Improved Watershed Transform for Medical Image Segmentation Using Prior Information

V. Grau*, A. U. J. Mewes, M. Alcañiz, *Member, IEEE*, R. Kikinis, and S. K. Warfield, *Member, IEEE*

***Abstract*—The watershed transform has interesting properties that make it useful for many different image segmentation applications: it is simple and intuitive, can be parallelized, and always produces a complete division of the image. However, when applied to medical image analysis, it has important drawbacks (oversegmentation, sensitivity to noise, poor detection of thin or low signal to noise ratio structures). We present an improvement to the watershed transform that enables the introduction of prior information in its calculation. We propose to introduce this information via the use of a previous probability calculation. Furthermore, we introduce a method to combine the watershed transform and atlas registration, through the use of markers. We have applied our new algorithm to two challenging applications: knee cartilage and gray matter/white matter segmentation in MR images. Numerical validation of the results is provided, demonstrating the strength of the algorithm for medical image segmentation.**

***Index Terms*—Biomedical imaging, image segmentation, morphological operations, tissue classification, watersheds.**

I. Introduction

A. The Watershed Transform

The watershed transform [1] is a popular segmentation method coming from the field of mathematical morphology. The intuitive description of this transform is quite simple: if we consider the image as a topographic relief, where the height of each point is directly related to its gray level, and consider rain gradually falling on the terrain, then the watersheds are the lines that separate the "lakes" (actually called *catchment basins)* that form. Generally, the watershed transform is computed on the gradient of the original image, so that the catchment basin boundaries are located at high gradient points.

The watershed transform has been widely used in many fields of image processing, including medical image segmentation, due to the number of advantages that it possesses: it is a simple, intuitive method, it is fast and can be parallelized (in [2], an almost linear speedup was reported for a number of processors up to 64), and it produces a complete division of the image in separated regions even if the contrast is poor, thus avoiding the need for any kind of contour joining. Furthermore, several researchers have proposed techniques to embed the watershed transform in a multiscale framework, thus providing the advantages of these representations [3]–[5]. Some important drawbacks also exist, and they have been widely treated in the related literature. Among the most important are as follows.

- *Oversegmentation:* When the watershed transform infers catchment basins from the gradient of the image, the result of the watershed transform contains a myriad of small regions, which makes this result hardly useful. The use of a marker image [6], [7] to reduce the number of minima of the image and, thus, the number of regions, is the most commonly used solution. Also interesting is the utilization of a scale space approach to select the interesting regions, using different filters (morphological operations [8], or nonlinear diffusion [9].)
- *Sensitivity to noise:* Local variations of the image can change dramatically the results. This effect is worsened by the use of high pass filters to estimate the gradient, which amplify the noise. Anisotropic filters have been used to minimize this problem [9], [10].
- *Poor detection of significant areas with low contrast boundaries:* If the signal to noise ratio is not high enough at the contour of interest, the watershed transform will be unable to detect it accurately. Furthermore, the watershed transform naturally detects the contours with higher value between markers, which are not always the contours of interest. A clear example is white matter-gray matter surface detection, where the proximity of other, higher contrast surfaces such as gray matter-CSF or CSF-bone will make the task difficult for the plain watershed transform.
- *Poor detection of thin structures:* When the watershed transform is applied on the gradient image, the smoothing associated with gradient estimation, together with usual approach of storing gradient values only at the image pixel positions rather than with sub-pixel accuracy, make it difficult to detect thin catchment basin areas. Often this is critical for successful segmentation of medical images.

Lately, much interest has been raised in the medical imaging community about segmentation algorithms that use active contours or surfaces. *Snake* models [11] were first used, while more recently another technique has received much attention:

Manuscript received June 16, 2003; revised November 4, 2003. This work was supported in part by the National Institutes of Health (NIH) under Grant P41 RR 13218, Grant P01 CA67165, Grant R01 RR1747, Grant R01 CA86879, and Grant R21 MH67054; in part by a research grant from the Whitaker Foundation; and in part by Acciones de Apoyo a la Investigación-Generalitat Valenciana and Programa de Incentivo a la Investigación-Universidad Politécnica de Valencia. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was S. Pizer. *Asterisk indicates corresponding author.*

*V. Grau is with the Surgical Planning Laboratory, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02138 USA and also with the MedicLab, Universidad Politécnica de Valencia, 46022 Valencia, Spain (e-mail: vgrauc@bwh.harvard.edu).

A. U. J. Mewes, R. Kikinis, and S. K. Warfield are with the Surgical Planning Laboratory, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02138 USA.

M. Alcañiz is with the MedicLab, Universidad Politécnica de Valencia, 46022 Valencia, Spain.

Digital Object Identifier 10.1109/TMI.2004.824224

the *level set* approach [12], which represents a deformable two-dimensional contour as a level curve of a three-dimensional (3-D) function that is deformed by internal and external forces. Though these approaches offer very nice behavior in many different applications, the watershed transform presents some advantages [13].

- The watershed lines always correspond to the most significant edges between the markers. So this technique is not affected by lower-contrast edges, due to noise, that could produce local minima and, thus, erroneous results, in energy minimization methods.
- Even if there are no strong edges between the markers, the watershed transform always detects a contour in the area. This contour will be located on the pixels with higher contrast.

B. Overview of the Work Presented

In this paper, we present a novel modification of the watershed transform that significantly improves the watershed transform performance by allowing the introduction of functions based on prior information. After a presentation of the standard watershed definition and the marker imposition technique in Sections II-A and II-B, the proposed improvement in this definition is introduced in Section II-C, along with an efficient algorithm for its calculation. We propose, as general-purpose functions applicable to a majority of medical image segmentation applications, the use of differences in probability maps, introduced in Section II-D.

Furthermore, we also propose an additional introduction of prior knowledge in watershed segmentation through the use of a statistical atlas. Details about this technique are given in Section II-E.

We explain the use of the newly proposed technique for two challenging medical image segmentation applications: knee cartilage and white matter/gray matter segmentation. Details about how the newly proposed techniques can be applied to these cases are given in Sections II-F1 and II-F2.

In Section III, results obtained for these two applications are presented, including exhaustive validation of the knee cartilage segmentation approach by comparison with repeated manual segmentations, and a comparison of the brain results with those obtained with other methods especially designed for this particular application. Section IV is dedicated to the conclusions.

II. Methods

A. Definition of the Watershed Transform for Discrete Images

Several definitions of the watershed transform have been promulgated: an excellent review can be found in [14], where additional details can be found. We are going to focus on the ones based on *topographical distance*, when applied to discrete images. For this purpose, we need first to introduce two important definitions: the *lower slope* and the *lower neighbors*.

Let f be a digital gray value image of arbitrary dimensionality, which we assume to be *lower complete*, that is, it has no plateaus outside the minima. The behavior of the method at the plateaus will be explained later. The *lower slope* of f at a pixel p, $\mathrm{LS}(p)$ is defined

$$\mathrm{LS}(p) = \max_{q \in N_G(p) \cup p} \left(\frac{f(p) - f(q)}{d(p,q)} \right) \tag{1}$$

where $N_G(p)$ is the set of neighbors of p on the grid G, and $d(p,q)$ is the Euclidean distance between pixels p and q. Note that the term inside the brackets is an approximation to the directed gradient to the pixel p. We define $0/0 = 0$ so that, for $p = q$, this term is zero. In this way, we keep $\mathrm{LS(p)} \geq 0$, even when p is a local minimum. The lower slope is necessary to define a steepest slope relation between voxels, which will be used to calculate the watershed transform.

The concept of lower neighbors is derived directly from the lower slope: for each image pixel p, its set of lower neighbors, denoted as $\Gamma(p)$, is

$$\Gamma(p) = \left\{ p' \in N(p) \;\middle|\; \frac{f(p) - f(p')}{d(p,p')} = \max_{p' \in N_G(p) \cup p} \left(\frac{f(p) - f(p'')}{d(p,p'')} \right) = \mathrm{LS}(p) \right\} \tag{2}$$

so the set of lower neighbors is the subset of neighboring pixels for which the directed gradient to the pixel p equals its lower slope. If we consider only a first-order neighborhood (so all distances are equal)

$$\Gamma(p) = \{ p' \in N(p) \,|\, f(p') = \min_{p'' \in N_G(p) \cup p} (f(p'')) \}. \tag{3}$$

Note that, though not considered explicitly in the derivation of this equation, it also covers the case of p being a local minimum, in which $\mathrm{LS}(p) = 0$ and consequently the set of lower neighbors is empty.

