SEGMENTATION EVALUATION BY FUSION WITH A GENETIC ALGORITHM

S. Chabrier, C. Rosenberger and B. Emile

Laboratoire Vision et Robotique - UPRES EA 2078
ENSI de Bourges - Université d'Orléans
10 boulevard Lahitolle, 18020 Bourges Cedex, France
phone: +33 2248484000, fax: +33 2248484040
email: sebastien.chabrier@ensi-bourges.fr

ABSTRACT

The goal of this work is to be able to quantify the quality of a segmentation result without any *a priori* knowledge. We propose in this article to fusion different unsupervised evaluation criteria. In order to identify the best ones to fusion, we compared six unsupervised evaluation criteria on a database composed of synthetic gray-level images. Vinet's measure is used as an objective function to compare the behavior of the different criteria. A new criterion is derived by linearly combining the best ones. The linear coefficients are determined by maximizing the correlation factor with the Vinet's measure by a genetic algorithm. We present in this article some experimental results of evaluation of natural gray-level images.

1. INTRODUCTION

Segmentation is a fundamental stage in image processing since it conditions the quality of interpretation. Many segmentation methods have been proposed in the literature [6], [3] but it still remains difficult to evaluate their efficiency. In order to make an objective comparison of different segmentation results, some evaluation criteria have already been defined and some literature is available. Briefly stated, there are two main approaches.

On the one hand are evaluation methods based upon the computation of a dissimilarity measure between a segmentation result and a ground truth (due to the use of synthetic images or derived by an expert). These methods are of widely use for example in medical applications.

On the other hand are unsupervised evaluation criteria providing to quantify the quality of a segmentation result by computing different statistics without any *a priori* knowledge. In [9], a comparative study of evaluation methods of segmentation results of gray-level images is developed. This article focuses on this kind of approach. Most of tested criteria are not adapted for textured images. The problem is that most of real images are composed of textured regions. In order to solve this problem, Mac Cane [4] showed that it is necessary to use the maximum of criteria and to combine different of them.

An evaluation criterion can be used for different applications. One application is the comparison of different segmentation results of a single image. This enables us to compare the behavior of different segmentation methods in order to choose the most appropriate one for a given application. Another application is to facilitate the choice of the parameters of a segmentation method. Image segmentation needs generally the definition of some input parameters, which are usually defined by the user. This task, that is sometimes arbitrary, can be automated by determining the best parameters with the evaluation criterion.

In the first part of this article, we realize a comparative study of six unsupervised evaluation criteria. We use a database of synthetic images and the Vinet's measure as an objective function. In the second part, we define a new criterion by combining the best criteria in order to improve the quality of evaluation. Finally, we present some experimental results of evaluation on two natural images and we show the efficiency of the proposed method.

2. DEVELOPED METHOD

The idea of this work is to improve the quality of the evaluation of segmentation result by combining different criteria. First, we realize a comparative study of evaluation criteria from the literature. We use a genetic algorithm to find out the optimal linear combination of the best criteria.

2.1 Unsupervised Evaluation criteria

We selected from the state of art [9] six unsupervised evaluation criteria of a gray-level image segmentation result and one supervised criterion (used as an objective function) :

- Zeboudj's contrast (Zeboudj) : This contrast takes into account the internal and external contrast of the regions measured in the neighborhood of each pixel. If we note $W(s)$ the neighborhood of the pixel s, $f(s)$ the pixel intensity and L the maximum intensity, the contrast inside (I_i) and with outside (E_i) the regions R_i are respectively:

$$I_i = \frac{1}{A_i} \sum_{s \in R_i} \max\left\{c(s,t), t \in W(s) \cap R_i\right\} \tag{1}$$

$$E_i = \frac{1}{l_i} \sum_{s \in F_i} \max \left\{ c(s,t), t \in W(s), t \notin R_i \right\} \qquad (2)$$

where A_i is the surface and F_i is the border (of length l_i) of the region R_i. The contrast of R_i is :

$$C(R_i) = \begin{cases} 1 - \frac{I_i}{E_i} & \text{if } 0 < I_i < E_i \\ E_i & \text{if } I_i = 0 \\ 0 & \text{otherwise} \end{cases} \qquad (3)$$

