Combination of Hidden Markov Random Field and Conjugate Gradient for Brain Image Segmentation

EL-Hachemi Guerrout[1], Samy Ait-Aoudia[1], Dominique Michelucci[2] and Ramdane Mahiou[1]
[1]*Laboratoire LMCS, Ecole nationale Supérieure en Informatique, Oued-Smar, Algiers, Algeria*
[2]*Laboratoire LE2I, Université de Bourgogne, Dijon, France*
{*e_guerrout, s_ait_aoudia, r_mahiou*}*@esi.dz, dominique.michelucci@u-bourgogne.fr*

Keywords: Brain image segmentation, Hidden Markov Random Field, The Conjugate Gradient algorithm.

Abstract: Image segmentation is the process of partitioning the image into different regions, meaningful and easier to analyze. Nowadays, the segmentation became a necessity in many practical medical imaging like locating tumors and pathologies. The Hidden Markov Random Field model is one of several techniques used in image segmentation. It provides an elegant way to model the segmentation problem. This modeling leads to the minimization of an objective function. The Conjugate Gradient algorithm (CG) is one of the most well known optimization techniques. In this paper we show the combination of Hidden Markov Random Field model and the Conjugate Gradient algorithm to achieve a good segmentation of brain images.

1 INTRODUCTION

Image segmentation is the process to partition a given image into different meaningful regions easier to analyze.

Nowadays, the segmentation became a necessity in many practical medical imaging like locating tumors, pathologies, measure tissue volumes, diagnosis, study of anatomical structure, surgery planning, virtual surgery simulation, intra-surgery navigation, etc.

There are several techniques to achieve the segmentation process. We can classify them into thresholding methods (Kumar et al., 2007; Natarajan et al., 2012; Zhao et al., 2015), clustering methods (Pham et al., 2000; Wu and Leahy, 1993; Chuang et al., 2006), edge detection methods (Perona and Malik, 1990; Senthilkumaran and Rajesh, 2009; Canny, 1986), region-growing methods (Lin et al., 2012; Roura et al., 2014), watersheds methods (Benson et al., 2015; Masoumi et al., 2012), model-based methods (Chan et al., 2001; Ho et al., 2002; McInerney and Terzopoulos, 1996; Wang et al., 2014) and Hidden Markov Random Field methods (Zhang et al., 2001; Held et al., 1997; Panjwani and Healey, 1995).

Threshold-based techniques are based on image histogram: they find one or more intensity thresholds to separate the different image classes. If the image contains n distinctive classes, $n-1$ thresholds are necessary. For example, to segment the image into two classes, foreground and background, one threshold is necessary. The disadvantage of the threshold-based techniques is that they are very noise sensitive.

Region-based methods assemble neighboring pixels of the image in non overlapping regions according to some homogeneity criterion.

In model-based segmentation, a model is built for a specific anatomic structure by incorporating a priori information concerning shape, location, and orientation. The presence of noise degrades the segmentation quality. That is why denoising phase is generally an essential prior.

In classification methods, pixels are classified according to some properties or criteria: gray level, texture or color.

Hidden Markov Random Field (HMRF) (Geman and Geman, 1984) provides an elegant way to model the segmentation problem based on the MAP (Maximum A Posteriori) criterion (Wyatt and Noble, 2003). The MAP estimation leads to the minimization of an energy function (Szeliski et al., 2008). Therefore, optimization techniques are necessary to compute a solution.

The Conjugate Gradient Algorithm (Møller, 1993; Powell, 1977; Shewchuk, 1994) is one of the most popular optimization methods.

This paper presents a segmentation method based on the combination of Hidden Markov Field model and Conjugate Gradient algorithm.

Brain MR images segmentation has attracted a

particular attention in medical imaging. Thus our tests focus on BrainWeb [1] (Cocosco et al., 1997) and IBSR[2] databases where the ground truth is known.

The segmentation quality is evaluated using Dice Coefficient (DC) (Dice, 1945) criterion. DC measures how much the segmentation result is close to the ground truth.

