 - \frac{D_H(R^1, R^2) + D_H(R^2, R^1)}{2 \times Card(X)}$$

D_H is the directional Hamming distance. It is defined as the total area under the intersections between all regions of R^2 and their non-maximal intersected regions from R^1 .

$$D_H(I_{R^1}, I_{R^2}) = \sum_{i=1}^{n_2} \sum_{k=1, k \neq i}^{n_1} card(R_i^2 \cap R_k^1)$$

- Kara-Falah [4]

This is an extension of the measure proposed by Baddeley [7]. This measure calculates the dissimilarity between two segmentations regions I_{R^1} and I_{R^2} . It is based on an average distance between pixels misclassified and the boundaries of the regions closest.

$$KAR(R^1, R^2) = \left[\frac{1}{card(X)} \sum_{(x,y) \in X} |f_{I_{R^1}, I_{R^2}}(x, y)|^p \right]^{\frac{1}{p}}$$

- Martin [5]

Is to measure the error between two segmentations, one is a reference and the other is result in each pixel. It is based on the error $E(R^1, R^2, x)$ which represents the error of the segmentation result compared to the field $E(R^2, R^1, x)$ which represents the truth error of the ground truth from the segmentation result. Dissimilarity can be expressed either by a local consistency error calculation:

$$MAR1(R^1, R^2) = \frac{1}{card(X)} \sum_{x \in X} \min\{E(R^1, R^2, x), E(R^2, R^1, x)\}$$

Either by an error global coherence calculation:

$$MAR2(R^1, R^2) = \frac{1}{card(X)} \min \left\{ \sum_{x \in X} E(R^1, R^2, x), \sum_{x \in X} E(R^2, R^1, x) \right\}$$

b) Evaluation of segmentation contours

Generally, these methods are based on calculation of distance between the contour of the pixels of image segmented and the reference image.

The first steps proposed in this category to measure the quality of detection boundaries are based on the calculation of the

difference between the theoretical boundaries and borders obtained.

The mistake of over-detection:

$$SUR(I_F, I_{ref}) = \frac{card(I_F) - card(I_F \cap I_{ref})}{card(I) - card(I_{ref})}$$

The sub-error detection:

$$SOUS(I_{F1}, I_{F2}) = \frac{card(I_{ref/F}^{cont})}{card(I_{ref}^{cont})}$$

The tracking error:

$$LOC(I_F, I_{ref}) = \frac{card(I_{ref/F} \cup (I_F / ref))}{card(I)}$$

One of the first measures divergence proposed also is the RMS (Root Mean Square) [45].

$$RMS(I_1, I_2) = \left[\frac{1}{card(X)} \sum_{(i,j) \in X} (g_{I_1}(i, j) - g_{I_2}(i, j))^2 \right]^{\frac{1}{2}}$$

This measure can be extended:

$$L_q(I_1, I_2) = \left[\frac{1}{card(X)} \sum_{(i,j) \in X} (g_{I_1}(i, j) - g_{I_2}(i, j))^q \right]^{\frac{1}{q}}$$

These distance measurements may be supplemented by different distances from the probabilistic interpretation of images: distances and Kullback Bhattacharyya and Jensen [46].

$$KUL(I_1, I_2) = \frac{1}{card(X)} \sum_{(i,j) \in X} (g_{I_1}(i, j) - g_{I_2}(i, j)) \times \log \left(\frac{g_{I_1}(i, j)}{g_{I_2}(i, j)} \right)$$

$$BHA(I_1, I_2) = -\log \left(\frac{1}{card(X)} \sum_{(i,j) \in X} \sqrt{g_{I_1}(i, j) \times g_{I_2}(i, j)} \right)$$

$$JEN(I_1, I_2) = J \left(\frac{I_1 + I_2}{2}, I_1 \right)$$

Criteria	Publication date
Fram and Deutsh	1975
Pratt	1978
Peli and Malah	1982
Straster and Gerbrands	1991
Baddeley	1992
Huttenlocher	1993
Hausdorff	1998
Barranco Lopez	1998
Nguyen and Ziou	2000
Correia et Pereira	2000
Roman-Roldan	2001
Odet	2002

- Hausdorff[6]

This distance is a measure of the spatial distance between two sets of pixels.

$$HAU(I_F, I_{ref}) = \max(h(I_F, I_{ref}), h(I_{ref}, I_F))$$

- Baddeley[7]

This is a variation of the distance Hausdorff for binary images. Thereafter, he proposed an extension to the images in grayscale. This criteria has a high sensitivity to noise.

$$BAD1(I_{F1}, I_{F2}) = \max_{(i,j) \in (I_{F1} \cup I_{F2})} (d((i,j), I_{F1}), d((i,j), I_{F2}))$$

