Automatic segmentation of clinical structures for RTP: Evaluation of a morphological approach

Bueno, Gloria , Fisher, Mark , Burnham, K. , Mills, John A. and Haas, O.C.L.

Automatic segmentation of clinical structures for RTP: Evaluation of a morphological approach

Gloria Bueno[a], Mark Fisher[b], Keith Burnham[c], John Mills[d] and Olivier Haas[c*]

[a]LSIIT, Universit Louis Pasteur - CNRS-UPRES-A 7005, 4. Bd. Sbastien Brant, 67400 Illkirch, FR
[b]School of Information Systems, University of East Anglia, Norwich NR4 7TJ, UK
[c]Control Theory and Applications Centre, Coventry University, Priory Street, Coventry CV1 5FB, UK
[d]Walsgrave Hospital NHS Trust, Clifford Bridge Road, Walsgrave, Coventry CV2 2DX, UK

Abstract. This paper describes a morphological segmentation technique developed to assist clinicians and radiologists in conformal RTP. The system uses a watershed transformation to give an initial over-segmentation of a 2D CT; this is then refined by homotopy modification with markers identified automatically from the image histogram and the final segmentation is represented by a region adjacency graph which may be interactively edited by the user. To evaluate the system, 77 CT images from 11 patients have been segmented and the results compared against those produced by 5 clinical experts. The approach has enabled the quality of the automatic segmentation to be assessed in terms of the overall expert intra/inter variability. It has been proved good agreement between computer generated and expert manual delineations.

1 Introduction

Image segmentation is a key problem in computer vision, considered by many authors, [1] and an important stage in radiotherapy treatment planning (RTP) that is often performed manually, turning into a tedious and repetitive task [2]. Nevertheless, at present there is no universally accepted segmentation method that is proven to work on a large representative image database. Medical practice and the need for clinical accountability have driven researchers to develop interactive tools for medical segmentation [3], however these systems are still susceptible to intra/inter clinician variability and human error. Moreover, although quantitative evaluation of medical segmentation algorithms is an important step towards establishing the validity and clinical applicability of an algorithm few researchers subject their algorithms to rigorous testing [4]. The problems often associated with evaluation are: lack of a definitive ground truth, difficulty in defining a metric, lack of standardized statistical protocol and tedious data collection. Some methodologies have recently been proposed [4] but there is still a compelling need for evaluation and comparison using standardized protocols.

We address these problems by developing a method that is applicable to X-ray CT segmentation of regions of interest (ROI) in the pelvic area, that is, the bladder, the rectum and the seminal vesicles (sem.ves.) for RTP. Additionally, we have attempted to evaluate the system by direct comparison with manual delineations by clinical experts at Walsgrave NHS Trust, Coventry, UK. In the evaluation of our system we considered both the shape of the delineated region, compared using measures derived from the Hausdorff distance and geometric signature waveforms, and its gross position, assessed by calculating a ratio of common area. The results were evaluated by Receiver Operating Characteristics (ROC) analysis, Kappa statistic and the overlapping area detected.

2 The Segmentation Tool

The watershed transformation was used to produce an initial segmentation of the X-ray CT (Fig.1a). This transformation has its foundation in the field of mathematical morphology and has been extensively described in the literature, [5]. The idea of the watershed transform is to treat the image as a topographical surface, $\mathcal{T}(p)$, such that grey level image intensities are altitudes. The transformation attributes a geodesic *influence zone* to each regional minima. It is an attractive segmentation operator since it is efficient, robust, always producing closed boundaries and may be easily extended to be applied on 3D image data, [6]. In order to segment the edge contours of the image, the watershed is applied on its gradient i.e. $\mathcal{T}(p) = |\nabla I(x, y)|$. When the watershed transformation is applied to an image an over-segmentation arises from the many regional minima present. Many researchers have addressed the over-segmentation problem, [7] and an elegant solution from mathematical morphology is homotopy modification [8]. Homotopy modification uses markers to pre-select pixels within homogeneous regions of the image which are subsequently used to suppress the small minima which produce the over-segmentation, thus enabling the image to be defined in terms of a much simpler region adjacency graph (RAG). The selection of marker pixels

---

[*]email for correspondence: mhf@sys.uea.ac.uk and bueno_gloria@yahoo.com

is a difficult task that is usually done interactively. In this work we have used an automatic marker extraction algorithm which derives markers from an analysis of the image histogram, as the pelvic CT images exhibit a similar characteristic set of peaks, corresponding to the different organ regions; shown in Fig.1b. Our algorithm identifies marker pixels within each peak and back-projects them into the image iteratively until the over-segmentation is reduced to an acceptable level (Fig.1c). Other marker extraction methods have been reported in the literature, these are based on a hierarchical process using morphological filters by reconstruction which look for flat regions. Our somewhat simpler approach works successfully on our database of CT images derived from X-ray transmission. This approach is consistent with the view expressed by some researchers who have predicted that image quality from future generations of CT scanners will be so good that they may be processed solely in terms of flat zones [9].