This paper is organized as follows. We begin by introducing the concept of Hidden Markov Field in the section 2. A short section 3 is devoted to the well known Conjugate Gradient algorithm. The section 4 is devoted to the experimental results. Finally, section 5 concludes the paper.

2 HIDDEN MARKOV RANDOM FIELD (HMRF)

Let $S = \{s_1, s_2, \ldots, s_M\}$ be the sites or positions set. Both the image to segment and the segmented image are formed of M sites. Each site $s \in S$ has a neighborhood set $V_s(S)$.

A neighborhood system $V(S)$ has the following properties:

$$\begin{cases} \forall s \in S, s \notin V_s(S) \\ \forall \{s,t\} \in S, s \in V_t(S) \Leftrightarrow t \in V_s(S) \end{cases} \quad (1)$$

A r-order neighborhood system $V^r(S)$ is defined by the following formula:

$$V_s^r(S) = \{t \in S | \text{ distance}(s,t)^2 \leq r^2 \wedge s \neq t\} \quad (2)$$

The distance only depends on the pixel position *i.e.,* it is not related to the pixel value.

A clique c is a subset of S where all sites are neighbors to each other. For a non single-site clique, we have:

$$\forall \{s,t\} \in c, s \neq t \Rightarrow (t \in V_s(S) \wedge s \in V_t(S)) \quad (3)$$

A p-order clique noted C_p contains p sites i.e. p is the cardinal of the clique.

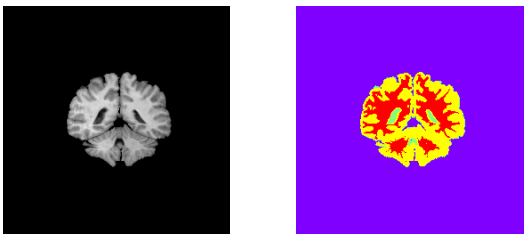
Let $y = (y_1, y_2, \ldots, y_M)$ be the pixels values of the image to segment and $x = (x_1, x_2, \ldots, x_M)$ be the pixels classes of the segmented image. y_i and x_i are respectively pixel value and class of the site s_i.

The image to segment y and the segmented image x are seen respectively as a realization of a Markov Random Field $Y = (Y_1, Y_2, \ldots, Y_M)$ and $X = (X_1, X_2, \ldots, X_M)$. The Markov Random Field is a family of random variables.

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

[2] https://www.nitrc.org/projects/ibsr

The random variables $\{Y_s\}_{s \in S}$ take their values in the gray level space $E_{obs} = \{0, \ldots, 255\}$ (obs for observed). The random variables $\{X_s\}_{s \in S}$ take their values in the discrete space $E = \{1, \ldots, K\}$. K is the number of classes or homogeneous regions in the image. Figure 1 shows an example of image to segment.



y: The image to segment x: The segmented image

Figure 1: An example of segmentation with $K = 4$.

The segmentation of the image y consists to seek a realization x of the hidden field X. HMRF models this problem by maximizing the probability $P[X = x \mid Y = y]$.

$$x^* = \underset{x \in \Omega, \Omega = E^M}{\arg\max} \{P[X = x \mid Y = y]\} \quad (4)$$

From the Bayes rule, we get:

$$P[X = x \mid Y = y] = \frac{P[Y = y \mid X = x] \times P[X = x]}{P[Y = y]} \quad (5)$$

Based on the conditional independence we have:

$$P[Y = y \mid X = x] = \prod_{s \in S} P[Y_s = y_s \mid X_s = x_s] \quad (6)$$

By the assumption that $P[Y_s = y_s \mid X_s = x_s]$ follows a normal distribution with mean μ_{x_s} and standard deviation σ_{x_s}, we will have:

$$P[Y_s = y_s \mid X_s = x_s] = \frac{1}{\sqrt{2\pi\sigma_{x_s}^2}} exp\left(\frac{-(y_s - \mu_{x_s})^2}{2\sigma_{x_s}^2}\right) \quad (7)$$