To reduce the influence of noise, it replaces the max operator by an average:

$$BAD2(I_{F1}, I_{F2}) = \left[\frac{1}{card(X)} \sum_{(i,j) \in (I_{F1} \cup I_{F2})} |d((i,j), I_{F1}), d((i,j), I_{F2})|^p \right]^{\frac{1}{p}}$$

- Pratt[8]

This is an empirical measure of comparison between a map boundaries I_F to assess and a map borders reference I_{Ref} .

$$PRA(I_F, I_{ref}) = \frac{1}{MP} \sum_{k=1}^{card(I_F^{Cont})} \frac{1}{1 + \alpha d^2(I_F^{Cont}(k), I_{ref}^{Cont})}$$

- Odet[9]

To evaluate a result of binary segmentation and different levels of error, four divergence measures have been proposed by C. Odet:

$$ODI_n = \frac{1}{No} * \sum_{k=1}^{No} \left(\frac{d_o(k)}{d_{TH}} \right)^n$$

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

$$UDI_n = \frac{1}{Nu} * \sum_{k=1}^{No} \left(\frac{d_u(k)}{d_{TH}} \right)^n$$

$$UDP_n = \frac{1}{Nu} * \sum_{k=1}^{No} \left(\frac{d_u(k)}{d_{TH}} \right)^n * sign(d_u(k))$$

- Huttenlocher[10]

This criterion is used to compare binary images. It is a good measure, but it is sensitive to noise and its computing time is important.

$$h_k(I_F, I_{ref}) = K_{a \in I_F^{Cont}}^{th} \min_{b \in I_{ref}^{Cont}} ||a - b||$$

- Peli and Malah[11]

This criterion is based on two static criteria, the average

$$PELI1(I_F, I_{ref}) = \frac{1}{card(I_F)} \sum_{k=1}^{card(I_F)} d(I_F^{Cont}(k), I_{Ref}^{Cont})$$

And the variance of the error detection

$$PELI2(I_F, I_{ref}) = \frac{1}{card(I_F)} \sum_{k=1}^{card(I_F)} d^2(I_F^{Cont}(k), I_{Ref}^{Cont})$$

- Strater and Gerbrands[12]

This is a modification of criteria PRATT defined as follows:

$$STR(I_F, I_{ref}) = \left\{ \begin{array}{ll} \frac{1}{C_{I_F/ref}} \sum_{k=1}^{C_{I_F/ref}} \frac{1}{1 + \alpha d^2(I_F^{Cont}(k), I_{ref}^{Cont})} & si C_{I_F/ref} > 0 \\ 1 & si C_{I_F/ref} = 0 \end{array} \right\}$$

- Correia and Pereira [13]

It is a spatial assessment based on four characteristics which are: loyalty shape, geometric similarity, the similarity of the contents of borders and the statistical similarity data.

$$COR = \alpha * forme + \beta * geom + \gamma frontiere + \delta * statistic$$

- Barranco Lopez [14] [17]

This measure is based on several specific segmentation characteristics: quality measurement, the divergence, the neighborhood, the empirical accuracy, the model low error and errors of interpretation.

$$BAR(I_F, I_{ref}) = K \left[w \sum_{(i,j) \in I_{ref}^{Cont}} \frac{1 + bn_b(i,j)}{1 + pn_e(i,j) + i_{bh} n'_h(i,j)} + \sum_{(i,j) \in I_{ref}^{Cont}} \frac{1 + hn_h(i,j)}{1 + C_{Euler} i_{bh} n'_b(i,j)} \right]$$

- Fram and Deutsch [15]

Fram and Deutsch were the first to propose an evaluation for detectors borders. It is based on two quantitative parameters. The first is a ratio of the pixels detected in the boundary due to the detected signal and the pixels at the boundary due to the signal and noise. The second parameter is a measure of the distribution of the real detected boundary pixels along the border.

- Nguyen and Ziou [16]

The idea is to vary the characteristics of a synthetic image and the parameters of a detector boundary and subsequently compare the results with theoretical borders and to make an assessment.

2) Evaluation with traces made by experts

Criteria	Publication date
Zhang	1994
Yang	1995
Heipke	1997
Shufelt	1999
Bowey	2001
Letournel	2002
Everingham	2002

- Bowyer [18]

According to Bowyer, a ground truth given by experts is based on three types of pixels: the pixels belonging to a border, pixels belonging to no border and the pixels belonging to an unknown area.

If a pixel belongs to a boundary of the ground truth and the border of the result then evaluate this pixel is called "true positive" (VP). If a pixel belongs to a border in earnings estimate but not in the ground truth then it is called "false positive" (FP). If a pixel belongs to a neutral region then it is not VP or FP. This test uses the ROC (Receiver Operating Characteristic) curves while adjusting settings edge detector. The "true positives" and "false positive" report should be maximum. An average ROC curve is obtained for the average rating of edge detector.