The mapping $\Gamma(p)$ introduces a new relation between pixels, which will be used for the watershed transform calculation. At this point, we can define the concept of steepest descent.

> A path $\pi = (p_0, \ldots, p_l)$ between $p_0 = p$ and $p_l = q$ is called a path of steepest descent if $p_{i+1} \in \Gamma(p_i) \forall i = 0, \ldots, l-1$.

A pixel q is said to belong to the *downstream* of pixel p if there exists a path π of steepest descent between p and q. Conversely, a pixel q is said to belong to the *upstream* of p if p belongs to the downstream of q. Using this concept, one can define the catchment basins in the following way [14]:

> For each regional minimum m_i in the image, its associated catchment basin, $\mathrm{CB}(m_i)$ is the set of points in the upstream of m_i.

Those pixels which are in the upstream of at least two minima, i.e., for which there are at least two paths of steepest descent which lead to different minima, do not belong to any catchment basin and are, thus, *watershed pixels*. Note that, in the discrete case, two catchment basins will not always be separated by a connected surface of watershed pixels. Actually, these appear only at locations where the boundary is symmetric.

In most applications, the watershed transform is not calculated directly on the image that we want to segment, but rather on the absolute value of its gradient, which has high values at the objects' contours. Gradient estimation at the center of the pixels worsens the original resolution of the image, causing problems when the regions to segment are very thin. Calculation of the

watershed transform using arc weights in a graph [15], [16] allows the substitution of the usual gradient estimation by absolute differences of gray level calculated directionally during the flooding process, achieving in this way a higher resolution. This idea was used in [17] for brain segmentation. The lower slope is, thus, redefined as

$$\mathrm{LS}(p) = \max_{q \in N_G(p) \cup p} \left(\frac{g(p,q)}{d(p,q)} \right) \tag{4}$$

where $g(p,q)$ is a new function, that is calculated for the link (p,q) (in [17], they choose $g(p,q) = |f(p) - f(q)|$, thus calculating the gradient between two pixels). We can now calculate an equivalent function for the simplified lower neighbors

$$\Gamma(p) = \{p' \in N(p) \,|\, g(p,p') = \max_{p'' \in N_G(p) \cup p} g(p,p'')\}. \tag{5}$$

This modified condition allows the use of a *directional* function in substitution of the absolute value of the gradient, thus improving the results, as shown in [17].

B. Marker Imposition

One of the most important drawbacks associated to the watershed transform is the oversegmentation that commonly results. The usual way of predetermining the number and approximate location of the regions provided by the watersheds technique consists in the modification of the homotopy of the function to which the algorithm is applied. This modification is carried out via a mathematical morphology operation, g*eodesic reconstruction* [18], by which the function is modified so that the minima can be imposed by an external function (the marker function). All the catchment basins that have not been marked are filled by the morphological reconstruction and so transformed into nonminima plateaus, which will not produce distinct regions when the final watersheds are calculated.

In [6], several algorithms to calculate the geodesic reconstruction of an image are presented. Among them, it is especially interesting in our case the one that uses queues, as it can be easily integrated in our watershed calculation algorithm that will be presented in Section II-C. We consider the case in which the marker function is binary, i.e., we have a set of selected regions, which we want to belong to different objects in the final segmentation.

Let I be the original grayscale image, defined in the domain D_i, modified by assigning value 0 to the pixels corresponding to selected markers, J the grayscale marker image, constructed by assigning value 0 to the marker pixels, and ∞ to the rest of the image.

1. Initialization of the queue with the boundaries of the markers
   For every pixel $p \in D_i$:
   If $J(p) = 0$ and $\exists q \in N_G(p) \,|\, J(q) = \infty \Rightarrow$ insert pixel p in the queue
2. Propagation (flooding)
   While the queue is not empty:
   Extract pixel p from the queue
   For every pixel $q \in N_G(p)$, if $J(q) > J(p)$ and $I(q) \neq J(q)$,
   $J(q) = \max(J(p), I(q))$
   Add pixel q to queue

In Step 2), the conditions $J(q) < J(p)$ and $I(q) \neq J(q)$ only control collision of the flooding coming from different minima. Another possibility is performing an *ordered propagation*, in which the pixel p with smaller $I(p)$ value is extracted from the queue in the first place. In Section II-C, a watershed calculation algorithm which embeds the geodesic reconstruction process will be introduced.

C. Proposed Improvement of the Watershed Definition

In this section, we present an original modification of the classical watershed transform, which enables the introduction of prior knowledge about the objects.

In practical applications, we often have available prior information about the absolute or relative intensities of the objects. Unfortunately, the watershed definition based on (4) does not allow the introduction of this information in the calculation.

To be able to take advantage of this prior information, while maintaining the desirable properties of the watershed transform mentioned in Section I-A, we propose the utilization of a set of lower cost functions, one for each of the objects to be detected in the image

$$\mathrm{LS}_k(p) = \max_{q \in N_G(p) \cup p} \left(\frac{f_k(p,q)}{d(p,q)} \right) \tag{6}$$

where $f_k(p,q)$ is the function that quantifies the probability of having an edge between the pixels p and q, given that pixel p has previously received label k.

With this new definition, for example, if a bright object has to be segmented from a darker background, the function f_k can be selected so that it detects a high decrease in the pixel value when we travel from the inside to the outside of the object, and not the opposite. Other knowledge-based conditions can be easily introduced, depending on the application.

To be able to calculate the functions used in (6), it is assumed that the label of pixel p is already known before labeling pixel q. This can be achieved if we use a region-growing algorithm for watershed calculation: the labels of the markers (i.e., of the image minima after geodesic reconstruction) are known at the start of the process, and gradually propagated to the rest of the image, following the *upstream* of each minimum, as explained in Section II-A. The seeds of the region growing process are the markers m_i, where i is a unique label for each marker. The following algorithm, combining geodesic reconstruction and watershed calculation with our proposed modification, can be used.

1) *Initialization:* All pixels belonging to each marker m_i are identified and receive a particular label i. The interior voxels of each marker are assigned to the subset S. The pixels at the boundary of each marker, along with all pixels outside the markers, are assigned to the subset $\bar{S}$.
2) Select the pixel $x \in \bar{S}$ for which

$$\min_{(y \in S) \wedge (x \in N_G(y))} f_k(y,x) = \min_{(y \in S) \wedge (z \in \bar{S}) \wedge (z \in N_G(y))} f_k(y,z)$$

   where k is the label previously assigned to pixel y. This pixel is removed from $\bar{S}$: $\bar{S} \leftarrow \bar{S} - \{x\}$ and subsequently inserted into S. If $\bar{S}$ is empty, END; else go to Step 3).
3) For each neighbor w of x belonging to $\bar{S}$ without label do: $\mathrm{label}(w) = \mathrm{label}(x)$. Return to Step 2).

Note that Step 2) assures that the voxels are processed in the correct order. This is where our novel improvement expressed in (6) is introduced in the algorithm. The algorithm can be implemented in an efficient way using ordered queues. The use of ordered queues in the region growing algorithm above solves the problem of plateau labeling without the need for any additional preprocessing or postprocessing. Usually, pixels are processed in ascending order, as selected in Step 2). On plateaus, where no such order can be used, propagation is done by adding, at each step, an additional layer of pixels neighboring those already labeled. In case the plateau separates two catchment basins, labeling is done by alternating propagation from both sides. Thus, plateaus are partitioned according to pixel distances to the markers. An exhaustive analysis of watershed calculation algorithms and plateau handling can be found in [14].

For some pixels at the edge of the final regions, there is more than one class which provides the same value of $f_k(x, y)$. These pixels are labeled as watershed and, therefore, do not belong to any of the regions. In our experiments, a very low number of pixels were finally labeled as watersheds. To obtain a complete partition of the image in separated regions, we assigned to these pixels the label of the adjacent region whose average value was closer to the pixel value.

D. Probability-Based Functions for Medical Image Segmentation

The improvement in the watershed transform that we propose in Section II-C opens a wide range of possibilities for different applications, depending on the amount of knowledge available about the objects. For generic image segmentation, we propose the use of a function that measures the difference in class probability between two neighboring pixels. We assume normal distributions for the objects in the image, for which mean and variance are calculated using a set of seed voxels for each class. The seed voxels can be selected manually or using automatic techniques, and can be the same set used as markers in the following watershed calculation. Posterior probabilities for each class k at each pixel $p, P(k \mid I_p)$, are calculated using Bayes' rule as

$$P(k \mid I_p) = \frac{P(I_p \mid k)P(k)}{\sum_k P(I_p \mid k)P(k)} \tag{7}$$

where I_p is the intensity, at pixel p, of the image that we want to segment.

