3D BRAIN MRI SEGMENTATION BASED ON ROBUST HIDDEN MARKOV CHAIN

S. Bricq[*]*, Ch. Collet*

Université Strasbourg I
LSIIT - UMR CNRS 7005
Pôle API, Bd S. Brant,
F-67412 Illkirch - France

J.-P. Armspach

Université Strasbourg I
LINC - UMR CNRS 7191
4 rue Kirschleger
F-67205 Strasbourg - France

ABSTRACT

In this paper, we present a robust method to estimate parameters of Hidden Markov Chains (HMC) in order to segment brain MR images. Indeed, parameter estimation can be very sensitive to the presence of outliers in the data. We propose to use the Trimmed Likelihood Estimator (TLE) to extract such outliers and to accurately estimate the parameters of different tissue classes in a robust way. Moreover neighborhood information is included in the model by using Hidden Markov Chains. Experimental results on 2D synthetic data and on 3D brain MRI are included to validate this approach.

Index Terms— Image segmentation, Hidden Markov models, robustness, Magnetic Resonance Imaging

1. INTRODUCTION

Segmentation is an important step for quantitative analysis of 3D brain images. Manual tracing of cerebral structures in MR images by a human expert is a time-consuming process and it is prone to intra- and inter-observer variability, which deteriorates the significance of the resulting segmentation analysis. Due to issues such as partial volume effects, noise or acquisition artifacts, segmenting brain MRI remains a challenging task. Furthermore the presence of pathological abnormalities (such as tumors or lesions) can bias parameter estimation.

In this paper, we present a robust method to estimate parameters to segment brain MR images using the Hidden Markov Chain (HMC) model. For this aim, we use the Trimmed Likelihood Estimator (TLE) to extract outliers and to estimate the parameters of the different classes in a robust way.

The paper is organized as follows: next section introduces the Trimmed Likelihood Estimator (TLE) and the FAST-TLE algorithm. In section 3, we present how this robust estimator can be used to estimate HMC parameters in the presence of outliers in the data. In section 4, results obtained on 2D synthetic images and on 3D brain MRI are shown. Finally in section 5, conclusions are drawn and future developments are suggested.

[*]Thanks to Region of Alsace for funding.

2. TRIMMED LIKELIHOOD ESTIMATOR

The Trimmed Likelihood Estimator (TLE) was introduced by Neykov and Neytchev [1] and developed to estimate mixture of multivariate normals and generalized linear models in a robust way [2, 3]. The optimization scheme used to compute this estimator derives from the optimization scheme of the Least Trimmed Squares (LTS) estimators of Rousseeuw and Leroy [4]. This algorithm was used to segment brain MRI by Aït-Ali in the frame of gaussian mixtures [5]. The main idea lies in finding h observations from N samples for which the likelihood is maximum and thus in removing the $N-h$ observations whose values would be highly unlikely to occur if the fitted model was true. We will aply this estimator to estimate parameters in a Hidden Markov Chain framework.

2.1. Trimmed Likelihood Estimator

Let us consider N i.i.d observations $y_n \in \mathbb{R}^q$ for $n = 1, \ldots, N$ with probability density $f(y;\theta)$ depending on an unknown parameter $\theta \in \Theta^p \subset \mathbb{R}^p$. The Trimmed Likelihood Estimator (TLE) [6] is defined as:

$$\hat{\theta}_{TLE} = \arg\min_{\theta\in\Theta^p} \sum_{n=1}^{h} \psi(y_{\nu(n)};\theta) \tag{1}$$

where for a fixed θ, $\psi(y_{\nu(1)};\theta) \leq \psi(y_{\nu(2)};\theta) \leq \ldots \leq \psi(y_{\nu(N)};\theta)$ and $\psi(y_n;\theta) = -\log f(y_n;\theta)$. Furthermore $\nu = (\nu(1), \ldots, \nu(N))$ denotes the corresponding permutation of the indices, which depends on θ and h is the trimming parameter corresponding to the amount of values including in parameter estimation. This leads to:

$$\hat{\theta}_{TLE} = \arg\max_{\theta\in\Theta^p} \prod_{i=1}^{h} f(y_{\nu(i)};\theta) \tag{2}$$

General conditions for the existence of a solution of (Eq. 1) are proved in [7]. Convergence and asymptotic properties are studied in [8, 9].

