

A Novel Bayesian Approach to Adaptive Mean Shift Segmentation of Brain Images

Qaiser Mahmood, Artur Chodorowski, Andrew Mehnert, Mikael Persson

Department of Signals and Systems, MedTech West, Chalmers University of Technology, Sweden
{qaiserm, artur.chodorowski, andrew.mehnert, mikael.persson}@chalmers.se

Abstract

We present a novel adaptive mean shift (AMS) algorithm for the segmentation of tissues in magnetic resonance (MR) brain images. In particular we introduce a novel Bayesian approach for the estimation of the adaptive kernel bandwidth and investigate its impact on segmentation accuracy. We studied the three class problem where the brain tissues are segmented into white matter, gray matter and cerebrospinal fluid. The segmentation experiments were performed on both multi-modal simulated and real patient T1-weighted MR volumes with different noise characteristics and spatial inhomogeneities. The performance of the algorithm was evaluated relative to several competing methods using real and synthetic data. Our results demonstrate the efficacy of the proposed algorithm and that it can outperform competing methods, especially when the noise and spatial intensity inhomogeneities are high.

1. Introduction

Automated brain segmentation is an active area of research in the field of medical imaging. The resulting segmentation yields a patient-specific demarcation of individual tissues/structures. This permits not only the quantitative characterization of these tissues (e.g. in the study of Alzheimer's disease and multiple sclerosis), but also the construction of patient specific models of tissue conductivity (e.g. for application in EEG source localization in epilepsy patients) or biomechanical properties (e.g. for application in surgical simulation). The majority of methods that have been proposed for automated segmentation of brain tissues are based on statistical parametric models. The MPM-MAP (Maximizer of the posterior marginals- Maximum *a posteriori*) algorithm [1] exemplifies this approach. It implements Bayesian segmentation based on non-rigid registration of the atlas. The algorithm uses Expectation-Maximization (EM) for estimation of

model parameters and Hidden Markov Random Fields (HMRF) for spatial coherences. Two other examples are KVL (K. Van Leemput) [2] and CGMM (Constrained Gaussian mixture model) [3] which use maximum *a posteriori* probability (MAP) or maximum likelihood (ML) method for the estimation of model parameters. A drawback with these approaches is that it is difficult to integrate pixel spatial information with multi-valued pixel information (e.g. when several different MR scans have been acquired). This is because the HMRF is itself hard to implement in high dimensional feature space.

Mean-shift (MS) segmentation overcomes this drawback. Mean-shift [4, 5] is a non-parametric technique used to estimate the modes of the multivariate distribution underlying a feature space. It does not require any prior information to initialize the position of the clusters and also does not constraint the shape of the clusters. Mean-shift segmentation involves concatenating both the spatial and range domains of an image and identifying modes in this multidimensional joint spatial-range feature space. The only free parameter is the kernel size (called the bandwidth). If the chosen value is too small then insignificant modes are detected (over-clustering). If it is too large then significant modes can be merged (under-clustering). Several methods [6] are available for the estimation of a single fixed bandwidth. However, the use of a single fixed bandwidth has the drawback that it can yield under- or over-clustering when the feature space has significantly different local characteristics across the space. Variable or adaptive bandwidth methods have been proposed [7] to overcome this drawback. Such methods involve determining the bandwidth value for each feature point by using the pilot density estimate. In [8] it was shown that adaptive mean-shift clustering, in which the pilot density estimate is based on the nearest neighborhood, performs better than a single fixed bandwidth in high dimensional space. Nevertheless the bandwidth value defined in terms of the Euclidean distance to the k th farthest feature point

from the kernel center can be biased by outliers [9]. Herein, we propose an adaptive approach that does not have this drawback. It is based on a novel variation on the Bayesian approach initially proposed in [9] for the estimation of the global fixed kernel bandwidth. The approach includes modeling the distributions of variances of k -nearest neighbor (kNN) features and fitting Gamma distribution functions to these. It is suitable when there is no knowledge of the underlying feature distribution. In our variation we used this idea locally for adaptive bandwidth estimation. However, we adopt the adaptive mean-shift frame work proposed in [10] by integrating the Bayesian based adaptive bandwidth in the mean-shift algorithm for the segmentation of brain into three tissues; white matter (WM), gray matter (GM) and cerebrospinal fluid (CSF) and evaluate its impact on the final segmentation results.

