

Robust Brain Segmentation Using Histogram Scale-Space Analysis and Mathematical Morphology

J.-F. Mangin¹, O. Coulon^{1,2}, and V. Frouin¹

¹ Service Hospitalier Frédéric Joliot, Département de Recherche Médicale, Direction des Sciences du Vivant, Commissariat à l'Énergie Atomique
E-mail: mangin@shfj.cea.fr, P-mail: place du Général Leclerc, 91406 Orsay, France

² Département Signal et Image, ENST Paris

Abstract. In this paper, we propose a robust fully non-supervised method dedicated to the segmentation of the brain in T1-weighted MR images. The first step consists in the analysis of the scale-space of the histogram first and second derivative. We show first that the crossings in scale-space of trajectories of extrema of different derivative orders follow regular topological properties. These properties allow us to design a new structural representation of a 1D signal. Then we propose an heuristics using this representation to infer statistics on grey and white matter grey level values from the histogram. These statistics are used by an improved morphological process combining two opening sizes to segment the brain. The method has been validated with 70 images coming from 3 different scanners and acquired with various MR sequences.

1 Introduction

Brain segmentation in magnetic resonance (MR) T1-weighted images has been one of the most addressed applications of image analysis in the field of medical imaging. Indeed, brain segmentation is usually a first step before detection of embedded anatomical or pathological structures (hemispheres, tissues, cortical folds, deep nuclei, ventricles, tumors, vascular lesions...), registration with other modalities or atlases, and 3D visualisation (neurosurgical planning, brain mapping...). A lot of methods have been proposed, but none of them has reached a level of robustness sufficient to be used routinely by clinicians. Usual methods rely on a few parameters which have to be tuned manually to different sets of acquisition parameters. In this paper we propose a new fully non supervised approach designed to assure a high quality result for a large range of MR sequences.

In the following, we assume that the only image which has been acquired is a 3D T1-weighted MR image covering most of the head. Situations where T2-weighted or proton-density images can also be used could lead to further developments integrating their additional discriminative power. We do not address such developments, first because such images are not always acquired due to timing considerations, second because slice thickness of T1-weighted images is usually much lower, finally because T1-weighted images are sufficient for our brain segmentation purpose.

The wide variety of methods proposed in literature to segment the brain in T1-weighted images can be browse according to the few recurrent schemes involved. It should be noted that a simple classification scheme based for instance on texture analysis is a priori not sufficient to solve the problem. Therefore, the more frequent approach consists in using mathematical morphology or region growing algorithms to discriminate the brain from the surrounding structures in a preclassified binary image [9,6]. Multi-resolution or scale-space analysis appears to be an interesting alternative to the previous scheme [11]. Other approaches relying on the deformable model paradigm seem to be more difficult to master mainly because of initialisation difficulties [10].

Our paper proposes several improvements of the mathematical morphology based scheme, which has been the most successful one. These improvements lead to a segmentation method robust for a large range of images. This method has been validated on a set of 70 images acquired on three different scanners using gradient-echo and inversion-recovery sequences with various parameters. In the next section, a rapid survey of the mathematical morphology based scheme will help us to catch some of its intrinsic weaknesses which will be overcome throughout the paper.

2 The classical scheme

2.1 Histogram analysis

The first step consists in binarizing the initial grey level image. All voxels which could belong to the brain are selected according to a range [Tlow, Thigh] of grey level values supposed to include grey matter and white matter tissues (cf. Fig. 1). This binarization aims at disconnecting as much as possible the brain from surrounding structures. The low threshold eliminates the cerebrospinal fluid (CSF) and the skull while the high threshold eliminates the fat and the vascular system. Since the MR signal do not correspond to an absolute measure of some physical property, both thresholds have to be tuned for each image. This adaptation is difficult to perform automatically because of large variations of the image contrast according to the MR sequence parameters (spin-echo, gradient-echo, inversion-recovery, echo time, repetition time, excitation number, slice thickness...) and according to the subjects (anatomy, pathologies...) (cf. Fig. 2).

