34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Brain anatomical structure segmentation by adaptive bandwidth density estimation

Guadalupe Desirée López Palafox[1] and Juan Ramón Jimenéz Alaníz[1]

***Abstract*— Determination of region in a space of multimodal features of brain MR images requires kernel estimation tecniques with bandwidths that are adapted locally. The bandwidth selection is a critical aspect at the filtering stage of image segmentation. This work presents two methods for determinate the adaptive bandwidth in the application of density estimation, in the segmentation of regions at the feature space of an MRI. Two adaptive methods: sample point and k-nearest neighbors, where applied for real and synthetic data, achieved similarity indexes of 0.68 and 0.71 for gray matter and white matter respectively.**

I. INTRODUCTION

Image segmentation is a very important step in image processing. It plays a crucial role in extraction of useful information and attributes from images for all medical imaging applications. It is one of the important steps leading to image understanding, analysis and interpretation. The principal goal of image segmentation is to partition an image into regions (or classes) that are homogeneous with respect to one or more characteristics or features under certain criteria [1].

Automated MRI segmentation classify brain voxels into one of three main tissue types: gray matter (GM), white matter (WM) and cerebro-spinal fluid (CSF) [1]; the segmentation of structures of MRI is applied in the study of many disorders, such as multiple sclerosis, schizophrenia, epilepsy, Parkinson's disease, Alzheimer's disease, cerebral atrophy, etc [2]. Additionally, MRI segmentation is important in image processing to identify anatomical areas of interest for diagnosis, treatment, surgical planning, image registration and functional mapping [1]. For this task is been used the supervised approaches, where intensity values of labeled voxel samples from each tissue must be provided during the learning phase. In a subsequent classification phase, the unlabeled voxels are classified using a selected classifier. This method requires human interaction to select the samples and is therefore semi-automatic [1].

Statistical segmentation techniques define a parametric model representing the tissue, assuming particular distribution forms on the selected feature space; this assumption can introduce artefacts implied by the density model choice. The use of nonparametric approach is to let the data guide a search for the function which fits them best without the restrictions imposed by the parametric model [3]. A clustering technique which does not requires prior information of the numbers of clusters, and does not restricts their shape of the density distribution is the Mean Shift (MS) procedure. This procedure is an iterative technique which estimates the modes of the multivariate distribution underlying the feature space; the number of cluster is obtained automatically by finding the centers of the densest regions in the space [4]. However one of the limitations of mean shift is the specification of parameter named the kernel width or bandwidth. In this paper it is shown that using an adaptive bandwidth in the MS obtained in two different ways, improve the results in the filtering stage in the segmentation process. The adaptive bandwidth was applied to real and synthetic data of brain MRI and the results achieved show that the adaptive bandwidth is better than fixed bandwidth both in real and synthetic data.

The paper is organized as follows. In Section 1 is introduced the fixed bandwidth, the limitations of the fixed bandwidth are reviewed in Section 2. Section 3 presents the two types of adaptive bandwidth used. In Section 4 the results obtained will be shown with the adaptive bandwidth for real and synthetic data. Finally, conclusions are presented in Section 5.

II. FIXED BANDWIDTH DENSITY ESTIMATION

The multivariate kernel density estimation with kernel K and window width h is defined by [3], [5]:

$$\hat{f}(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \tag{1}$$

The kernel function $K(x)$ is a function defined for d-dimensional vectors X_i, $i = 1, \ldots, n$ that are the given multivariate data set whose underlying density f is unknown and must be estimated. The kernel is taken to be a radially symmetric, non-negative function, centered at zero and integrating to one, an example, the multivariate Epanechnikov that is an optimal kernel for minimize the error at the density estimation function [5]:

$$K_e(x) = \begin{cases} \frac{1}{2}c_d^{-1}(d+2)(1-x^T x) & if\ x^T x < 1 \\ 0 & otherwise \end{cases} \tag{2}$$

[1]G.D. López-Palafox is with the Neuroimaging Laboratory, Department of Electrical Engineering, Universidad Autónoma Metropolitana-Iztapalapa, Av. San Rafael Atlixco 186, Col. Vicentina, D.F., 09340, México (e-mail:desslopa@gmail.com)