- Shufelt [19]

He added for the two concepts mentioned in Bowyer, the "true negative" notions indicating pixels belonging to the bottom of the VT and for the result to evaluate and "false negative" indicating the pixels belonging to the object in the VT and bottom of segmentation to evaluate.

- Letournel [20]

Letournel proposes to take in consideration a number of criteria such as the index of elongation, uniformity, consistency and contrast borders. The evaluation is achieved by combining different measures providing segmentation result close to segmentation results expressed by experts.

- Heipke [21]

This criterion includes the indices Shufelt to quantify the quality of segmentation method roads.

- Everingham [22]

Everingham proposed a method called Pareto front which consists in combining functions accurately and cost functions in a function to optimize overall accuracy.

- Yang [23]

Yang proposes an evaluation method of segmentation of elements in simple shape (approximating a circle or an ellipse) from a reference obtained by an expert path.

- Zhang and Gerbrands [24]

They proposed an objective evaluation of segmentation, quantitative and general to be valid for all segmentation techniques.

3) Psychovisual evaluation

It requires human intervention to judge the result of the segmentation. It is a non-objective solution.

- Mike Heath [25]

Mike Heath proposed a method based solely on human judgment. It is used in the case of treatment of synthetic images.

- Cian W. Shaffrey [26]

It is a method close to the method Mike Heath but takes a segmented image entirely unlike the other method that takes a particular object to be segmented.

- Arbelaez [27]

Arbelaez provides an assessment based on the Precision-Recall method. This assessment is particularly relevant methods of image segmentation based on distance measurements (measurements of energy along the path of the image or on ultrametric).

B. Unsupervised Evaluation

Unlike the supervised evaluation, non supervised criteria used to judge the image quality of the result of an image without any a priori knowledge processing algorithm.

1) Evaluation of segmentation regions

Criteria	Publication date
Weszka and Rosenfeld	1978
Cochran	1985
Levine and Nazif	1985
Sahoo	1988
Zeboudj	1988
Liu and Yang	1994
Borsotti	1998
Rosenberger	1999
Philipp-Foliguet	2005
Chabrier	2006

- Levine and Nazif [28]

Measure based on the inter-region contrast between a given region and its neighboring regions. Two neighboring regions tell if part of their borders is common.

$$LEV1(I_R) = 1 - \sum_{k=1}^{NR} \frac{\sum_{(i,j) \in R_k} [g_I(i,j) - \sum_{(m,n) \in R_k} g_I(m,n)]^2}{card(I) \times \left(\max_{(i,j) \in R_k} (g_I(i,j)) - \min_{(i,j) \in R_k} (g_I(i,j)) \right)^2}$$

- Weszka and Rosenfeld [29]

Measurement based on a quantization noise. It is obtained by calculating a matrix of co-occurrence of gray levels of the image.

- Cochran [30]

Cochran offers a confidence index of homogeneity regions. Calculation of this index is based on one or more attributes of pixels belonging to a region.

$$COC(I_R) = \frac{\max_{k=1} \sigma_I^2(R_k)}{\sum_{k=1}^{NR} \sigma_I^2(R_k)}$$

- Zeboudj [31]

This criterion is based on the inside contrast and outside contrast in order to maximize the contrast between regions and reduce intra-regional contrasts.

$$ZEB(I_R) = \frac{1}{card(I)} \sum_{k=1}^{NR} card(R_k) \times C(R_k)$$

- Borsotti [32]

This criterion has been proposed to improve the standard of Liu and Yang which penalizes segmentations with too many regions.

$$BOR(I_R) = \frac{\sqrt{NR}}{10^4 \times card(I)} \sum_{k=1}^{NR} \left[\frac{e_k^2}{1 + \log(card(R_k))} + \left(\frac{X(card(R_k))}{card(R_k)} \right)^2 \right]$$

- Sahoo [33]

It is a variant of the criterion of Levine and Nazif.

$$SAH(I_R) = 1 - \frac{LEV2(I_R)}{C}$$

where

$$LEV2(I_R) = \sum_{k=1}^{NR} \sum_{(i,j) \in R_k} \left[g_I(i,j) - \frac{1}{card(R_k)} \sum_{(m,n) \in R_k} g_I(m,n) \right]^2$$

- Rosenberger [34]

This criterion evaluates segmentation with L classes of an image I. This criterion is based on the homogeneity maximizations regions (interior contrasts) and the contrast between regions (outside contrast).

$$ROS(I_L) = \frac{C_{Intra}(I_L) + 1 - C_{Inter}(I_L)}{2}$$

- Liu and Yang [35]

They proposed a function that takes into account areas which must be uniform and homogeneous.