Structures we wish to segment typically exhibit a significant spatial homogeneity, and so it is desirable to achieve this behavior in the obtained probability values. Markov Random Fields [19], [20] provide a convenient way to model local correlations between pixels. In our implementation, we use the Potts' model, a generalization of the classical Ising model for more than two states, which defines the class probability for class k at pixel I as

$$P_i(k) = \frac{1}{Z} \exp\left(-\beta \frac{\sum_{j \in N_G(i)} \delta(k, k_j)}{d(i, j)} \right) \tag{8}$$

where $N(i)$ are the neighbors of the voxel i, k_j is the class assigned to pixel j and $d(i, j)$ is the Euclidean distance between voxels i and j. The function $\delta(k_1, k_2)$ quantifies the relation between two neighboring voxels k_1 and k_2, in the following way:

$$\delta(k_1, k_2) = \begin{cases} -1, & \text{if } k_1 = k_2 \\ +1, & \text{if } k_1 \neq k_2 \end{cases}. \tag{9}$$

To estimate the model, we have used iterative conditional modes (ICM), a method first proposed by Besag [20], in which (7) and (8) are iteratively solved. Though the method does not provide an exact solution, it has the advantage of its fast convergence: in the original paper by Besag, he states that few if any changes occur after the sixth cycle. In practical applications, we have found that for our needs it is usually enough with the first iteration of ICM. (Note that in this step we do not want to classify the voxels, but only to obtain an estimate of its probability values)

Given these definitions, we have found a useful set of contrast functions for the new watershed transform, $f_k(p, q)$, is the following:

$$f_k(p, q) = P(k \mid I_p) - P(k \mid I_q) \tag{10}$$

where $P(k \mid I_p)$ and $P(k \mid I_q)$ denote the posterior probabilities of class k at pixels p and q.

E. Automatic Calculation of Markers From an Atlas

In order to benefit from the geodesic reconstruction technique described in Section II-B, the markers must be correctly located. In the past, either interactive selection or ad-hoc techniques have been usually applied for this purpose. In order to provide fully automatic segmentation, we propose here for the first time the automatic generation of markers from a statistical atlas. We describe a robust method for taking a prior model of the anatomy we expect to segment in the form of a statistical atlas and projecting marker positions into the image to be segmented.

A probabilistic atlas is used, in which each voxel of the image has associated probability values for each one of the image classes. In this way, these probability values can be assigned, after registration, to the voxels of the image to be segmented. For this purpose, the nonrigid registration technique described in [21] was used. This is an iterative method consisting of two stages: correction of intensity differences between the images and registration of the corrected images. Note that these are probabilities associated only with the spatial position of the voxels, and not with their intensities. In Fig. 1, the probability values for gray matter $(P_{s,\mathrm{gm}})$, white matter $(P_{s,\mathrm{wm}})$, and CSF $(P_{s,\mathrm{csf}})$, are shown for a registered brain MR image.

To obtain accurate results, it is absolutely necessary that all the marker voxels for a particular object are located inside this object in the images we want to segment: even a single false positive can produce significant errors in the final result. To assure that, we use a skeletonization followed by an outlier removal.

- *Skeletonization:* Object probability images are thresholded and skeletons are calculated on individual slices using mathematical morphology techniques [22]. A sample result of the obtained skeleton for WM $(\mathrm{Sk}_{\mathrm{wm}})$, GM $(\mathrm{Sk}_{\mathrm{gm}})$ and CSF $(\mathrm{Sk}_{\mathrm{csf}})$ is shown in Fig. 1.
- *Outlier removal:* The skeletons obtained in the previous step are used to calculate the parameters of normal distributions, as described in Section II-D. Skeleton voxels whose probability of belonging to the associated class is

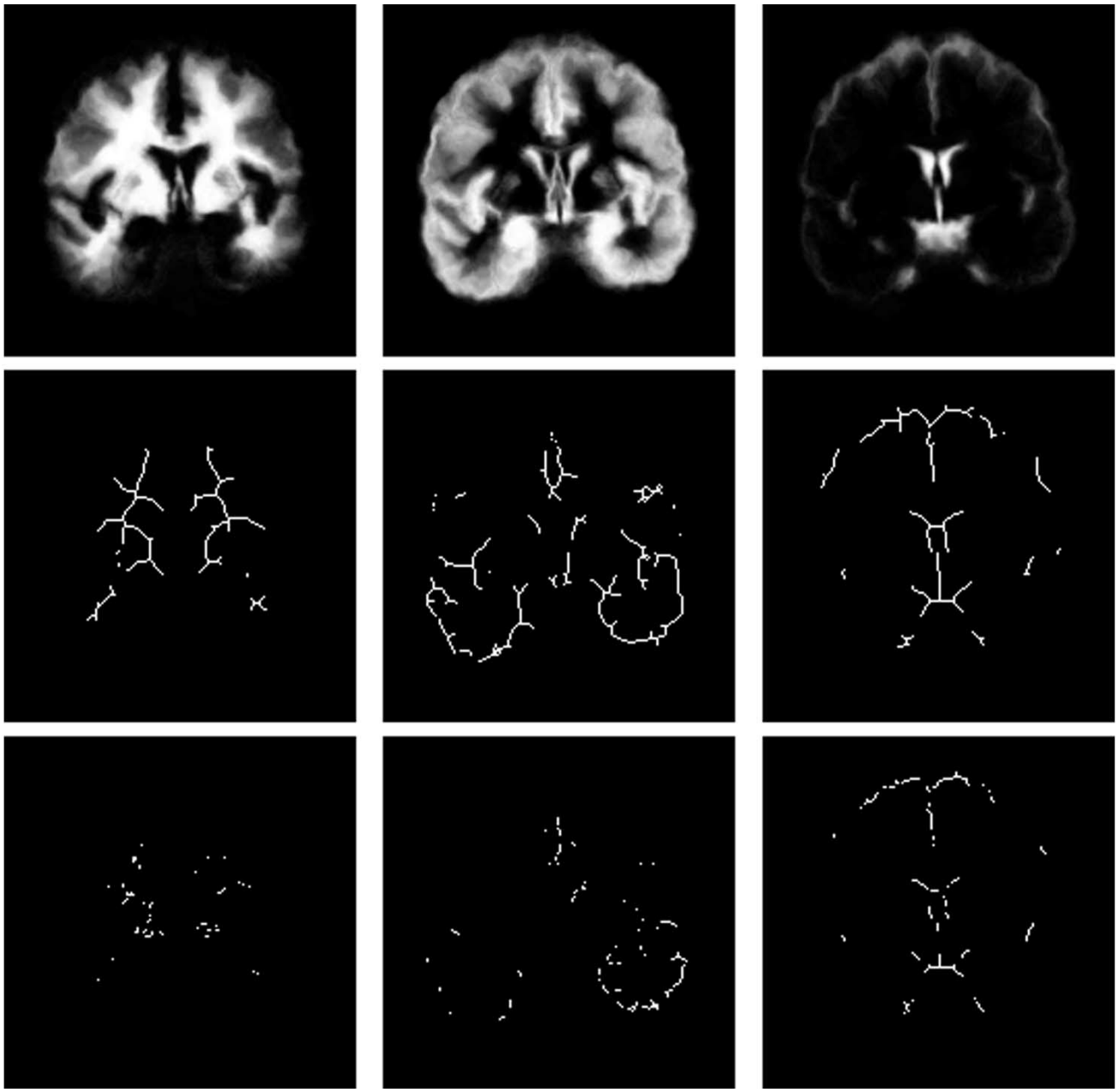

Fig. 1. Top: Probabilistic atlas aligned with a particular case. Results for white matter, gray matter and CSF are shown. Middle: skeleton calculated for the atlas results. Bottom: Modified skeleton after elimination of outliers. Note that a very conservative threshold has been used, to ensure that all marker voxels are inside their corresponding objects. Our improvements in the watershed transform eliminate the need of a very accurate marker to obtain a good segmentation, thus enabling the utilization of an automatic marker procedure such as the one shown.

smaller than a certain threshold are considered outliers and removed, generating the corrected skeleton Sk_corr

$$\mathrm{Sk\_corr}_k = \{x \in Sk_k \,|\, P(k \,|\, I_x) > T_{\mathrm{Sk}}\} \tag{11}$$

where $P(k \,|\, I_x)$ is the Bayesian probability as calculated in (7), and T_{Sk} is the threshold selected for the skeleton. A conservative threshold should be used in this case, to assure that no false positives remain in the marker set. The final results for this case are shown in Fig. 1.