According to equation 6 and 7 we get:

$$P[Y = y \mid X = x] = \prod_{s \in S} \frac{1}{\sqrt{2\pi\sigma_{x_s}^2}} exp\left(\frac{-(y_s - \mu_{x_s})^2}{2\sigma_{x_s}^2}\right) \quad (8)$$

$$\Leftrightarrow P[Y = y \mid X = x] = \text{Cst}_1 \; exp\left(-\Psi_1(x,y)\right) \quad (9)$$

where $\text{Cst}_1 = (2\pi)^{\frac{-M}{2}}$ and M is the image pixel number.

$$\Psi_1(x,y) = \left(\sum_{s \in S} \left[\ln(\sigma_{x_s}) + \frac{(y_s - \mu_{x_s})^2}{2\sigma_{x_s}^2} \right] \right)$$

According to Hammersley-Clifford theorem (Hammersley and Clifford, 1971) which establishes the equivalence between Markov field and Gibbs, we get:

$$P[X = x] = \frac{exp\left(\frac{-U(x)}{T}\right)}{\sum_{\xi \in \Omega} exp\left(\frac{-U(\xi)}{T}\right)} \tag{10}$$

where T is a control parameter called temperature.

The energy $U(x)$ is defined by Potts model (Swendsen and Wang, 1987) as follows:

$$U(x) = \beta \sum_{c_2 = \{s,t\}} (1 - 2\delta(x_s, x_t)) \tag{11}$$

where β is a constant and δ is the Kronecker's delta:

$$\delta(a,b) = \begin{cases} 1 & \text{if} \quad a = b \\ 0 & \text{if} \quad a \neq b \end{cases} \tag{12}$$

$P[Y = y]$ is a constant, so pose:

$$P[Y = y] = \text{Cst}_2 \tag{13}$$

By replacing the equations (9), (10) and (13) in the equation (5), we will have:

$$\begin{aligned} P[X = x \mid Y = y] &= \frac{\text{Cst}_1 \exp\left(-\Psi_1(x,y)\right) \exp\left(\frac{-U(x)}{T}\right)}{\text{Cst}_2 \sum_{\xi \in \Omega, \Omega = E^M} \exp\left(\frac{-U(\xi)}{T}\right)} \\ &= \text{Cst}_3 \exp\left(-\Psi(x,y)\right) \end{aligned}$$

where

$$\text{Cst}_3 = \frac{\text{Cst}_1}{\text{Cst}_2 \sum_{\xi \in \Omega, \Omega = E^M} \exp\left(\frac{-U(\xi)}{T}\right)}$$

$$\Psi(x,y) = \Psi_1(x,y) + \frac{U(x)}{T} \tag{14}$$

Maximizing the probability $P[X = x \mid Y = y]$ is equivalent to minimizing the function $\Psi(x,y)$.

$$x^* = \underset{x \in \Omega, \Omega = E^M}{\arg \min} \{\Psi(x,y)\} \tag{15}$$

The computation of the exact segmentation x^* is impossible (Geman and Geman, 1984). Therefore optimization techniques are necessary to compute an approximate solution $\hat{x}$.

Let $\mu = (\mu_1, \ldots, \mu_j, \ldots, \mu_K)$ be the means and $\sigma = (\sigma_1, \ldots, \sigma_j, \ldots, \sigma_K)$ be the standard deviations of K classes in the segmented image $x = (x_1, \ldots, x_s, \ldots, x_M)$ *i.e.,*

$$\begin{cases} \mu_j = \frac{1}{|S_j|} \sum_{s \in S_j} y_s \\ \\ \sigma_j = \sqrt{\frac{1}{|S_j|} \sum_{s \in S_j} (y_s - \mu_j)^2} \\ \\ S_j = \{s \mid x_s = j\} \end{cases} \tag{16}$$

Our way to minimize the function $\Psi(x,y)$ is to minimize instead the function $\Psi(\mu)$. We can always compute x through μ by classifying y_s into the nearest mean μ_j *i.e.,* $x_s = j$ if the nearest mean to y_s is μ_j. Thus instead of looking for x^*, we look for μ^*.

$$\begin{cases} \mu^* = \arg_{\mu \in [0 \ldots 255]^K} \min \{\Psi(\mu)\} \\ \\ \Psi(\mu) = \sum_{j=1}^{K} f(\mu_j) \\ \\ f(\mu_j) = \sum\limits_{s \in S_j} [\ln(\sigma_j) + \frac{(y_s - \mu_j)^2}{2\sigma_j^2}] + \frac{U(x)}{T} \end{cases} \tag{17}$$

where $U(x)$ and S_j are defined in equations (11) and (16).

