RELATIVE ANATOMICAL LOCATION FOR STATISTICAL NON-PARAMETRIC BRAIN TISSUE CLASSIFICATION IN MR IMAGES

Eduard Solanas, Valerie Duay, Olivier Cuisenaire and Jean-Philippe Thiran

Signal Processing Laboratory (LTS)
Swiss Federal Institute of Technology (EPFL)
CH-1015 Lausanne, Switzerland
WWW home page: http://ltswww.epfl.ch/~brain
{eduardo.solanas, valerie.duay, olivier.cuisenaire, jp.thiran}@epfl.ch

ABSTRACT

We propose a statistical non-parametric classification of brain tissues from an MR image based on the voxel intensities and on the relative anatomical location of the different tissues. Classically, the overlap of the tissue probability distribution functions for voxel intensities can be reduced by using multi-component (T1w,T2w,Pd,...) MR images, but at a much higher cost for image acquisition. Instead, we generate an artificial image component as the distance from the edges of the segmented brain. The non-parametric k-Nearest Neighbors rule (k-NN) is used since it requires no a priori on the probability distribution of this distance component. The k-NN rule is also tested using different metrics (Euclidean, weighted Euclidean, Mahalanobis) in the classification space to define what "nearest neighbors" are.

The results are twofold: firstly we show that all metrics perform well in ideal conditions, but that the Mahalanobis (and to some extent the weighted Euclidean) metric is more robust in case of under-training of the classifier. Secondly we show that using the relative anatomical location in combination with the intensity information improves the classification of the tissues.

1. INTRODUCTION

Pattern classification in medical imaging is a challenging task, sometimes almost impossible without intensive manual interaction in pathology cases. In order to cope with this complexity, a promising direction is the introduction of a priori information about the problem to be solved. In this work, our goal is to develop a classification method that will help in the automation of brain MR images analysis by providing a segmentation tool able to classify either normal or abnormal anatomy. Our method combines the information given by the intensity values of the voxels in the image and its relative distance with respect to the brain surface. With this new information we are able to localize the different tissues in the joint histogram even if they overlap intensity values thanks to the discriminative action of the relative distance information. This generates a classification of the data which is more reliable than those methods based only in the intensity spectra.

Different classification techniques can be found in the literature. Parametric techniques like C-Means, K-Means, finite Mixture of Gaussian probabilistic distribution modelization, and others, are optimal for classification when the distribution of the data is known. For brain MRI, it is well known that the voxel intensity histogram corresponds to a mixture of Gaussians except for the background noise which corresponds to a Rayleigh distribution [1]. Once the background is extracted the different tissues can be easily classified by one of the previous methods. In our case, we do not have any *a priori* information about the relative distance probabilistic distribution, which is the main reason to use a classification rule which makes no assumption on the probabilistic distribution of the samples. We decided to use a multiple feature k-Nearest Neighbors (k-NN) classification rule with an appropriate distance metric. Therefore, we tested different distance metrics to find the most reliable one. Tested metrics were: the Euclidean distance, and two others which are sensitive to inter-variable changes, a weighted Euclidean distance and the Mahalanobis distance [2].

2. METHOD

Our algorithm combines the gray level values of the MR images with the Euclidean distance (ED) from each of the voxels to the brain surface to build up a joint histogram to identify the different tissues.

Initialization for the algorithm consists of image acquisition and tissue class prototype selection. Next the ED map is calculated in order to build up each voxel feature vector. Then a classification is performed based upon these features using the k-NN rule . Next sections describe the method and the different metrics that were tested for the classification.

2.1. Fast Euclidean distance map for anisotropic volumes

We applied the method proposed by Saito and Toriwaki [3] in 1994, based on the exact Euclidean metric for an n-dimensional picture to compute the distance from any voxel in the brain to its surface. To achieve this they use a serial composition of one-dimensional filters. Cuisenaire [4] has shown in 1999 that this is the fastest algorithm for 3D images with a typical MRI size, i.e. 256x256xS (where S is the number of slices) or smaller. In terms of complexity, calculating the distance map of an image of NxNxN voxels takes $o(N^3)$ memory access and $o(N^4)$ CPU operations.

The basic idea consists in the following two items:

(1) Minimize the square of the ED instead of the exact distance in the process of transformations.

(2) Implement the transformation by decomposing the procedure into serial execution of the three one-dimensional transformations.