[1]J.R. Jiménez-Alaníz is with the Neuroimaging Laboratory, Department of Electrical Engineering, Universidad Autónoma Metropolitana-Iztapalapa, Av. San Rafael Atlixco 186, Col. Vicentina, D.F., 09340, México (e-mail:jajr@xanum.uam.mx)

where C_d is the volume of the unit d-dimensional sphere. Substituting (2) in (1) and after some algebraic manipulations, it is obtained what is know as mean shift (MS):

$$M_h(x) = \frac{1}{n_x} \sum_{X_i \in S_h(x)} X_i - x \tag{3}$$

where the size of the region $S_h(x)$ is a function of the bandwidth h and n_x is the number of observations X_i falling within $S_h(x)$. The MS vector can be improved weighting each pixel within a region by a confidence edge, such that the pixels which are situated near one edge less influence on the determination of the new cluster center. The equation (3) modified by the inclusion of the weighting of the edge confidence is set as:

$$M_h(x) = \frac{1}{\sum(1-\phi_i)} \sum_{X_i \in S_h(x)} (1-\phi_i)\, X_i - x \tag{4}$$

where ϕ_i is the edge confidence associated to X_i. The terminology fixed bandwidth is due to the fact that h is held constant across $x \in R^d$. The fixed bandwidth procedure estimates the density by taking the average of identically scaled kernels.

The most widely used way of placing a measure on the global accuracy of $\hat{f}$ as an estimator of f is the mean integrated square error (MISE). The MISE allows to observe that one of the fundamental problems of density estimation includes the bias and variance. As the bias is proportional to h^4, so for this amount decrease, is necessary that the value of h should be small. However, if h is small, means an increase in variance, since the latter is proportional to $(nh)^{-1}$. This is known as the trade-off between bias and variance, and it is a mathematical quantification of the critical role played by the bandwidth in the estimation of the density function.

III. ADAPTIVE BANDWIDTH

A. *Variable bandwidth by sampling point method*

For multivariate data, a correct choice of bandwidth is a very complex problem,- One of the most used method for local bandwidth adaptation, consider the bandwidth proportional to the inverse of the square root of a first approximation of the local density. The bandwidth h can varied on each data point $h = h(X_i)$. For each point X_i, can be obtained the sample point of density estimator [3]:

$$\hat{f}_{sp} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h(X_i)^d} K\left(\frac{x - X_i}{h(X_i)}\right) \tag{5}$$

for which the estimate of f at x is the average of differently scaled kernels centered at each data point. The sample point estimators are themselves densities, being non-negative and integrating to one. The bias becomes proportional to h^4, while the variance remains unchanged, $h_(X_i)$ is taken as:

$$h(X_i) = h_0 \left[\frac{\lambda}{f(X_i)}\right]^{\frac{1}{2}} \tag{6}$$

where h_0 represents the fixed bandwidth and λ is a proportional constant. It is important to say that $f(X_i)$ is unknown and must be estimated (called pilot) from the data in first stage, and is denoted by $\tilde{f}$. The strategy used for this sample point method is:

1) Find the pilot estimate $\tilde{f}(X_i)$ that satisfies $\tilde{f}(X_i) > 0$ for all i.
2) Define the factor bandwidth λ by:

$$log\lambda = n^{-1} \sum log\tilde{f}(X_i) \tag{7}$$

In the first step, to build the pilot estimator can be used the mean shift procedure with fixed bandwidth.

B. *Variable bandwidth by k-nearest neighbors*

Another method that define the bandwidth used by the pilot density estimation is the nearest neighbors [6], [3] [7]. The nearest neighbor rule relies on a metric function between patterns. A metric, is a function that provides a generalized distance between two patterns. The Minkowski metric is one general class of metrics for d-dimensional patterns, where the number of k-neighbors must be chosen large enough to ensure that there is an increased density where all the kernels have bandwidths h_i. Be $X_{i,k}$, the k-nearest neighbor to point X_i so is taken:

$$h_i = \| X_i - X_{i,k} \|_1 \tag{8}$$

where L_1 norm is used [6].

