

Multimodal MR Brain Segmentation Using Bayesian-based Adaptive Mean-Shift (BAMS)

Qaiser Mahmood^{1,2}, Mohammad Alipoor^{1,2}, Artur Chodorowski^{1,2}, Andrew Mehnert^{1,2}, and Mikael Persson^{1,2}

¹ Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden

² MedTech West, Sahlgrenska University Hospital, Gothenburg, Sweden

{qaiserm, alipoor, artur.chodorowski, andrew.mehnert, mikael.persson}@chalmers.se

Abstract. In this paper, we validate our proposed segmentation algorithm called Bayesian-based adaptive mean-shift (BAMS) on real multimodal MR images provided by the MRBrainS challenge. BAMS is a fully automatic unsupervised segmentation algorithm. It is based on the adaptive mean shift wherein the adaptive bandwidth of the kernel for each feature point is estimated using our proposed Bayesian approach [1]. BAMS is designed to segment the brain into three tissues; white matter (WM), gray matter (GM) and cerebrospinal fluid (CSF). The performance of the algorithm is evaluated relative to the manual segmentation (ground truth). The results of our proposed algorithm show the average Dice index 0.8377 ± 0.036 for the WM, 0.7637 ± 0.038 for the GM and 0.6835 ± 0.023 for the CSF.

Keywords: Magnetic Resonance, brain segmentation, mean shift, Bayesian

1 Introduction

Automated segmentation of MR images of the brain is an active area of research in the field of neuro imaging. The resulting segmentation yields a patient-specific labeling of individual tissues which makes possible for the quantitative characterization of these tissues (for example, in the study of Alzheimer's disease and multiple sclerosis).

The majority of unsupervised methods that have been proposed for automated segmentation of brain tissues are based on statistical parametric models. These methods assume some distributional form for the underlying probability distribution of the data and seek to estimate its parameters. Some of these [2, 3] are purely voxel wise intensity based clustering methods. A downside of these is that they may give poor tissue classifications in the presence of additive noise and the multiplicative bias field [4]. Some of the parametric methods [5, 6] using a Markov random field (MRF) statistical spatial model to improve the smoothness of segmentation. A drawback with these approaches is that MRF algorithm is computationally expensive and requires critical parameters settings at higher dimensional feature space [4].

An alternative unsupervised approach that doesn't require many parameters, incorporates the spatial information easily into a higher dimensional feature space (multimodal MR images) is mean shift (MS) clustering. In MS, only the parameter that influences the clustering is called bandwidth of the kernel. A couple of MS methods [4, 7] based on the adaptive bandwidth have been proposed for brain tissue segmentation in MR images. The adaptive bandwidth estimator [8] used in [4], is based on the k nearest neighbour (k NN) distance. A downside is that this approach is known to be biased by outliers for Euclidean distance [9]. Drawback of adaptive bandwidth estimator used in [7] include that it requires an initial density estimate (called the pilot). These collective limitations motivated the development of new algorithm presented in [1]. In [1], we applied our proposed algorithm only on single modality (T1) real MR data sets. In this paper we evaluate the performance of our proposed algorithm on multimodal MR data sets obtained from the MRBrainS challenge.

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 algorithm is presented in section 4. Finally, the experimental results are presented in section 5.

2 Adaptive mean shift

Let $\{\mathbf{x}_i \in \mathbb{R}^d | i = 1, \dots, n\}$ denote a set of feature vectors in d -dimensional space. The adaptive kernel density estimate of the underlying multivariate probability function at point \mathbf{x} is given by

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

where $h_i > 0$ is the adaptive bandwidth and $k : [0, 1] \rightarrow \mathbb{R}$ is the kernel profile of the spherically symmetrical kernel K with bounded support defined as

$$K(\mathbf{x}) = c_{k,d} k\left(\|\mathbf{x}\|^2\right) \quad \|\mathbf{x}\| \leq 1 \quad (2)$$

and $c_{k,d}$ is a normalizing constant ensuring that the kernel K integrates to 1. In this work, the Gaussian kernel is applied which is defined as

$$K_G(\mathbf{x}) = c_{k,d} \exp\left(-\frac{1}{2} \|\mathbf{x}\|^2\right) \quad (3)$$