To apply unconstrained optimization techniques, we redefine the function $\Psi(\mu)$ for $\mu \in \mathbb{R}^K$ instead of $\mu \in [0 \ldots 255]^K$. For that, the new function $\Psi(\mu)$ becomes as follows:

$$\Psi(\mu) = \begin{cases} \sum_{j=1}^{K} f(\mu_j) \text{ if } \mu \in [0 \ldots 255]^K \\ +\infty \text{ otherwise} \end{cases} \tag{18}$$

3 THE CONJUGATE GRADIENT (CG) ALGORITHM

In practice, we used the GNU Scientific Library implementation of Polak-Ribière Conjugate Gradient method (Polak and Ribière, 1969; Grippo and Lucidi, 1997) (gsl_multimin_fdfminimizer_conjugate_pr).

To use Conjugate Gradient Algorithm, we need the first derivative. Since no mathematical expression is available, it is approximated with finite differences (Eberly, 2003) as follows:

$$\begin{cases} \Psi'(\mu) = \left(\frac{\partial \Psi}{\partial \mu_1}, \ldots, \frac{\partial \Psi}{\partial \mu_n}\right) \\ \\ \frac{\partial \Psi}{\partial \mu_i} = \frac{\Psi(\mu_1, \ldots, \mu_i + \varepsilon, \ldots, \mu_n) - \Psi(\mu_1, \ldots, \mu_i - \varepsilon, \ldots, \mu_n)}{2\varepsilon} \end{cases} \tag{19}$$

The good approximation of the first derivative relies on the choice of the value of the parameter ε. Through the tests conducted, we have selected 0.01 as the best value.

4 EXPERIMENTAL RESULTS

To evaluate the segmentation quality we have used Dice Coefficient (Dice, 1945).
Dice Coefficient (DC) measures how much the result is close to the ground truth. Let the resulting class be $\hat{A}$ and its ground truth be A^*. The Dice Coefficient is given by the following formula:

$$DC = \frac{2|\hat{A} \cap A^*|}{|\hat{A} \cup A^*|} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \qquad (20)$$

where TP stands for true positive, FP for false positive and FN for false negative. DC equals 1 in the best case *i.e.,* $\hat{A}$ and A^* are identical and equals 0 in the worst case *i.e.,* there is an empty intersection between $\hat{A}$ and A^*.

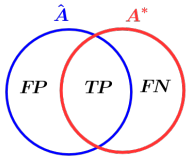


Figure 2: TP, FP and FN.

We have compared the combination method HMRF-CG to some methods: MRF-Classical (Yousefi et al., 2012), MRF-ACO-Gossiping (Yousefi et al., 2012) and MRF-ACO (Salima and Mohamed, 2003).

Table 1 shows the mean DC values of the three classes: GM (Grey Matter), WM (White Matter) and CSF (Cerebro Spinal Fluid). The images are obtained form the IBSR database. The slices used are: 1-24/18, 1-24/20, 1-24/24, 1-24/26, 1-24/30, 1-24/32 and 1-24/34. The parameters used by HMRF-CG are: Temperature $T = 10$ and the initial point $\mu^0 = (1,5,140,190)$. The parameters used by the others methods are given in (Yousefi et al., 2012; Salima and Mohamed, 2003).

Table 2 shows mean DC values for the three classes: GM (Grey Matter), WM (White Matter) and CSF (Cerebro Spinal Fluid) of HMRF-CG and LGMM (Liu and Zhang, 2013) methods. The images are chosen from Brainweb databases with Modality= T1, Slice thickness = 1mm. The slices used are: 85, 88, 90, 95, 97, 100, 104, 106, 110, 121 and 130. The parameters used by the LGMM method are given in

Table 1: Mean DC values (the best results are given in bold type).

