Unsupervised evaluation of image segmentation application to multi-spectral images

Conference Paper · September 2004

DOI: 10.1109/ICPR.2004.1334206 · Source: IEEE Xplore

5 authors, including:

Sébastien Chabrier
University of French Polynesia

Bruno Emile
Université d'Orléans

Hugo Laurent
University of Ottawa

Christophe Rosenberger
National Graduate School of Engineering and Research Center (Caen)

Unsupervised Evaluation of Image Segmentation Application to Multi-spectral Images

S. Chabrier, B. Emile, H. Laurent, C. Rosenberger and P. Marché
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 2248484000
email: sebastien.chabrier@ensi-bourges.fr

Abstract

We present in this article a study of some unsupervised evaluation criteria of an image segmentation result. The goal of this work is to be able to automatically choose the parameters of a segmentation method best fitted for an image or to fusion different segmentation results. We compared six unsupervised evaluation criteria on a database composed of 100 synthetic gray-level images segmented by four methods. Vinet's measure is used as an objective function to compare the behavior of the different criteria. We finally apply these criteria to evaluate segmentation results of multi-components images. We present in this article some experimental results of evaluation of gray-level and multi-components natural 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 [7], [4] but it still remains difficult to evaluate their efficiency. Actually, many works have been performed to solve the more general problem of the evaluation of an image segmentation result.

Two main approaches exist. 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. This article focus on this kind of approach. In [10], a comparative study of evaluation methods of segmentation results of gray-level images is developed. Mac Cane [5] showed that in order to evaluate a segmentation method, it is necessary to use the maximum of criteria and to combine the results. Most of tested criteria are not adapted for textured images. The problem is that most of real images are composed of textured regions.

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 generalize this approach to multi-components images. We present some experimental results of evaluation of natural images and we show the efficiency of the proposed method.

2. Developed Method

2.1. Unsupervised Evaluation criteria

We selected from the state of art [10] 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 outside (E_i) of the region R_i are respectively :

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

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

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 :

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

The global contrast is :

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

- Levine and Nazif's interclass contrast (Inter) [6] : This criterion (C_{Inter}) computes the sum of contrasts of the regions (R_i) balanced by their surfaces (A_i). The contrast of a region is calculated starting from contrasts with the regions which are contiguous to it :

$$C_{Inter} = \frac{\sum_{R_i} A_i c_i}{\sum_{R_i} A_i} \tag{5}$$

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) [2] : This measure is based on the number, the surface and the variance of the regions.

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

Where, m_i is the mean of the region R_i, and $R(A_i)$ is the number of regions whose surface is equal to A_i.

- Rosenberger's criterion (Rosenberger) [9] :
  This method enables to evaluate a segmentation result without any *a priori* knowledge. 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.

2.2. Image database

The database used for our tests includes 100 synthetic images composed of textured and uniform regions. Each image contains five regions : two textured regions from the Brodatz's album [3], two strongly noisy regions and one uniform region (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 [4] :

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

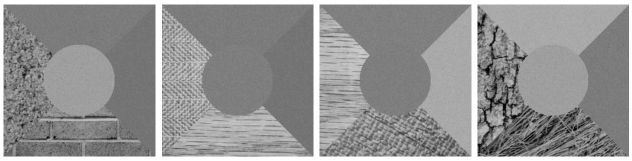


Figure 1. Examples of test images.

2.3. Discussion

We studied the correlation factor of each criterion as indicator of similarity (see Table 1). It was computed considering 400 segmentation results (100 images and 4 methods). The last column shows the correlation factor of each criterion with the Vinet's measure. The absolute value of the correlation factor of two variables is near zero when they are complementary and near 1 when they are similar. The criteria which obtain the best correlation factor with the Vinet's measure (ground truth) are Zeboudj, Intra-inter and Rosenberger. In order to complete the previous analysis, we use five synthetic segmentation results (see Figure 2) to evaluate the behavior of each criterion. Table 2 shows the values of the different criteria for each synthetic segmentation result of the Figure 2. It is clear that the ground truth VT5 (composed of 5 regions) is the best result. However three evaluation criteria do not give this segmentation result as being the best. That is due to the fact that no *a priori* knowledge is used and that they privilege the homogeneity intra-region as well as a low number of region. We can notice that the Rosenberger and Zeboudj criteria seem to give the best results. We can always notice that the Rosenberger's criterion is more general-purpose than Zeboudj and better takes

0-7695-2128-2/04 $20.00 (C) 2004 IEEE

into account complex segmentation results. Zeboudj's contrast gives better results on very simple segmentation results.

