CLOUDS: A MODEL FOR SYNERGISTIC IMAGE SEGMENTATION

*Paulo A. V. Miranda, Alexandre X. Falcão**

State University of Campinas
Institute of Computing
Campinas, Brazil
{pavm,afalcao@ic.unicamp.br}

Jayaram K. Udupa

Univ. of Pennsylvania
Dept. of Radiology
Philadelphia, PA, USA

ABSTRACT

Image segmentation consists of recognizing the object in the image and precisely delineating its spatial extent. We present a model, called *clouds*, that exploits the synergism which commonly exists between recognition and delineation for more effective segmentation. The model can reduce user's intervention to simple corrections or even eliminate it altogether, achieving high accuracy. We evaluate the method in the task of 3D MR image segmentation of the brain in isolating automatically: brain without medulla and spinal cord; just the cerebellum; and the brain hemispheres without medulla, spinal cord, and cerebellum. These structures are connected in several parts, which poses a serious challenge for simplistic segmentation strategies. The entire process takes a few seconds on modern PCs and provides accurate results. The applications for *clouds* go beyond medical imaging, opening new vistas in a variety of areas served by segmentation.

***Index Terms*—** MR image segmentation, image foresting transform, model-based and image-based segmentation, graph-cut measures, medical image processing.

1. INTRODUCTION

Image segmentation involves effective object *recognition* and *delineation*. Recognition is the task of determining an object's approximate whereabouts and location in the image. Delineation completes segmentation by defining the precise spatial extent of the object. Humans usually outperform computers in object recognition, but reverse is true for delineation. While the user can often solve the recognition problem by a simple (seed) point selection or by an appropriate initialization action, perfectly repeatable delineation is challenging because of intra and inter operator subjectivity. On the other hand, computers can perform repeatable delineation, but the absence of global information makes computer recognition of objects a difficult task. This explains why some successful interactive approaches combine recognition by the user with delineation by the computer in a synergistic way, for more effective and foolproof segmentation [1, 2].

Segmentation methods can be divided into *model-based* and *image-based* approaches. Model-based methods create statistical models by employing supervised learning. A training set of images/objects is provided with appropriate human interaction and these data are registered into a common reference space to form the model. Active shape models [3] (ASM) and atlas-based approaches [4] are examples of model-based methods that have been used for MR image segmentation of the anatomic structures of the brain [5, 6]. Accurate registration is a separate problem in these methods which is also required for effective object recognition. In ASM, landmarks have to be selected on the surface of the training objects and their correspondence provides a statistical model of possible variations in shape. The registration between the image and the model is required for recognition. Delineation sometimes ignores important image information, by the act of forcing the results to fit with the model. Brain atlases are usually created by registration of training images based on certain landmarks and with no segmentation. In the reference space, image structures suffer from different degrees of distortion and the matching among corresponding voxels is imprecise. However, a brain atlas can help in the recognition task, leaving delineation for other approaches [6]. Image-based methods exploit image properties for more effective delineation, but their lapses in global information makes recognition an insurmountable problem [7, 8]. In order to reduce/eliminate the need for user intervention, it is then important to combine recognition by model-based approaches with delineation by effective image-based methods [9, 10].

From interactive segmentation strategies, we know that the human operator plays the role of a model while the computer performs delineation, and both operate in a synergistic way. The hybrid approach presented in this paper, called *clouds*, illustrates this strategy. Some advantages of this method are: the model is very simple to create from a set

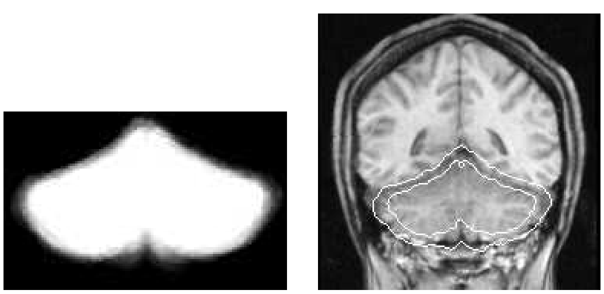
*The authors thank FAPESP, and CNPq for the financial support, and Dr. Fernando Cendes (FCM-UNICAMP) for the images.

of training objects; image registration can help, but it is not essential either among training objects or for recognition; and delineation can be done by any accurate image-based method. The next sections present the principles underlying *clouds*, its application to MR image segmentation of the brain with evaluation for several structures, and our conclusions.