Parameters of the normal distributions presented in Section II-D were also calculated using the atlas registration results. Though, for this purpose, the effect of a few outliers would not be as critical, we found in our experiments that, for thin structures such as gray matter or CSF, atlas registration produces nonzero results in an area that spreads significantly out of the corresponding object. For this reason, we have used only skeleton voxels to calculate these parameters.

F. Clinical Applications

In this section, we describe the application of the presented algorithms to two clinical applications: knee cartilage and white matter/gray matter segmentation.

1) Knee Cartilage Segmentation: A number of diseases and injuries affect the knee joint by damaging the cartilage. Hence,

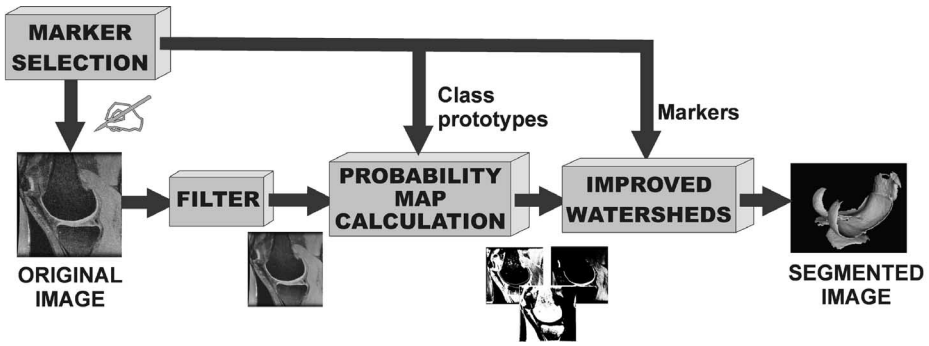

Fig. 2. Block diagram of the segmentation system used for knee cartilage segmentation.

an accurate measurement of the volume and thickness of cartilage is often crucial to obtain an adequate diagnosis and to assess evolution of the patients under treatment.

Several studies have been carried out in order to determine the accuracy and reliability of MR images for cartilage disease assessment. Direct two-dimensional measurement on single slices is not feasible: as their locations and orientations cannot be repeated exactly in longitudinal studies, it is impossible to use them to measure changes of cartilage over time. On the other hand, current MR protocols have shown enough accuracy and reliability for the detection of cartilage lesions [23], [24].

An accurate segmentation of the cartilage is necessary to calculate its total volume. Manual segmentation, used in some studies [24], [25], is a time-consuming task that can only be carried out by a trained expert. For these reasons, the development of segmentation algorithms that minimize user interaction is a very interesting challenge.

Several groups have attempted to develop fully automatic or semi-automatic knee cartilage segmentation algorithms. Heuristic, strongly user-guided systems have been often proposed, as in [26], where a simple thresholding, followed by simple user-selected operations, is used. In [27] and [28], B-splines are fitted to a set of points, manually selected along each articular contour curve: user interaction is, thus, fundamental. In [29], a spline is drawn on a selected slice, and then adjusted to the cartilage contour by optimization of a cost function. An acknowledged challenge is the decisive role played by the user when selecting the end points for femoral cartilage. In [30], a more complex approach is used, including an Active Shape Model which contains the possible variations in shape. It is important to note that all the abovementioned methods perform the segmentation slice by slice and not on the whole volume at once, thus not assuring consistency in the direction normal to the slices. An interesting approach, that actually deals with the 3-D volume, is presented in [31]. In this case, the user is required to perform an interactive registration of a knee template to the subject, and then selects points from several anatomical structures. An iterative classification algorithm, modified by the registered template, produces the final segmentation.

Our goal was to segment the overall cartilage volume of the knee joint from the MR images, in order to allow cartilage volume and thickness measures. We used the algorithm presented in Sections II-C and II-D to segment the knee cartilage in MR images, working directly in 3-D. The use of an atlas in this application is restricted by the movement of the knee joint, which makes the available registration methods not usable. For this reason, we applied our improved watershed transform using a set of manually selected voxels as markers for each of three classes: cartilage, bone and other tissues. Future development of a reliable knee atlas will enable the complete automation of the process by using the algorithm introduced in Section II-E. This last class represents the ligament and muscle around the cartilage. In order to reduce noise without compromising the contrast at significant edges, the images were first filtered using the anisotropic diffusion filter presented in [32]. In Fig. 2, the complete block diagram for this application is shown.

2) Brain Segmentation: Accurate segmentation of cortical gray matter is critically important in understanding structural changes associated with central nervous system diseases such as multiple sclerosis and Alzheimer's disease, and also the normal aging process. Measures of change in cortical gray matter volume and thickness are suggested to be important indicators of atrophy or disease progression.

Many approaches have been tried for gray matter segmentation. Classification has often been used, with sophisticated approaches such as the one presented in [33], [34], where the expectation-maximization technique is used to correct the bias field, and combined with an atlas, during the classification process. In [31], a classification step is combined with spatial information, and partial volume effects are also considered in [35]. Also relying on a classification procedure, good results are achieved in [36] or [37], at the expense of introducing many heuristic corrections.

Another important research line used for cortex segmentation uses deformable models. In [38], the *active ribbon* is introduced: two coupled bidimensional contours that evolve to find the gray matter/white matter and gray matter/CSF interfaces in a single slice. In [39], a similar approach is carried out using coupled level sets in three dimensions. Classification and deformable models have also been combined [40].

In this case, we can make use of a probabilistic atlas to obtain the initial set of markers, as presented in Section II-E. Details about the atlas creation are given in [41]. As in the knee cartilage segmentation, an anisotropic diffusion filter [32] is initially applied. The complete algorithm is, thus, very similar to the one presented in Fig. 2, with the substitution of manual seed selection by the atlas-based algorithm.

Two-channel images (T1- and T2-weighted) are used in this case. This does not affect the segmentation algorithm at all, as long as (9) can be calculated.

The whole brain segmentation process is carried out in three steps: brain stripping, white matter/gray matter interface detection, and gray matter/CSF interface detection.

a) Brain Stripping: Most gray matter/white matter segmentation systems need a previous, independent step to remove the external nonbrain tissues from the images. Several methods to extract this surface have been presented. In [42], a review of these methods is presented, classifying them in three types: manual segmentation, threshold with mathematical morphology operations, and surface-model-based. Manual segmentation takes a huge amount of expert time, making it unusable in practical scenarios. Mathematical morphology techniques can be used to correct the results obtain by an initial thresholding step [43]–[45], but usually, these methods involve highly ad-hoc procedures, difficult to automate when applied to images acquired with different protocols. Deformable models [36], [42] seem to be more robust, but they still require the inclusion of many heuristic procedures in the method, and they are difficult to apply to images with a range of contrast types.

We have used our improved watershed transform to perform the brain stripping operation. The markers used were, for the brain, the combined skeletons of gray and white matter and CSF, obtained as explained in Section II-E, and for the external volume, the voxels for which the spatial probability obtained after registration with the atlas is zero:

$$\begin{aligned} \mathrm{Marker}_{\mathrm{ICC}} &= \mathrm{Sk\_corr}_{\mathrm{wm}} + \mathrm{Sk\_corr}_{\mathrm{gm}} + \mathrm{Sk\_corr}_{\mathrm{csf}} \\ \mathrm{Marker}_{\mathrm{ext}} &= (P_{s,\mathrm{gm}} = P_{s,\mathrm{wm}} = P_{s,\mathrm{csf}} = 0) \end{aligned} \tag{12}$$

where $\mathrm{Marker}_{\mathrm{ICC}}$ and $\mathrm{Marker}_{\mathrm{ext}}$ represents the markers for the intracranial cavity (ICC) and the external objects, respectively.