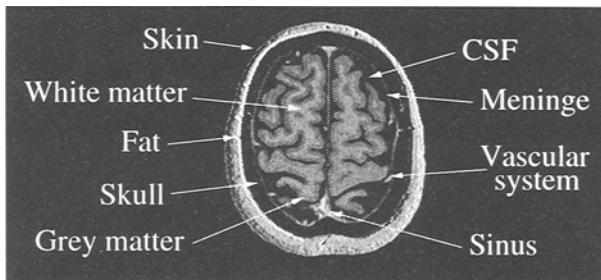


Fig. 1. A simple description of the various head tissues observed in T1-weighted MR images.

The only a priori knowledge on the MR signal nature which seems to be invariant across T1-weighted sequences is the relative positions of tissues along

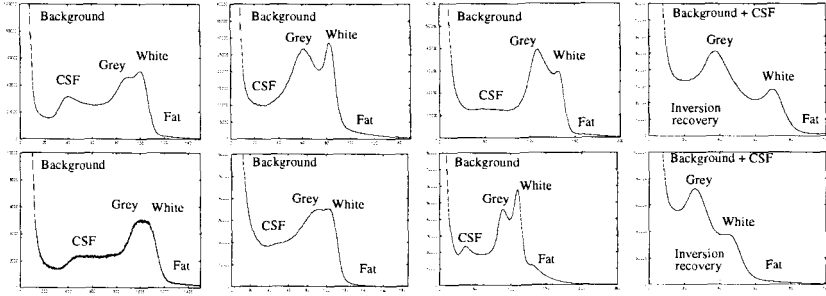


Fig. 2. The histograms of 3D T1-weighted MR images turn out to be especially variable.

the grey level scale. The first part of this paper proposes a fully non supervised process using this structural information to analyse the scale-space of the image histogram. This analysis gives statistical parameters on brain tissue grey levels from which is computed the binarization. Because the classical “finger print” analysis based on first derivative extrema fails for some configurations where grey and white matter are especially mixed, we have designed a new histogram mode detection method using the structure of the two first derivative extrema in the scale-space. Some crossings of extremum trajectories follow simple topological properties which help to analyse the scale-space structure according to a priori knowledge.

2.2 Morphological processing

Unfortunately, a reasonable estimation of the two thresholds is not always sufficient to assure a good behaviour of the segmentation process. Indeed, the quality of the result can be significantly lowered by a slight modification of the thresholds, especially the lowest one. Moreover, for some images, which can appear visually very good, not a pair of thresholds give acceptable segmentation. This unstability is easy to understand from a description of the process (cf. Fig. 3).

In order to achieve a whole disconnection between the brain and the surrounding structures preserved by the binarization (with “good” thresholds: scalp, meninges, eyes, sinuses), different methods have been proposed. They all rely on the effect of the morphological “opening” operation. The binary image is eroded to cut the remaining connections. Then the largest or the more central 3D connected component is selected as a seed of the brain. Finally the brain shape is recovered by a geodesic dilation of the seed conditionally to the initial binary image (which is often related to region growing).

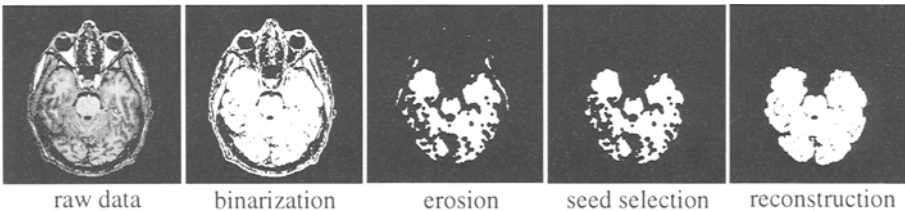


Fig. 3. The classical morphological scheme to segment the brain in T1-weighted MR images.