0-7803-6725-1/01/$10.00 ©2001 IEEE

2.2. *k*-NN classification rule

The k-Nearest Neighbors (*k*-NN) rule is a non-parametric technique used for supervised pattern classification. Duda and Hart [5] in 1973, provide an excellent description of the method and its properties though Fix and Hodges [6] in 1951, appear to be the first who made the formulation of this rule. Given a training data set, P, consisting of N prototype patterns (vectors) of dimension D and the corresponding correct classification of each prototype into one of C classes, a pattern v of unknown class is classified as class c if most of the k closest prototype patterns are from class c (Cover and Hart [7], 1967). Distance is measured with a distance metric appropriate to the problem domain.

Fix and Hodges [8] in 1952, also established the consistency of the rule, under the assumption of normal statistics, for sequences such that $k \to \infty$ and $k/N \to 0$. On the other hand, as the rule makes no assumption on the probabilistic distribution its probabilistic error R must be at least as large as the Bayes probability of error R^* (due to the overlap of the distributions). Cover and Hart [7] show that the conditional risk for the 1-NN, under certain statistical assumptions, is $R \leq 2R^*$. The risk for the k-NN is bounded by $(1+1/k)R^*$.

2.3. Possible distance metrics

The general distance metric definition between the vector samples x_i and the prototypes p_j is given by:

$$d_{ij}^2 = (x_i - p_j)^T \Sigma^{-1} (x_i - p_j)$$

Where d_{ij}^2 represents the squared distance; superscript T denotes (vector) transpose; and Σ^{-1} is the inverse of the applied distance metric matrix.

In the case of the Euclidean distance (ED) $\Sigma = Id$, where Id stands for the identity matrix. For the variance weighted Euclidean distance metric (wED) $\Sigma = \Sigma_v$, where Σ_v represents the variance matrix of the prototypes of class c to which p_j belongs. And for the Mahalanobis distance metric (MD) $\Sigma = \Sigma_c$, where Σ_c is the covariance matrix of the prototypes of class c to which p_j belongs.

While ED metric is the same for all classes, wED and MD are metrics which take the sample variability into account, so these metrics depend on which class the sample belongs to. In general, wED and MD are a very useful way of determining the similarity of a set of values from an unknown sample to a set of values measured from a collection of known samples. Instead of treating all values equally when calculating the distance, they weight the differences by the range of variability in the direction of the sample point. This metrics construct a space that weights the variation in the sample along the axis of elongation less than in the shorter axis of the group ellipse. In terms of measurements, sample A will have a substantially smaller distance to the mean than sample B since it lies along the axis of the group that has the largest variability. Therefore, sample A is far more likely to be classified as the same class as the group. One of the main reasons the MD method is used is that it is very sensitive to inter-variable changes in the training data. MD looks at not only variations (variance) between the responses at the same variable,which is the case of the wED, but also at the inter-variable variations (covariance). In other words, the difference between the wED and MD falls in the direction of the axis of the ellipse.While for the wED the axis are parallel to the variables axis, ellipse axis for the MD can have any direction. The Mahalanobis group defines a multidimensional space whose boundaries determine the range of variation that are acceptable for unknown samples to be classified as members.

2.4. Compensated geometrical relative location for classification

2.4.1. Training vector prototypes

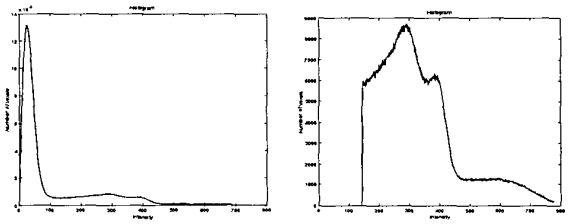
Prototypes are voxels selected by an operator familiar with the anatomy of the subject to be segmented and the operation of the classification process. These voxels are representative of particular tissue classes which are used to model the probability density function of the features of these classes. In the vector prototype selection process we obtain the spectral information of the prototype (the value of the prototype in each of the features) with this information are able to build up a joint histogram which models the probability density function of our classes. Typically the number of prototypes to be selected for each class has to be big (ideally $k \to \infty$) to have a small error R (see 2.2) but much smaller than N such that $k/N \to 0$. In our case we use the voxel intensity and the ED to the surface of the brain values to construct our feature space. Fig. 2 shows the obtained joint probability distribution function for the different classes. As we can see classes are well separated, even if their intensities overlap one another.

Concerning the time of training prototypes acquisition, it takes about one minute for an experienced operator.