2. THE *CLOUDS* MODEL

For a given structure of interest (e.g., cerebellum), a set of training objects must be provided. These objects should capture among them shape variations of that structure in order to teach the computer how to recognize it in the image. Instead of registering them as in ASM, we only translate them onto a common reference point (geometric center) and separate them into *groups* (clusters) of high similarity in shape, rotation and scale (texture may be used as well). In medical imaging and other applications (e.g., license plate recognition), it is possible to acquire images as per a disciplined regimen so that a small number of groups will suffice (one in many instances).

The average of the objects in each group creates (i) an interior region consisting of voxels that belong to all objects of the group, (ii) an exterior region with voxels that do not belong to any object in the group, and (iii) an uncertainty region composed of voxels that belong to some but not to all objects in the group. The fuzzy appearance of the resulting image resembles a *cloud* (Figure 1a). The *cloud model* then is a triple consisting of a set of *clouds* (one *cloud* per group), a delineation algorithm (image-based approach), and a functional. To segment a new image, each *cloud* moves over the image and, for each position, delineation is done inside the uncertainty region to obtain a candidate object. The functional is evaluated to obtain a matching score for recognition, by taking into account local and global object properties (e.g., shape and texture). The desired segmentation is expected to be the one with maximum score among those obtained from all *clouds*.



(a) (b)

Fig. 1. (a) A coronal slice of the 3D *cloud* of the cerebellum. (b) The uncertainty region over a slice of a test image.

Note that delineation is constrained in the uncertainty region, which is defined by the model, but the delineation method exploits image properties according to the image-based approach. Recognition is based on the functional, but it is applied to the delineated objects. Thus, the model employs recognition and delineation in a tightly coupled manner.

2.1. Model

An image $\hat{I}$ is a pair $(\mathcal{I}, \vec{I})$ where $\mathcal{I} \subset Z^n$ is the image domain and $\vec{I}(p)$ assigns a set of m scalars $I_i(p)$, $i = 1, 2, \ldots, m$, to each voxel $p \in \mathcal{I}$. This definition applies to multi-dimensional and multi-parametric images. We are interested in $n = 3$ and $m \geq 1$. The subindex i is removed when $m = 1$. In a binary image $\hat{I}$, $I(p) = 1$ for object voxels and $I(p) = 0$ for background voxels. The *clouds* are obtained by grouping and averaging the training set of binary images, with all objects translated to a fixed reference point (geometric center). The result is a *cloud image* $\hat{C} = (\mathcal{I}, \vec{C})$, where $C_i(p) \in [0, 1]$, $i = 1, 2, \ldots, m$. For any *cloud* i, $C_i(p) = 1$ in its interior, $0 < C_i(p) < 1$ in its uncertainty region, and $C_i(p) = 0$ in its exterior. The *cloud model* consists of a *cloud* image $\hat{C}$, a delineation algorithm A, and a functional F.

2.2. Implementation

Many approaches can be taken for grouping objects into *clouds*, for delineating candidate objects in a test image, and for recognizing the desired object by a functional. The grouping can be done by mapping the training objects as nodes of a complete graph, whose arcs between objects are weighted by their similarity values. The similarity between objects can be measured, for example, by Dice similarity. The graph can be partitioned into maximal clicks (groups), where all pairs of objects have similarity above a threshold. In this work, we found that a single *cloud* is sufficient for each structure.