Once completed the filtering image data applying the MS procedure using the variable bandwidth method: sampling point and *k*-nearest neighbors; it is obtained an image with intensity homogeneous regions, each of them corresponding to one mode found. After filtering, the next stage is the fusion of regions, where regions with some similarity are joined with the aim of decreasing the over-segmentation. A pruning phase is used to eliminate those very small regions that by themselves can not be a brain structure. The last stage is the classification of regions obtained after the pruning stage, so we use a priori information contained in probabilistic maps of white matter, gray matter and cerebrospinal fluid [8] for the image segmentation in these three classes.

The segmentation methodology applied can be summarized as follows:

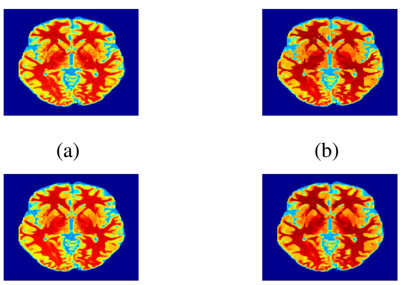
1) Estimation of the Edge Confidence Map from the data.
2) Filtering of the data by the process of MS using both methods of adaptive bandwidth.
3) Fusion of regions through Region Adjacency Analysis and Pruning of regions.
4) Region classification achieved by the a posteriori probability calculated for each probabilistic map.

IV. RESULTS

The performance of the segmentation process by mean shift was evaluated with real and synthetic images. This images were taken from the Internet Brain Segmentation Repository [10] and the BrainWeb: Simulated Brain Database (SBD) [9], respectively. The synthetic data corresponds to a brain digital phantom of T1 weighted images with a $181 \times 217 \times 181$ size, 1 mm^3 voxel resolution, 3% noise, and 20% intensity inhomogeneity. For this volume the ground truth was available for quantitative comparison. The real data consist of 20 normal brain MR of T1 weighted data sets and their corresponding manual segmentations provided by the Center for Morphometric Analysis at Massachusetts General Hospital and available at [10]. The voxel resolution is 1 mm axial axes, 3 mm coronal axes and 1 mm sagittal axes. For the classification stage was used a priori information contained in probabilistic maps available in the software SPM [8].

A. Synthetic Data

To process the data with the fixed bandwidth is used a bandwidth of spatial and intensity $h_s = 6$ and $h_i = 9$ respectively. Figure 1 shows the results obtained at the filtering stage using the three types of bandwidth (fixed, sample point and k-nearest neighbors). The fixed bandwidth is consider the initial estimation of the sample point bandwidth $h(X_i)$, that is used in the information processing with adaptive bandwidth and whose result is shown in Figure 1 at the bottom left. For the method of k-nearest neighbors, which is the third form of data processing, we used a $k = 200$, and the result obtained is shown in the lower right corner of Figure 1.



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

Fig. 1: a) Synthetic data slice 70 at the stage of MS filter. b) Filter image using fixed bandwidth with $h_s = 6$ and $h_i = 9$. c) Filter image using sample point bandwidth and d) Filter image using nearest neighbors bandwidth for $k = 200$

The Figure 2 shows the filtered image segmentation by classifying each of the regions using the prior information of the probabilistic maps. The classified images can be seen more easily in homogeneous intensity regions, that in the original image, because the MS has removed image noise, resulting in homogeneous intensity regions, and subsequently refining the probabilistic classification. The segmented image, using the two types of variable bandwidth, are very similar quantitatively, as shown Tanimoto indexes of Table I, but as can be seen from this table, the segmentation using the estimated bandwidth from sample point has a small improvement in the rates of white and gray matter compared to the estimation method for nearest neighbor.