The derivative of the adaptive density estimator in eq. 1 leads to

$$\begin{aligned} \hat{\nabla} f_K(\mathbf{x}) \equiv \nabla \hat{f}_K(\mathbf{x}) &= \frac{2c_{k,d}}{nh_i^{d+2}} \sum_{i=1}^n (\mathbf{x} - \mathbf{x}_i) k'\left(\left\|\frac{\mathbf{x} - \mathbf{x}_i}{h_i}\right\|^2\right) \\ &= \frac{2c_{k,d}}{nh_i^{d+2}} \left[\sum_{i=1}^n g\left(\left\|\frac{\mathbf{x} - \mathbf{x}_i}{h_i}\right\|^2\right) \right] \times \left[\frac{\sum_{i=1}^n \mathbf{x}_i g\left(\left\|\frac{\mathbf{x} - \mathbf{x}_i}{h_i}\right\|^2\right)}{\sum_{i=1}^n g\left(\left\|\frac{\mathbf{x} - \mathbf{x}_i}{h_i}\right\|^2\right)} - \mathbf{x} \right] \end{aligned} \quad (4)$$

where $g(x) = -k'(x)$. The right-most factor (in square brackets) in eq. 4 is called the mean shift vector. It points toward the direction of maximum increase in density and is defined as

$$M_{h_i, G}(\mathbf{x}) = \frac{\nabla \hat{f}_K(\mathbf{x})}{\hat{f}_G(\mathbf{x})} = \frac{\sum_{i=1}^n \mathbf{x}_i g\left(\left\|\frac{\mathbf{x} - \mathbf{x}_i}{h_i}\right\|^2\right)}{\sum_{i=1}^n g\left(\left\|\frac{\mathbf{x} - \mathbf{x}_i}{h_i}\right\|^2\right)} - \mathbf{x} \quad (5)$$

where G represents the kernel and defined as

$$G(\mathbf{x}) = c_{g, d} g(\|\mathbf{x}\|^2) \quad (6)$$

Adaptive mean shift is an iterative hill climbing procedure in which the kernel G starts from an initial position \mathbf{y}_1 and moves towards the position closer to the higher dense region. This process is continued until the position in higher dense region is achieved which represents a mode of the density. The update rule of kernel position is given by

$$\mathbf{y}_{j+1} = \frac{\sum_{i=1}^n \mathbf{x}_i g\left(\left\|\frac{\mathbf{y}_j - \mathbf{x}_i}{h_i}\right\|^2\right)}{\sum_{i=1}^n g\left(\left\|\frac{\mathbf{y}_j - \mathbf{x}_i}{h_i}\right\|^2\right)}, \quad j = 1, 2, \dots \quad (7)$$

where $\{\mathbf{y}_j\}_{j=1,2,\dots}$ represents the successive locations of the kernel G . The feature points that converge to the same mode constitute a cluster. In order to perform the clustering using spatial information of voxel with its range (intensity) value, a joint spatial-range domain kernel $K_{h_s, h_i}(\mathbf{x})$ is used. It is defined as a product of spatial and range domain kernels and is given by

$$K_{h_s, h_i}(\mathbf{x}) = \frac{C}{h_s^p h_i^d} k\left(\left\|\frac{\mathbf{x}^s}{h_s}\right\|^2\right) k\left(\left\|\frac{\mathbf{x}^r}{h_i}\right\|^2\right) \quad (8)$$

where \mathbf{x}^s represents a vector of voxels spatial coordinates, \mathbf{x}^r represents a vector of voxels range (intensity) values and h_s and h_i are their corresponding kernel bandwidths, and C is a normalization constant.

3 Bayesian approach for adaptive bandwidth estimation

Our proposed approach is based on a novel variation on the Bayesian approach initially proposed in [9] for the global bandwidth estimation of the kernel. In our variation we used this idea locally for adaptive bandwidth estimation.

The bandwidth is modeled by the *a posteriori* probability density function $p(s|\mathbf{x})$ of local data spread s given the feature point \mathbf{x} . Let $M < n$ (total number of feature points) be the number of nearest neighborhoods to a feature point \mathbf{x} . We can then define the pseudolikelihood

$$P(s|\mathbf{x}) = \prod_{j=1}^N P(s|\mathbf{x}_{M_j}) \quad (9)$$

where $P(s|\mathbf{x}_{M_j})$ is the probability of local data spread s depending on the nearest neighborhood samples to \mathbf{x}_{M_j} and $\{M_j \mid j = 1, 2, \dots, N\}$ is the set of N such neighborhoods of various sizes. The evaluation of probabilities over the entire range of number of neighborhoods M_j is given as

$$P(s|\mathbf{x}_{M_j}) = \int P(s|M_j, \mathbf{x}_{M_j})P(M_j|\mathbf{x}_{M_j})dM_j \quad (10)$$