For delineation, we propose Algorithm 1 based on the image foresting transform [11] (IFT). A 3D gradient image $\hat{I} = (\mathcal{I}, I)$ is interpreted as a graph $(\mathcal{I}, \mathcal{A})$, whose nodes are the voxels $p \in \mathcal{I}$ and the arcs $(p, q) \in \mathcal{A}$ are defined between 6-neighbors. The uncertainty region is a set $\mathcal{U} \subset \mathcal{I}$ of voxels (Figure 1b). The interior and exterior regions contain boundary voxels, which have at least one voxel in $\mathcal{U}$ as a 6-neighbor. These boundary voxels form one internal set $\mathcal{S}_i$ and one external set $\mathcal{S}_e$ of seeds for the IFT. Each arc $(p, q) \in \mathcal{A}$ is weighted by the mean gradient value $\frac{I(p)+I(q)}{2}$. Any sequence of adjacent voxels forms a path and the cost of a path is the maximum arc weight along it. It is expected that the arc weights within $\mathcal{U}$ are higher on the object's boundary than inside and outside it. The seed sets $\mathcal{S}_i$ and $\mathcal{S}_e$ compete for voxels in $\mathcal{U}$, such that a voxel receives label $L(p) = 0$ if the minimum-cost path from the seed sets comes from $\mathcal{S}_e$, and label $L(p) = 1$ otherwise.

By centering $\mathcal{U}$ at each voxel of a search region, a score F is obtained for each candidate object computed by the delineation algorithm. As the boundary between voxels with label

1 and voxels with label 0 represents a cut in the graph $(\mathcal{I}, \mathcal{A})$, we propose to use the mean-cut measure [12] as functional F. When $\mathcal{U}$ contains the object's boundary (Figure 1b), we expect the object to be defined by the union between the interior of the *cloud* and the voxels with labels $L(p) = 1$ in $\mathcal{U}$.

Algorithm 1 performs delineation and functional computation simultaneously for any given image location. It can take time proportional to the number of voxels in $\mathcal{U}$ (sublinear), when the priority queue Q is implemented as suggested in [1].

Algorithm 1 – DELINEATION ALGORITHM

INPUT:	Gradient image $\hat{I}$, adjacency $\mathcal{A}$, seed sets $\mathcal{S}_i$ and $\mathcal{S}_e$, and uncertainty region $\mathcal{U}$.
OUTPUT:	Label map L initially zeroed and mean cut $F = 0$ initially.
AUXILIARY:	Cost map c initially zeroed, variables cst and cut size $sz = 0$ initially, priority queue Q initially empty, and status map s to indicate when a voxel has been inserted in Q (1), has never been inserted in Q (0), and has been removed from Q (2).

For all $p \in \mathcal{U}$, *set* $c(p) \leftarrow +\infty$ *and* $s(p) \leftarrow 0$.
For all $p \in \mathcal{S}_i$, *set* $L(p) \leftarrow 1$, $s(p) \leftarrow 1$, *and insert* p *in* Q.
For all $p \in \mathcal{S}_e$, *set* $L(p) \leftarrow 0$, $s(p) \leftarrow 1$, *and insert* p *in* Q.
While Q *is not empty, do*
    Remove from Q *a voxel* p *such that* $c(p)$ *is minimum.*
    Set $s(p) \leftarrow 2$.
    For each q *such that* $(p, q) \in \mathcal{A}$, *do*
        If $c(q) > c(p)$, *then*
            Compute $cst \leftarrow \max\{c(p), \frac{I(p)+I(q)}{2}\}$.
            If $cst < c(q)$, *then*
                If $s(q) = 1$, *remove* q *from* Q.
                Set $c(q) \leftarrow cst$, $L(q) \leftarrow L(p)$.
                Insert q *in* Q *and* $s(q) \leftarrow 1$.
        Else
            If $s(q) = 2$ *and* $L(q) \neq L(p)$, *then*
                Set $F \leftarrow F + \frac{I(p)+I(q)}{2}$.
                Set $sz \leftarrow sz + 1$.
Set $F \leftarrow F/sz$.