2.2. FAST-TLE algorithm

The FAST-TLE algorithm was developed in [10]. It can be described as follows: given the subset $H^{old} = \{y_{j_1}, \ldots, y_{j_N}\} \subset \{y_1, \ldots, y_N\}$,

- Compute $\hat{\theta}^{old} := MLE$ (Maximum Likelihood Estimator) based on H^{old} using Eq. 1.
- Define $Q^{old} = \sum_{i=1}^{k} \psi(y_{j_i}, \hat{\theta}^{old})$.
- Sort $\psi(y_i, \hat{\theta}^{old})$ for $i = 1, \ldots, N$ in ascending order: $\psi(y_{\nu(i)}, \hat{\theta}^{old}) \leq \psi(y_{\nu(i+1)}, \hat{\theta}^{old})$ and get the permutation $\nu = (\nu(1), \ldots, \nu(n))$.
- Define $H^{new} = \{y_{\nu(1)}, \ldots, y_{\nu(N)}\}$.
- Compute $\hat{\theta}^{new} := MLE$ based on H^{new} using Eq. 1.
- Define $Q^{new} = \sum_{i=1}^{k} \psi(y_{\nu(i)}, \hat{\theta}^{new})$.

We propose to adapt this algorithm to the estimation of Hidden Markov Chain parameters in next section.

3. ROBUST HIDDEN MARKOV CHAIN SEGMENTATION

3.1. Hidden Markov Chain

To segment Brain MRI, we propose to use Hidden Markov Chains (HMC) by using a 3D Hilbert-Peano scan of the data cube [11]. HMC have been widely used to segment 2D images [12]. The interest of Markov Chain methods for image segmentation compared to 3D Markov Random Fields (MRF) models is that being based on 1D modeling, they result in lower computing costs. The first step of segmentation algorithms based on HMC consists in transforming the image into a vector. Once all the processing has been carried out on the vector, the inverse transformation is applied on the segmented chain to obtain the final segmented image.

Let us now consider two sequences of random variables $X = (X_n)_{n \in S}$ the hidden process, and $Y = (Y_n)_{n \in S}$ the observed one, with S the finite set corresponding to the N voxels of the image. Each X_n takes its value in a finite set of K classes $\Omega = \{\omega_1, \ldots, \omega_K\}$ and each Y_n takes its value in $\mathbb{R}$. X is a Markov Chain if $P(X_{n+1} = \omega_{k_{n+1}} | X_n = \omega_{k_n}, \ldots, X_1 = \omega_{k_1}) = P(X_{n+1} = \omega_{k_{n+1}} | X_n = \omega_{k_n})$. Thus X will be determined by the initial distribution $\pi_k = P(X_1 = \omega_k)$ and the transition matrix $a_{kl}^n = P(X_{n+1} = \omega_l | X_n = \omega_k)$. We assume the homogeneity of the Markov Chain which means that the transition matrix is independent of the location n: $a_{kl}^n = a_{kl}$, for $1 \leq n < N$. The dependency graph of a HMC is presented in Fig. 1.

In the case of brain MRI segmentation, each X_n takes its value in a set of $K = 3$ classes $\Omega = \{WM, GM, CSF\}$, where WM, GM and CSF correspond respectively to white matter, gray matter and cerebrospinal fluid. The likelihood

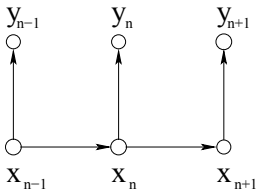


Fig. 1. *Dependency graph of Hidden Markov Chain*

$f_k(y_n, \theta) = P(Y_n = y_n | X_n = \omega_k)$ of the observation y_n conditionnaly to $X_n = \omega_k$ is assumed to be a Gaussian density with mean $\mu = (\mu_1, \ldots, \mu_K)$, variance $\sigma^2 = (\sigma_1^2, \ldots, \sigma_K^2)$ and $\theta = (\mu, \sigma^2)$. One of the interests of Hidden Markov Chains is the possibility of computing exactly the posterior marginals at each location. To obtain a labelling $\hat{x}$ of the image, we use the MPM (Mode of Posterior Marginals) estimator [13]:

$$\hat{x}_n = \arg\max_{\omega_k \in \Omega} P(X_n = \omega_k | Y = y) \quad (3)$$

$$= \arg\max_{\omega_k \in \Omega} \alpha_n(k) \beta_n(k) \quad (4)$$

with $\alpha_n(k) = P(X_n = \omega_k, Y_1, \ldots, Y_n)$ forward probability and $\beta_n(k) = P(Y_{n+1}, \ldots, Y_N | X_n = \omega_k)$ backward probability [14]. These probabilities can be computed recursively. This recursive computation is detailed in Sec 3.2 in the robust case.