The global contrast is :

$$C_z = \frac{1}{A} \sum_i A_i C(R_i) \qquad (4)$$

- Levine and Nazif's interclass contrast (Inter) [5] : This criterion (C_{Inter}) computes the sum of contrasts of the regions (R_i) weighted by their surfaces (A_i). The contrast of a region is calculated starting from contrasts with the regions which are contiguous to it :

$$C_{Inter} = \frac{\sum_{R_i} A_i c_i}{\sum_{R_i} A_i} \qquad (5)$$

with $c_i = \sum_{R_i} \frac{l_{ij}|m_i - m_j|}{m_i + m_j}$, m_i : the mean gray-level of the region R_i, l_{ij} the length of the frontier between R_i and R_j, l_i the perimeter of the region R_i.
- Levine and Nazif's intra-class uniformity (Intra) : This criterion computes the sum of the normalized standard deviation of each region.
- Combination of intra-class and inter-class disparity (Intra-inter) : This indicator combines similar versions of the Levine and Nazif interclass and intra-class contrast.
- Borsotti criterion (Borsotti) [1] : This measure is based on the number, the surface and the variance of the region :

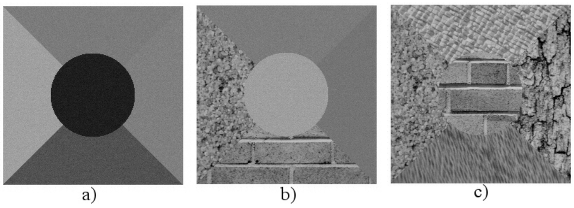
$$C_B = \frac{\sqrt{N}}{1000 \times A} \sum_{i=1}^{N} \left(\frac{\sum_{R_i} (f(s) - m_i)}{1 + \log(A_i)} + \frac{R(A_i)^2}{A_i^2} \right) \qquad (6)$$

where, m_i is the average value of the grey-levels in the region R_i, and $R(A_i)$ is the number of regions whose surface is equal to A_i.
- Rosenberger's criterion (Rosenberger) [7] :
The originality of this method lies in its adaptive computation according to the type of region (uniform or textured). In the textured case, the dispersion of some textured parameters is used and in the uniform case, gray-levels parameters are computed.
- Vinet's measure (Vinet) : it is a supervised evaluation criterion. It computes the correct classification rate by comparing the result with a ground truth. Since we work in this study on a database composed of synthetic images, the Vinet's measure is used as a point of comparison.

The database used for our tests includes 300 synthetic images composed of textured and uniform regions. Each image contains five regions of different types : texture extracted from the Brodatz's album [2] or uniform gray-level with low noise. One database called Unif, is composed of 5 uniform regions, the Mixed one is composed of 2 textured and 3 uniform regions and finally the Texture one contains only textured regions (see figure 1).

Each image is segmented by a classification method (fuzzy K-means) with a number of clusters equal to 5 and with 3 different types of parameter settings and the EDISON algorithm [3] :

- Segmentation adapted to uniform images : a 5x5 pixels analysis window and moments from order 1 to 4,
- Segmentation adapted to slightly textured images : a 9x9 pixels analysis window, moments from order 1 to 4 and attributes from the cooccurrence matrix,
- Segmentation adapted to strongly textured images : a 15x15 pixels analysis window, moments from order 1 to 4, attributes from the cooccurrence matrix and the normalized autocorrelation,
- Segmentation by the EDISON algorithm with defaults parameters (with the weight map option).



a) b) c)

Figure 1: Example of one image in each database a) Unif, b) Mixed and c) Texture.

We used the correlation factor of each criterion as indicator of similarity (see Table 1) between two criteria. It was computed on the 1200 segmentation results (300 images and 4 methods). The absolute value of the correlation factor of two criteria is near zero when they are complementary and near 1 when they are linearly dependent.

The criteria which obtain the best correlation factor with the Vinet's measure (reference) are Zeboudj, Borsotti and Rosenberger. In the uniform case (database Unif), the two criteria Zeboudj and Borsotti give the best results. When some regions in the image are textured (databases Mixed and Texture), the Rosenberger criterion is more efficient.