In this case, as the regions to segment are not homogeneous, it is not possible to use differences in probability calculations for the improved watershed transform. Instead, we use directional intensity differences, defined as:

$$\begin{aligned} f_{(k=\mathrm{ICC})}(p,q) &= ((\max(I_{\mathrm{SPGR}}(p) - I_{\mathrm{SPGR}}(q),0))^2 \\ &\quad + (\max(I_{T2}(p) - I_{T2}(q),0))^2)^{\frac{1}{2}} \\ f_{(k=\mathrm{ext})}(p,q) &= ((\max(I_{\mathrm{SPGR}}(q) - I_{\mathrm{SPGR}}(p),0))^2 \\ &\quad + (\max(I_{T2}(q) - I_{T2}(p),0))^2)^{\frac{1}{2}} \end{aligned} \tag{13}$$

Note that here the improvement proposed for the watershed transform allows substitution of the usual gradient calculation by an asymmetric gradient, thus avoiding the effect of many spurious contours.

b) White Matter/Gray Matter and Gray Matter/CSF Surface Detection: After eliminating the nonbrain voxels, the white matter/gray matter surface is detected, using as markers $\mathrm{Sk\_corr}_{\mathrm{wm}}$ for the white matter, and $\mathrm{Sk\_corr}_{\mathrm{gm}} + \mathrm{Sk\_corr}_{\mathrm{csf}}$ for the gray matter. Probability maps are calculated as explained in Section II-D, and (10) is applied.

In the last step, the gray matter/CSF surface is obtained in the same way, this time using the segmented white matter, along with $\mathrm{Sk\_corr}_{\mathrm{gm}}$, as inside markers, and $\mathrm{Sk\_corr}_{\mathrm{csf}}$ as outside markers.

III. Results

In order to demonstrate the advantages of the proposed technique, it was tested on the two applications presented in Sections II-F1 and II-F2: delineation of the knee cartilage and full segmentation of white matter and gray matter of the brain. In both cases, we used MR images, one channel in the knee application, and two (T1- and T2-weighted) for the brain. Processing was carried out totally in 3-D, both for classification and for watershed calculation. A significant effort was realized to validate the obtained results: in the knee images, a total of 43 whole volume manual segmentations of 7 cases were used, thus allowing for a study of intra- and interobserver variation obtained using our method, along with a direct comparison between manual and automatic segmentation. In the brain images, a standard data set, the BrainWeb [46] was used, to enable a direct comparison with other systems designed specifically for that purpose.

Numeric validation was done globally in all cases, by quantifying the accuracy of the segmentation in the whole dataset. For specific applications, sometimes more local validation of results would also be desirable. Here, a generic validation has been deemed adequate for our generic segmentation method.

In the following sections, the particular details of the segmentation algorithm for these applications, along with the results of the validation tests, are presented. The results demonstrate an accurate behavior of the algorithm on two different applications, and suggest its applicability on several other medical image segmentation tasks.

A. Knee Cartilage Segmentation

MR scans of the knee were acquired using a high-resolution fat-saturated 3-D SPGR sequence, with $(0.23 \times 0.23 \times 1.5\ \mathrm{mm}^3)$ resolution. Then the procedure described in Section II-F1 was used for segmentation. In all cases, around 50 marks for each class (cartilage, bone and other tissues) were manually selected, taking 5 to 10 min for a whole dataset. Results, including a three-dimensional reconstruction of the segmented cartilage, are shown in Fig. 3. The images show the difficulties associated to the application (thin structures, overlapping gray levels), and the ability of our system to obtain accurate segmentation of cartilage tissue.

1) Validation: MR scans from four volunteers were used for the tests. Subjects 1, 2 and 3 were scanned once, and the images were segmented manually, slice by slice, by an expert, taking typically more than 2 h per scan.

Subject 4 was scanned four times. Scan 1 was taken in the same conditions as the previous subjects. Then a half-voxel shift was applied and Scan 2 was acquired. After moving the volunteer out of the scanner for 20 min and repositioning the knee, Scans 3 and 4 were acquired, again with a half-voxel shift applied after Scan 3. In this way, we wanted to check the consistency of our segmentation algorithm against the partial volume and repositioning effects. These four scans were segmented five times each by two different experts (a total of 40 segmentations) to check intraobserver and interobserver variations.

a) Intraobserver and Interobserver Assessment: In order to use the segmentation for monitoring of disease progression,

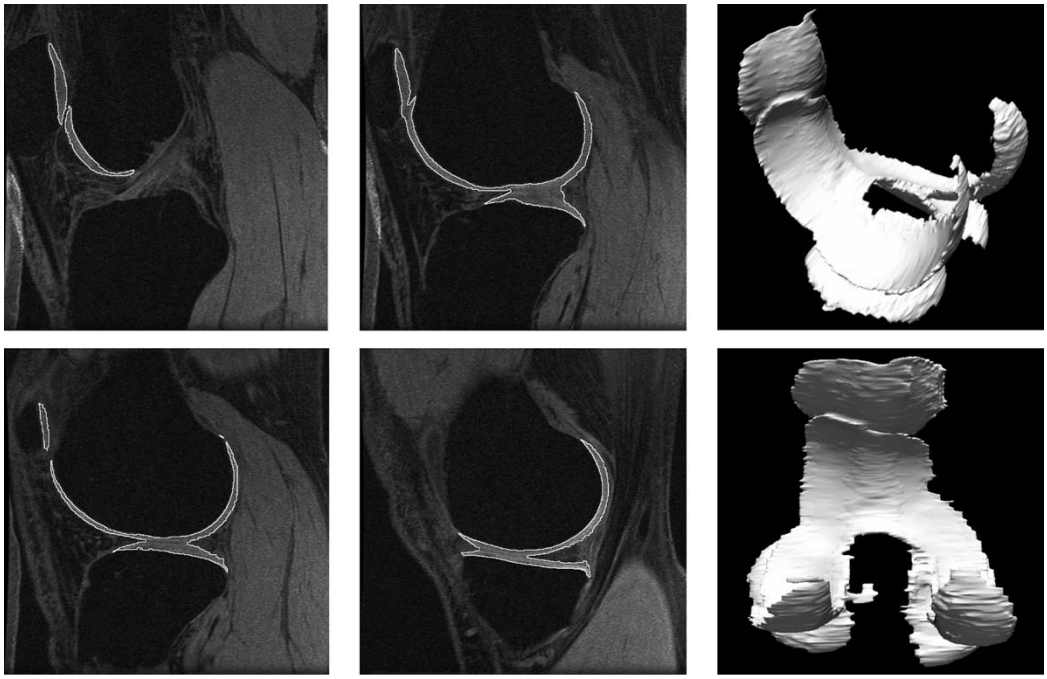

Fig. 3. Sample segmentation results from subject 3. In the right side, 3-D reconstructions from two different positions are shown. As can be seen, our algorithm is able to segment accurately the whole cartilage volume.

TABLE I
MEAN SENSITIVITY, SPECIFICITY AND DICE COEFFICIENTS, IN PERCENTAGE, OF THE REPEATED SEGMENTATIONS FOR EACH OF THE FOUR SCANS OF THE FOURTH SUBJECT, COMPARED TO RESULTS USING OUR ALGORITHM. EACH SCAN WAS SEGMENTED FIVE TIMES BY EACH EXPERT AND FIVE MORE TIMES BY OUR METHOD WITH DIFFERENT SEED SETS

	Intra-observer Assessment						Inter-observer Assessment			Our algorithm		
	Expert 1			Expert 2								
Scan	Sens	Spec	Dice	Sens	Spec	Dice	Sens	Spec	Dice	Sens	Spec	Dice
1	89.30	99.97	92.95	90.49	99.97	93.65	92.27	99.92	92.71	96.46	99.98	97.50
2	91.41	99.97	94.40	91.08	99.97	94.14	94.04	99.90	92.78	97.93	100	98.75
3	91.75	99.97	94.57	89.62	99.97	93.18	93.31	99.88	91.75	99.05	99.99	99.11
4	92.45	99.98	95.10	90.78	99.97	93.90	94.99	99.86	91.78	96.33	99.99	97.87

it is extremely important that the intra- and interobserver variations are kept small: only in this way, an accurate assessment of the progression of the degenerative diseases can be done.

For each scan of the Subject 4, the ground truth cartilage was determined by the STAPLE method [47], [48]. The STAPLE algorithm estimates from a collection of segmentations both the unknown ground truth and performance parameters for each segmentation. For intraobserver assessment, mean values of sensitivity, specificity and the Dice similarity coefficient (ratio between intersection and sum of the volumes compared) of the five manual segmentation repetitions of each scan carried out by each expert, compared to the ground truth provided by STAPLE, are shown in Table I. These values were obtained by measuring the overlap between each of the five segmentations and the ground truth. To determine interobserver variation, a comparison was performed between the ground truth results for the two segmenters: the related parameters for this interobserver assessment are also shown in Table I.

Finally, the variability of our method was estimated. Five segmentations were done using different sets of starting points, and then combined using STAPLE. The results of comparison between each of the five segmentations and the combined one, shown in Table I, demonstrate the superior consistency of the presented method, compared to manual segmentation. The small variability of the results when using different marker sets demonstrates a small dependence on the marker location, minimizing user guidance and opening the possibility of using ad-hoc techniques for automatic marker selection. This performance can be crucial to understand the smallest real detectable change in the assessment of new treatments or drugs.