Methods	Dice Coefficient			
	GM	WM	CSF	Mean
Classical-MRF	0.771	0.828	0.253	0.617
MRF-ACO	0.778	0.827	0.263	0.623
MRF-ACO-Gossiping	0.778	0.827	0.262	0.623
HMRF-CG	**0.859**	**0.855**	**0.381**	**0.698**

LGMM (Liu and Zhang, 2013). The parameters used by HMRF-CG are: Temperature T is given in the Table 2 and the initial point $\mu^0 = (1,45,110,150)$. The first column (N,I) gives the noise and the intensity non-uniformity.

Table 2: Mean DC values (the best results are in bold type).

(N,I)	HMRF-CG					LGMM			
	Dice Coefficient					Dice Coefficient			
	T	GM	WM	CSF	Mean	GM	WM	CSF	Mean
(0%,0%)	10	**0.970**	**0.990**	**0.961**	**0.974**	0.697	0.667	0.751	0.705
(3%,20%)	4	**0.940**	**0.965**	**0.940**	**0.949**	0.905	0.940	0.897	0.914
(5%,20%)	1	**0.918**	**0.952**	**0.924**	**0.931**	0.912	0.951	0.893	0.918

Figure 3 shows a sample of images to segment from IBSR databases.

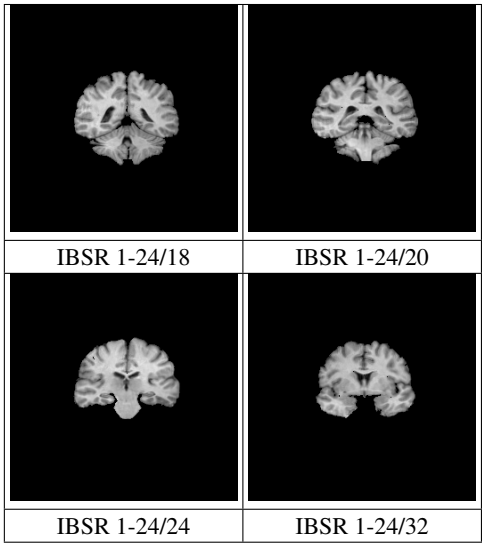


Figure 3: A sample of images to segment from IBSR database.

Figure 4 shows a sample of segmented images using HMRF-CG method.

Figure 5 shows a sample of ground truths images from IBSR database.

Figure 6 shows a sample of images to segment from Branweb databases and segmented images using HMRF-CG. The first column (N,I) gives the noise, the intensity non-uniformity and the temperature.

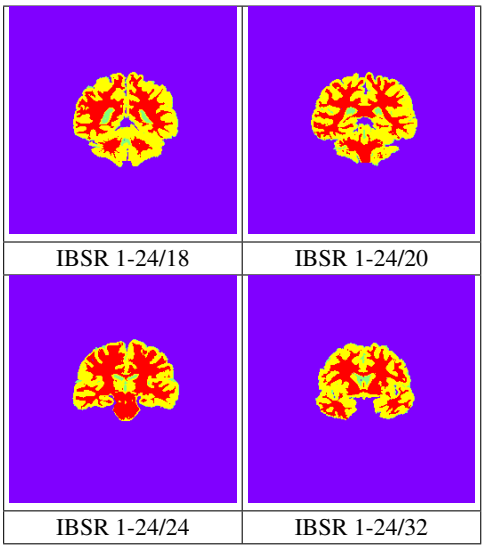


Figure 4: A sample of segmented images using HMRF-CG.

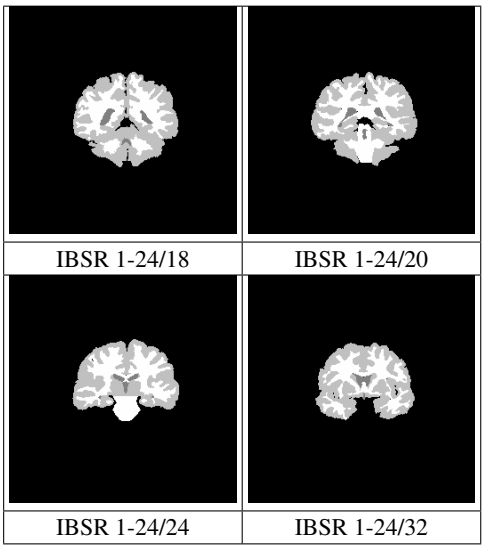


Figure 5: A sample of ground truths images from IBSR database.