The interactive segmentation tools are the two primary edit operations, *cut* and *merge*, applied to the RAG. The RAG is a kind of lattice induced by a grid and defined by a 4-tuple $\mathcal{G} = (V_{\mathcal{G}}, E_{\mathcal{G}}, \mathcal{L}, \alpha)$. Although the over-segmentation is greatly reduced by the homotopy modification described earlier, some ROIs remain shattered, or on the other hand, they may have been merged, as it is impossible to automatically pre-select a perfect set of markers. This is particularly evident in the rectum and sem.ves., which sometimes require manual interaction by using the editor.
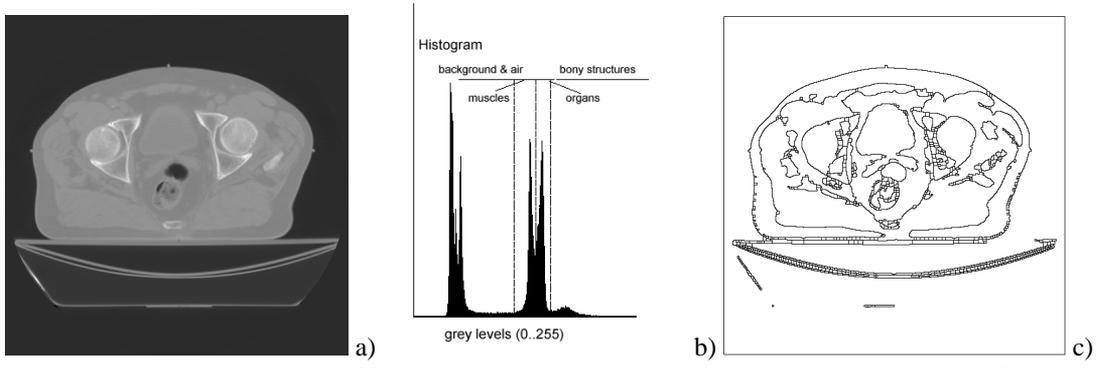

Figure 1. Watershed segmentation. a) CT from a study patient, b) image histogram, c) *unsupervised* segmentation.

3 Evaluation Method

A real image database comprising 11 pelvic CT studies each having 7 slices was used to evaluate the computer generated delineations produced by the segmentation tool. Five clinicians from the Radiotherapy Department of Walsgrave Hospital were asked to delineate the PTA (Prescribed Target Area) ensuring that absorb the CTA (Clinical Target Area) corresponding to the bladder, rectum, and sem.ves., including the GTA (Gross Tumor Area), [2]. The data set was delineated twice by each of the 5 clinical experts in order to address the problem of inter/intra-observer variability. Then the multiple expert observers' outlines were averaged to generate the ground truth (GT). The procedure used to average the contours is similar to the one described by Chalana *et al.* [4] and is based on establishing one-to-one correspondence between the points constituting the different curves to be averaged. The segmentation was also assessed by a statistical analysis of the ratio of detected pixels and the distance between boundaries. This is described as follows:

3.0.1 Area Measurement

Let GT denote the segmented area representing *ground truth*, and $\overline{GT}$ its complement, AS the segmented area obtained by the semi-automatic approach. ROC analysis was used to assess the sensitivity and specificity of the method by the fraction of True Positive Fractions (TPF) and False Positive Fractions(FPF) detected [10]. Ratios of overlapping areas were also assessed by applying the similarity Kappa Index, (KI), [11] and the overlap, [12]:

$$ TPF = \frac{AS \cap GT}{GT} \qquad FPF = \frac{|AS - GT|}{\overline{GT}} \qquad KI = 2 \times \frac{GT \cap AS}{GT + AS} \qquad overlap = \frac{GT \cap AS}{GT \cup AS} \qquad (1) $$

Table 1, rows 1, 3 and 5 show the average result of this measurements when validating the computer generated segmentation against the GT for all the CT images considered (7×11) and for the different ROIs. Rows 2, 4 and 6 show the average results of the manual delinations, ($2 \times 5 \times 7 \times 11$), against the GT.