Zeboudj	1	0.81	0.02	-0.57	-0.09	0.06	**-0.21**
Inter		1	-0.06	-0.39	-0.03	0.04	0.01
Intra			1	-0.30	0.77	0.18	0.13
Intra-inter				1	-0.22	0.12	**0.22**
Borsotti					1	-0.07	-0.07
Rosenberger						1	**0.25**
Vinet							1

Table 1. Correlation factor of each evaluation criterion.

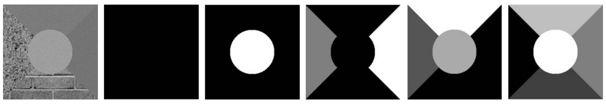


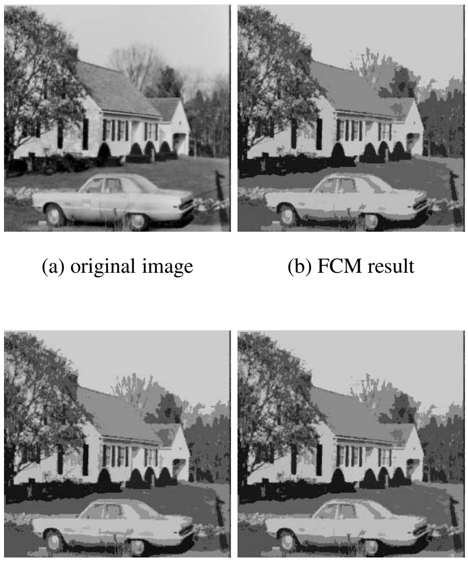
Figure 2. original image and 5 synthetic segmentation results.

	VT1	VT2	VT3	VT4	VT5
Zeboudj	0	0.14	0.09	0.24	**0.34**
Inter	0	**0.20**	0.02	0.12	0.13
Intra	0.08	0.05	0.16	0.11	**0.17**
Intra-inter	0.46	**0.55**	0.48	0.52	0.51
Borsotti	0.01	0.008	**0.019**	0.011	0.013
Rosenberger	0	0.58	0.42	0.60	**0.80**
Vinet	0.20	0.40	0.62	0.79	**1**

Table 2. Evaluation of synthetic segmentations results.

2.4. Application to natural gray-level images

These various evaluation criteria were also tested on real images, for different applications (aerial pictures, outdoor scene,..). These images were segmented by three different segmentation methods : FCM, PCM [1] and EDISON [4]. We only present in this article the segmentation result of one image (see Figure 3). It is quite difficult to determine visually the best segmentation result. If we focus on the trees, many regions are detected by the Edison algorithm. One notices that the criteria Intra, Borsotti and Intra-inter support regions of regular form, of average size, with a good uniformity. On the contrary, the criteria of Zeboudj, Inter and Rosenberger support more complex regions with a strong inter-region contrast.



(a) original image (b) FCM result

(c) PCM result (d) EDISON result

Figure 3. Three segmentation results of image "CAR"

	FCM	PCM	Edison
Zeboudj	**0.64**	0.63	0.53
Inter	0.27	**0.28**	0.23
Intra	4.39	4.44	**4.66**
Intra-inter	0.55	0.55	**0.54**
Borsotti	3.95	1.68	**0.21**
Rosenberger	0.60	**0.61**	0.50

Table 3. Evaluation of the segmentation results of the Figure 3.