It is based on the number, the average intensity and the area of the various regions obtained by segmenting an image in RGB color.

$$LIU(I_R) = \frac{1}{1000 \times (card(I))} \sqrt{NR} \sum_{k=1}^{NR} \frac{e_k^2}{\sqrt{card(R_k)}}$$

- Philipp-foliguet [36]

It is a multi-criteria approach. The principle is to consider the region as a piece of a model. Thus, a measure of total energy is calculated on the whole of the model for assessing the result of the segmentation.

$$PHI_k(I_R) = \sum_{i=1}^{NR} E_D(R_i) + k \times E_c(R_i)$$

- Glory [37]

Glory et al. provide an improved test Borsotti avoiding minimize the number of regions.

Hui Zhang [38] Jason E.Fritts A.Goldman and Sally [39] proposed improved criterion Borsotti adding new entropy. According to them, the intra-region uniformity should be maximum when the inter-regional consistency should be minimized.

- Coquerez and Devars [30]

Coquerez Devars and uniformity is an intra-regions where the quality measurement is based on the variation of the gray level test.

- Chabrier [47]

Two approaches have been proposed to improve the efficiency of the best criteria if unsupervised. The first is to learn to detect the type of an image to fit the selection criteria used to evaluate the results of the image segmentation. The second offers a fusion of several evaluation criteria.

2) Evaluation of segmentation contours

Criteria	Publication date
Kitchen and Rosenfeld	1984
Levine and Nazif	1985
Canny	1986
Tan, Gelfand and Delp	1992
Demigny	1997
Han and Kim	2002

- Kitchen and Rosenfeld [40]

It is a measure based on the consistency of two criteria: the continuity and thickness of the border.

- Tan, Gelfand and Delp [41]

Tan, Gelfand and Delp add a cost function which is the sum of all local costs for each pixel, which takes into account the dissimilarity, the curvature, the number of pixels, the fragmentation and thinness.

- Levine and Nazif [28]

Levine and Nazif provide an evaluation of two types of borders. Computing the intra-contrast regions enables detection of boundaries between two objects when calculating the average gradient along a line image, allows the extraction of the lines passing through the objects of the image.

- Han and Kim [42]

Han and Kim define a distance ambiguity measured using a fuzzy approach. It allows us to quantify the accuracy of the boundary obtained. Measuring overall segmentation ambiguity can be obtained through the calculation of three indices: the ambiguities of existence and location directly derived from a modeling approach to performing fuzzy outline, and ambiguity layout representing a measure connectivity of the various points of the boundaries.

- Demigny [43]

Demigny extended the evaluation criteria proposed by Canny [44] from the continuous case to the discrete case. These criteria were used in the development of shunts optimal filters for edge detection filters such as Canny, Deriche, and Shen-Castan ...

3) Evaluation mixte

Mixed evaluation criteria are criteria not making distinction between the region segmentation or contour segmentation.

C. Generic criteria

These are criteria that evaluate segmentation results regardless of its type, whether in regions, classes, borders or even how unsupervised or supervised. One example is the criterion Gerbrands and Zhang [24].

III. COMPARATIVE STUDY

Several comparative [47] studies have been conducted to find the best criteria for each category. Based on these studies we can conclude that:

- In case supervised by using synthetic images
 - **Segmentation classes:** Vinet and Martin
 - **Segmentation contours:** Pratt
- In case unsupervised
 - **Uniform Image:** Zeboudj
 - **Mixed Image:** Levine and Nazif
 - **Textured Image:** Rosenberger

