Multimodal Evaluation Method for Medical Image Segmentation

Rubén Cárdenes[1], Meritxell Bach[2], Ying-Veronica Chi[5], Ioannis Marras[4], Rodrigo de Luis[1], Mats Anderson[3], Peter Cashman[5] and Matthieu Bultelle[5]

[1] LPI, University of Valladolid, Spain
[2] EPFL, Lausanne, Switzerland
[3] MI, Linkoping University, Sweden
[4] Aristotle University of Thessaloniki, Greece
[5] Imperial College, London, UK

Abstract. This paper is a joint effort between several institutions that propose a new evaluation method for segmentation techniques using multimodal information. We propose new similarity measures based on the location and the intensity values of the misclassified voxels and also based on the connectivity and the boundaries of the segmented data and we show how the combination of these measures can improve the quality of the evaluation. The study that we show here has been carried out using four different segmentation methods from four different labs applied to a MRI simulated dataset of the brain. We claim that our new measures improve the robustness of the evaluation and provides better understanding about the difference between segmentation methods.

Key words: Evaluation, Segmentation, Multimodal evaluation, Similarity Measures, Brain tissue segmentation

1 Introduction

The goal of medical image segmentation is to obtain a labeled image where each label corresponds to the real anatomy of the patient. Several technical factors make this goal hard if not impossible to achieve with the current technology. The data acquisition process always introduces some noise, and the formation of the image is limited in resolution, therefore each pixel would correspond to more than one tissue, situation known as partial volume effect.

A crucial aspect of segmentation techniques is their reliance on contextual information for them to be effective. An important source of contextual information for medical data is the medical knowledge collected on the problem. Turning this medical knowledge into a set of criteria adapted to computer vision is one of the most difficult aspects of the development of computerized segmentation routines. It follows that segmentation techniques are best suited to specific applications and classes of data, for example to a type of data where an underlying assumption is true. No segmentation is better than the others for any purpose. Thus for a particular problem we have to figure out what available method fits

best into our needs in terms of accuracy, speed, reproducibility and user interaction.

It is sometimes difficult to assess the accuracy of a method and if it is good enough for a given application. The common way to do this is to compare its results with the results obtained by other techniques against reference data, known as ground truth or gold standard. This process is known as validation or evaluation. Comparison between the result of an algorithm and the reference data is achieved by computing a distance (metric) between the two datasets, see [1–3] for examples.

2 State of the Art on Evaluation Techniques

Many works to evaluate segmentation methods has been reported in the last two decades. A good survey about segmentation evaluation can be found in [4]. This author distinguishes the evaluation methods between empirical (based on the study of the results) and analytical (based only on intrinsic features of the methods). The empirical methods are divided into goodness and discrepancy methods, where the former are based on the study of the results themselves, and the latter compare the results with a reference or ground truth. Among the discrepancy methods, there exist several features reported to measure the quality of the segmentation: number of misclassified voxels, position of misclassified voxels, number of objects in the image, feature values of segmented objects and other miscellaneous quantities.

Most of the methods in the literature for segmentation evaluation are based on classic discrepancy methods, limited to the computation of the number of voxels of the segmented classes in the results and in a gold standard. Other authors has introduced the location of the misclassified voxels as a feature to measure the discrepancy between segmented images, for example, Yasnoff [5], Straters [6] and later Pichon [7] proposed to use an error distance from the misclassified voxels to the gold standard. Huttenlocher [8] use the partial Hausdorff distance between set of voxels, and also [1] proposed an overlap distance using fuzzy set theory to take into account fractional labels coming from multiple test images. Other work proposed by Cardoso [2] presents a general distance between segmentation partitions to measure the quality of a given segmentation.

One interesting work about segmentation evaluation is the one published by Udupa [3] who proposed a methodology based not only on the accuracy of the segmentation. They also present measures of precision (reproducibility) and efficiency (time taken), and they stated that the combination of those factors are essential in the assessment of the performance of any segmentation method.

Some other methods has been proposed to perform segmentation evaluation without a ground truth, see for instance [9–11].