3.2. Robust parameter estimation

Parameter estimation using the EM algorithm [15] can be sensitive to outliers in the data. To estimate parameters in a robust way, we thus adapt the FAST-TLE algorithm presented in Sec. 2.2 to HMC. This leads to [14]:

1. Compute $\hat{\theta}^{(p-1)} := MLE$ using EM, based on the whole dataset.
2. Sort residus $r_n = -\log f(y_n; \hat{\theta}^{(p-1)})$ for $n = 1, \ldots, N$

$$f(y_n; \hat{\theta}^{(p-1)}) = \sum_{\omega_k} P(Y_n = l, X_n = \omega_k, \hat{\theta}^{(p-1)})$$

$$= \sum_{\omega_k} P(X_n = \omega_k) f_k(y_n, \hat{\theta}^{(p-1)})$$

3. Define $H^{(p)} = \{y_{\nu(1)}, ..., y_{\nu(h)}\}$ the subset containing the h vectors with the lowest residus for $\hat{\theta}^{(p-1)}$.
4. Compute $\hat{\theta}^{p} := MLE$ using EM, based on $H^{(p)}$. We assign the likelihood of data considered as outliers to one, i.e. $f_{x_n}(y_n) = 1$. On the location where the data is considered as an outlier, only prior distribution takes place in the labelling process. Calculation of the different probabilities becomes:

- Forward probabilities:
  - $\alpha_1(k) = \pi_k f_k(y_1, \hat{\theta}^p)$
  - $\alpha_n(k) = \sum_{l=1}^{K} \alpha_{n-1}(l) a_{lk} f_k(y_n, \hat{\theta}^p)$ with $f_k(y_n, \hat{\theta}^p) = 1$ if y_n is considered as an outlier.
- Backward probabilities:
  - $\beta_N(k) = 1$
  - $\beta_n(k) = \sum_{l=1}^{K} \beta_{n+1}(l) a_{kl} f_l(y_{n+1}, \hat{\theta}^p)$ with $f_l(y_{n+1}, \hat{\theta}^p) = 1$ if y_{n+1} is considered as an outlier.
- *a posteriori* joint probabilities: $\xi_n(i,j) = \frac{\alpha_{n-1}(j) a_{ji} f_i(y_n, \hat{\theta}^p) \beta_n(i)}{\sum_k \alpha_n(k)}$
- *a posteriori* marginal probabilities: $\gamma_n(i) = \frac{\alpha_n(i)\beta_n(i)}{\sum_j \alpha_N(j)}$
- $\mu_i = \frac{\sum_{n_1} \gamma_{n_1}(i) y_{n_1}}{\sum_{n_1} \gamma_{n_1}(i)}$ with y_{n_1} belonging to the subset $H^{(p)}$.
- $\sigma_i = \frac{\sum_{n_1} \gamma_{n_1}(i)(y_{n_1} - \mu_i)(y_{n_1} - \mu_i)^t}{\sum_{n_1} \gamma_{n_1}(i)}$ with y_{n_1} belonging to the subset $H^{(p)}$.

5. Back to step 2 until convergence.

We then apply this method to the segmentation of 2D synthetic images and 3D brain MR images.

4. VALIDATION

4.1. Synthetic 2D images

To validate this approach, tests have been carried out on a 64×64 synthetic image composed of 3 classes corrupted by Gaussian noise which parameters are described in Tab. 1. Moreover intensity value of 30 pixels of class 2 has been set to 200. As the mean of class 2 is 100, these pixels will simulate outliers in the data. We test our robust HMC approach for different values of the trimming parameter h.

Class	Mean μ	Variance σ^2
Class 1	75	9
Class 2	100	9
Class 3	125	9

Table 1. Parameters of the synthetic image.