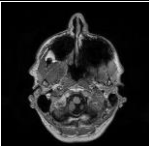
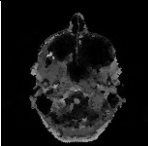
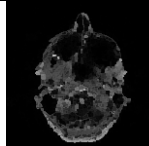
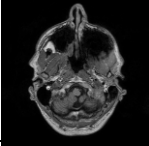
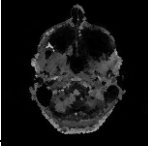
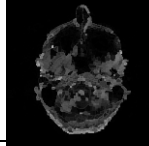
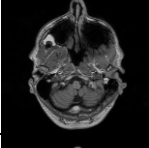
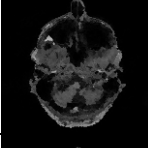
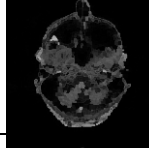
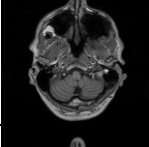
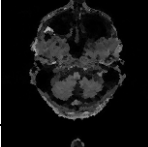
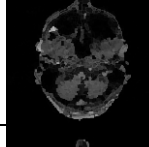
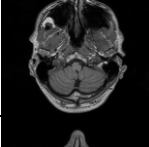
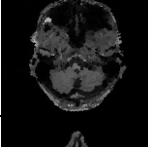
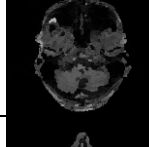
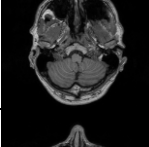
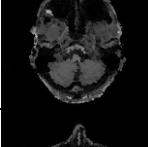
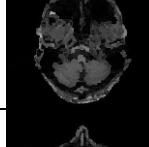
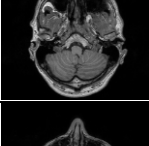
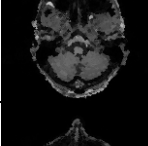
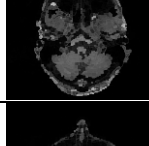
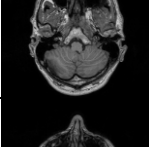
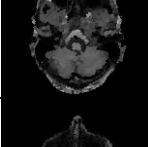
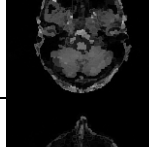
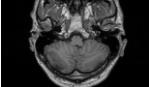
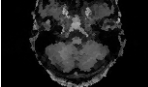
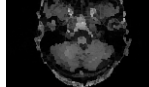
The supervised evaluation gives better results compared to the unsupervised assessment unless it requires a priori knowledge which is not always available.

IV. EXAMPLES OF USE

In this section, we provide examples of using some unsupervised evaluation criteria that will be applied to images results from segmentation by a topological operator "watershed".

When applying the operator segmentation (watershed), a setting of the connection is required. The value of the connection is either 4 or 8. We will apply the evaluation criteria Zeboudj, Rosenberger, Borsotti Nazif and Levine on the segmented image.

After we applied these indicators of 3 images taken randomly varying their sizes (128x128, 256x256, 512x512 and 1024x1024). We note the variance value of the indicator proportional to the size.

NUMBER IMAGE	Original image	Segmented image (connectivity 4)	Segmented image (Connectivity 8)
01			
02			
03			
04			
05			
06			
07			
08			
09			

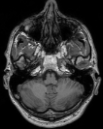
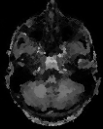
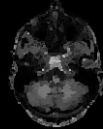
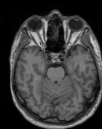
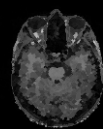
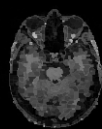
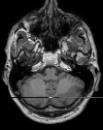
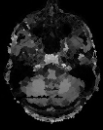
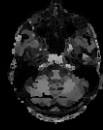
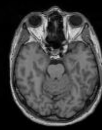
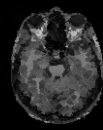
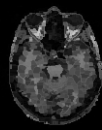
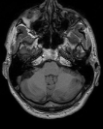
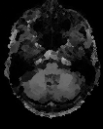
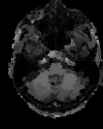
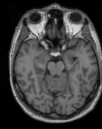
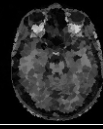
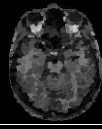
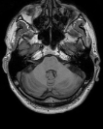
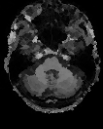
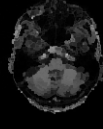
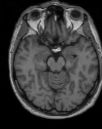
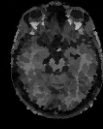
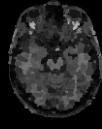
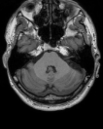
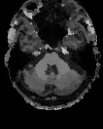
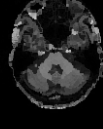
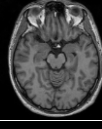
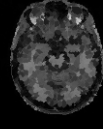

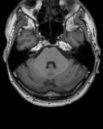
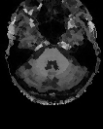
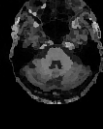
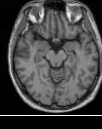
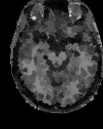
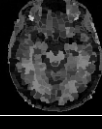
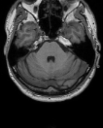
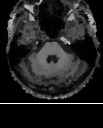
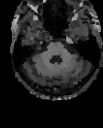
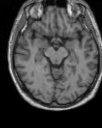
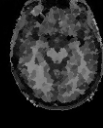
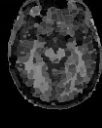
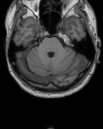
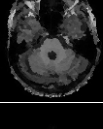
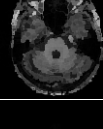
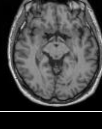