The main goal of this paper is to introduce new similarity measures to combine them for segmentation evaluation, in terms of accuracy, using a known ground truth. In order to demonstrate this methodology we will compare four segmentation techniques for a specific application: brain tissue segmentation.

There is of course, a problem inherent to this way of evaluation, because it is quite difficult to obtain a reliable reference segmentation dataset. The most used approach is to use manual segmentation, or a combination of several manual segmentations, from several experts if possible. There is however the possibility to validate brain tissue segmentation methods on a brain *simulated* data set as the one proposed by the *Brain Web* MR simulator [12]. Their data is very well-suited for this purpose since a ground-truth classification is known while different types of MR modalities and image resolution and artifacts can be reproduced.

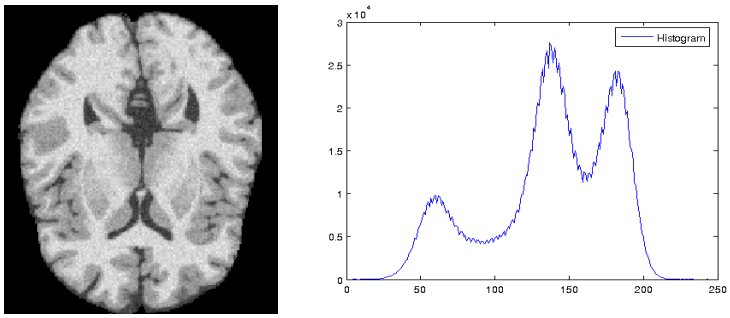
3 Segmentation Methods

The methods used in this work for the evaluation study has been provided by four different institutions involved in the Work Package dedicated to medical applications (WP10), of the Similar network of excellence. These methods are:

- **Mean-Shift initialized Level Set method (MSiLS)** The mean-shift is a non parametric clustering technique, that in this case has been implemented in combination with the classic geometric active contour introduced by Caselles [13] and Malladi [14]. For this application a low-resolution mean-shift method was leveraged to denoise the original image and then to initialize deformable contours around CSF and WM homogeneous regions. Then the contours evolve converging toward the implicit fuzzy boundaries between CSF, WM, and GM driven by the standard geometric active contour energy. Since energy-based deformable contour methods are known to work better on large images, the original images were augmented 10-fold before segmentation.
- **Statistical Parametric Classification using Gaussian Hidden Markov Random Field Model, (GHMRF)** This is a non-supervised parametric classification technique, that assumes that the data is modeled by a mixture of Gaussian distributions, and which includes a local spatial prior modeled by the Markovian theory. The segmentation is done using an adaption of the EM algorithm, and the probability distributions are initialized by the k-means algorithm. See [15] for details.
- **k Nearest Neighbors, (kNN)** This is a supervised non-parametric technique, that in its classic approach, classifies each voxel independently, by searching in a set of voxels selected and classified manually by a user. In order to speed up the search, we implement the algorithm proposed by [16]. Due to that this method is not context dependent, we have applied an initial non linear filtering [17] in order to remove some noise.
- **Split and Merge Segmentation, (SM)** This technique is a generalization of the classic split-merge algorithm in the sense that during the splitting procedure the volume is not subdivided into sub-volumes of the same type (parallelepipeds) but different splitting configurations are tested. The method is based on two parameters (T1 for the split part, and T2 for the merge part), see [18].

4 Evaluation Study

As we have said, the images used in this study comes from the digital brain phantom from McConnell Brain Imaging Center [12]. In this work, all the methods have been applied to that dataset with noise 5% and no RF on the T1-weighted modality. The volume used has been preprocessed to remove non brain tissues, so only the intracranial cavity has been used in the experiments, using a volume of size 161x187x161 voxels, with isotropic 1 mm voxel size. One axial slice as well as the volume histogram of this data set are shown in Fig. 1.