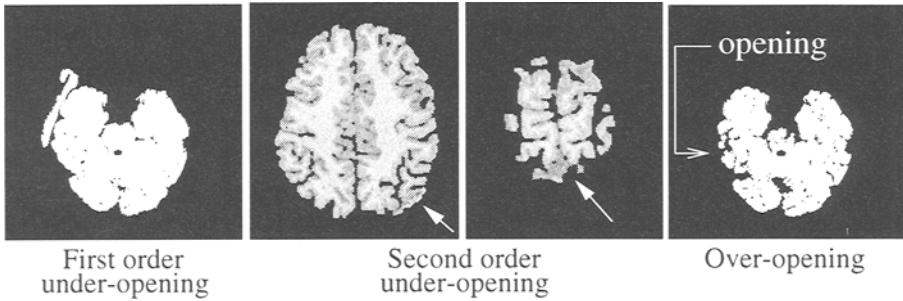


Fig. 4. The classical morphological scheme (cf. Fig. 3) presents intrinsic weaknesses which can lead to three kinds of problems.

The erosion and dilation structuring element diameters, like the two thresholds, influence largely the quality of the final result. Three kinds of failures can occur, sometimes simultaneously and with various amplitudes (cf. Fig. 4). First, the eyes or a large part of the scalp can be included in the brain segmentation which means that the erosion has not done its job, either because the binarization was too permissive or because the erosion diameter was too low. These situations will be called first order under-opening problems. Second, a part of the brain like a gyrus can be removed because of the opening effect of the erosion-dilation sequence. This can happen when the binarization is too selective or when the erosion diameter is too large. We will talk about over-opening problems. The unstability of the process stems from the two conflictual constraints imposed to the parameter choice by these two opposite kinds of problems. The domain of acceptable parameters can be especially narrow or even empty. Finally, the last kind of errors called second order under-opening problems correspond to connections with some small pieces of sinus or meninges. These problems are the more difficult to get completely rid of because they often occur when previous problems have been solved.

In this paper, the classical morphological process is modified in order to obtain a correct behaviour for a large parameter domain. A first modification relies on a simple regularization of the binarization. Then a two step geodesic reconstruction process restricts the opening effect potentially removing gyri to the smallest one. It should be noted that because of frequent non stationarities in MR images, simple thresholding operations are not always sufficient to get good binarizations for the whole brain. In the following, we assume that the amplitude of these non stationarities is slight, otherwise a preprocessing could be required which is beyond the scope of the paper [15].

3 1D scale-space analysis

In this section, we described the automatic analysis of the histogram scale-space which leads to an estimation of mean and standard deviations of grey and white matter grey levels. Each tissue class is represented by a specific histogram mode. The more important ones, which are related to background, CSF, grey matter, white matter, fat and vascular system, always appear with the same order along the grey scale (cf. Fig. 2). Various approaches have been proposed to deal with the detection of these modes, including K-means [6] and fit with a sum of Gaus-

sians or with other a priori models [13]. Our own unsuccessful experiments with such methods when dealing with a wide set of sequence parameters led us to search for a more robust approach.

3.1 Cascades of singularities

Linear scale-space analysis is an appealing approach when dealing with 1D signals because of the surprisingly simple geometry of extremum trajectories induced by the causality property [16,5]. Indeed, extrema of the signal and its first derivatives often have direct semantic interpretations, which make them good candidates for deriving structural descriptions. Usual approaches rely only on extrema of one single derivative. For instance, finger prints stem from first derivative extrema [16] while blob based primal sketches stem from signal (or Laplacian) extrema [7,2]. Finger print based analysis has been proposed to initialise the decomposition of 1D signals in mixtures of Gaussians [1,4]. In the case of MR image histograms, the Gaussian hypothesis may appear justified at first glance. In fact such an hypothesis does not take into account partial volume effect, signal non stationarities, Gibbs artefacts, subject motions and other MRI specific artefacts. Therefore, to get a more robust approach, we make the histogram mode detection rely directly on histogram scale-space.

The goal of our analysis is the detection of grey and white matter modes. Because these two neighboring modes can be especially mixed, histogram and first derivative extrema are not sufficient in all cases to detect them (cf. Fig. 5). Simple experiments with sums of Gaussians will show that higher derivatives have a better detection power in such situations.