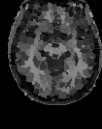
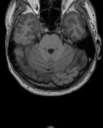
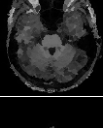
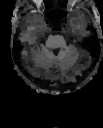
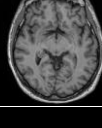
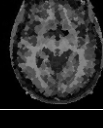
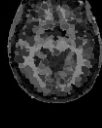
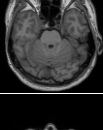
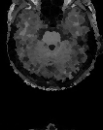
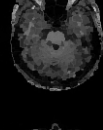
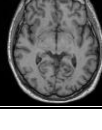
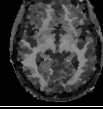
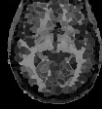
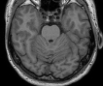
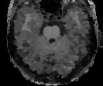
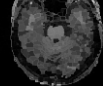
10				21			
11				22			
12				23			
13				24			
14				25			
15				26			
16				27			
17				28			
18				29			
19				30			
20							

Table 1: 30 MRI images of brain and their segmentation by “watershed”

IMAGE	CONNECTIVITY 4				CONNECTIVITY 8			
	ZEBOUDJ 3	LN	ZEBOUDJ 7	ROSENBERGER	ZEBOUDJ 3	LN	ZEBOUDJ 7	ROSENBERGER
01	0.1582	0.5475	0.3750	0.3258	0.2744	0.5448	0.4046	0.3498
02	0.2990	0.5451	0.4918	0.3162	0.3907	0.5424	0.5068	0.3684
03	0.2649	0.5453	0.4765	0.3319	0.3803	0.5461	0.4956	0.3845
04	0.2753	0.5451	0.4806	0.3208	0.3727	0.5457	0.4870	0.3490
05	0.2684	0.5479	0.4649	0.3268	0.3711	0.5467	0.4836	0.3603
06	0.2463	0.5387	0.4530	0.3337	0.3648	0.5446	0.4850	0.3525
07	0.2814	0.5500	0.4856	-	0.3993	0.5446	0.5083	0.3474
08	0.2864	0.5478	0.4777	0.3598	0.3902	0.5470	0.5038	-
09	0.2573	0.5479	0.4574	0.3416	0.3723	0.5513	0.4902	-
10	0.2748	0.5406	0.4640	0.3845	0.3717	0.5412	0.4803	-
11	0.2920	0.5421	0.4714	-	0.4082	0.5474	0.5050	0.3617
12	0.2351	0.5421	0.4382	0.2935	0.3502	0.5441	0.4621	0.3796
13	0.2628	0.5430	0.4528	0.3401	0.3642	0.5423	0.4769	0.3688
14	0.2709	0.5494	0.4622	-	0.3609	0.5486	0.4666	-
15	0.2541	0.5423	0.4531	-	0.3373	0.5493	0.4561	0.3554
16	0.2797	0.5458	0.4641	-	0.3408	0.5478	0.4565	0.3588
17	0.2745	0.5470	0.4659	-	0.3343	0.5460	0.4470	0.3604
18	0.2598	0.5462	0.4450	0.3449	0.3387	0.5519	0.4430	0.3639
19	0.2617	0.5465	0.4565	-	0.3533	0.5495	0.4654	-
20	0.2800	0.5502	0.4650	0.2738	0.3645	0.5514	0.4721	0.3110
21	0.2753	0.5459	0.4587	0.3576	0.3624	0.5505	0.4668	-
22	0.2655	0.5446	0.4553	-	0.3532	0.5474	0.4597	-
23	0.2698	0.5524	0.4525	-	0.3619	0.5573	0.4645	-
24	0.2438	0.5470	0.4375	-	0.3557	0.5475	0.4535	0.4153
25	0.2782	0.5490	0.4522	0.3593	0.3607	0.5504	0.4595	-
26	0.3004	0.5570	0.4749	0.2936	0.3919	0.5497	0.4846	-
27	0.3024	0.5435	0.4728	0.3504	0.4060	0.5288	0.4966	-
28	0.2950	0.5415	0.4784	0.3193	0.4070	0.5287	0.4986	0.3906
29	0.2944	0.5266	0.4754	0.3037	0.3889	0.5244	0.4792	0.3807
30	0.2909	0.5460	0.4717	0.3679	0.3855	0.5220	0.4829	-

Table 2: Results of the evaluation criteria for 30 segmented images

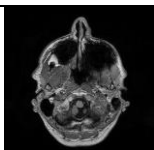
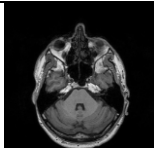
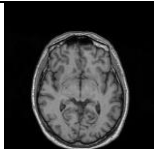
IMAGE	CONNECTIVITY 4				CONNECTIVITY 8			
	128x128	256x256	512x512	1024x1024	128x128	256x256	512x512	1024x1024
	0.5511	0.5475	0.5473	0.5446	0.5521	0.5448	0.5422	0.5396
	0.5506	0.5423	0.5449	0.5433	0.5519	0.5493	0.5467	0.5438
	0.5405	0.5460	0.5453	0.5394	0.5241	0.5220	0.5237	0.5220

Table 3: Results of the evaluation criteria Levine and Nazif for 3 images with different size

V. CONCLUSION

We have presented a state of the art on supervised and unsupervised evaluation criteria.