(a) (b)

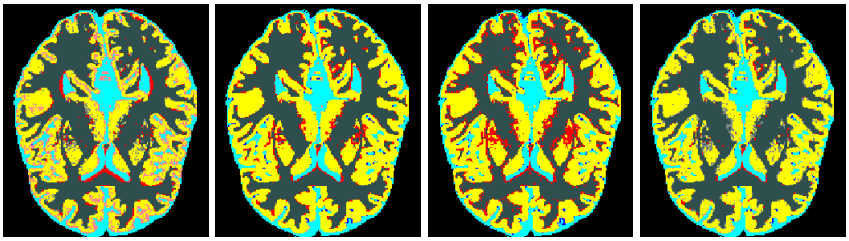
Fig. 1. *Axial slice of the brainweb simulated MRI (a) and volume histogram (b)*

For the segmentation using GHMRF method, the value of β is fixed empirically to 1.2, $U(x, \beta)$ follows the Potts model, and instead of computing Z ,the conditional probabilities at a given point $P(x_i|x_{N_i})$ are force to sum up 1 among all possible labels.

The kNN segmentation have been carried out using a training set of 194 points, using $K = 9$, and choosing $\tau = 100$, $\sigma = 2$, and 5 iterations for the non linear filtering.

For the SM method, the values that have been used for the input parameters were $T_1 = 10$, $T_2 = 38$ and $N = 3$. However the method is fairly insensitive to small deviations ($\pm 20\%$) from the values of T_1, T_2.

We show in Fig. 2, the segmentation results for the axial slice chosen, using blue for CSF, yellow for GM and dark green for WM. In these images we also show in grey and pink the voxels that overlap between GM and WM, and in blue and red the voxels that overlap between CSF and WM. There are no voxels in this slice belonging to the overlap between CSF and WM, because there is little overlapping between those classes and no overlap in this particular slice.



(a) (b) (c) (d)

Fig. 2. *Segmented images of the axial slice of Fig. 1 with error voxels overlapped, using MSiLS (a), GHMRF (b), kNN (c) and SM (d)*

4.1 Classic Similarity Measures

One classic approach to determine how good are the segmentations, are similarity measurements based on region overlap. One of the most common measures is the construction of the confusion tables, whose values represent the overlapping between two classes with respect to the number of voxels of the class in the gold standard

$$M_{ij} := \frac{|X_i \cap Y_j|}{|Y_j|} \tag{1}$$

where the subindices represent the classes and $||$ stands for the number of elements. Other common measures used are the Jaccard (JC), Dice Similarity (DS), Tanimoto (TN), and Volume Similarity (VS) coefficients. All of them take values between 0 and 1. If X is the set of voxels segmented as class c in one volume, Y is the set of voxels of the same class in the other volume, $a = |X \cap Y|$, $b = |X \backslash Y|$, $c = |Y \backslash X|$, and $d = |\overline{X \cup Y}|$, we can define these measures with the following expressions,

$$JC := \frac{|X \cap Y|}{|X \cup Y|} = \frac{a}{a+b+c} \tag{2}$$

$$DS := \frac{2|X \cap Y|}{|X| + |Y|} = \frac{2a}{2a+b+c} \tag{3}$$

These two coefficients are equal to one if X and Y are the same region, and zero if they are disjoint regions. In fact, they are related by $DS = 2JC/(JC+1)$.

$$TN := \frac{|X \cap Y| + |\overline{X \cup Y}|}{|X \cup Y| + |\overline{X \cap Y}|} = \frac{a+d}{a+2b+2c+d} \tag{4}$$

This case is one if X is equal to Y, and zero if they are disjoint regions and they occupy all the image.

$$VS := 1 - \frac{||X| - |Y||}{|X| + |Y|} = 1 - \frac{|b-c|}{2a+b+c} \tag{5}$$

This is one if the number of elements of X is equal to the number of elements of Y, and zero if one of them is empty. In Fig. 3 we show the results of these similarity measures computed over the segmented volumes obtained with each method, and using the gold standard.