Note that, before changing position of $\mathcal{U}$, the maps and auxiliary variables can be reinitialized in sublinear time, such that the search for a desired object can be done more efficiently.

3. APPLICATION TO BRAIN SEGMENTATION

The brain structures segmented in this work are: (S1) the brain (GM + WM) without medulla and spinal cord, (S2) the cerebral hemispheres without medulla, spinal cord, and cerebellum, (S3) the cerebellum, (S4) the right hemisphere without medulla, spinal cord, and cerebellum, and (S5) the left hemisphere without medulla, spinal cord, and cerebellum.

Object	Mean (%)	Std dev (%)	Time(s)
S1	96.95	0.55	18.7
S2	97.17	0.55	12.2
S3	93.13	1.20	1.9
S4	96.11	0.58	4.1
S5	96.02	0.59	4.1

Table 1. Mean times and mean and standard deviation of the Dice similarity.

The CSF is eliminated by removing voxels below the Otsu's threshold computed in the original images. The hemispheres are connected through the corpus callosum. The cerebellum is connected to the rest of the brain through the spinal cord and through its top due to partial volume. The absence of a clear boundary between these structures poses a challenge for segmentation.

Five *clouds* are created, one for each structure, by using interactive segmentation based on differential IFTs [2]. The MR-T1 images are interpolated to the same cubic dimensions ($0.98mm^3$) and aligned by the mid-sagittal plane [13] (MSP). This approach is fast (a few seconds), free of parameters, and independent of templates. Interpolation and alignment allow a single *cloud* for each structure and the MSP reduces the search region as follows. The center of $\mathcal{U}$ moves inside the MSP to segment S1, S2, and S3. S4 and S5 are obtained by reducing the search to the right and left sides of the MSP within S2.

Our strategy for brain segmentation follows the pipeline shown in Figure 2. First the method segments S1, then S2 is separated from S3. Although S3 should be the residue of S1 minus S2, the method is repeated by constraining the search into the intersection between this residue and the MSP, to delineate S3 more independently of possible errors in S2. The same happens when the method segments S4 and S5 on the right and left sides of S2.

We also used a multiscale search to speed up the recognition task. A three-level Gaussian pyramid was computed. We started the search at the lowest resolution and refined the best detected locations in the higher resolutions.

4. RESULTS

We have evaluated the method on the MRI datasets of 18 normal subjects from both sexes, in the age range from 25 to 49 years. The images were acquired with a 2T Elscint scanner and voxel size of $0.98 \times 0.98 \times 1.00\ mm^3$. We used the leave-one-out approach to compute the mean and standard deviation of the Dice similarity measure between the interactive (ground truth) and automatic segmentation results, and the mean execution times for each segmentation using a 3GHz Pentium IV PC. Table 1 shows that accurate results can be obtained in a few seconds.

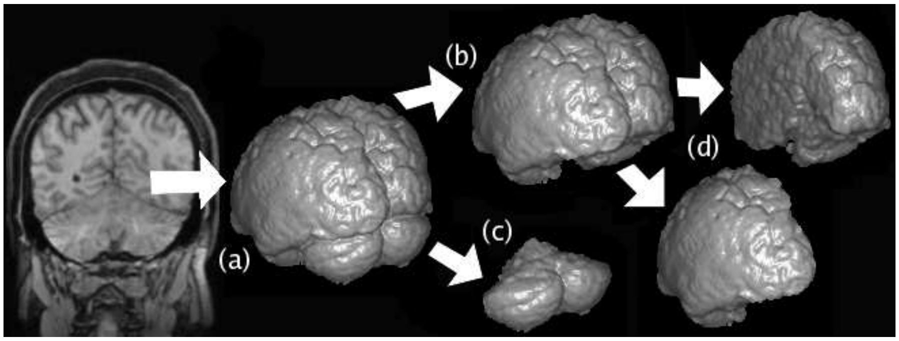


Fig. 2. (a) The brain is extracted without medulla and spinal cord (S1), then (b) both hemispheres are extracted (S2). (c) The cerebellum (S3) is obtained by constraining the search within the intersection of the MSP and the residue of S1 minus S2. (d) The hemispheres S4 and S5 are separated from each other by constraining the search to the right and left sides of the MSP.