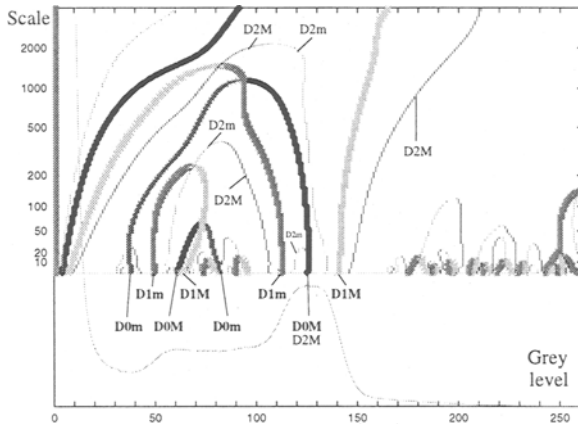
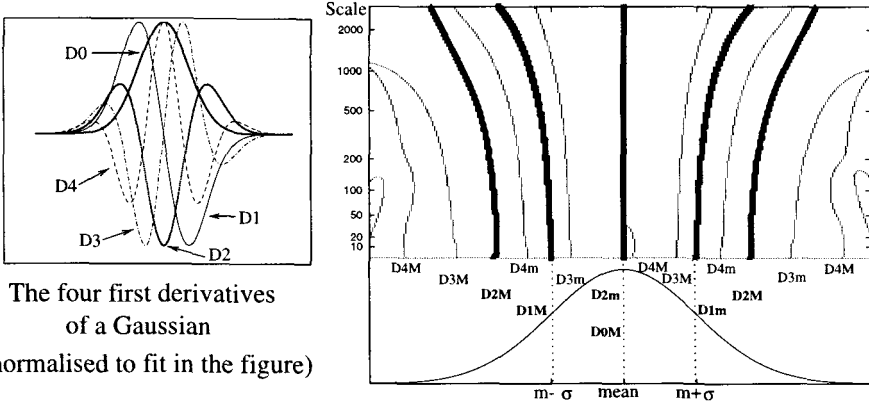


Fig. 5. The derivative extrema of a histogram (D_0) and its two first derivatives (D_1 and D_2) in the scale-space. DiM denotes a maximum while Dim denotes a minimum. Grey and white matter modes are too mixed to be detected from D_0 and D_1 extrema. In return, D_2 extrema reveal the mixture.

Straightforward considerations on 1D functions with zero values on their domain bounds lead to the recursive property that N extrema in the i^{th} derivative implies at least $(N + 1)$ nested extrema in the $(i + 1)^{th}$ derivative. Thus one single Gaussian leads to one signal extremum, two first derivative extrema, three second derivative extrema, etc (cf. Fig. 6). These extrema can be tracked



The four first derivatives of a Gaussian (normalised to fit in the figure)

Fig. 6. The derivative extrema of a binarized Gaussian in the scale-space computed from heat equation. D_iM denotes a maximum while D_im denotes a minimum of the i^{th} derivative. Extrema which do not reach scale 100 have not been tracked.

throughout the scale-space (cf. Fig. 6). The sum of two Gaussians lead to different extremum patterns according to their relative localizations, amplitudes, means and standard deviations. Intuitively, when the two Gaussians are sufficiently distant from each other, the pattern of the sum extrema corresponds to the juxtaposition of two single Gaussian extrema patterns. When the Gaussians are brought closer to each other, some extrema disappear. For very closed Gaussians, the sum extrema pattern differs from the single Gaussian extrema pattern only for high derivative extrema.

Because of the nature of the linear scale-space, which amounts to convolutions with Gaussians (the Gaussian is the Green function of the heat equation), a sum of two Gaussians remains a sum of two Gaussians throughout the whole scale-space. Therefore, since smoothing the sum is similar to bringing the two Gaussians “closer” to each other (while increasing their standard deviations), the structure of the sum extrema trajectories in the scale-space corresponds to a sequence of extremum extinction (cf. Fig. 7). These extinctions, which are the well known singularities related to scale-space bifurcations [7], follow a simple rule giving rise to what we will call singularity cascades.