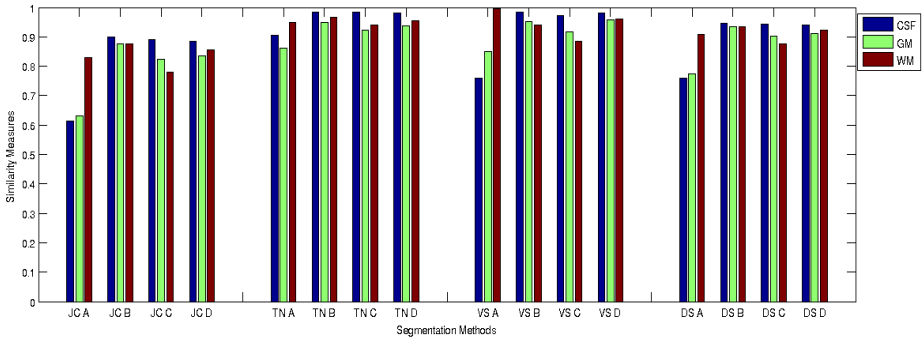


Fig. 3. *Classic similarity measures (JC, TN, VS and DS) computed for all methods, A: MSiLS, B: GHMRF, C: kNN, and D: SM*

Looking at Fig. 3, we can have a rough idea about the accuracy of the different methods. However, some values like the TN coefficients differs from the values obtained by the other coefficients (for instance, the classes are ordered different than the other three coefficients) and give values that hardly can differentiate the methods. This is because it depends on the number of voxels outside X and Y, that can be very large in our case, therefore leading to values near one, even if there is not too much overlapping. The VS coefficients present results not realistic (notice an almost perfect classification of WM in MSiLS method), that is because it depends only on the number of voxels of X and Y, and it can be one even if there exists no overlapping at all. And finally the JC and VS coefficients show values equivalent as expected. For those reasons, we will use the JC coefficient for our evaluation study.

4.2 Distance Based Similarity Measures

The error measures described above are based only on the number of voxels of the classes in the segmented image and in the gold standard, their union, and their intersection. We propose in this section to include the voxels location to improve qualitatively the error measurements. As Pichon [7], and also Crum [1] recently proposed, it is important to use the distances from the misclassified voxels to the ground truth in order to improve the similarity measures. We can define the distances from the misclassified voxels as in [7]

$$d(r) := \begin{cases} 0, r \in X \cap Y \\ min_{x \in X} ||r - x||, r \in Y \backslash X \\ min_{y \in Y} ||r - y||, r \in X \backslash Y \end{cases} \quad (6)$$

Other popular distance is the Hausdorff distance, defined as

$$H(X, Y) := \max\{\max_{x \in X} \min_{y \in Y} ||x - y||, \max_{y \in Y} \min_{x \in X} ||x - y||\} \quad (7)$$

which is the maximum distance one set has to move its boundaries so that it would enclose the other set. With any of the distance definitions mentioned, it is possible to obtain more reliable similarity measures. Some of those measures can be the Yasnoff discrepancy measure [5], the Factor of Merit [6], or just the mean (μ) and standard deviation (σ).

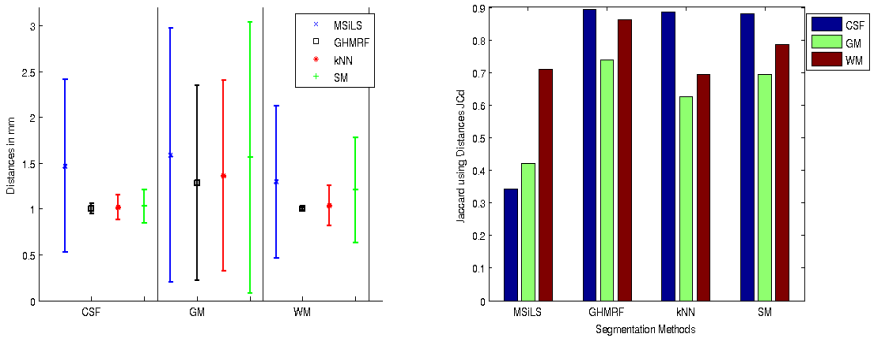
We propose to use the distance defined in (6), to define a new similarity measure that takes values between 0 and 1. The idea is to penalize more those voxels that are more distant from their corresponding class in the gold standard, i.e. to weight every misclassified voxel by its Euclidean distance to the nearest voxel of the class it should belong to. To compute those Euclidean distances, it is enough to simply compute the Distance Transformation (DT) from a given class in the gold standard to the rest of the image, and look at the voxels of the DT at the positions of the misclassified voxels. We will use the squares of the distances to penalize more to very distant voxels.