2.2 Fusion of criteria

In order to improve the previous evaluation criteria, we suggest to fusion them. A possible method consists in linearly combining the three previous criteria. The goal is so to determine the optimal linear coefficients (a,b,c) such as the linear combination of the selected criteria has the

	Unif	Mixed	Texture
Borsotti	-0.845	-0.264	-0.207
Zeboudj	0.909	-0.003	-0.142
Inter-region	0.200	0.227	0.181
Intra-region	0.158	0.030	0.256
Intra-inter	0.226	0.096	0.010
Rosenberger	0.277	0.143	0.355

Table 1: Correlation factors for each evaluation criterion with Vinet's measure for each image database.

most similar behavior to the Vinet's measure. We propose here to use again the correlation factor with the Vinet's measure computed on the 1200 segmentation results. The optimization method we use is a genetic algorithm. This method is appropriate because the objective function to maximise is non-linear and this approach gave good results in previous applications [8].

Genetic algorithms determine solutions of functions by simulating the evolution of a population until survival of best fitted individuals. Survivors are individuals obtained by crossing-over, mutation and selection of individuals from the previous generation. A genetic algorithm is defined by considering five essential data :

1. *genotype* : a set of characteristics of an individual such as its size. A vector of parameters (a,b,c) is considered as an individual,
2. *initial population* : a set of individuals characterized by their genotypes. It is composed of a set of random values of parameters,
3. *fitness function* : this function provides to quantify the fitness of an individual according to the environment. We take the correlation factor with the Vinet's measure over the 1200 segmentation results,
4. *operators on genotypes* : they define alterations on genotypes in order to evoluate the population during generations. There exists three types of operators :
   - individual mutation : genes of an individual are modified in order to better adapt to the environment. We use the Non-Uniform mutation process which randomly selects one chromosome j, and sets it equal to a non-uniform random number.
   - selection of an individual : individuals that are not adapted to the environment do not outlive to the next generation. We used the normalized geometric ranking selection method which defines a probability P_i for each individual i to be selected.
   - crossing-over : two individuals can reproduce themselves by combining their genes. We use the arithmetic crossover which produces two complementary linear combinations of the parents.
5. *stopping criterion* : this criterion allows to stop the evolution of the population. We choose to consider the stability of the standard deviation of the evaluation criterion of the population.

Given these five information, the execution of the genetic algorithm is realized in four steps :

1. definition of the initial population and computation of the fitness function of each individual,
2. selection and crossing-over of individuals of the population,
3. evaluation of individuals in the population,
4. back to the step 2 if the stopping criterion is not satisfied.

3. EXPERIMENTAL RESULTS

We use the previous genetic algorithm to combine the three selected evaluation criteria : Borsotti, Zeboudj and Rosenberger. We use a population of 100000 individuals and 200 iterations in the genetic algorithm. We obtain for each case (textured, uniform, mixed or unknown) optimal linear coefficients. In order to quantify the benefit of fusion, we present in the Table 3, the correlation factor on each database (Unif, Mixed, Texture and global) obtained by the best criterion and after fusion.

	Global	Unif	Mixed	Texture
best criterion	0,7068	0,9009	0,2638	0,3554
After fusion	0,7711	0,9483	0,2934	0,4195

Table 2: Comparison of the correlation factor with the Vinet's measure with the best criterion and after fusion.

We remark that the fusion process provides an increase of performance of nearly 10% compared to the best criterion for each database. We illustrate the efficiency of the approach on two real images of different types : the first one is considered as mixed and the second one as textured.

a) Outdoor image

The well known image ”CAR” (see figure 2) is an outdoor scene with uniform (sky,...) and textured regions (tree,...). It is considered as mixed, so we used the linear coefficients for this case.

Three evaluation criteria (Intra-Inter, Intra et Borsotti) consider the EDISON segmentation result as the best (see Table 3). As for us, this segmentation result seems to be effectively visually the best one. Some evaluation criteria, such as the Rosenberger's and Zeboudj's ones, prefer another segmentation result. But, if we focus on the fusion criterion, the EDISON segmentation result is preferred.

b) Radar image

The second image is a radar one (see Figure 3). As the image is very noisy, this image can be considered as textured.