The rest of the paper is organized as follows. In Section 2 we summarize the adaptive mean-shift algorithm. In Section 3 we present the Bayesian approach for adaptive bandwidth estimation. The proposed segmentation framework is presented in Section 4, and the experimental results on simulated and real data are presented in Section 5. Finally, the conclusions are drawn in Section 6.

2. Adaptive mean shift segmentation

The set of feature vectors in d - dimensional space is defined as $\mathbf{x}_i \in R^d, i = 1, \dots, n$. The estimation of the data density by the Parzen window Kernel density estimator at point \mathbf{x} is given by

$$\hat{f}_K(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_i^d} k\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h_i}\right\|^2\right) \quad (1)$$

where $k(\mathbf{x})$ is the kernel profile of the kernel K, n is the number of the data points in d -dimensional space and $h(\mathbf{x}_i) \equiv h_i$ is the bandwidth to each data point \mathbf{x}_i known as the adaptive bandwidth [7]. In our proposed mean shift algorithm, the bandwidth is estimated using the Bayesian approach explained in Section 3 and the Epanechnikov kernel K is used to compute the density of each feature point \mathbf{x}_i . The main property of mean shift that provides the basis of brain image clustering is the mean shift vector. It is computed by taking the derivative of equation (1) and is expressed as

$$m_G(\mathbf{x}) = C \frac{\nabla \hat{f}_K(\mathbf{x})}{\hat{f}_G(\mathbf{x})} = \frac{\sum_{i=1}^n \frac{1}{h_i^{d+2}} \mathbf{x}_i g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h_i}\right\|^2\right)}{\sum_{i=1}^n \frac{1}{h_i^{d+2}} g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h_i}\right\|^2\right)} - \mathbf{x} \quad (2)$$

where $C > 0$ is a constant, $g(\mathbf{x}) = -k'(\mathbf{x})$ and $m_G(\mathbf{x})$ is the mean shift vector which is

proportional to the normalized gradient of the kernel density estimator of the data. It points from the data sample \mathbf{y}_j towards the data sample \mathbf{y}_{j+1} defined

$$\mathbf{y}_{j+1} = \frac{\sum_{i=1}^n \frac{1}{h_i^{d+2}} \mathbf{x}_i g\left(\left\|\frac{\mathbf{y}_j-\mathbf{x}_i}{h_i}\right\|^2\right)}{\sum_{i=1}^n \frac{1}{h_i^{d+2}} g\left(\left\|\frac{\mathbf{y}_j-\mathbf{x}_i}{h_i}\right\|^2\right)}, j = 1, 2 \dots \quad (3)$$

which lies in a higher density region. The mean-shift vector iteratively climbs the higher density region until the stationary data point is reached. This point, called the convergence or mode, is where the gradient of the kernel estimator becomes zero. Another important property of mean shift is that it facilitates clustering in a joint spatial-range domain. This is attractive for the brain segmentation problem where we have multiple values for each voxel. The joint-domain kernel $K_{h_s, h_{r_i}}$, which is the product of spatial and range domain symmetric kernels, is given by

$$K_{h_s, h_{r_i}}(\mathbf{x}) = \frac{c}{h_s^p h_{r_i}^d} k\left(\left\|\frac{\mathbf{x}^s}{h_s}\right\|^2\right) k\left(\left\|\frac{\mathbf{x}^r}{h_{r_i}}\right\|^2\right) \quad (4)$$

where h_s^p and $h_{r_i}^d$ are the p -dimensional spatial and d -dimensional range domain kernel bandwidths. In our case, the spatial bandwidth h_s is fixed but the range bandwidth h_{r_i} can vary depending on the local data; i.e. it is adaptive.

A drawback of the mean shift algorithm is that as the dimensionality of the feature space increases so too does the computational complexity. In particular the computation of all neighborhood queries in each iteration of equation (3) is prohibitive. In order to overcome this problem, we implemented fast adaptive mean-shift (FAMS) [8] using the kd-tree structure [11] to efficiently compute the k -nearest neighborhood queries ($k < n$) in equation (3).