TABLE II
ANALYSIS OF THE RESULTS OBTAINED WITH OUR ALGORITHM AND WITH THE STANDARD WATERSHED TRANSFORM. SENSITIVITY, SPECIFICITY AND DICE COEFFICIENT WERE CALCULATED BY USING MANUAL SEGMENTATIONS AS GROUND TRUTH. THE IMPORTANCE OF THE PROPOSED MODIFICATIONS IS DEMONSTRATED BY THE CLEARLY SUPERIOR RESULTS. AS IS SHOWN, OUR ALGORITHM IS ABLE TO SEGMENT DIFFERENT SCANS WITH CONSISTENT ACCURACY

	Standard watersheds			Our algorithm		
Subject	Sens.	Spec.	Dice	Sens	Spec.	Dice
1	56.41	99.80	65.17	88.52	99.86	87.45
2	43.20	99.77	53.20	94.33	99.84	89.37
3	52.21	99.61	56.51	94.54	99.77	91.19
4 - scan 1	47.15	99.39	47.53	89.65	99.87	89.88
4 - scan 2	45.23	99.86	57.78	86.49	99.90	89.07
4 - scan 3	53.25	86.13	8.32	87.63	99.90	89.84
4 - scan 4	45.77	99.77	55.67	89.05	99.88	89.78

b) Comparison With Manual Segmentations: We compared the results obtained using our algorithm with the ground truth provided by manual segmentations. For the fourth subject, the ground truth results provided by STAPLE, as described above, were used. Values of sensitivity, specificity and Dice coefficient are shown in Table II. Results for the standard watershed transform, computed on the gradient of the original image using the same set of markers, are also presented: comparison with our new algorithm demonstrates the importance of our improvements.

c) Robustness With Repositioning and Partial Volume Effect: Clinical assessment of cartilage loss progression typically involves repeated scans with uncontrolled repositioning. For this reason, it is necessary to check the consistency of our method when the same knee is scanned in different positions. Scans from subject 4, acquired as described above, were used for this purpose. Total volumes were calculated for the segmentations carried out by experts 1 and 2 and by our program, after determining the ground truth for each of them using the STAPLE method. Table III shows the coefficient of variation for each expert, for the combined segmentations of both experts, and for the algorithm presented. Our method shows a performance comparable or superior to the manual segmenters.

B. Brain Segmentation

Two channel MR brain images were obtained with resolution ($0.9375 \times 0.9375 \times 1.5$ mm^3), and segmented as explained in Section II–F2. Sample results are given in Fig. 4, where the accuracy of our method can be appreciated. An aspect of special difficulty for a majority of brain segmentation algorithms is the detection of thin sulci; the observation of results in Fig. 4 shows how our proposed system is capable of their accurate segmentation.

Due to the low contrast contour between white and gray matter, compared to other contours with higher contrast in the neighborhood, the application of the plain watershed transform here would yield clearly inaccurate results. In order to determine which parts of the algorithm (marker generation from the atlas, substitution of the gradient by probability maps, the proposed improvement in the watershed transform) were responsible for the quality of the results shown in Fig. 4, additional segmentations were performed. In Fig. 5, results are shown for comparison. The images at the left column were obtained using the same atlas-generated markers, but with a conventional watershed transform applied on the gradient image. Those at the central column were obtained with the same markers, substituting the gradient by differences in the probability maps, but still using the classic watershed transform. The right column shows the results obtained using our algorithm. From the observation of these results, it is evident that the watershed improvement proposed here is essential to obtain accurate results. As mentioned above, the detection of thin structures, usually a major drawback of the watershed transform, is significantly improved through the use of our novel algorithm.

TABLE III
COEFFICIENT OF VARIATION FOR THE VOLUMES OF THE 4 SCANS OF SUBJECT 4, CALCULATED FOR EACH EXPERT, FOR BOTH EXPERTS COMBINED, AND FOR OUR METHOD

Expert 1	Expert 2	Experts	Our method
1.79	1.70	2.26	1.91

Validation: Quantifying the quality of the segmentation is difficult in real patient images, as the ground truth is unknown. We have used the Brain Web MR simulation program [46], for which both MR images and ground truth segmentation are available online. We obtained T1–T2 images, with 3% noise and 20% inhomogeneity. It is important to note that the atlas used is absolutely independent from the BrainWeb dataset, thus keeping algorithm development and validation totally separated. Our results, for white matter and gray matter, are shown in Table IV. It is interesting to compare the performance of our system to other groups that have tried this operation on the same phantom recently, as Van Leemput *et al.* [49], or Zeng *et al.* [39]. Two issues must be mentioned to fully understand the results. In Van Leemput *et al.*, results were obtained using 1 and 3 channels, while in our case 2 channels were used. Both results are shown for comparison. In Zeng *et al.*, the segmentation algorithm is designed specifically to detect the cerebral cortex, and so it does not achieve acceptable results in the brain stem and the cerebellum. For this reason, the authors provide the gray matter validation results only for selected sets of 49 frontal coronal slices and 56 top axial slices. Results shown for our algorithm correspond to the whole volume, including brain stem and cerebellum. As can be seen, though our system is a general-purpose one, the results are comparable to specialized brain segmentation systems.

IV. CONCLUSION

We have presented a new segmentation algorithm, based on an improved watershed transform, which enables the use of different prior information-based difference functions for different objects, instead of the usual gradient calculation. In the particular applications presented, these functions were calculated from the probability values for each voxel and each class.

For applications where atlases are available, another improvement was presented, by using the atlas to generate markers for