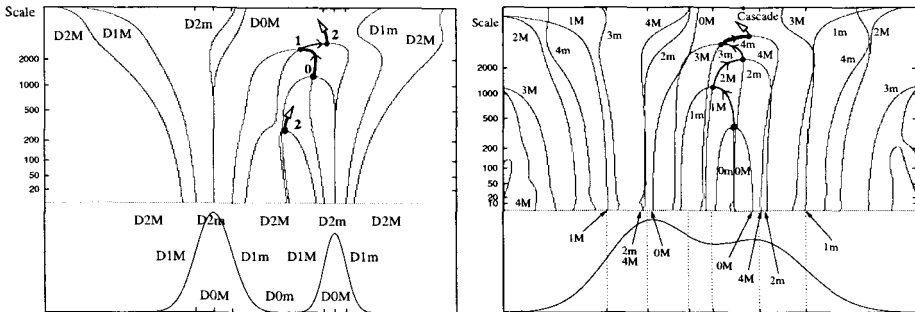


Fig. 7. The derivative extrema of the sum of two Gaussians in the scale-space giving rise to cascades of singularities (left: two cascades of order 0 and 2 are shown, right: the cascade of order 0 is followed until the singularity of order 4).

The simple syntactic rule leading to the nice nested structure of cascades stems from the fact that the order i singularities are the only locations in scale-space where the trajectories of the i^{th} and $(i+1)^{\text{th}}$ derivative extrema can “meet” each other. This can be understood first if one look at the expression of the drift velocity of a i^{th} derivative extremum under scale-space smoothing. Let $D_i : \mathbb{R}_+ \times \mathbb{R}_+ \rightarrow \mathbb{R}$ be the scale-space representation of the histogram i^{th} derivative. In non-degenerate situations, an estimate of a D_i extremum velocity $v_i(g, s)$ (where g denotes the grey level and s denotes the scale) along its trajectory defined by $D_{i+1}(g, s) = 0$, is given by the implicit function theorem [7]:

$$v_i(g, s) = -\frac{1}{2} \frac{D_{i+3}(g, s)}{D_{i+2}(g, s)}, D_{i+1}(g, s) = 0. \quad (1)$$

Hence, a crossing between $D_{i+1}(g, s) = 0$ and $D_{i+2}(g, s) = 0$ would lead to a degenerate situation where the drift velocity tends to infinity. Therefore singularities of order i are the only locations where such trajectories can meet.

Second, since at least one extremum of the $(i+1)^{\text{th}}$ derivative has to exist between a pair of minimum and maximum of the i^{th} derivative at every scale where this pair exists, one $(i+1)^{\text{th}}$ derivative extremum trajectory is bound to cross a pair extinction. Then, the structure of extinctions of Fig. 7 is invariant whatever the two Gaussian parameters are, namely one cascade of order 0 linking singularities of order 0, 1, 2 and more (cf. Fig. 7), one cascade of order 2 linking singularities of order 2 and more, and higher order cascades which do not appear in the figures.

Previous observations on singularity cascades lead to the conclusion that the mixture of two histogram modes can be untangled if high derivative extrema are used. Indeed, the mixture of very close modes will simply lead to cascades of higher order than usual. With MRI histograms, second derivative extrema turn out to be sufficient to deal with the grey and white matter mode detection for all the configurations in our database.

If we assume now that a MRI histogram mode is relatively symmetric, a good estimation of its mean value is provided by the related second derivative minima. Then, an estimation of standard deviation can be derived from the neighboring first derivative extrema (cf. Fig. 6). Therefore, we have devised a method using the cascade property to detect simultaneously first and second derivative extremum trajectories related to grey and white matter modes.

3.2 From cascades to modes

In the following an algorithm is described which uses the notion of cascade to construct a structural representation of the histogram scale-space based on first and second derivative extremum trajectories. Then, an heuristics uses this representation first to discriminate inversion-recovery MR sequences from other ones, second to detect trajectories related to grey and white matter modes.

Whatever the histogram is, if one looks high enough in the scale-space, only one second derivative minimum remains alive (cf. Fig. 6). The first step of the process consists in computing the scale-space of the histogram first and second derivatives until the first scale where only one second derivative minimum exists. This is done using the usual discretization of heat equation [7]. Then all extrema

We assume that two types of families exist, which are related to the two cascades of Fig. 7. Those which are related to the order 0 cascade include the $D2m$ trajectory of one specific histogram mode. They can be full or childless. The other families, related to order 2 cascade of Fig. 7 are just bringing the $D2M$ trajectories required to complete the signatures. They are necessarily childless. A family volume corresponds to the integral of the histogram over a specific range which is illustrated in Fig. 8.