When parameter estimation is not robust ($h = 100$), variance of class 3 is very high: indeed "outliers" pixels are considered to belong to class 3. As a consequence, the border between class 2 and 3 is not well segmented (Fig. 2 c). When h decreases, i.e. when the number of pixels considered as outliers and thus not included in parameter estimation increases, variance of class 3 decreases. If h is too low, the variance of class 1 decreases and mean of class 1 increases slightly: the pixels with the lowest intensities are considered to be the farest from the model and thus rejected. If the outliers correspond to potential lesions or pathological abnormalities, they can be extracted in order to be post-processed (Fig. 2 f).

h	μ_1	σ_1^2	μ_2	σ_2^2	μ_3	σ_3^2
100	75	8.47	100.08	8.95	126.21	182.37
99	75.08	7.96	100.08	9.08	124.95	8.89
95	75.86	5.40	100.08	9.08	124.95	8.89

Table 2. Results obtained on the synthetic image. Parameters estimation for different values of h are reported.

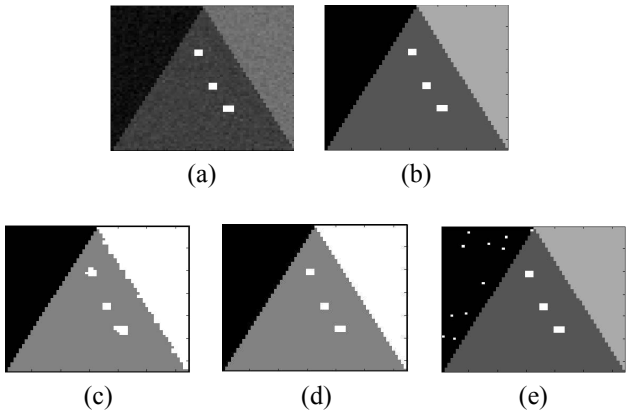


Fig. 2. Results obtained on a synthetic 2D image. (a) corresponds to the noisy image and (b) to the ground truth. (c) and (d) correspond respectively to the results obtained using the robust HMC segmentation with value of h respectively 100 and 99. In (e) outliers obtained with $h = 99$ are shown in white.

4.2. Tests on the Brainweb database

We applied the robust HMC model presented in previous section to brain MRI segmentation. This method has been tested on the Brainweb database[1] [16] which offers a large amount of different phantoms of MR brain images with varying noise levels from 0% to 9% and varying levels of non-uniformities from 0% to 40%. From these phantoms, the tissue classification in white matter (WM), gray matter (GM) and cerebrospinal fluid (CSF) is known. To evaluate the performance of our algorithm, we use the Kappa index (KI):

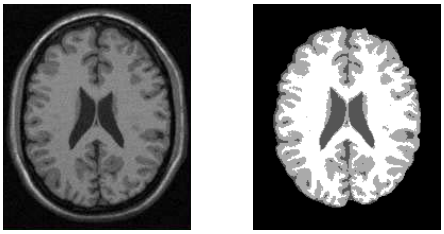
$$KI = 2\frac{SEG \bigcap GT}{SEG + GT} \tag{5}$$

where GT stands for the ground truth and SEG for the segmentation obtained. The method was tested with different values of the trimming parameter h. Results obtained are presented in Tab. 3 and in Fig. 3 .

[1] http://www.bic.mni.mcgill.ca/brainweb/

	9% noise					
	0% inhomogeneity			20% inhomogeneity		
h	WM	GM	CSF	WM	GM	CSF
100	91.93	89.67	90.41	90.52	88.49	89.96
99	91.93	89.76	90.58	90.52	88.57	90.12
95	91.93	89.33	89.86	90.52	88.07	89.35

Table 3. Results obtained on two Brainweb images with 9% of noise and 0 and 20% inhomogeneity. Kappa index (in %) is reported for white matter (WM), gray matter (GM) and cerebrospinal fluid (CSF) for different values of h.



(a) (b)

Fig. 3. Results obtained on a Brainweb image. (a) corresponds to the noisy image and (b) to the segmentation with $h = 99\%$.