3. Bayesian adaptive bandwidth

In our adaptive MS algorithm we compute the adaptive bandwidth by incorporating the Bayesian approach proposed in [9]. The bandwidth is modeled by a *posteriori* probability density function $P(s|\mathbf{x})$ of local data spread s , where \mathbf{x} is a data point. If we let $K < n$ be the number of nearest neighborhoods to a data sample \mathbf{x}_i then we can arrange all the other data samples in ascending order with respect to Euclidean distance to the data point \mathbf{x}_i thus

$$\|\mathbf{x}_{i,(1)} - \mathbf{x}_i\| < \dots < \|\mathbf{x}_{i,(K)} - \mathbf{x}_i\| < \dots < \|\mathbf{x}_{i,(n-1)} - \mathbf{x}_i\| \quad (5)$$

where $\mathbf{x}_{i,(j)} \neq \mathbf{x}_i$ and $j = 1, \dots, n - 1$. The probability density function of the bandwidth is then defined

$$P(s|\mathbf{x}) = \prod_{j=1}^K P\left(s \mid \mathbf{x}_{K_j}\right) \quad (6)$$

where $P(s|\mathbf{x}_{K_j})$ is the probability of local data spread s based on the K_j nearest neighbor data samples to \mathbf{x}_{K_j} and $\{K_j, j = 1, \dots, N\}$ is the set of N neighborhoods of various sizes. The evaluation of probabilities over the entire range of K_j is expressed as

$$P(s|\mathbf{x}_{K_j}) = \int P(s|K_j, \mathbf{x}_{K_j}) P(K_j|\mathbf{x}_{K_j}) dK_j \quad (7)$$

and by using the Bayesian rule, it leads to

$$P(K_j|\mathbf{x}_{K_j}) = \frac{P(\mathbf{x}_{K_j}|K_j)P(K_j)}{P(\mathbf{x}_{K_j})} \quad (8)$$

where $P(\mathbf{x}_{K_j}|K_j)$ represents the probability of the data sample population depending on the K_j nearest neighborhoods and $P(K_j)$ is considered to have a uniform distribution which is bounded as $K_j \in [K_1, K_2]$, $j = 1, 2, \dots, N$ and the number of nearest neighbors K_j is defined by

$$K_j = K_1 + j \frac{K_2 - K_1}{N} \quad (9)$$

The local variance s is given by the following expression

$$s_j = \frac{\sum_{l=1}^{K_j} \|\mathbf{x}_{(l)} - \mathbf{x}_i\|^2}{K_j - 1}, i = 1, 2, \dots, n, j = 1, 2, \dots, N \quad (10)$$

where $\mathbf{x}_{(l)}$ are the nearest neighbors to the feature point \mathbf{x}_i . The distribution of variances is modeled as Gamma distribution which is a generalization of the chi-square distribution and is more convenient to model the bandwidth distribution for the kernels of non-Gaussian distribution. The Gamma distribution is defined as

$$P(s|\alpha, \beta) = \frac{s^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} e^{-\frac{s}{\beta}}, s \geq 0, \alpha, \beta > 0 \quad (11)$$

where α and β defines the shape and the scale of the Gamma distribution respectively and these parameters are estimated using the maximum likelihood method. The adaptive bandwidth is defined to be the mean of the Gamma distribution and is given by

$$\hat{h}(\mathbf{x}_i) = \hat{\alpha} \hat{\beta}, i = 1, 2, \dots, n \quad (12)$$

4. Segmentation framework

The proposed algorithm is a variation on the algorithm presented in [10]. In particular we use the Bayesian estimation of the adaptive bandwidth instead of estimating it based on the distance to the k -nearest neighbours. The algorithm can be applied to both single-valued and multi-valued MR (from multi-modal MR acquisitions) volumes for the segmentation of three tissues (WM, GM and CSF) in the brain. Before applying the algorithm, several pre-processing steps are required: extraction of the brain by using the BET tool in FSL [12]; correction of spatial inhomogeneities using the non-parametric model in the SPM tool [13]

(with a minimising function based on the entropy of the image intensity histogram); and normalization of intensity values to ensure the same dynamic range values 0 to 1 for all the multi-modal MR data. The proposed segmentation algorithm comprises the following steps:

1. The joint spatial-range bandwidth $h_i = [h_{r_i}, h_s]$ is determined in which the range bandwidth h_{r_i} for each feature point \mathbf{x}_i is computed by applying the Bayesian approach described in Section 3 and the size of the spatial bandwidth is defined to be $h_s = 5$.
2. Adaptive mean shift is applied to find the convergence of all data points by iterating equation (3) and the k -nearest neighbourhood queries are done by using the kd-tree structure.
3. The output from step 2 (which typically contains a large number of modes) is post-processed to merge clusters that are close to each other in the range domain (a fixed radius window is applied for pruning).
4. The desired number of tissues is obtained by applying voxel weighted k-means clustering algorithm to the output (prune modes) from step 3.

5. Experimental results

In order to evaluate the performance of BAMS* as compared to the kNN-AMS* algorithm and other competing methods, we used both synthetic 3D multi-modal MR data and real single-valued MR data. We kept the same values for the parameters used in the pre and post-processing steps (such as bias correction, the pruning modes and voxel weighted k-means) for all methods. In the case of kNN-AMS, the bandwidth was estimated for each feature point \mathbf{x}_i by considering 120 numbers of neighbours as proposed in [10]. In the case of BAMS we bound the nearest neighbors sizes $K_j \in [100, 330]$. The synthetic MR multi-modal data was obtained from the Brainweb simulated brain database (SBD)¹ for additive noise levels ranging from 1% to 9%. In particular spatially aligned T1, T2 and proton density PD data were obtained each with 1 mm^3 voxel resolution. To compare the performance of the competing methods, we selected the same six slices in each modality containing the thin sulcal CSF structure. The real data was obtained from the IBSR database². It comprises twenty T1-weighted volumes with a coronal slice thickness of 3.1 mm and voxel size of $1 \times 3.1 \times$

*Bayesian based adaptive mean-shift (BAMS) (our proposed method)

*k-nearest neighbor adaptive mean-shift (kNN-AMS) (presented in [10])

¹<http://www.bic.mni.mcgill.ca/brainweb>

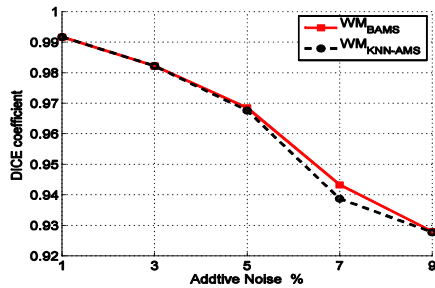
²<http://www.cma.mgh.harvard.edu/ibsr/>

1 mm^3 . Notably the data suffers from low contrast and high spatial inhomogeneities. The ground truth (i.e. expert labelling of the tissue types) was available for both the synthetic and real data. Segmentation accuracy was measured quantitatively using the Dice and Tanimoto coefficients. These measure the degree of overlap between the ground truth and the automated segmentations. They are defined

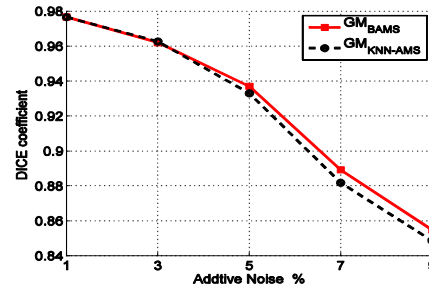
$$Dice(i) = \frac{2V_{ae}^i}{(V_a^i + V_e^i)}, Tc(i) = \frac{V_{ae}^i}{(V_a^i + V_e^i - V_{ae}^i)} \quad (13)$$

where V_{ae}^i is the number of voxels that are assigned to tissue i by both the segmentation algorithm and the ground truth, V_a^i and V_e^i denote the number of voxels to tissue i assigned by the automated algorithm and ground truth respectively.