The two biggest families are selected. Then the two scales $s1$ and $s2$ where occur their respective order 2 singularities are compared. If $s2/s1 > 0.25$ where $s1 > s2$, the histogram is supposed to stem from an IR sequence. Indeed, this ratio discriminates situations where grey and white matter modes merge into a "brain mode" which survive during a long scale range (standard sequences) versus situations where the brain mode has a very short life time (IR sequences) (cf. Fig. 9 and 10). This simple rule led to a 100% successful discrimination with our database (20 IR sequences, 50 standard sequences).

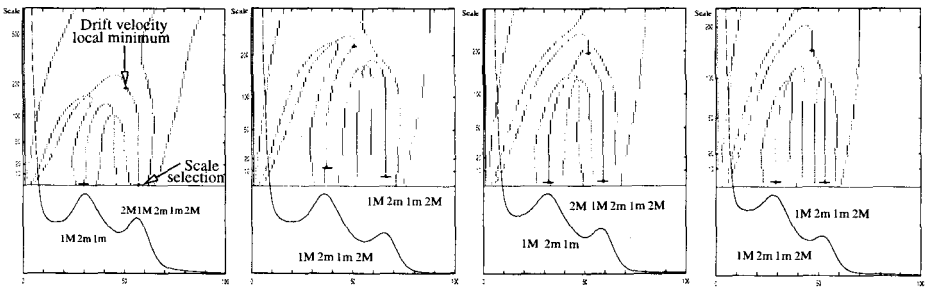


Fig. 9. The signatures and scale selections of grey and white matter modes in four different histograms stemming from Inversion Recovery MR sequences.

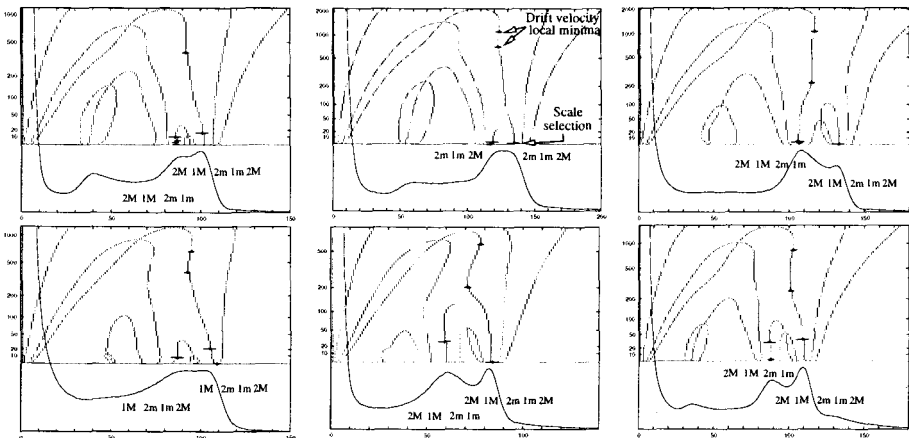


Fig. 10. The signatures and scale selections of grey and white matter modes in six different histograms stemming from standard $T1$ -weighted MR sequences.

Then, in the IR situation, the pattern matching algorithm previously described is applied to the trajectory set made up by the infinity family and the

two biggest families. This analysis provides the grey and white matter signatures (cf. Fig. 9).

In the other situations, a first signature analysis is applied to the trajectory set made up by the infinity family and the family endowed with the highest order 2 singularity in scale-space. This first analysis provides the “brain mode” signature. Then two new families are added to the trajectory set. The first one is selected as the highest family (once again highest means highest order 2 singularity in scale-space) which $loc_0(D2m)$ is located in the range $[loc_0(D1M_{brain}), loc_0(D1m_{brain})]$, where $D1i_{brain}$ denotes order 1 trajectories belonging to the brain mode signature and loc_0 denotes location at 0 scale. The second one is selected as the highest family which $loc_0(D2m)$ is located in the range $[loc_0(D1m_{background}), loc_0(D1M_{brain})]$. Finally a second analysis is applied with this four family set which provides the grey and white matter signatures (cf. Fig. 10).