The new measure we propose is called JCd, and is defined by substituting the values b and c from (2), by $\sum_i d(x_i)^2$ and $\sum_i d(y_i)^2$ respectively, where x_i are misclassified voxels of X that should be classified as Y, y_i are voxels of Y that should be classified as X, and $d()$ is the distance defined in (6). Obviously we can use any of the other measures definitions of equations 4,5,3, to construct the new measure, but we will use the JC coefficient for the reasons commented in sect. 4.1.

The mean and standard deviation of the distances for every segmented class and for every method is shown in Fig. 4 (a). In Fig. 4 (b), we show the values of JCd.

4.3 Intensity Based Similarity Measures

In this section we introduce another similarity measure, but this time instead of using Euclidean distances in the image, we use the intensity values. The idea is to penalize more the misclassified voxels that should belong to a given class c, when it is close to the theoretic mean of that class c, because the voxels that are near the theoretic mean, should be easy to classify. Therefore, we will define a weighting function, dependent on the theoretic mean and variance of each class. Those parameters are easy to obtain from the gold standard and the original data, by computing the mean and variance values of the original voxels indexed by each class in the gold standard segmentation. Then, we can construct three probability density functions for each class, Y_{csf}, Y_{gm} and Y_{wm}, that can be used to define the weighting function F, that we can express as



(a) (b)

Fig. 4. *Average distances and standard deviation for the misclassified voxels (a) and distance based similarity measures, JCd, computed for all methods (b)*

$$F = H(1 + Y_{csf} + Y_{gm} + Y_{wm}) \tag{8}$$

where H is a constant that increases the penalization effect at each misclassified voxel if it increases. We are not using the number of voxels of each class to weight each density function because we just want to weight each misclassified voxel by its distance to be theoretic mean. We show the weighting function F, in Fig. 5 (a).

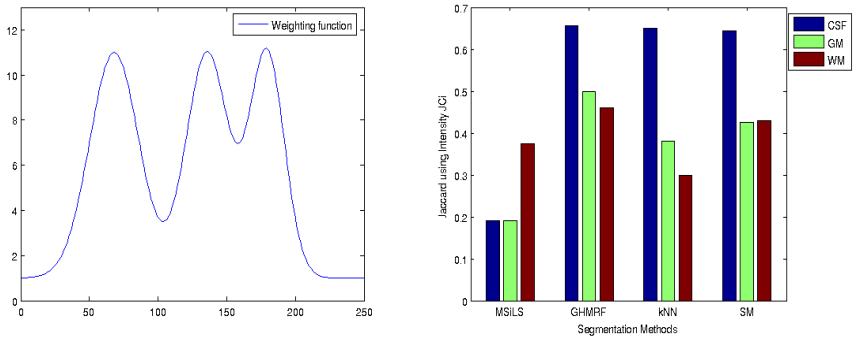
Using this function, we can define a new similarity measure, that we will call JCi changing b and c by $\sum_i F(x_i)$ and $\sum_i F(y_i)$ respectively in (2). Again, we obtain a measure constrained between 0 and 1, and the results obtained are shown in Fig. 5 (b), using $H = 10$.