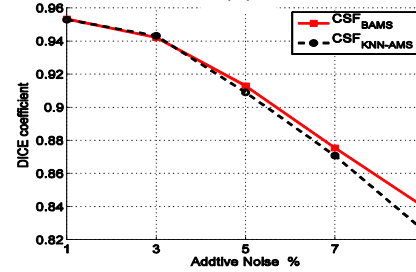
The quantitative analysis of segmentation accuracy (BAMS and kNN-AMS only) for the multi-modal synthetic data is shown in Fig. 1 (a-c). Here we observed that at the 1% and 3% levels of noise, the performances of both algorithms for the classification of WM, GM and CSF are almost same. However in the range 5% to 9% the results of the BAMS algorithm are better than kNN-AMS. A possible explanation is that the kNN based adaptive bandwidth might be more sensitive to noise. It is clear that by increasing the noise levels from 5% to 9%, the misclassification of voxels in the case of the kNN-AMS algorithm becomes higher especially in the case of GM and CSF segmentation. Table 1 shows the comparison of average values of Dice for WM, GM and CSF for both algorithms. For all three tissues BAMS is more accurate.



(a)



(b)



(c)

Fig.1. Quantitative analysis, Brainweb input: T1, T2, Pd data, Dice coefficient for (a) WM (b) GM and (c) CSF.

Table 1 Average value for Dice coefficient

Algorithm	WM	GM	CSF
BAMS	0.963	0.924	0.905
kNN-AMS	0.962	0.921	0.900

The quantitative analysis of segmentation accuracy (BAMS, kNN-AMS, as well as several competing methods published on the IBSR² site) for the real data (twenty real T1 weighted volumes) is shown in Fig.2 (a-c). The trends in the graph show that BAMS outperforms kNN-AMS for the GM classification especially in the first six volumes. These volumes are acknowledged in the literature to be difficult to segment [10] because of the high noise and spatial inhomogeneities they contain. It can also be seen that both algorithms have the same performance in the volume range from 13 to 20 (these volumes are of better quality, particularly with respect spatial inhomogeneities). A possible reason for the lower performance of the kNN-AMS algorithm compared to BAMS for the six difficult volumes might be biased estimation of Euclidian based bandwidth because of high spatial inhomogeneities (especially for the case of GM segmentation in which most of the GM voxels are falsely classified as WM and CSF). Table 2 shows the comparison of average Tanimoto coefficients for BAMS, our implementation of kNN-AMS, and the other competing methods published on the IBSR² site. From the table it can be seen that BAMS is more accurate (has higher coefficient values) than the competing methods for the WM, GM and CSF

classification. In comparison to kNN-AMS, it is comparable in the case of WM classification but is more accurate for the GM and CSF classification.

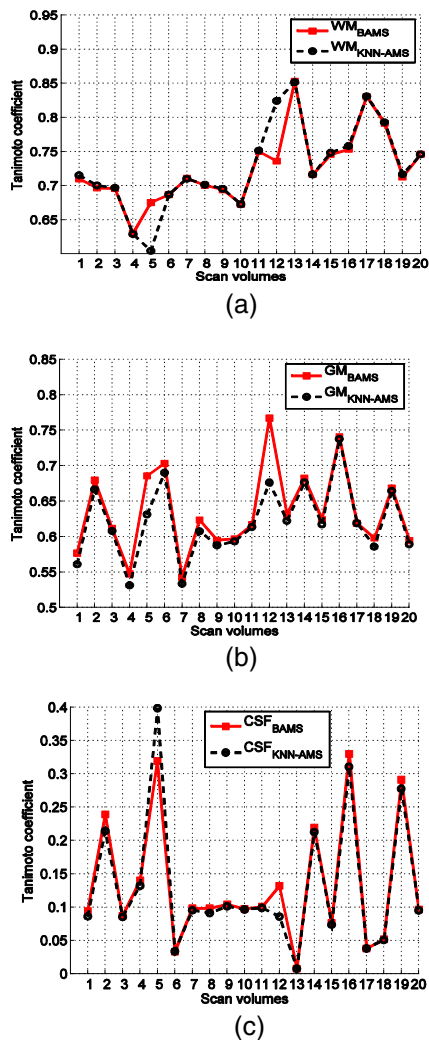


Fig. 2. Quantitative analysis, IBSR input: twenty real T1 data, Tanimoto coefficient for (a) WM (b) GM (c) CSF.