REFERENCES

- [1] W. A. Yasnoff, J. K. Mui, et J. W. Bacus. Error measures for scene segmentation. *Pattern Recognition*, 9 : 217–231, 1977.
- [2] L. Vinet. Segmentation et mise en correspondance de régions de paires d'images stéréoscopiques. Thèse de Doctorat, Université de Paris IX Dauphine, Juillet 1991.
- [3] Q. Huang et B. Dom. Quantitative Methods of Evaluating Image Segmentation. *International Conference on Image Processing (ICIP'95) Washington DC, USA*, 3 : 53–56, 1995.
- [4] R. Kara-Falah, P. Bolon . Mesure de dissimilarité entre deux segmentations. In *Quatorzième colloque GRETSI*, p. 763-766, Juan les Pins, septembre 1993 .
- [5] D. Martin, C. Fowlkes, D. Tal, et J. Malik. A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics. 8th Int'l Conf. *Computer Vision*, pages 416–423, July 2001.
- [6] M. Beauchemin, KP. B. Thomson, et G. Edwards. On The Hausdorff Distance Used For The Evaluation Of Segmentation Results. *Canadian Journal of Remote Sensing*, 24(1) : 3–8, 1998.
- [7] A. J. Baddeley. An Error metric for binary images. *Robust Computer Vision*, pages 59–78, 1992.
- [8] W. Pratt, O. D. Faugeras, et A. Galgalowicz. Visual discrimination of stochastic texture fields. *IEEE Transactions on Systems, Man, and Cybernetics*, 8(11) : 796–804, 1978.
- [9] C. Odet, B. Belaroussi, et H. Benoit-Cattin. Scalable Discrepancy Measures for Segmentation Evaluation. *ICIP*, 1 : 785–788, septembre 2002.
- [10] D. Huttenlocher et W. Rucklidge. A multi-resolution technique for comparing images using the Hausdorff distance. *IEEE Conference on CVPR*, pages 705–706, 1993.
- [11] T. Peli et D. Malah. A study of edge detection algorithms. *Computer Graphics and Image Processing*, 20 : 1–21, 1982.
- [12] K. Strasters et J. J. Gerbrands. Three-Dimensional Image Segmentation Using a Split, Merge and Group Approach. *Pattern Recognition Letters*, 12 : 307–325, 1991.
- [13] P. Correia et F. Pereira. Objective Evaluation of Relative Segmentation Quality. *International Conference on Image Processing*, pages 308–311, September 2000.
- [14] V. Barranco López, R. Román Roldán, J.F. Gómez Lopera, et J. Martínez Aroza. A measure of quality for evaluating methods of segmentation and edge detection. *International Computer Symposium, Workshop on Image Processing and Character Recognition ICS'98*, pages 221–226, 1998.
- [15] J. R. Fram et E. S. Deutsch. On the quantitative evaluation of edge detection schemes and their comparison with human performance. *IEEE Trans. on Computers*, 24 : 616–628, 1975.
- [16] T.B. Nguyen et D. Ziou. Contextual and non-contextual performance evaluation of edge detectors. *Pattern Recognition Letters*, 21 : 805–816, 2000.
- [17] R. Roman-Roldan, J. F. Gomez-Lopera, C. Atae-allah, J. Martinez-Aroza, et P. L. Luque-Escamilla. A measure of quality for evaluating methods of segmentation and edge detection. *Pattern Recognition*, 34 : 969–980, 2001.
- [18] K. Bowyer. Edge Detector Evaluation Using Empirical ROC Curves. *Computer Vision and Understanding*, 84 : 77–103, 2001.
- [19] J. A. Shufelt. Performance Evaluation and Analysis of monocular building extraction from aerial imagery. *Transaction on Pattern Analysis and Machine Intelligence*, 21(4) : 311–326, 1999.
- [20] V. Letournel. Contribution à l'évaluation d'algorithmes de traitement d'images. thesis, ENST Paris, décembre 2002.
- [21] C. Heipke, H. Mayer, C. Wiedemann, et O. Jamet. Evaluation of automatic road extraction. *International Archives of Photogrammetry and Remote Sensing*, Vol. 32 Part 3-2W3 : 47–56, 1997.
- [22] M. R. Everingham, H. Muller, and B. T. Thomas, editors. *Evaluating Image Segmentation Algorithms Using the Pareto Front*, 2002.
- [23] L. Yang, F. Albrechtsen, T. Lonnestad, and P. Grottum. A supervised approach to the evaluation of image segmentation methods. *Computer Analysis of Image and Pattern*, 970:759–765, 1995.
- [24] Y.J. Zhang and J.J. Gerbrands. Objective and quantitative segmentation evaluation and comparison. *Signal Processing*, 39:43–54, 1994.
- [25] Mike Heath, Sudeep Sarkar, Thomas Sanocki, et Kevin Bowyer. Comparison of Edge Detectors : A Methodology and Initial Study. *Computer vision and image understanding*, 69(1) : 38–54, january 1996.
- [26] C. W. Shaffrey, Ian H. Jermyn, et N. G. Kingsbury, editors. *Psychovisual Evaluation of Image Segmentation Algorithms*. ACIVS (Advanced Concepts for Intelligent Vision Systems), september 2002.
- [27] P. A. Arbeláez Escalente. Une approche métrique pour la segmentation d'images, Thèse de Doctorat, Université Paris Dauphine, novembre 2005.
- [28] M. D. Levine et Ahmed M. Nazif. Dynamic Measurement of Computer Generated Image Segmentations. *IEEE Transaction on Pattern Analysis and Machine intelligence*, PAMI-7(2) : 155–164, March 85.
- [29] J. S. Weszka et A. Rosenfeld. Threshold evaluation techniques. *IEEE Trans., SMC-8* : 622–629, 1978.
- [30] J-P. Cocquerez et J. Devars. Détection de Contours dans les Images Aériennes : Nouveaux Opérateurs. *Traitement du signal*, 2(1) : 45–65, 1985.
- [31] R. Zeboudj. Filtrage, Seuillage Automatique, Contraste et Contours : du Pré-Traitement à l'Analyse d'image. Thèse de Doctorat, Université de Saint Etienne, 1988.
- [32] M. Borsotti, P. Campadelli, et R. Schettini. Quantitative evaluation of color image segmentation results. *Pattern Recognition Letters*, 19 : 741–747, 1998.
- [33] P. K. Sahoo, S. Soltani, A. K. C. Wong, et Y. C. Chen. A survey of thresholding techniques. *CVGIP*, 41 : 233–260, 1988.