Bayes rule yields

$$P(M_j|\mathbf{x}_{M_j}) = \frac{P(\mathbf{x}_{M_j}|M_j)P(M_j)}{P(\mathbf{x}_{M_j})} \quad (11)$$

where $P(\mathbf{x}_{M_j}|M_j)$ is the probability of the feature point \mathbf{x}_{M_j} given the M_j nearest neighborhood. Hereinafter $P(M_j)$ is considered to have uniform distribution on the interval $[M_1 M_2]$. Several values are selected for M_j in this interval according to

$$M_j = M_1 + j \frac{M_2 - M_1}{N} \quad (12)$$

For a given M_j the local variance s_j is computed as

$$s_j = \sum_{l=1}^{M_j} \frac{\|\mathbf{x}_{(i,l)} - \mathbf{x}_i\|^2}{M_j - 1}, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, N \quad (13)$$

where $\mathbf{x}_{(i,l)}$ is the l -th nearest neighbor to the feature point \mathbf{x}_i . The distribution of variances is modeled as the Gamma distribution defined as

$$P(s|\alpha, \beta) = \frac{s^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} e^{-\frac{s}{\beta}}, \quad s \geq 0, \quad \alpha, \beta > 0 \quad (14)$$

where α and β define the shape and the scale of the Gamma distribution, respectively.

These parameters are estimated using the maximum likelihood approach. Finally the adaptive bandwidth is estimated by the product of these parameters, identically the mean of Gamma distribution, i.e.

$$\hat{h}(\mathbf{x}_i) = \hat{h}_i = \hat{\alpha}\hat{\beta}, \quad i = 1, 2, \dots, n \quad (15)$$

4 Proposed segmentation algorithm

The pre-processing steps includes: (1) extraction of brain mask which is done by employing the BET2 [2] tool on a T1 weighted image with parameter settings; fractional intensity threshold=0.2 and threshold gradient=0.05. These parameters optimal values are determined by using the 5 training data sets. The resultant brain mask is then used to extract the brain of multimodal MR images. (2) Normalization of intensity values of each modality to the interval [0 1] which is done by applying the linear histogram stretching.

The proposed algorithm BAMS involves the following steps for segmenting the

brain into WM, GM and CSF.

1. BAMS takes feature vectors \mathbf{x}_i as input, which represent the thick slice multimodal MR brain images. The modes or clusters of the feature points are determined using the eq.7. The clustering is done in the joint spatial-range domain using the joint-domain kernel defined in eq.8. The spatial bandwidth is set as $h_s = 9$ and the adaptive range bandwidth h_i is estimated by employing our proposed Bayesian approach described in section 3 with parameters settings $M_1 = 100$, $M_2 = 750$ and $N = 10$. These parameters optimal values and the spatial bandwidth optimal value are determined empirically by using the 5 training datasets. The output of BAMS is a large number of modes or clusters.
2. Finally the three tissues are obtained by merging the clusters (output from step 1) using the voxel weighted k -means clustering algorithm [4]. Cluster initialization is performed making use of prior information of tissue intensity ranges in the multimodal MR images.

The schematic procedure of the proposed algorithm is shown in Fig.1.

5 Results

Our proposed unsupervised algorithm BAMS is applied on 12 test data sets, provided from the MRBrainS challenge. Each data set consists of thick slice multimodal (T1-weighted, T1-weighted inversion recovery (T1_IR) and FLAIR) MR images with voxel resolution $0.958\text{mm} \times 0.958\text{mm} \times 3\text{mm}$. The segmentation performance of BAMS for each data set is evaluated relative to manual segmentation. The evaluation is performed by the MRBrainS challenge organizers. The average segmentation performance of proposed method BAMS is presented in Table 1.

The proposed algorithm BAMS is implemented in MATLAB (R2010a) and it takes approximately one and a half hours for segmenting the multimodal MR images on a single desktop PC running 64-bit Ubuntu 10.04 with 8 GiB ram memory and 8 Intel Core i7 CPU 870 at 2.93 GHz.

Structure	Dice (%)	Mod. HD (mm)	Abs. volume diff. (%)
Gray matter	76.37 ± 3.79	5.77 ± 3.27	9.57 ± 7.25
White matter	83.77 ± 3.61	5.00 ± 3.57	24.74 ± 10.80
Cerebrospinal fluid	68.35 ± 2.33	3.73 ± 0.58	11.95 ± 9.22
Brain	93.24 ± 1.52	7.19 ± 3.90	5.45 ± 5.17
All intracranial structures	93.81 ± 1.09	6.27 ± 1.50	3.99 ± 1.97

Table 1. Average segmentation performance of proposed algorithm BAMS in terms of Dice, modified Hausdorff distance (Mod. HD) and absolute volume difference (Abs. volume diff.).