3.0.2 Distance Measurement

Hausdorff Distance (HD) [4,9] was used to assess the distance between the computer generated contour, $\mathcal{X}_{AS}$, and the GT, $\mathcal{X}_{GT}$. This is defined as the maximun of the distance to the closest point's between the two curves. If the two curves are represented as sets of points, $\mathcal{X}_{AS} = a_1, a_2, ..., a_m$ and $\mathcal{X}_{GT} = b_1, b_2, ..., b_m$, the distance to the closest point, a_i to the curve $\mathcal{X}_{GT}$, $d(a_i, \mathcal{X}_{GT})$, and the HD between $\mathcal{X}_{AS}$ and $\mathcal{X}_{GT}$, $e(\mathcal{X}_{AS}, \mathcal{X}_{GT})$ is given by:

$$d(a_i, \mathcal{X}_{GT}) = min_j \parallel b_j - a_i \parallel, \qquad e(\mathcal{X}_{AS}, \mathcal{X}_{GT}) = max(max_i\{d(a_i, \mathcal{X}_{GT})\}, max_j\{d(b_j, \mathcal{X}_{AS})\}) \qquad (2)$$

Radial signature of the delineated boundaries were also calculated. First we calculate a common centroid of the boundaries from which radial lines are drawn projecting outward. The intersection of these radial lines with $\mathcal{X}_{AS}$ and $\mathcal{X}_{GT}$ define the corresponding points' distributions. Points were placed in $10°$ sectors so 36 points were analysed. However, this measure is only valid if each point of the boundary is visible from the centroid. This is usually the case for the bladder and rectum, but for the sem.ves., we made a partition between right and left. To assess the significance of the variations in the radial signatures an unpaired Students t test was carried out under the *null hypothesis* of equal distribution means. This is computed by:

$$|t| = \frac{|\bar{a}_{AS} - \bar{b}_{GT}|}{\sigma_D}, \qquad \sigma_D = \sqrt{\frac{\sum_{i \in AS}(a_i - \bar{a}_{AS})^2 + \sum_{i \in GT}(b_i - \bar{b}_{GT})^2}{N_{AS} + N_{GT} - 2}\left(\frac{1}{N_{AS}} + \frac{1}{N_{GT}}\right)} \qquad (3)$$

where: N_{AS} and N_{GT} are the number of points of the distributions, $\mathcal{X}_{AS} = \{a_{36}\}$ and $\mathcal{X}_{GT} = \{b_{36}\}$, with $(N_{AS} + N_{GT} - 2) = 70$ degrees of freedom. The range of values within which the distribution mean is likely to lie are the confidence intervals, given by $((\bar{a}_{AS} - \bar{b}_{GT}) \pm (|t|_\alpha \times \sigma_D))$. with α being the significance. If α is small, conventionally < 0.05, the null hypothesis is rejected as implausible, and the observed difference is very significant. The results may be expressed in terms of *confidence levels*, $100(1 - \alpha)\%$, that is the minimum and maximum values of the range of plausible or believable values. Usually confidence statements are uttered at levels of 95% and 99%. The average values of $|t|$ for the different ROI for the 11 study patients are shown in Table 2 rows $1, 3$, and 5 for the computer and rows $2, 4$ and 6 for the manual segmentation.

4 Results

Structures	TPF	FPF	Kappa Index	Overlaping
Bladder	0.85 ± 0.22	0.17 ± 0.12	0.93 ± 0.14	0.87 ± 0.17
	0.87 ± 0.12	0.25 ± 0.19	0.89 ± 0.22	0.92 ± 0.22
Rectum	0.77 ± 0.17	0.27 ± 0.12	0.84 ± 0.25	0.73 ± 0.22
	0.89 ± 0.27	0.17 ± 0.16	0.92 ± 0.14	0.96 ± 0.14
Vesicles	0.68 ± 0.15	0.35 ± 0.23	0.91 ± 0.31	0.85 ± 0.27
	0.79 ± 0.24	0.21 ± 0.22	0.87 ± 0.23	0.91 ± 0.21

Table 1. Area Detected Validation for Computer and Manual Segmentation.