3.3 Scale selection

In this section, we describe how mean and standard deviation of a mode are estimated from its signature. Scale selection is one major problem of linear scale-space analysis, mainly because the hypothesis that an extremum trajectory in scale-space arises from a single event is rarely true. Therefore, the simple “select at high scale and track until lowest scale” strategy is far from being sufficient. Intuitively, at highest scales extrema are related to a mixture of modes (like the brain) while at lowest scales they are related to noise or minor histogram modes. We propose an heuristics based on local minima of the second derivative minimum drift absolute velocity. This particular points turn out to be especially similar from histogram to histogram which make them good candidates to develop a new kind of primal sketch (cf. Fig. 9 and 10), but such an approach is beyond the scope of this paper. Our heuristics stems from the idea that extremum trajectories alternate periods of high drift velocity with periods of stability which could be related to scale ranges where they are caught by some underlying event. It should be noted that drift velocity minima are also strongly related to crossings between order i and order $(i+2)$ trajectories, which can be understood from Eq. 1 (cf. Fig. 7).

Using Eq. 1 to detect velocity minima leads to bad results induced by trajectory discretization effect. Therefore we have developed a more robust approach which inferred drift velocity estimations with discrete computations. The trajectory is considered has a sequence of segments of constant abscisse. Then each segment is endowed with a velocity inversely proportional to its length. Finally, using also velocity sign, all absolute velocity minima are detected.

This method provides the points pointed out in Fig. 9 and 10. Then, the scale s_{mode} from which is estimated a mode mean value is given by the first local minimum of the drift velocity appearing along the $D2m_{mode}$ trajectory under the lowest order 2 singularity of the whole family set. Finally, the two “standard deviations” are estimated from the $D1M_{grey}$ and $D1m_{white}$ trajectories, which always exist.

The whole histogram analysis algorithm has been applied with success on the 70 histograms of our database. The analysis takes less than one second on a standard sparc station.

4 Morphological process

Now, we propose briefly an improved morphological process qualitatively robust to small variations of the values extracted from the histogram.

4.1 Regularized binarization

The first step of the process consists in a regularized binarization. The regularizing effect is classically obtained from the classical Markovian approach relying on Ising model [3, 9]. The closing effect on the binarization allows us to use a relatively selective low threshold $m_G - 2\sigma_G$. This improves the effect of the following erosion and reduces the risk of first order under opening (cf. Fig. 4). It should be noted that the regularization is restricted to a range of grey levels which limits the risk of creating thick connections between the brain and the surrounding tissues (meninges, sinus, skull marrow).

4.2 Brain segmentation

In order to reduce over-opening effects often induced by the standard morphological process (cf. Fig. 3, 4), we propose to combine two morphological openings related to two different structuring element diameters. The largest opening \mathcal{O}_l will have to fully disconnect the brain from surrounding tissues, leading to the brain seed. The smallest opening \mathcal{O}_s will have to define the final brain shape, which means mainly removing meninges and noise. The whole process can be summarized as follows:

- Apply \mathcal{O}_s opening to binarization B to get object $\mathcal{O}_s(B)$ (structuring element: 26-neighbourhood);
- Apply 3mm erosion to $\mathcal{O}_s(B)$ (structuring element: 3mm radius ball) [8];
- Select largest 26-connected component as brain seed S ;
- Compute geodesic chamfer distance to S conditionally to $\mathcal{O}_s(B)$ (denoted $d(S))_n$;
- Threshold the previous distance to get brain complement $T_{>8mm}(d(S))$;
- Dilate brain complement conditionally to $T_{>4mm}(d(S))$ until convergence;
- Removes the dilation result from $T_{<=8mm}(d(S))$ to get the final brain segmentation.

The largest opening \mathcal{O}_l is related to the 3mm radius ball structuring element, which is defined from a chamfer distance adapted to the voxel anisotropy [8]. This choice makes the opening effect of \mathcal{O}_l independent of the voxel geometry. The 26-neighborhood has been chosen for \mathcal{O}_s to assure a minimum opening effect in all directions whatever the voxel geometry is. Geodesic chamfer distances are efficiently implemented using a thick front propagation [14].