TABLE I: Tanimoto Coefficient for Synthetic Data

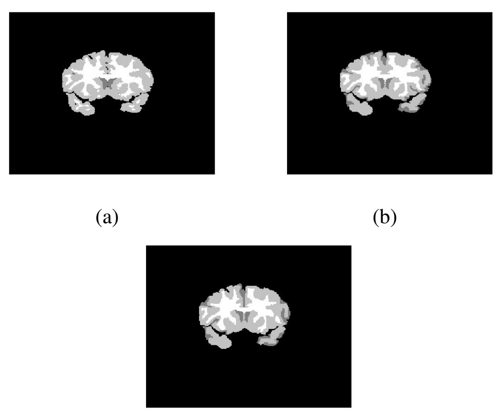
Method	Background	CSF	Gray	White
Fixed Bandwidth	0.9976	0.4726	0.5681	0.4330
Sample Point	0.9979	0.4454	0.6958	0.5760
k-Nearest Neighbor	0.9956	0.4646	0.6731	0.5595

(a) (b) (c)

Fig. 2: a) Synthetic data slice 70, (b)Segmented image using sample point and (c)Segmented image using k-nearest neighbor

B. Real Data

The same procedure was applied to real data, by processing the 20 data sets [10]. This 20 volumes have the expert manual segmentation and this segmentation is considered the reference to be used to compare our procedures: fixed bandwidth (Fixed), variable bandwidth by sample point (S. Point) and variable bandwidth by k-nearest neighbor (k-nearest) with the methods available at the IBSR, this methods are: adaptive MAP (AMP), biased MAP (BMAP), fuzzy c-means (FUZZY), maximun aposteriori probability (MAP) and maximum likelihood (MLC). We calculate the Tanimoto average index for each class, for all real data and for the previously mentioned methods, and this values are shown on Table II. For the case of the background is not shown the graphic because it is very similar for all data. Figure (3) shows the reference image and the images processed with adaptive bandwidth can be noted a high qualitative similarity between them. On Figure 4, 5, 6 are plot the Tanimoto coefficient, and we observed that Figure 4 has better results using adaptive mean shift than the other method reported on [10] and [1]. The coefficients obtained with fixed bandwidth are minor compared to those obtained with variable bandwidth. Brain studies are plotted in order of difficulty to segment, being the number 1 the most difficult and least difficult the number 20.



(a) (b) (c)

Fig. 3: a) Real data expert segmentation slice 37, (b)Segmented image using sample point and (c)Segmented image using k-nearest neighbor

V. CONCLUSIONS

The results of image segmentation using a statistical estimation technique, are influenced primarily by the bandwidth

TABLE II: Average Tanimoto Coefficient for Real Data

Method	Background	CSF	Gray	White
Fixed Bandwidth	0.9857	0.5126	0.4420	0.5014
Sample Point	0.9918	0.6363	0.6832	0.7105
k-Nearest Neighbor	0.9903	0.6362	0.6200	0.6560
AMP	0.999	0.069	0.564	0.567
BMAP	0.999	0.071	0.558	0.562
Fuzzy	0.999	0.048	0.473	0.567
MAP	0.999	0.071	0.550	0.554
MLC	0.999	0.062	0.477	0.571

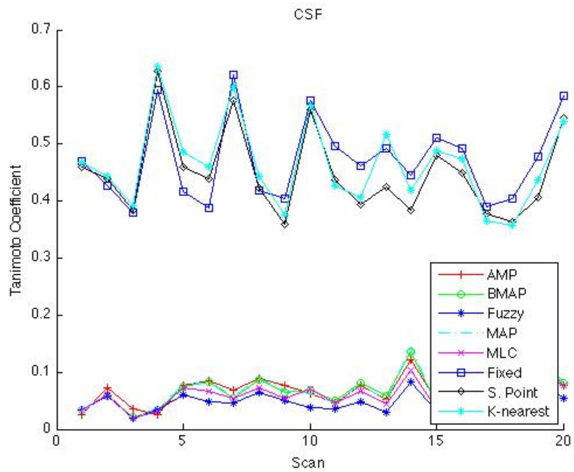


Fig. 4: Tanimoto coefficient for CSF for real data segmentation using the fixed and adaptive bandwidth