- [34] C. Rosenberger. Mise en Oeuvre d'un Système Adaptatif de Segmentation d'Images. Thèse de Doctorat, Université de Rennes 1, décembre 1999.
- [35] J. Liu et Y.-H. Yang. Multiresolution Color Image Segmentation. *IEEE Trans on PAMI*, 16(7) : 689–700, 1994.
- [36] S. Philipp-Foliguet et L. Guigues. Critères multi-échelles d'évaluation de la segmentation. GRETSI, 2005.
- [37] Glory, E., Meas-Yedid, V., Pinset, C., Olivo-Marin, J.-C. et Stamon, G. Évaluation automatique de la qualité de segmentations couleur en imagerie cytologique. Conférence de Reconnaissance des Formes et Intelligence Artificielle, 2006.
- [38] Y.J. Zhang. Evaluation and comparison of different segmentation algorithms. *Pattern Recognition Letters*, 18: 963–974, 1997.
- [39] Hui Zhang, Sharath Cholleli, Sally Goldman, and Jason Fritts, Meta-Evaluation of Image Segmentation Using Machine Learning, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1138-1145, June 2006.
- [40] L. Kitchen et A. Rosenfeld. Scene analysis using region-based constraint filtering. *Pattern Recognition*, 17(2): 189–203, 1984.
- [41] H. L. Tan, S. B. Gelfand, et Delp. A cost minimization approach to edge detection using simulated annealing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14 : 3–18, 1992.
- [42] J. H. Han et T. Y. Kim. Ambiguity distance: an edge evaluation measure using fuzziness of edges. *Fuzzy Sets and Systems*, 126 : 311–324, 2002.
- [43] D. Demigny and T. Kamlé. A discrete expression of canny's criteria for step edge detector performances evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(11):1199–1211, november 1997.
- [44] J.Canny, A computational approach to edge detection, *IEEE Pami*, 8(6):679-698, novembre 1986.
- [45] D. Coquin, P. Bolon, et Y. Chehadeh. Evaluation quantitative d'images filtrées. GRETSI97, 2 : 1351–1354, 1997.
- [46] M. Basseville. Distance measures for signal processing and pattern recognition. *Signal processing*, 18 (4) : 349–369, 1989.
- [47] S. Chabrier. Contribution à l'évaluation de performances en segmentation d'images. Supported thesis 14 décembre 2005 Bourges.