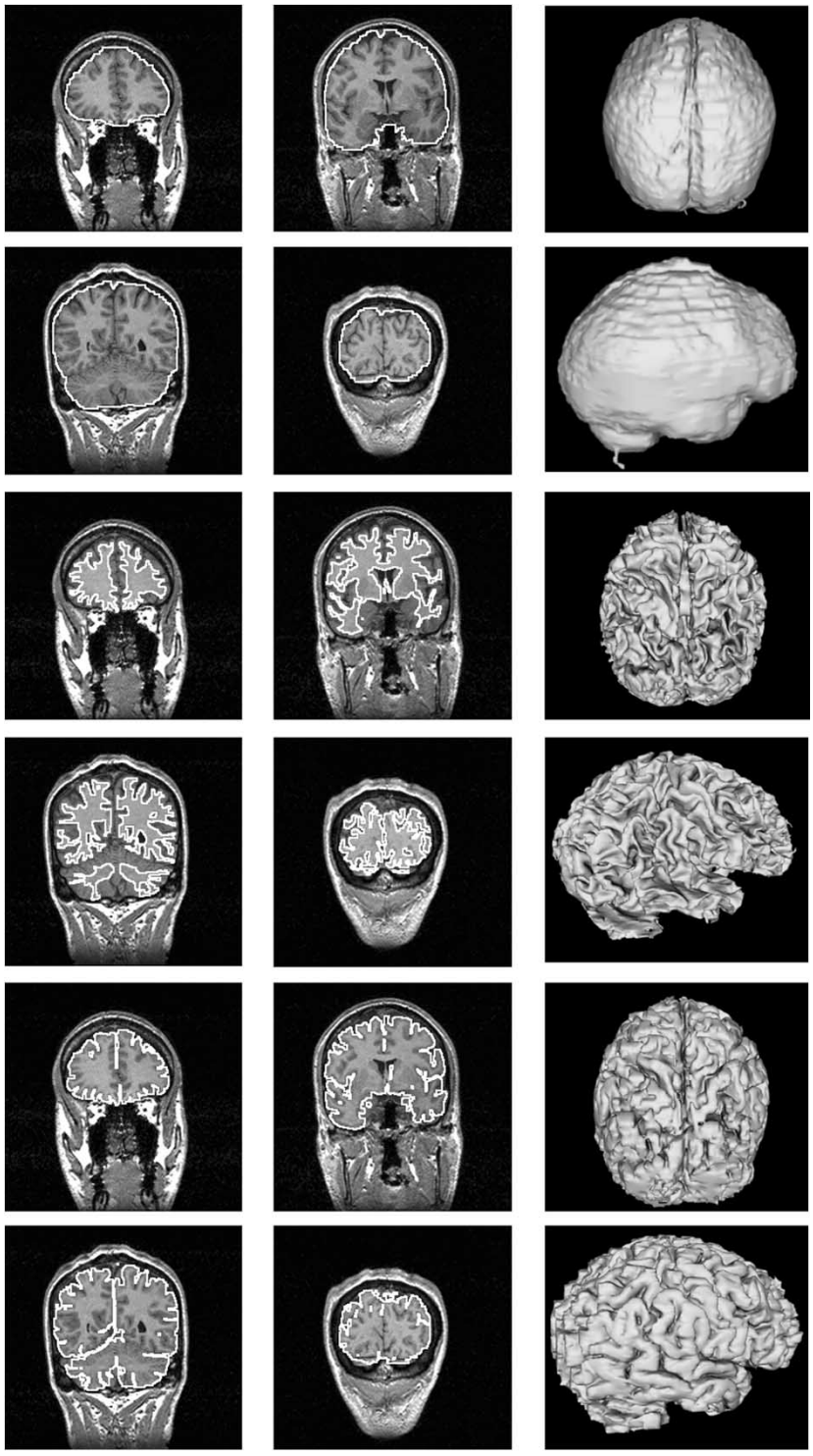

Fig. 4. Results for brain segmentation on four sample slices, and 3-D reconstructions. Top: result of the brain stripping process. Middle: Gray matter/white matter surface. Bottom: Gray matter/CSF surface. Note the accurate performance of the algorithm when dealing with very thin structures.

the watershed transform. For this purpose, after the registration step, skeletonization and outlier elimination were performed.

We applied the algorithm to two important applications: knee cartilage segmentation and white matter/gray matter segmentation in MR images. In knee cartilage validation, seven scans were used; our algorithm demonstrated similar or superior performance to that of manual segmentation by experts. Brain segmentation was performed using the BrainWeb simulator. In this case, an atlas was used, so the process was totally automatic. The results, using our general algorithm, were similar to those of state of the art systems, specifically designed for this application.

The method presented showed convincing accuracy on the two applications in which it was tested. These results also suggest that the system could be used in many different medical image segmentation problems.

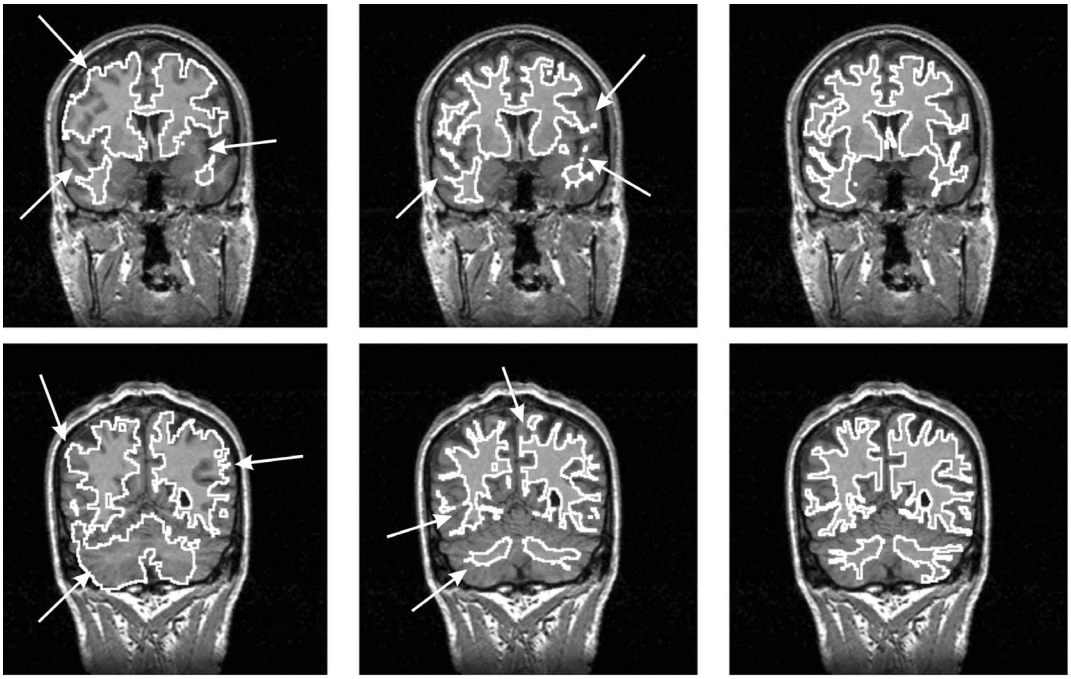

Fig. 5. Left column: results obtained using a conventional watershed algorithm on the gradient image. Middle column: results using conventional watersheds on the difference of probabilities. Right column: results using our proposed algorithm. In all cases, the same atlas-generated markers were used. Arrows were used to highlight some specific regions where significant differences with the proposed algorithm are evident. The images show the importance of our improvements on the watershed transform in the quality of the final results.

TABLE IV
DICE COEFFICIENTS OBTAINED BY OUR SYSTEM ON THE BRAINWEB PHANTOM, COMPARED WITH OTHER APPROACHES. (NOTE: IN VAN LEEMPUT *et al.*, RESULTS SHOWN WERE OBTAINED USING 1 AND 3 CHANNELS, WHILE IN OUR CASE, 2 CHANNELS WERE USED. IN ZENG *et al.*, GRAY MATTER SEGMENTATION IS VALIDATED ONLY ON SELECTED FRONTAL AND AXIAL AREAS, WHILE OUR SYSTEM SOLVED THE MORE CHALLENGING PROBLEM OF WHOLE BRAIN SEGMENTATION.)

	Our algorithm	Van Leemput et al	Zeng et al.
White matter	94.56%	90% (1 channel) 92% (3 channels)	94.4%
Gray matter	89.02%	89% (1 channel) 93% (3 channels)	93.1% (frontal slices) 93.8% (top slices)

ACKNOWLEDGMENT

The authors would like to thank Dr. M. Brem and Dr. G. Welsch for manually segmenting most of the knee scans.

REFERENCES

[1] S. Beucher and F. Meyer, "The morphological approach to segmentation: The watershed transform," in *Mathematical Morphology in Image Processing*, E. R. Dougherty, Ed. New York: Marcel Dekker, 1993, vol. 12, pp. 433–481.

[2] A. N. Moga and M. Gabbouj, "Parallel image component labeling with watershed transformation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, pp. 441–450, May 1997.

[3] J. M. Gauch, "Image segmentation and analysis via multiscale gradient watershed hierarchies," *IEEE Trans. Image Processing*, vol. 8, pp. 69–79, 1999.

[4] O. F. Olsen and M. Nielsen, "Multi-scale gradient magnitude watershed segmentation," in *ICIAP'97–9th Int. Conf. on Image Analysis and Processing*, ser. Lecture Notes in Computer Science. Berlin, Germany: Springer-Verlag, 1997, vol. 1310, pp. 6–13.

[5] E. Dam and M. Nielsen, "Non-linear diffusion for interactive multi-scale watershed segmentation," in *MICCAI 2000: Fourth International Conference on Medical Image Computing and Computer-Assisted Intervention*, ser. Lecture Notes in Computer Science. Berlin, Germany: Springer-Verlag, 2000, vol. 1935, pp. 216–225.

[6] J. L. Vincent, "Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms," *IEEE Trans. Image Processing*, vol. 2, pp. 176–201, 1993.

[7] S. Beucher, "Watershed, hierarchical segmentation and waterfall algorithm," in *Mathematical Morphology and Its Applications to Image Processing*, Dordrecht, The Netherlands: Kluwer, 1994, pp. 69–76.

[8] P. Jackway, "Gradient watersheds in morphological scale-space," *IEEE Trans. Image Processing*, vol. 5, pp. 913–921, June 1996.

[9] J. Weickert, "Fast segmentation methods based on partial differential equations and the watershed transform," in *Proc. DAGM Symp.*, 1998, pp. 93–100.

[10] J. Sijbers, P. Scheunders, M. Verhoye, A. Van der Linden, D. Van Dyck, and E. Raman, "Watershed-based segmentation of 3D MR data for volume quantization," *Magn. Reson. Imag.*, vol. 15, pp. 679–688, 1997.

[11] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Comput. Vis.*, vol. 1, pp. 321–331, 1987.

[12] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *Int. J. Comput. Vis.*, vol. 22, pp. 61–79, 1997.

[13] H. T. Nguyen, M. Worring, and R. V. D. Boomgaard, "Watersnakes: Energy-driven watershed segmentation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 25, pp. 330–342, March 2003.