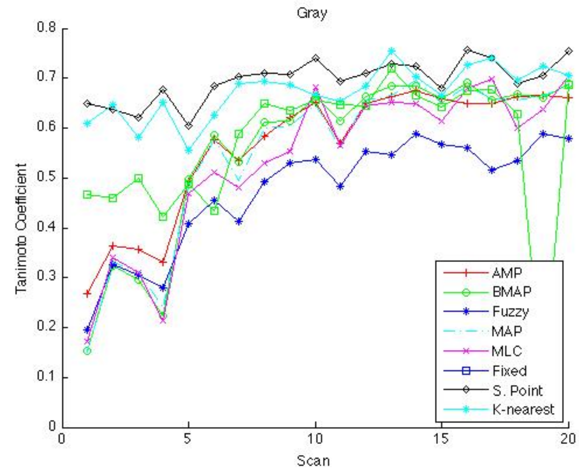


Fig. 5: Tanimoto coefficient for Gray Matter for real data segmentation using the fixed and adaptive bandwidth

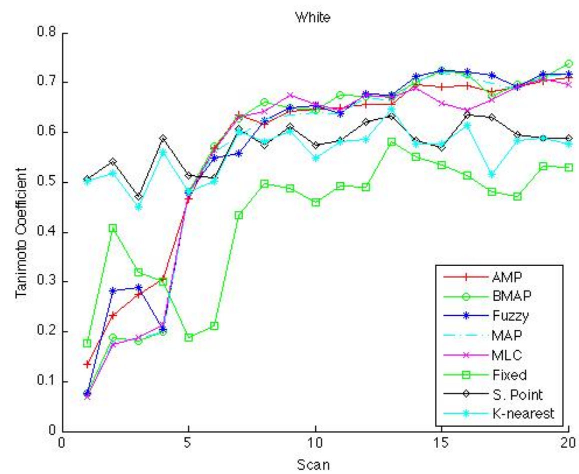


Fig. 6: Tanimoto coefficient for White Matter for real data segmentation using the fixed and adaptive bandwidth

of the kernel, because the quality depends strongly on the bandwidth used. We can verify that selection of fixed bandwidth in the estimation of the pilot function for sample point is not relevant [2]. In the classification stage can be seen that both methods have good classification rates and the variation between them is minimal, so can be selected any of both methods, one disadvantage of the *k*-nearest neighbors method is the run time, that is much higher compared to the sample point method. Using a priori information improves the classification results of both synthetic and real data, also we found that the CSF classification is better than results presented in [1], since they do not use a priori information. We conclude that using both apriori information and adaptive bandwidth, sample point or *k*-nearest neighbors, improves the segmentation.

VI. ACKNOWLEDGEMENTS

This work was financially supported by CONACyT Grant number 288132.

REFERENCES

[1] A.Mayer and H. Greenspan. An adaptive mean shift frame work for mri brain segmentation. *International Symposium Biomedical Imaging*, pages 319–322, 2006.

[2] J.Jimenez Alaniz, V. Medina Banuelos, and O. Yanez. Data drivem brain mri segmentation supported on edge confidence and a priori tissue information. *IEEE Transactions on Medical Imaging*, 25(1):74–83, 2006.

[3] J.Jimenez Alaniz, M.Pohl Alfaro, V. Medina Banuelos, and O. Yanez. Segmenting brain mri using adaptive mean shift. *Proceedings of the 28th IEEE EMBS Annual International Conference*, pages 3114–3117, 2006.

[4] D. Comaniciu and P.Meer. Mean shift analysis and applications. *IEEE Int. Conf. Computer Vision (ICCV'99)*, pages 1197–1203, 1999.

[5] B.W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall / CRC, 1998.

[6] I. Shimshoni, B. Georgescu, and P. Meer. Mean shift based clustering in high dimensions: A texture classification example. *Proceeding of the Ninth IEEE International Conference on Computer Vision*, pages 456–463, 2003.

[7] R.O. Duda, P.E. Hart, and D.G.Stork. *Pattern Classification*. John Wiley and Sons, 2 edition, 2001.

[8] J. Ashburner, K. Friston, A. Holmes, and J.B. Poline. Statistical parametric mapping spm99b. `http://www.fil.ion.ucl.ac.uk/spm/`, December 2011.

[9] Centre for Morphometric Analysis. Internet brain segmentation repository. `http://www.cma.mgh.harvard.edu/ibsr/`, December 2011.

[10] McConnell Brain Imaging Centre. Brain web: Simulated brain database. `http://www.bic.mni.mcgill.ca/brainweb/`, December 2011.