Table 2 Average values for Tanimoto coefficient

Algorithm	WM	GM	CSF
BAMS	0.725	0.635	0.133
kNN-AMS	0.727	0.621	0.129
FixedwMS	0.628	0.594	-
AMAP	0.570	0.560	0.070
BMAP	0.563	0.559	0.071
tskmeans	0.572	0.478	0.049

An example of the segmentation results (BAMS and kNN-AMS) for the multi-modal synthetic data at the

noise level 9% is presented in Fig. 3(d, e). It is difficult to judge the difference between both algorithms results visually. Therefore we have marked the kNN-AMS output image with several red circles to illustrate misclassifications of CSF which are not present in the BAMS output image. Fig. 3(f) shows the ground truth image.

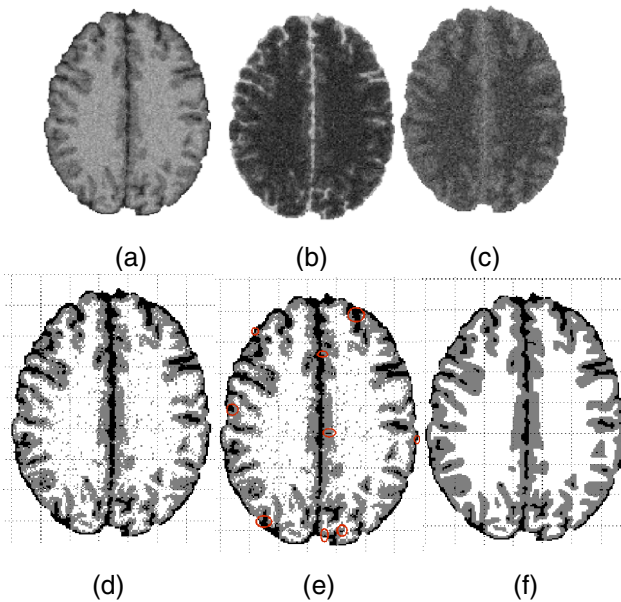


Fig. 3. A sample slice from Brainweb. (a-c) Input slices (T1, T2, Pd) at 9% noise levels (d) BAMS segmentation (e) kNN-AMS segmentation (f) Ground truth (white matter (WM) in white, gray matter (GM) in gray, CSF in black).

An example of the segmentation results (BAMS and kNN-AMS) for scan volume 12 of the real data is shown in Fig. 4(c, d, f, g). It can be clearly seen that BAMS yields better results than kNN-AMS for the GM and CSF classification because in the kNN-AMS output image most of the GM voxels are misclassified as CSF (this corroborates the results in Fig. 2 (b-c)). Fig. 4(e, h) shows the ground truth slices.

The result for the estimation of bandwidth for one data point x_i is illustrated in Fig. 5 where the histogram of local neighborhoods distances for $K_j \in [100, 330]$ is fitted to the Gamma distribution.

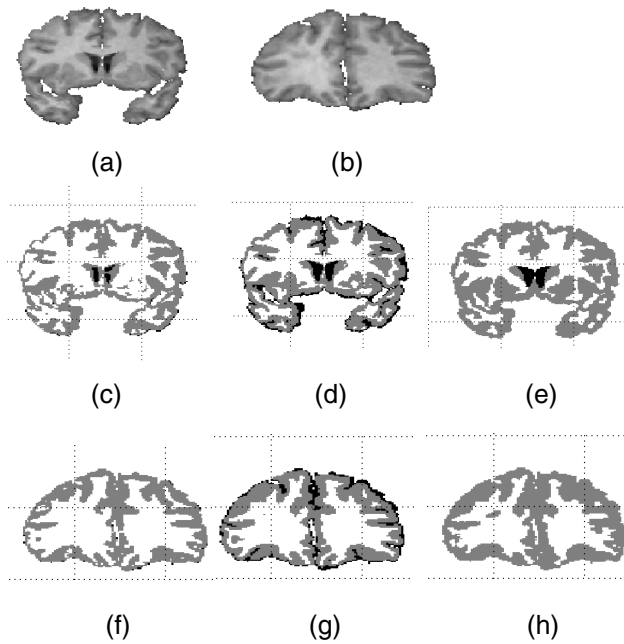


Fig. 4. Two slices from IBSR. (a, b) Input slices: T1 (c, f) BAMS segmentation outputs (d, g) kNN-AMS segmentation outputs (e, h) Ground truth image slices (white matter (WM) in white, gray matter (GM) in gray, CSF in black).