4.4 Connectivity Coefficient

Other similarity measure can be defined using the connectivity of labeled images. The connectivity of a region X, in a 3D regular grid is defined using a morphological dilation operator $\mathcal{D}_s$, with s a structuring element. We say that X is connected with other region Y, if

$$\mathcal{D}_s(X) \cap Y \neq \emptyset \tag{9}$$

We use a 3x3x3 structuring element, thus defining a connectivity taking the 26 closest neighbors of each voxel. The number of connected components for each class N_{X_c} in the segmented volume can be compared with the number of connected components for the same class in the gold standard N_{Y_c}. The definition of a connectivity coefficient CC that takes values between 0 and 1 can be expressed as



(a) (b)

Fig. 5. *Weighting function F (a), and intensity based similarity measure computed for all methods (b)*

$$CC_c := \frac{\min\{N_{X_c}, N_{Y_c}\}}{N_{X_c} + N_{Y_c}} \tag{10}$$

4.5 Similarity Measures on the Boundaries

It is also interesting to use the segmented boundaries to measure the similarity between the ground truth and the segmentations. A measure between 0 and 1 can be defined using the JC for boundaries. Given the boundary of one segmented class, c, ∂X_c, and the boundary of that class in the ground truth, ∂Y_c, the boundary JC coefficient is defined as

$$BJC_c := \frac{|\partial X_c \cap \partial Y_c|}{|\partial X_c \cup \partial Y_c|} \tag{11}$$

Sometimes, the segmented images may contains many small groups of isolated voxels. Of course, those erroneous voxels are significant on our error measures, but we want a measure definition that does not take into account those voxels, because counting scattered voxels placed outside the main boundaries or in holes inside them, will decrease this similarity measure even if the boundary of the ground truth really fits with the boundary of the segmented image, and that issue will be addressed by the connectivity coefficient, CC. Therefore, we will use the modified boundary for every class in the segmented image $\partial X'(c)$. We can express $\partial X(c)$ as the union of non connected sets

$$\partial X_c = \bigcup_i \partial X_c^i \quad \text{where} \quad \mathcal{D}_s(\partial X_c^i) \cap \partial X_c^j = \emptyset \ \forall \ i, j, \ i \neq j \tag{12}$$

where $\mathcal{D}_s$, is the morphological dilation operator, as defined before. The new boundary $\partial X'(c)$ is then defined as

$$\partial X'_c = \bigcup_k \partial X_c^k \quad \text{where} \quad \partial X_c^k \cap \partial Y_c \neq \emptyset \ \forall \ k \tag{13}$$

and the modified measure is:

$$BJC'_c = \frac{|\partial X'_c \cap \partial Y_c|}{|\partial X'_c \cup \partial Y_c|} \tag{14}$$

4.6 Global Multimodal Similarity Measure

A global definition of a similarity measure is also needed. We propose to use the above definitions to combine different features to obtain more objective and reliable results. In this work we state that, as well as in human vision, an intelligent system should employ several features to decide between different segmentation results. An intelligent similarity measure, will emerge from the combination of the measures proposed before, a multimodal similarity measure. The Fig. 6 illustrate better our idea. In those figures we have plotted the results of two similarity measures, one of them representing the x axis and the other one the y axis. Using this representation we can see more clearly the differences between several methods than in unidimensional plots. In the figures we have also plotted some circles placed at the middle point of each method, by averaging the values of all classes, and using a radius proportional to the standard deviation. Notice that better measures correspond to smaller circles and closer to the $(1, 1)$ point.

Finally, in order to define a global similarity measure that include all the measures described here, let's construct a vector of similarity measures for a given class

$$\mathbf{v_c} = [JC_c, JCd_c, JCi_c, CC_c, BJC'_c] \tag{15}$$

which is a vector composed of the classic Jaccard coefficient JC, the distance based JC, the intensity JC, the connectivity coefficient, and the modified boundary JC. The global similarity measure for a given class c, will be

$$G_c := [\mathbf{v_c} \mathbf{K} \mathbf{v_c}^T]^{1/2} \tag{16}$$

where $\mathbf{K}$ is a matrix whose elements K_{ij} weights the different measures between them. For simplicity, we will use $\mathbf{K}$ with values different from zero only in the diagonal, $K_{ii} \neq 0$, all of them taking the same value, which is natural because all the measures used are defined between 0 and 1, being 0 the worst case and 1 the perfect case. Depending on the application, it could be useful to increase or decrease some of the values in the diagonal of $\mathbf{K}$, to give more importance to some of the measures. To obtain a final value for the entire method, we propose to combine the values obtained for each class, using the number of voxels of each class in the gold standard $|Y_c|$ as the weights:

$$G := \frac{\sum_c G_c |Y_c|}{\sum_c |Y_c|} \tag{17}$$

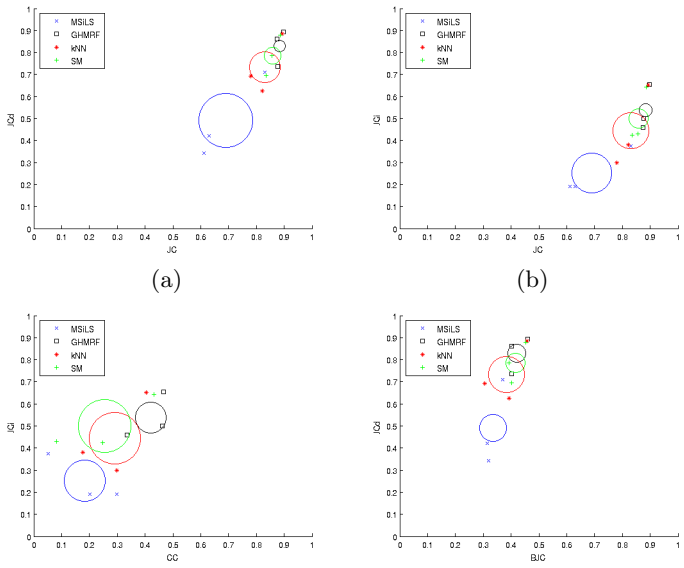


Fig. 6. *Joint similarity measures, JC vs JCd (a), JC vs JCi (b), CC vs JCi (c) and BJC vs JCd (d)*

We show in Fig. 7, the values for the global similarity measures per class and for the whole segmentation, for the four different methods studied.

5 Conclusion and Future Works

The main contribution in this evaluation study is the combination of several features to obtain the maximum reliability in the evaluation of several segmentation methods. As far as we know, this is the first that multiple similarity measures are combined to evaluate the accuracy of several segmentation methods. We have shown that classic similarity measures such as JC, VS, TN and DS produce similar values that could arise in erroneous decisions. Therefore, we have proposed a set of new similarity measures, using different criteria than the sizes of volume overlapping. We propose a new measure based on the distances from the misclassified voxels to the ground truth, then we propose to use the intensity of the misclassified voxels, as well as the connectivity and the boundaries of the segmented images. As a result we have proposed a new global multimodal similarity measure that combines the similarity measures proposed.

We have also presented 2D plots of pairs of similarity measures that show how the combination of several measures improves the visual representation of the difference between several methods, and motivate the validity of the multidimensional or multimodal global measure proposed here. Needless to say, the

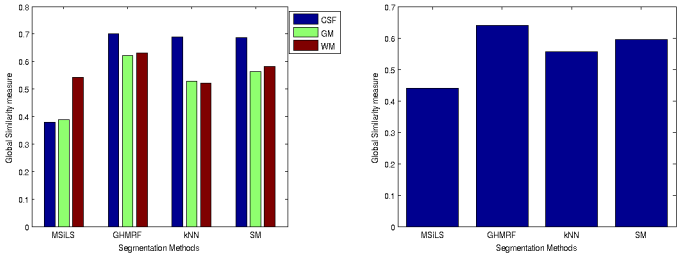


Fig. 7. *Global similarity measures per class (a) and averaged (b)*

coefficients of the matrix **K** should be selected appropriately for every application in order to provide an objective global similarity measure.

The correspondence between visual inspection (see Fig. 2), and the numeric values of our global measure fits quite well, resulting in a classification in order of decreasing quality: GHMRF, SM, kNN and SMiLS. This is a natural result, because GHMRF method is designed specifically for this particular application. SM and kNN methods produce fairly good results, and the SMiLS method performs also good, taking into account that it is not optimized for this task.