5. CONCLUSION

We introduced *clouds*, a new approach for synergistic image segmentation, which weaves delineation and recognition as tightly coupled tasks for more effective segmentation. The method consists of a fuzzy shape model, a delineation algorithm (image-based approach), and a functional. We implemented it using an IFT algorithm for delineation and the mean graph-cut measure as functional. The method was evaluated in a difficult segmentation task involving brain structures in MR-T1 images and presented highly accurate results obtained within a few seconds per structure. *Clouds* seems to be a simple yet powerful model-based strategy that achieves automation in both recognition and delineation. Our future goals will be to expand the purview of the model, to segment other GM structures, and to carry out more extensive evaluation.

6. REFERENCES

[1] A.X. Falcão, J.K. Udupa, and F.K. Miyazawa, "An ultra-fast user-steered image segmentation paradigm: Live-wire-on-the-fly," *IEEE TMI*, vol. 19, no. 1, pp. 55–62, Jan 2000.

[2] A.X. Falcão and F.P.G. Bergo, "Interactive volume segmentation with differential image foresting transforms," *IEEE TMI*, vol. 23, no. 9, pp. 1100–1108, 2004.

[3] T. Cootes, C. Taylor, D. Cooper, and J. Graham, "Active shape models – their training and application," *CVIU*, vol. 61, no. 1, pp. 38–59, 1995.

[4] K.A. Ganser, H. Dickhaus, R. Metzner, and C.R. Wirtz, "A deformable digital brain atlas system according to Talairach and Tournoux," *Medical Image Analysis*, vol. 8, no. 1, pp. 3–22, Mar 2004.

[5] Nicolae Duta and Milan Sonka, "Segmentation and interpretation of MR brain images using an improved knowledge-based active shape model," in *IPMI*, 1997, pp. 375–380.

[6] V. Grau, A.U.J. Mewes, M. Alcaniz, R. Kikinis, and S.K. Warfield, "Improved watershed transform for medical image segmentation using prior information," *IEEE TMI*, vol. 23, no. 4, pp. 447–458, Apr 2004.

[7] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Intl. J. of Computer Vision*, vol. 1, pp. 321–331, 1987.

[8] S. Beucher and F. Meyer, "The morphological approach to segmentation: The watershed transformation," in *Mathematical Morphology in Image Processing*, chapter 12, pp. 433–481. Marcel Dekker, 1993.

[9] J.Liu and J.K. Udupa, "Oriented active shape models," in *SPIE on Medical Imaging*, 2006, vol. 6144, pp. 30–40.

[10] J.Liu, *Synergistic hybrid image segmentation: combining model and image-based strategies*, Ph.D. thesis, Dept. of Bioengineering, Univ. of Pennsylvania, Nov 2006, advisor: J. Udupa.

[11] A.X. Falcão, J. Stolfi, and R.A. Lotufo, "The image foresting transform: Theory, algorithms, and applications," *IEEE TPAMI*, vol. 26, no. 1, pp. 19–29, 2004.

[12] S. Wang and J.M. Siskind, "Image segmentation with minimum mean cut," in *ICCV*, 2001, vol. 1, pp. 517–525.

[13] F.P.G. Bergo, G.C.S. Ruppert, L.F. Pinto, and A.X. Falcão, "Fast and robust mid-sagittal plane location in 3D MR images of the brain," in *Intl. Conf. on Bio-inspired Systems and Signal Processing*, Funchal, Portugal, Jan 2008, to appear.