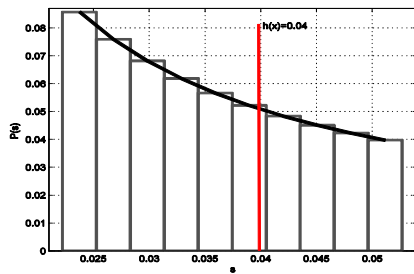


Fig.5. Bandwidth estimation (shown in red) for one data point x_i .

6. Conclusion

We have presented a novel Bayesian approach to adaptive bandwidth estimation in the mean-shift algorithm and evaluated its application to brain MR images. The results on the multi-modal synthetic data show that the proposed algorithm outperforms the kNN-AMS algorithm for high noise levels (from 5% to 9% additive Gaussian noise). In the case of the real data (not uni-modal), the results are not so clear cut. The proposed method outperforms all other competing methods published on the IBSR website except kNN-AMS. Compared to kNN-AMS, BAMS performs

similarly for GM and CSF classification and actually does better in several cases. For the WM classification the proposed method performs similarly to kNN-AMS across all volumes. In the future we plan to investigate other adaptive bandwidth selection methods. One possibility is variable bandwidth selection based on the covariance matrix which is most stable across scales [14].

7. References

- [1] J. L. Marroquín, B. C. Vemuri, S. Botello, and F. Calderon, "An accurate and efficient Bayesian method for automatic segmentation of brain MRI," *IEEE Trans. Med. Imag.*, vol.21, no.8, Aug. 2002, pp. 934–944.
- [2] K. Van Leemput, F. Maes, D. Vandeurmeulen, and P. Suetens, "Automated model-based tissue classification of MR images of the brain," *IEEE Trans. Med. Imag.*, vol.18, no.10, 1999, pp. 897–908.
- [3] H. Greenspan, A. Ruf and J. Goldberger, "Constrained Gaussian mixture model framework for automatic segmentation of MR brain images," *IEEE Trans. Med. Imag.*, vol. 25, no. 9, Sep. 2006, pp. 1233–1245.
- [4] K. Fukunaga and L. Hostetler, "The estimation of the gradient of a density function, with applications in pattern recognition," *IEEE Trans Inf. Theory*, vol.21, no.1, 1975, pp. 32-40.
- [5] D. Comaniciu and P. Meer, "Mean Shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Analysis Machine Intell.*, vol. 24, no. 5, 2002, pp. 603-619.
- [6] M. C. Jones, J.S Marron and S. J. Sheather, "A brief survey of bandwidth selection for density estimation," *Journal of the American Statistical Association*, vol.91, no.433, pp.401-407.
- [7] D. Comaniciu, V. Ramesh, P. Meer, "The variable bandwidth mean-shift and data-driven scale selection," in *ICCV, 2001*, pp.438-445.
- [8] B. Georgescu, I. Shimshoni, and P. Meer, "Mean-shift based clustering in high dimensions: A texture classification example," in *ICCV, 2003*, pp. 456–463.
- [9] A. G. Bors and N. Nasios, "Kernel bandwidth estimation for nonparametric modeling," *IEEE Trans.SMC*, 2009, pp.1543 – 1555.
- [10] A. Mayer and H. Greenspan, "An adaptive mean-shift framework for MRI brain segmentation," *IEEE Trans. Med. Imag.*, vol. 28, no. 8, Aug. 2009, pp. 1238–1249.
- [11] J. Friedman, J. Bentely, R. Finkel, "An Algorithm for finding best matches in logarithmic expected time," *ACM Transactions on Mathematical Software*, vol. 3, no.3, 1997, pp. 209-226.
- [12] S. M. Smith, "Fast robust automated segmentation," *Human Brain Mapping*, vol.17, no.3, 2002, pp. 143-155.
- [13] Mangin, J.-F, "Entropy minimization for automatic correction of intensity nonuniformity". *Proc. IEEE workshop on Mathematical Methods in Biomedical Image Analysis*, 2000, pp 162-169.
- [14] D. Comaniciu, "An algorithm for data-driven bandwidth selection", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 2, 2003, pp. 281–288.