For white matter, robust estimation does not improve the results. On the other hand, gray matter and cerebrospinal fluid results are slightly improved. Indeed, for a trimming parameter h of 99%, Kappa index is higher than for a value of h of 100% (which corresponds to a classical EM estimation). However, the main drawback of this method is that parameter h corresponding to the percentage of pixels used to estimate parameters has to be set to a value neither too high to reject all the outliers, nor too low in order to not lose information. Indeed for a value of h of $h = 95\%$, Kappa index is lower than in the non-robust case. In practice we can have an *a priori* on the size of lesions and on the percentage of outliers, thus the trimming parameter h can be chosen more precisely.

5. CONCLUSION AND OUTLOOK

In this paper, we have described and validated a robust method for tissue classification of brain MR images. We use HMC to include neighborhood information in the model. To estimate model parameter in a robust way, the Trimmed Likelihood Estimator was used. This method has been validated on 2D synthetic images and 3D brain phantoms. Future work will consist in applying this method to lesion detection.

6. REFERENCES

[1] N.M. Neykov and P.N. Neytchev, "A robust alternative of the MLE," *Compstat'90*, pp. 99–100, 1990.

[2] D.L. Vandev and N.M. Neykov, "Robust Maximum Likelihood in the Gaussian Case," in *New Directions in Data Analysis and Robustness*, S. Morgenthaler et al., Ed., Birkhäuser Verlag Basel, Switzerland, 1993, pp. 259–264.

[3] C.H. Müller and N.M. Neykov, "Breakdown points of the trimmed likelihood and related estimators in glms," *J. Statist. Plann. Inference*, vol. 116, pp. 503–519, 2003.

[4] P.J. Rousseeuw and A.M. Leroy, *Robust Regression and Outlier Detection*, Wiley, 1987.

[5] L.S. Aït-ali, S. Prima, P. Hellier, B. Carsin, G. Edan, and C. Barillot, "STREM: a robust multidimensional parametric method to segment MS lesions in MRI," in *8th International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI'2005*, J. Duncan and G. Gerig, Eds., Palm Springs, USA, October 2005, vol. 3749 of *Lecture Notes in Computer Science*, pp. 409–416, Springer.

[6] D.L. Vandev and N.M. Neykov, "About regression estimators with high breakdown point," *Statistics*, vol. 32, pp. 111–129, 1998.

[7] R. Dimova and N.M. Neykov, "Generalized d-fullness techniques for breakdown point study of the trimmed likelihood estimator with applications," *Theory and applications of recent robust methods*, 2004.

[8] P. Čížek, "Robust estimation in nonlinear regression and limited dependent variable models," *EconPapers*, 2002.

[9] P. Čížek, "Asymptotics of least trimmed squares regression," *EconPapers*, 2004.

[10] N.M. Neykov and C.H. Müller, "Breakdown point and computation of trimmed likelihood estimators in generalized linear models," in *Developments in robust statistics*, R. Dutter, P. Filzmoser, U. Gatter, and P.J. Rousseeuw, Eds., Physica-Verlag, Heidelberg, 2003, pp. 277–286.

[11] Y. Bandoh and A. Kamata, "An address generator for a 3-dimensional pseudo-Hilbert scan in acuboid region," in *Proc. of IEEE Int. Conf on Image Process.*, 1999.

[12] R. Fjortoft, Y. Delignon, W. Pieczynski, M. Sigelle, and F. Tupin, "Unsupervised Classification of Radar Images Using Hidden Markov Chains and Hidden Markov Random Fields," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 3, pp. 675–686, March 2003.

[13] A. Gelman, J. Carlin, H. Stern, and D. Rubin, *Bayesian data analysis*, Chapman and Hall - New York, 2005.

[14] S. Bricq, Ch. Collet, and J.-P. Armspach, "Triplet Markov Chain for 3D MRI brain segmentation using a probabilistic atlas," in *IEEE 2006 International Symposium on Biomedical Imaging, April 6-9, ISBI'06*, 2006.

[15] M.A. Tanner, *Tools for statistical inference : methods for the exploration of posterior distributions and likelihood functions*, Springer Verlag, 1993.

[16] R.K.-S. Kwan, A.C. Evans, and G.B. Pike, "An Extensible MRI Simulator for Post-Processing Evaluation," *Visualization in Biomedical Computing (VBC'96)*, vol. 1131, pp. 135–140, 1996.