The whole segmentation has been applied to the 70 images of our database (40 healthy subjects/30 epileptic subjects involved in a neurosurgery protocol). The process takes about one minute on standard sparc stations. In 68 cases, the process was globally successful apart from some second order under-opening errors in a few cases and slight over-opening errors in temporal lobes for one specific MR sequence with non-homogeneities (see. Fig. 4). For one of the remaining cases, the \mathcal{O}_l structuring element radius had to be increased to 4mm to get a good result. The last case was presenting large non-homogeneities which call for a preprocessing step [15].

5 Conclusion

We have proposed a robust fully non-supervised method to segment the brain in T1-weighted MR images. Hence, this method can now be considered as a reliable preprocessing before more sophisticated image analysis approaches. The statistics inferred from the histogram can be used for other segmentation purposes when the contrast between grey and white matter is of interest. The scale-space based approach proposed in this paper could be used to analyse other kinds of histograms. Finally, the extension of the cascade notion to higher dimension signals would be very interesting but the question appears rather involved at first glance.

References

1. M. J. Carlotto. Histogram analysis using a scale-space approach. *IEEE PAMI*, 9(1):121–129, 1987.
2. O. Coulon, I. Bloch, V. Frouin, and J.-F. Mangin. Multiscale measures in linear scale-space for characterizing cerebral functional activations. In *Scale-Space'97, Utrecht*, number 1252 in LNCS, Springer, pp. 188–199, 1997.
3. S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE PAMI*, 6(6):721–741, 1984.
4. A. Goshtasby and W. D. O'Neil. Curve fitting by a sum of Gaussians. *CVGIP: Graphical Models and Image Processing*, 56(4):281–288, 1994.
5. J.J. Koenderink. The structure of images. *Biol. Cybernetics*, 50:363–370, 1984.
6. F. Kruggel and G. Lohmann. Automatical adaption of the stereotactical coordinate system in brain MRI datasets. In *XVth IPMI, Poultney, USA*, pages 470–476, 1997.
7. T. Lindeberg. *Scale-space theory in computer vision*. Kluwer, 1994.
8. J.-F. Mangin, I. Bloch, J. Lopez-Krahe, and V. Frouin. Chamfer distances in anisotropic 3D images. In *VII Europ. Signal Proces. Conf.*, pp 975–978, 1994.
9. J.-F. Mangin, V. Frouin, I. Bloch, J. Regis, and J. López-Krahe. From 3D magnetic resonance images to structural representations of the cortex topography using topology preserving deformations. *J. Math. Imag. Vision*, 5(4):297–318, 1995.
10. T. McInerney and D. Terzopoulos. Deformable models in medical image analysis: a survey. *Medical Image Analysis*, 1(2):91–108, 1996.
11. W. J. Niessen, K. L. Vincken, and M. A. Viergever. Comparison of multiscale representations for a linking-based image segmentation model. In *IEEE/SIAM MMBIA, San Francisco*, pp. 263–272, 1996.
12. I. Pollak, A. S. Willsky, and H. Krim. Scale space analysis by stabilized inverse diffusion equation. In *Scale-Space'97, LNCS-1252, Springer*, pp. 200–211, 1997.
13. L. V erard, P. Allain, J.-M. Trav ere, J. C. Baron, and D. Bloyet. Fully automatic identification of AC and PC landmarks on brain MRI using scene analysis. *IEEE TMI*, pages 610–616, 1997.
14. B. J. H. Verwer, P. W. Verbeek, and S. T. Dekker. An efficient uniform cost algorithm applied to distance transforms. *IEEE PAMI*, 11(4):425–428, 1989.
15. W. M. Wells III, W. E. L. Grimson, R. Kikinis, and F. A. Jolesz. Adaptive segmentation of MRI data. *IEEE TMI*, 15(4):429–442, 1996.
16. A.P. Witkin. Scale-space filtering. In *International Joint Conference on Artificial Intelligence*, pages 1019–1023, 1983.