2.4.2. Pre-processing

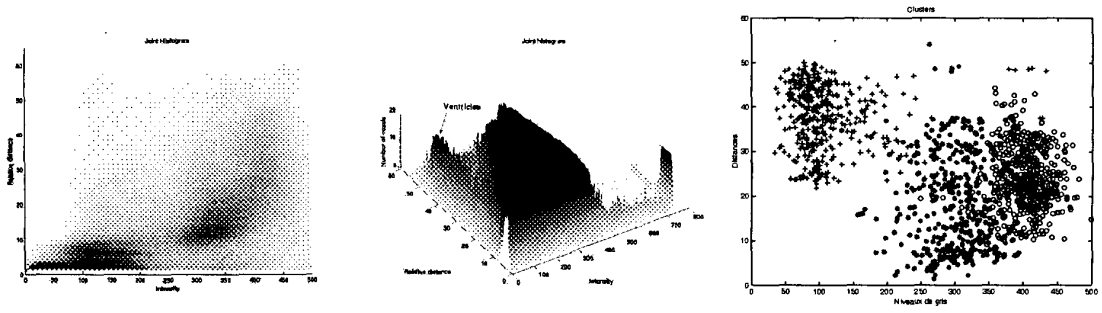
Prior to tissue classification we apply a procedure to automatically segment the brain from the rest of the head, based on histogram fitting and morphological operations [9]. Fig. 1 shows a typical histogram. The first maximum corresponds to the background which is fitted by a Rayleigh curve [1]. We first estimate the background threshold and compute a mask of the head within the VOI. Next we apply a series of mathematical morphology operations to obtain a coarse segmentation of the brain. These operations use the sphere as structuring element implemented applying distance transformations [4]. This pre-processing step needed to construct the relative distance map to the brain surface takes about 20 seconds depending on the number of operations.

Apart from the described pre-processing the MR data was not processed with any filter to try to improve the *SNR* or removing the bias field inhomogeneities that appear in the tissue areas due to the effect of equipment limitations and patient induced electrodynamic interactions.



(a) (b)

Fig. 1. Typical MR histogram before (a) and after (b) the pre-processing.



(a) (b) (c)

Fig. 2. Joint histogram (a), ventricle's region zoomed histogram (b) and training prototypes for classification (c).

2.4.3. Tissue classification

In this section we describe the method we propose for tissue classification after the pre-processing step.

Once we have applied our mask to the input volume we calculate the ED from each of the voxels to the border of the mask, which is approximately the same as calculating the distance to the surface of the brain, depending on the accuracy of our pre-processing step. Then we are able to extract the feature vector values of our training prototypes (see fig. 2). Our vector components consist of the intensity and relative distance values, which, as we will see, are reliable to achieve a good classification. Next, we have chosen the k-NN rule for voxel classification [10]. In next section we see some of the results we have obtained with this method.

3. RESULTS

In this section the application of relative anatomical location to segmentation problems involving normal anatomy is discussed. A visual and analytic comparison were done. In order to test the behavior of our method we compared it with the k-NN rule using only intensity values. An evaluation of different metrics such as ED, a weighted ED (wED) and the MD was also carried out (see sec. 2.3).

Three different sets of T1 weighted MR images of dimensions 256x256x128 were used to carry out the tests for normal anatomy. Their pixel size was 1 mm. and the distance between slices 1.25 mm.

The error ratio of the method was calculated using the method to classify the prototypes. All prototypes are used as training data except the one to be classified. The error ratio is defined as the percentage of mis-classified samples.

When observing the error ratio in the fig.4 for a given curve we notice a jagged behavior. This results from the fact that most classification decisions are the result of a vote of the k nearest samples, choosing between two classes. In such a dual vote, an even number of voters make a decision of lesser quality than the same number minus 1.

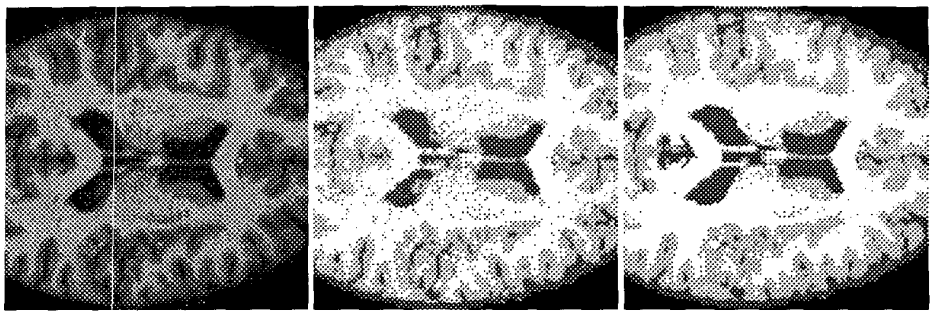
When analyzing the error probability distribution (Fig. 4) we notice that there are two groups of curves; one corresponding to the tests where the relative anatomical location was not taken into account, an the other for the tests using the relative distance information in combination with the intensity. For the first group the error probability is much higher than for the second one. As shown in fig. 2 this is caused by the overlapping intensity distributions of the gray and white matter, which is not the case when combining relative distance location and intensity values.