[14] J. B. T. M. Roerdink and A. Meijster, "The watershed transform: Definitions, algorithms and parallelization strategies," *Fundamenta Informaticae*, vol. 41, pp. 187–228, 2000.

[15] F. Meyer, "Color image segmentation," in *Proc. 4th Int. Conf. Image Processing*, 1992, pp. 523–548.

[16] R. Lotufo and A. Falcao, "The ordered queue and the optimality of the watershed approaches," in *Proc. Int. Symp. Mathematical Morphology*, 2000.

[17] J. P. Thiran, V. Warscotte, and B. Macq, "A queue-based region growing algorithm for accurate segmentation of multi-dimensional digital images," *Signal Processing*, vol. 60, pp. 1–10, 1997.

[18] C. Lantuejoul and F. Maisonneuve, "Geodesic methods in image analysis," *Pattern Recogn.*, vol. 17, pp. 117–187, 1984.

[19] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 6, pp. 721–741, 1984.

[20] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Statist. Soc. B*, vol. 48, pp. 259–302, 1986.

[21] A. Guimond, A. Roche, N. Ayache, and J. Meunier, "Three-dimensional multimodal brain warping using the demons algorithm and adaptive intensity corrections," *IEEE Trans. Med. Imag.*, vol. 20, pp. 58–69, Jan. 2001.

[22] J. Serra, *Image Analysis and Mathematical Morphology*. New York: Academic, 1982, vol. 1.

[23] D. T. Felson, R. C. Lawrence, M. C. Hochberg, T. McAlindon, P. A. Dieppe, M. A. Minor, S. N. Blair, B. M. Berman, J. F. Fries, M. Weinberger, K. R. Lorig, J. J. Jacobs, and V. Goldberg, "Osteoarthritis: New insights. Part 2: Treatment approaches," *Anna. Intern. Med.*, vol. 133, pp. 726–737, 2000.

[24] F. Eckstein, M. Schnier, M. Haubner, J. Priebsch, C. Glaser, K. H. Englmeier, and M. Reiser, "Accuracy of cartilage volume and thickness measurements with magnetic resonance imaging," *Clin. Orthop. Related Res.*, vol. 352, pp. 137–148, 1998.

[25] C. G. Peterfy, C. F. van Dijke, D. L. Janzen, C. C. Gluer, R. Namba, S. Majumdar, P. Lang, and H. K. Genant, "Quantification of articular cartilage in the knee with pulsed saturation transfer subtraction and fat-suppressed MR imaging: Optimization and validation," *Radiology*, vol. 192, pp. 485–491, 1994.

[26] M. A. Piplani, D. G. Disler, T. R. McCauley, T. J. Holmes, and J. P. Cousins, "Articular cartilage volume in the knee: Semiautomated determination from three-dimensional reformations of MR images," *Radiology*, vol. 198, pp. 855–859, 1996.

[27] Z. A. Cohen, D. M. McCarthy, S. D. Kwak, P. Legrand, F. Fogarasi, E. J. Ciaccio, and G. A. Athesian, "Knee cartilage topography, thickness, and contact areas from MRI: *In-vitro* calibration and *in-vivo* measurements," *Osteoarthritis Cartilage*, vol. 7, pp. 95–109, 1999.

[28] T. Stammberger, F. Eckstein, M. Michaelis, K. H. Englmeier, and M. Reiser, "Interobserver reproducibility of quantitative cartilage measurements: Comparison of B-spline snakes and manual segmentation," *Magn. Reson. Imag.*, vol. 17, pp. 1033–1042, 1999.

[29] J. A. Lynch, S. Zaim, J. Zhao, A. Stork, C. G. Peterfy, and H. K. Genant, "Cartilage segmentation of 3D MRI scans of the ostheoarthritic knee combining user knowledge and active contours," *Proc. SPIE*, vol. 3979, pp. 925–935, 2000.

[30] S. Solloway, C. E. Hutchinson, J. C. Waterton, and C. J. Taylor, "The use of active shape models for making thickness measurements of articular cartilage from MR images," *Magn. Reson. Med.*, vol. 36, pp. 943–952, 1997.

[31] S. K. Warfield, M. Kaus, F. A. Jolesz, and R. Kikinis, "Adaptive, template moderated, spatially varying statistical classification," *Med. Image Anal.*, vol. 4, pp. 43–55, 2000.

[32] K. Krissian, G. Malandain, and N. Ayache, "Directional anisotropic diffusion applied to segmentation of vessels in 3D images," in *Scale-Space Theory in Computer Vision*, ser. Lecture Notes in Computer Science. Berlin, Germany: Springer-Verlag, 1997, vol. 1252, pp. 345–348.

[33] W. M. Wells III, W. E. L. Grimson, R. Kikinis, and F. A. Jolesz, "Adaptive segmentation of MRI data," *IEEE Trans. Med. Imag.*, vol. 15, pp. 429–442, Aug. 1996.

[34] K. M. Pohl, W. M. Wells III, A. Guimond, K. Kasai, M. E. Shenton, R. Kikinis, E. Grimson, and S. K. Warfield, "Incorporating nonrigid registration into expectation-maximization algorithm to segment MR images," in *Proc. 5th Int. Conf. Medical Image Computing and Computer-Assisted Intervention (MICCAI 2002)*, 2002, pp. 564–571.

[35] D. W. Shattuck, S. R. Sandor-Leahy, K. A. Schaper, D. A. Rottenberg, and R. M. Leahy, "Magnetic resonance image tissue classification using a partial volume model," *Neuroimage*, vol. 13, pp. 856–876, 2001.

[36] A. M. Dale, B. Fischl, and M. I. Sereno, "Cortical surface-based analysis. I. Segmentation and surface reconstruction," *Neuroimage*, vol. 9, pp. 179–194, 1999.

[37] P. C. Teo, G. Sapiro, and B. A. Wandell, "Creating connected representations of cortical gray matter for functional MRI visualization," *IEEE Trans. Med. Imag.*, vol. 16, pp. 852–863, Dec. 1997.

[38] C. A. Davatzikos and J. L. Prince, "An active contour model for mapping the cortex," *IEEE Trans. Med. Imag.*, vol. 14, pp. 65–80, Mar. 1995.

[39] X. Zeng, L. H. Staib, R. T. Schultz, and J. S. Duncan, "Segmentation and measurement of the cortex from 3-D MR images using coupled surfaces propagation," *IEEE Trans. Med. Imag.*, vol. 18, pp. 100–111, Oct. 1999.

[40] C. Xu, D. Pham, and J. Prince, "Finding the brain cortex using fuzzy segmentation, isosurfaces, and deformable surfaces," in *Proc. XVth Int. Conf. Information Processing in Medical Imaging (IPMI 97)*, 1997, pp. 399–404.

[41] J. Rexilius, "Physics-Based Nonrigid Registration for Medical Image Analysis," masters thesis, Med. Univ. Luebeck, Luebeck, Germany, 2001.

[42] S. Smith, "Fast robust automated brain extraction," *Human Brain Map.*, vol. 17, pp. 143–155, 2002.

[43] L. Lemieux, G. Hagemann, K. Krakow, and F. G. Woermann, "Fast, accurate, and reproducible automatic segmentation of the brain in T1-weighted volume MRI data," *Magn. Reson. Med.*, vol. 42, pp. 127–135, 1999.

[44] M. Brummer, R. Mersereau, R. Eisner, and R. Lewine, "Automatic detection of brain contours in MRI data sets," *IEEE Trans. Med. Imag.*, vol. 12, pp. 153–166, June 1993.

[45] M. S. Atkins and B. Mackiewich, "Fully automatic segmentation of the brain in MRI," *IEEE Trans. Med. Imag.*, vol. 17, pp. 98–107, Feb. 1998.

[46] C. A. Cocosco, V. Kollokian, R. K.-S. Kwan, and A. C. Evans, "BrainWeb: Online interface to a 3D MRI simulated brain database," *Neuroimage*, vol. 5, no. 4 part 2/4 S245, 1997.

[47] S. K. Warfield, K. H. Zou, and W. M. Wells III, "Validation of image segmentation and expert quality with an expectation-maximization algorithm," in *Proc. 5th Int Conf. Medical Image Computing and Computer-Assisted Intervention (MICCAI 2002)*, 2002, pp. 298–306.

[48] S. K. Warfield, K. H. Zou, and W. M. Wells, "Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation," *IEEE Trans. Med. Imag.*, vol. 23, 2004, to be published.

[49] K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetens, "Automated model-based tissue classification of MR images of the brain," *IEEE Trans. Med. Imag.*, vol. 18, pp. 897–908, Oct. 1999.