5 DISCUSSION AND CONCLUSION

In this paper, we have described a method which combines Hidden Markov Random Field (HMRF) and Conjugate Gradient (GC). Segmentation quality measurement by Dice Coefficient was carried out on a sample of brain medical images from the well known databases IBSR and Brainweb.

The segmentation quality using HMRF-CG method depends on the choice of parameters (the initial point and temperature). This very sensitive task was conducted by performing numerous tests.

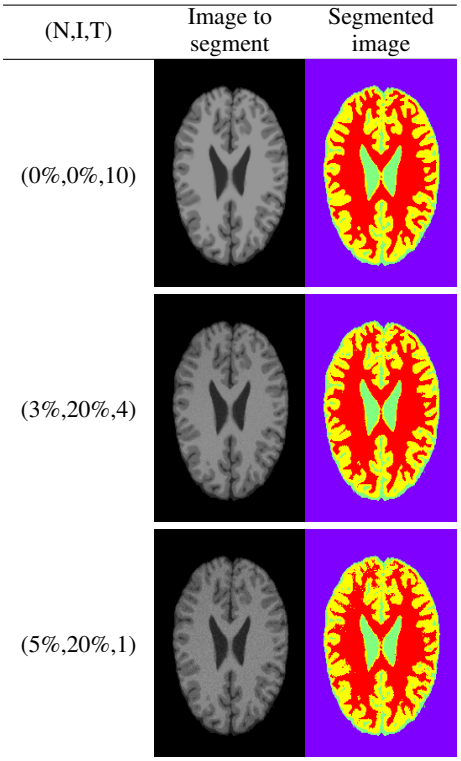


Figure 6: The slices number #97 with different noise and intensity non-uniformity from Brainweb databases and their segmentation using HMRF-CG.

From the results obtained, the HMRF-GC method outperforms the methods tested that are: LGMM, Classical MRF, MRF-ACO-Gossiping and MRF-ACO. Tests permit to find good parameters (initial points and temperature) for HMRF-CG, to achieve good segmentation results. Nevertheless, a statistical study is still needed to confirm the choice of parameters.

Acknowledgment

We would like to thank BRAINWEB and IBSR communities for granting the use of their databases in the evaluation of brain segmentation.

REFERENCES

Benson, C., Lajish, V., and Rajamani, K. (2015). Brain tumor extraction from MRI brain images using marker based watershed algorithm. pages 318–323.

Canny, J. (1986). A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (6):679–698.

Chan, T. F., Vese, L., et al. (2001). Active contours without edges. *Image processing, IEEE transactions on*, 10(2):266–277.

Chuang, K.-S., Tzeng, H.-L., Chen, S., Wu, J., and Chen, T.-J. (2006). Fuzzy c-means clustering with spatial information for image segmentation. *computerized medical imaging and graphics*, 30(1):9–15.

Cocosco, C. A., Kollokian, V., Kwan, R. K.-S., Pike, G. B., and Evans, A. C. (1997). Brainweb: Online interface to a 3d MRI simulated brain database.

Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302.

Eberly, D. (2003). Derivative approximation by finite differences. *Magic Software, Inc*.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (6):721–741.

Grippo, L. and Lucidi, S. (1997). A globally convergent version of the Polak-Ribière conjugate gradient method. *Mathematical Programming*, 78(3):375–391.

Hammersley, J. M. and Clifford, P. (1971). Markov fields on finite graphs and lattices.

Held, K., Kops, E. R., Krause, B. J., Wells III, W. M., Kikinis, R., and Muller-Gartner, H.-W. (1997). Markov random field segmentation of brain MR images. *Medical Imaging, IEEE Transactions on*, 16(6):878–886.

Ho, S., Bullitt, L., and Gerig, G. (2002). Level-set evolution with region competition: automatic 3-d segmentation of brain tumors. 1:532–535.