The brain classification study done here is not intensive, and it should be considered as a good example of how our proposed evaluation method can be applied. Notice also that new measures not related to accuracy, for instance, measures based on reproducibility, efficiency and user interaction, can be included in our model, as proposed by Udupa [3].

References

1. Crum, W., Camara, O., Hill, D.: Generalized overlap measures for evaluation and validation in medical image analysis. IEEE Transactions on Medical Imaging **25**(11) (2006) 1451–1461
2. Cardoso, J., L.Corte-Real: Toward a generic evaluation of image segmentation. IEEE Transactions on Image Processing **14**(11) (Nov. 2005) 1773–1782
3. Udupa, J., LaBlanc, V., Schmidt, H., Imielinska, C., Saha, P., Grevera, G., Zhuge, Y., Molholt, P., Jin, Y., Currie, L.: A methodology for evaluating image segmentation algorithms. In: SPIE Conference on Medical Imaging, San Diego, CA, USA (2002)
4. Zhang, Y.: A review of recent evaluation methods for image segmentation. In: International Symposium on Signal Processing and its Applications (ISSPA). (2001) 148–151
5. Yasnoff, W., Miu, J., Bacus, J.: "error measures for scene segmentation". Pattern Recognition **9** (1977) 217–231
6. Straters, K., Gerbrands, J.: Three-dimensional segmentation using a split, merge and group approach. Pattern Recognition Letters **12** (1991) 307–325

7. Pichon, E., Tannenbaum, A., Kikinis, R.: A statistically based flow for image segmentation. Medical Image Analysis **8** (2004) 267–274
8. Huttenlocher, D., Klanderman, G., Rucklidge, W.: Comparing images using the hausdorff distance. PAMI **15**(9) (1993) 850–863
9. Warfield, S., Zou, K., Wells, W.: Simultaneous truth and performance level estimation (staple): an algorithm for the validation of image segmentation. IEEE Transactions on Medical Imaging **23** (2004) 903–921
10. Bello, F., Colchester, A.: Measuring global and local spatial correspondence using information theory. In: MICCAI'98. LNCS, 1496 (1998) 964–973
11. Martin-Fernandez, M., Bouix, S., Ungar, L., McCarley, R., Shenton, M.E.: Two methods for validating brain tissue classifiers. In: MICCAI'05. LNCS, 3749, Palm Springs, CA, USA (Oct. 2005) 515–522
12. Collins, D., Zijdenbos, A., Kollokian, V., Sled, J., Kabani, N., Holmes, C., Evans, A.: Design and construction of a realistic digital brain phantom. IEEE Transactions on Medical Imaging **17**(3) (1998) 463–468 http://www.bic.mni.mcgill.ca/brainweb/.
13. Caselles, V., Catte, F., Coll, T., Dibos, F.: A geometric model for active contours. Numerische Mathematik **66** (1993) 1–31
14. Malladi, R., Sethian, J.A., Vemuri, B.C.: Evolutionary fronts for topology independent shape modeling and recovery. In: 3rd ECCV, Stockholm, Sweden (1994) 3–13
15. Bach-Cuadra, M., Cammoun, L., Butz, T., Cuisenaire, O., Thiran, J.: Comparison and validation of tissue modelization and statistical classification methods in t1-weighted mr brain images. IEEE Transactions on Medical Imaging **24**(12) (December 2005) 1548– 1565
16. Cuisenaire, O., Macq, B.: Fast k-nn classification with an optimal k-distance transformation algorithm. Proc. 10th European Signal Processing Conf. (2000) 1365–1368
17. Weickert, J., ter Haar Romery, B., Viergever, M.: Efficient and reliable schemes for nonlinear diffusion filtering. IEEE Transactions on Image Processing **7**(3) (march 1998) 398–410
18. Shu, Y., Bilodeau, G., Cheriet, F.: "segmentation of laparoscopic images: Integrating graph-based segmentation and multistage region merging". In: Second Canadian Conference on Computer and Robot Vision (CRV05). (2005)