It is important to remark the behavior of the ED metric when combining the relative distance and intensity informations for different N. We notice that it has the same behavior as the wED and MD until a certain k is reach, which is different for N=*300* or N=500, then the error increases. This is due to the fact that the condition $k/N \rightarrow 0$ is not valid anymore. This shows that the ED is valid while the measure is local. This is not the case of the wED and MD which are global metrics.

In our tests, the WED and MD have a similar behavior. It can be seen in 2 that the axis of the ellipses that model the clusters are parallel to the variables axis, which is the reason for the similar behavior of both metrics. In the case where the axis of the clusters were not parallel to the variables axis then the MD is likely to have a better performance.

4. CONCLUSIONS

We presented a robust and simple classification method for single channel MR which makes use of a new feature to achieve a good classification: the relative distance to the brain surface. Applying the k-NN rule to the *(intensity,distance)* vector yields a good discrimination of the different tissue classes. Since we did not use any anisotropic or bias inhomogeneities corrector filtering the obtained results could be improved by adding these filters in the pre-processing step. This will improve the *SNR* and help the samples to be closer between those belonging to the same cluster. We could also apply some post-processing (if needed) adding some spatial constraints to our classification in case any sample was misclassified. Due to the good discrimination between classes that we ob-



(a) (b) (c)

Fig. 3. Original image (a), classification for k=20 using only intensity values and the Euclidean metric (b) classification for k=20 using intensity and relative distance values and the Mahalanobis metric (c).

tained applying this method, we are currently doing some research for the localization of multiple sclerosis pathology.

5. REFERENCES

[1] M. E. Brummer, R. M. Mersereau, R. L. Eisner, and R. R. J. Lewine, "Automatic detection of brain contours in mri data sets," *IEEE Trans. on Medical Imaging*, vol. 12, pp. 153–166, 1993.

[2] L. Devroye, L. Gyorfi, and G. Lugosi, *A probabilistic theory of pattern recognition*, Springer Verlag, NY, 1996.

[3] T. Saito and J. I. Toriwaki, "New algorithms for euclidean distance transformation of an n-dimendional digitized picture with applications," *Pattern Recognition*, vol. 27, no. 11, pp. 1551–1565, 1994.

[4] O. Cuisenaire, *Distance tranformations: Fast algorithms and applications to medical image processing*, Ph.D. thesis, Catholic University of Louvain (UCL), 1999.

[5] R. O. Duda and P. E. Hart, *Pattern classification and scene analysis*, Wiley, New York, 1973.

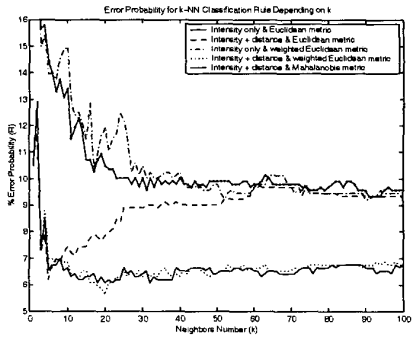
[6] E. Fix and J. L. Hodges, "Discriminatory analysis, nonparametric discrimination," Tech. Rep. Tex. Project 21-49-004, Rept. 4, Contract AF41(128)-11, USAF School of Aviation Medicine, Randolf Field, Tex, 1951.

[7] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.

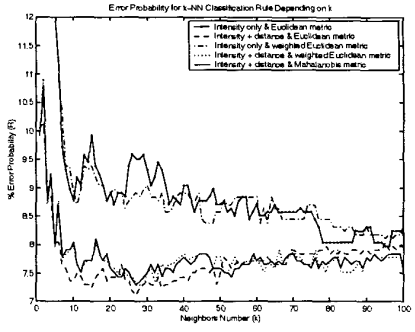
[8] E. Fix and J. L. Hodges, "Discriminatory analysis, small sample performance.," Tech. Rep. Tex. Project 21-49-004, Rept. 11, USAF School os Aviation Medicine, Randolf Field, 1952.

[9] P. Schroeter, *Unsupervised two-dimensional and three-dimensional image segmentation*, Ph.D. thesis, Swiss Federal Institute of Technology (EPFL), 1996.

[10] O. Cuisenaire and B. Macq, "Fast k-nn classification with an optimal k-distance transformation algorithm," *Proc. of the 10th European Signal Processing Conference*, pp. 1365–1368, 2000.



(a) N=300



(b) N=500

Fig. 4. Distribution of the error probability depending on k and the distance metric for N=300 (a) and N=500 (b); where N is the number of prototypes per class.