The segmentation result that can be visually considered as the best one is again the EDISON one. Table 4 show the

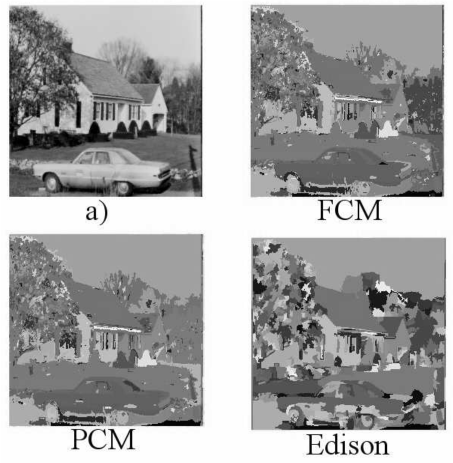


Figure 2: a) Original image and three segmentation results.

	FCM	PCM	EDISON
Borsotti	3.9530	1.6831	0.2165
Zeboudj	0.6424	0.6330	0.5305
Inter	0.2716	0.2725	0.2302
Intra	4.3969	4.4407	4.6605
Intra-Inter	0.5482	0.5517	0.5418
Rosenberger	0.615	0.6176	0.5043
Fusion	-3.8875	-1.8573	-0.4979

Table 3: Comparison of different segmentation results of an outdoor image by different evaluation criteria.

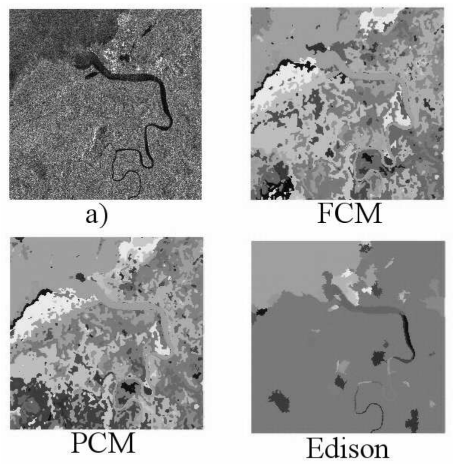


Figure 3: a) Radar image and three segmentation results.

values of each evaluation criterion. If we focus on the Rosenberger criterion, the value for the EDISON result is much better than the other segmentation results. On the contrary, the Zeboudj's criterion has a bad value for this segmentation result. If we consider now the fusion result, we see that the EDISON segmentation result is again correctly preferred.

	FCM	PCM	EDISON
Borsotti	0.1952	0.2793	0.0293
Zeboudj	0.1094	0.1172	0.0432
Inter	0.1401	0.1394	0.2559
Intra	6.2846	7.5824	1.1364
Intra-Inter	0.5196	0.5214	0.5419
Rosenberger	0.4699	0.4677	0.9074
Fusion	-0.1157	-0.1983	0.2381

Table 4: Comparison of different segmentation results of a radar image by different evaluation criteria.

4. CONCLUSIONS AND PERSPECTIVES

Segmentation evaluation is a great challenge and has lots of applications (comparison of segmentation methods, choice of parameters of a segmentation method for an image,...). We presented in this paper a brief comparative study of unsupervised evaluation criteria. We also showed the benefit of linearly combining the best ones to improve their performance. A genetic algorithm is used to determine the optimal coefficients. We are currently investigating on other fusion methods to automatically derive a more complex combination.

REFERENCES

[1] M. Borsotti, P. Campadelli, and R. Schettini. Quantitative evaluation of color image segmentation results. *Pattern Recognition Letters*, 19:741–747, 1998.

[2] P. Brodatz. *Textures, a photographic album for artists and designers*. Dover, New York, 1966.

[3] D. Comanicu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Machine Intell.*, 24:603–619, 2002.

[4] B. McCane. *On the evaluation of image segmentation algorithms*. Digital Image Computing: Techniques and Applications, 1997.

[5] A.M. Nazif and M.D. Levine. Low level image segmentation : an expert system. *IEEE Transaction on pattern analysis and machine intelligence*, 6:555–577, 1984.

[6] N. Paragios and R. Deriche. Geodesic active regions for supervised texture segmentation. *in Proceedings of 7th International Conference on Computer Vision*, 1:926–932, 1999.

[7] C. Rosenberger and K. Chehdi. Genetic fusion : Application to multi-components image segmentation. *IEEE ICASSP*, pages 2219–2222, Istanbul 2000.

[8] P. Wall. *A Genetic Algorithm for Resource-Constrained Scheduling*. PhD thesis, MIT, 1996.

[9] Y.J. Zhang. A survey on evaluation method for image segmentation. *Computer vision and pattern recognition*, 29:1335–1346, 1996.