2.5. Generalization to multi-components images

We try in this section to generalize this approach to multi-components images. The objective is to evaluate different segmentation results (obtained by using different parameters) by merging some evaluation criteria (among the six following methods : Zeboudj, Inter, Intra, Intra-inter, Borsotti and Rosenberger) computed for each band. Three simple fusion methods are used : the minimum, the maximum and the average value of the criterion computed on each band. In order to compare the different evaluation methods in the multi-components case, we used 20 synthetic images with 5 components. Each image is segmented with the MLBG method [8] using 32 different parameter settings. Vinet's measure is used again as an objective func-

0-7695-2128-2/04 $20.00 (C) 2004 IEEE

tion and allows us to sort each segmentation result. For each unsupervised evaluation method, each fusion method gives a sorting of the 32 segmentation results for each image. So judged, the best evaluation method associated with the best fusion process is the one for which the sorting is the most similar to the Vinet's measure for the 20 images. To compare two sorting of segmentation results, we sum each difference between the position in the sorting obtained by using the Vinet's measure and an other evaluation criterion.

3. Experimental results

Table 4 shows that there is no fundamental difference between the three fusion operators (mean, minimum, maximum). The best evaluation criterion in the multi-components case, in sense of our approach, is the Rosenberger's criterion with the fusion method based on the mean.

	Mean	Minimum	Maximum
Zeboudj	187	187	170
Inter	137	143	121
intra	187	187	187
intra-inter	209	209	209
Borsotti	149	145	149
Rosenberger	**51**	52	56

Table 4. Distance between criteria and Vinet with 3 fusions

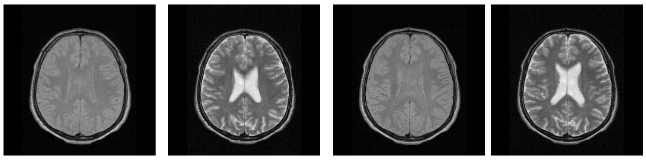
We applied this criterion in the multi-components case. Figure 4 presents three segmentation results of a MRI image with 4 bands obtained by the MLBG method with different parameters (windows size,...). The Rosenberger's criterion associated with Mean fusion can sort the different segmentation results. The presented result 3 is the best one (criterion : 0.731) before the result 2 (criterion : 0.66) and finally the result 1 (criterion : 0.649). This sorting of these segmentation results is difficult to validate with the visual perception even the last result seems to be more precise.

4. Conclusion and perspectives

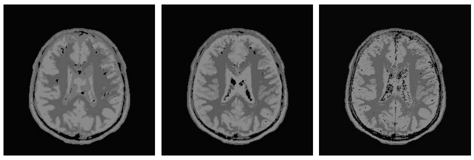
The article presented first a comparison of six unsupervised evaluation criteria of a segmentation result of gray-level images. This study reveals that two criteria seem to give better results than others : Zeboudj and Rosenberger criteria. Zeboudj's contrast supports more the small regions and is better adapted for uniform images. Rosenberger's criterion was favorably applied to textured images. For multi-components images, Rosenberger's criterion associated with mean fusion seems to give the closest to the results to the ground truth. We are working actually on more complex fusion operators to combine the values of evaluation criterion computed on each band.

Acknowledgement

The authors would like to thank financial support provided by the Conseil Régional du Centre and the European union(FSE).



(a) band 1 (b) band 2 (c) band 3 (d) band 4



(e) result1 (f) result2 (g) result3

Figure 4. Three segmentation results of a MRI image with 4 bands

References

[1] Vincent Barra. *Fusion d'Images 3D du Cerveau : Etude de Modèles et Applications*. PhD thesis, Université d'Auvergne, 2000.

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

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

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

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

[6] 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.

[7] 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.

[8] C. Rosenberger and K. Chehdi. Unsupervised segmentation of multi-spectral images. *International Conference on Advanced Concepts for Intelligent Vision Systems*, Ghent, Belgium, 2003.

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

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

0-7695-2128-2/04 $20.00 (C) 2004 IEEE