Kumar, S. et al. (2007). Skull stripping and automatic segmentation of brain MRI using seed growth and threshold techniques. pages 422–426.

Lin, G.-C., Wang, W.-J., Kang, C.-C., and Wang, C.-M. (2012). Multispectral MR images segmentation based on fuzzy knowledge and modified seeded region growing. *Magnetic resonance imaging*, 30(2):230–246.

Liu, J. and Zhang, H. (2013). Image segmentation using a local gmm in a variational framework. *Journal of mathematical imaging and vision*, 46(2):161–176.

Masoumi, H., Behrad, A., Pourmina, M. A., and Roosta, A. (2012). Automatic liver segmentation in MRI images using an iterative watershed algorithm and artificial neural network. *Biomedical Signal Processing and Control*, 7(5):429–437.

McInerney, T. and Terzopoulos, D. (1996). Deformable models in medical image analysis: a survey. *Medical image analysis*, 1(2):91–108.

Møller, M. F. (1993). A scaled conjugate gradient algorithm for fast supervised learning. *Neural networks*, 6(4):525–533.

Natarajan, P., Krishnan, N., Kenkre, N. S., Nancy, S., and Singh, B. P. (2012). Tumor detection using threshold operation in MRI brain images. pages 1–4.

Panjwani, D. K. and Healey, G. (1995). Markov random field models for unsupervised segmentation of textured color images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 17(10):939–954.

Perona, P. and Malik, J. (1990). Scale-space and edge detection using anisotropic diffusion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 12(7):629–639.

Pham, D. L., Xu, C., and Prince, J. L. (2000). Current methods in medical image segmentation 1. *Annual review of biomedical engineering*, 2(1):315–337.

Polak, E. and Ribière, G. (1969). Note sur la convergence de méthodes de directions conjuguées. *Revue française d'informatique et de recherche opérationnelle, série rouge*, 3(1):35–43.

Powell, M. J. D. (1977). Restart procedures for the conjugate gradient method. *Mathematical programming*, 12(1):241–254.

Roura, E., Oliver, A., Cabezas, M., Vilanova, J. C., Rovira, À., Ramió-Torrentà, L., and Lladó, X. (2014). Marga: multispectral adaptive region growing algorithm for brain extraction on axial MRI. *Computer methods and programs in biomedicine*, 113(2):655–673.

Salima, O. and Mohamed, B. (2003). MRF-based image segmentation using ant colony system. 2:012–24.

Senthilkumaran, N. and Rajesh, R. (2009). Edge detection techniques for image segmentation–a survey of soft computing approaches. *International journal of recent trends in engineering*, 1(2).

Shewchuk, J. R. (1994). An introduction to the conjugate gradient method without the agonizing pain. *Lecture available on internet*.

Swendsen, R. H. and Wang, J.-S. (1987). Nonuniversal critical dynamics in Monte Carlo simulations. *Physical review letters*, 58(2):86.

Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., Tappen, M., and Rother, C. (2008). A comparative study of energy minimization methods for Markov random fields with smoothness-based priors. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(6):1068–1080.

Wang, L., Shi, F., Li, G., Gao, Y., Lin, W., Gilmore, J. H., and Shen, D. (2014). Segmentation of neonatal brain MR images using patch-driven level sets. *NeuroImage*, 84:141–158.

Wu, Z. and Leahy, R. (1993). An optimal graph theoretic approach to data clustering: Theory and its application to image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 15(11):1101–1113.

Wyatt, P. P. and Noble, J. A. (2003). Map MRF joint segmentation and registration of medical images. *Medical Image Analysis*, 7(4):539–552.

Yousefi, S., Azmi, R., and Zahedi, M. (2012). Brain tissue segmentation in MR images based on a hybrid of MRF and social algorithms. *Medical image analysis*, 16(4):840–848.

Zhang, Y., Brady, M., and Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *Medical Imaging, IEEE Transactions on*, 20(1):45–57.

Zhao, M., Lin, H.-Y., Yang, C.-H., Hsu, C.-Y., Pan, J.-S., and Lin, M.-J. (2015). Automatic threshold level set model applied on MRI image segmentation of brain tissue. *Appl. Math*, 9(4):1971–1980.