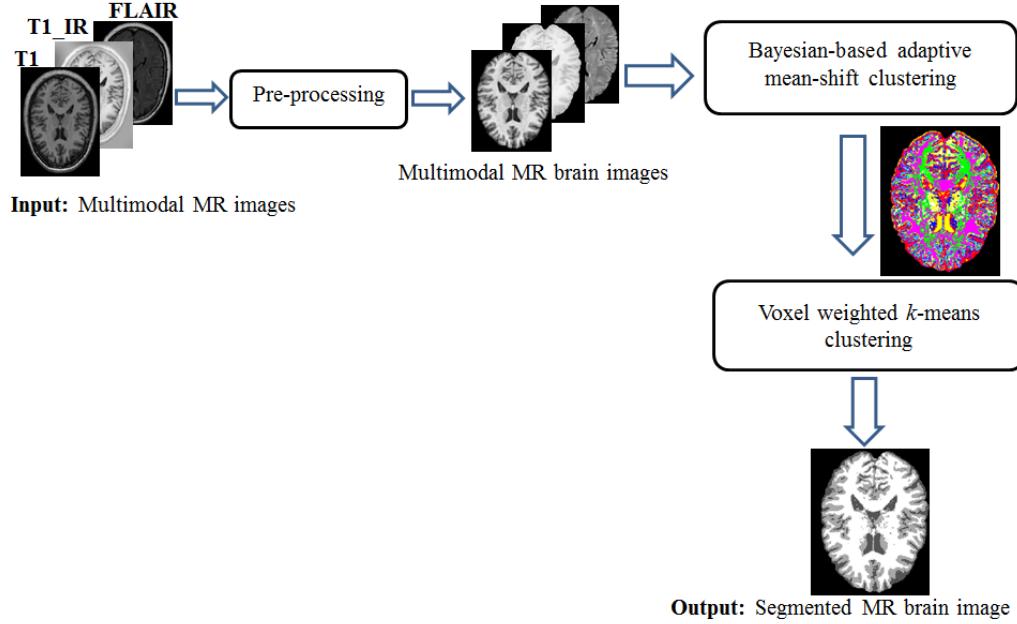


Fig. 1. Schematic procedure of proposed algorithm BAMS.

Acknowledgments. This work has been supported in part by the Chalmers University of Technology, Sweden and the Higher Education Commission (HEC) of Pakistan.

References

1. Mahmood, Q., Chodorowski, A., Mehnert, A., Persson, M.: A Novel Bayesian Approach to Adaptive Mean Shift Segmentation of Brain Images. In: 25th IEEE International Conference on Computer Based Medical Systems (CBMS), 1–6 (2012)
2. Van Leemput, K., Maes, F., Vandervelde, D., Suetens, P.: Automated model-based tissue classification of MR images of the brain. *IEEE Trans. Med. Imag.* 18, 897–908 (1999)
3. Greenspan, H., Ruf, A., Goldberger, J.: Constrained Gaussian mixture model framework for automatic segmentation of MR brain images. *IEEE Trans. Med. Imag.* 25, 1233–1245 (2006)
4. Mayer, A., Greenspan, H.: An adaptive mean-shift framework for MRI brain segmentation, *IEEE Trans. Med. Imag.* 28, 1238–1249 (2009)
5. Marroquin, J. L., Vemuri, B. C., Botello, S., Calderon, F.: An accurate and efficient Bayesian method for automatic segmentation of brain MRI, *IEEE Trans. Med. Imag.* 21, 934–944 (2002)
6. Zhang, Y., Brady, M., Smith, S.: Segmentation of brain MR images through a Hidden Markov Random Field Model and the Expectation Maximization algorithm, *IEEE Trans. Med. Imag.* 20, 45–57 (2001)

7. Jimenez-Alaniz, R. J., Pohl-Alfaro, M., Medina-Bafluelos, V., Yàñez Suárez, O.: Segmenting brain MRI using adaptive mean shift. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 3114–3117 (2006)
8. Comaniciu, D., Meer, P.: A robust approach toward feature space analysis., *IEEE Trans. Pattern Analysis Machine Intell.* 20, 603–619 (2002)
9. Bors, A. G., Nasios, N.: Kernel bandwidth estimation for nonparametric modeling, *IEEE Trans. SMC* 39, 1543–1555 (2009)