Structure	Radial Signature			Hausdorff Distance
	$|t|$	% Agreement		mm
	$(\mu \pm \sigma)$	$> 95\%$	$> 99\%$	$(\mu \pm \sigma)$
Bladder	0.9 ± 0.2	$82 \pm 6\%$	$92 \pm 4\%$	1.2
	0.9 ± 0.2	$81 \pm 6\%$	$92 \pm 2\%$	1.5 ± 0.2
Rectum	1.2 ± 0.4	$74 \pm 4\%$	$90 \pm 2\%$	1.7
	0.8 ± 0.1	$87 \pm 2\%$	$95 \pm 2\%$	0.8 ± 0.1
Vesicles	1.1 ± 0.3	$62 \pm 4\%$	$72 \pm 6\%$	1.5
	0.9 ± 0.2	$82 \pm 4\%$	$89 \pm 2\%$	1.1 ± 0.1

Table 2. Boundary Distance Validation for Computer and Manual Segmentation.

The tests show a varying degree of agreement between the automatic and manual boundaries for different confidence intervals and ROIs. Although there is a good agreement, the approach highlights some areas which are

consistently problematic in the sem.ves. and rectum. These cases were corrected by using the cut and merge operations. Fig.2.a) shows the sample image with 4 selected regions (mouse clicking) within the rectum. These regions are merged in order to segment the whole rectum Fig.2.b), the result may be compared with a typical manual delineation in Fig.2.c). The *unsupervised* segmentation of the whole image took 30 sec. and the selection of all ROIs including in the worst case where manual post-processing is necessary took less than 1 min. (on a Silicon Graphic for 512×512 images). This time is small compared with actual manual delineation. It is estimated that outlining slice-by-slice the PTA (target area and organs at risk) on a data set of 60 slices would take between 40-100 min.
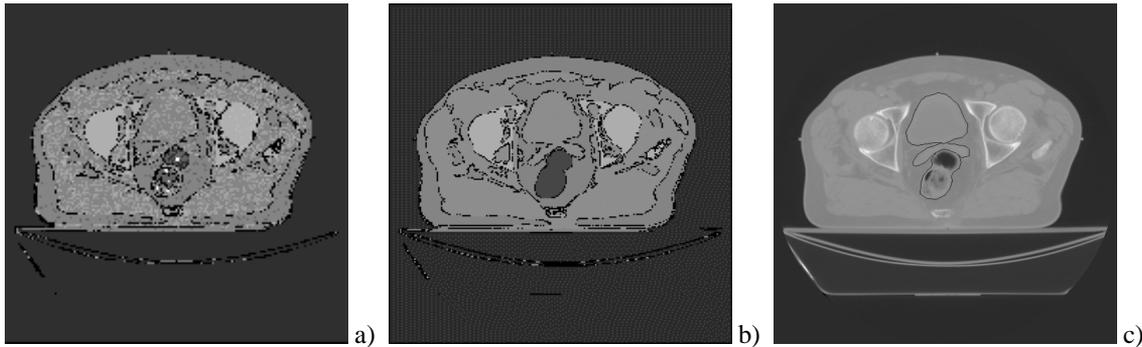

Figure 2. Post-processing. a) Selecting regions to merge. b) After the merging. c) Manual delineation of ROIs.

5   Conclusion

The paper has presented an unsupervised segmentation method and its quantitative evaluation applied to CT images for RTP. The segmentation has been implemented in a flexible way, combining two principal morphological techniques, homotopy modification by *markers* and the watershed transformation. Markers are extracted automatically using information provided by the image histogram. The method has been evaluated quantitatively on a database composed of 77 pelvic CT images of 11 patients by comparing the computer generated boundaries against those drawn manually by 5 clinical experts. The analysis showed there to be no significant difference between computer and manual segmentation. The segmentation tool may be used to successfully segment the PTA in RTP of the pelvic area, yielding significant gains in reproducibility and efficiency, in terms of time. The algorithms have been integrated into an experimental system developed during the course of a PhD research project which was supported by clinicians and radiologists at Walsgrave Hospital. We hope that the statistical evaluation techniques used in this pilot study may prove useful in devising methodologies for large scale trials.

References

1. J. Duncan & N. Ayache. "Medical image analysis: Progress over two decades and the challenges ahead." *IEEE Tran PAMI* **22(1)**, pp. 85–106, Jan 2000.
2. O. Haas. *Radiotherapy Treatment Planning. New System Approaches.* Springer-Verlag, 1998.
3. S. Dellepiane. "The active role of 2d and 3d images: Semi-automatic segmentation." In C. Roux & J.-L. Coatrieux (editors), *Contemporary Perspectives in Three-Dimensional Biomedical Imaging*, chapter 7, pp. 165–190. IOS Press, 1997.
4. V. Chalana & Y. Kim. "A methodology for evaluation of boundary detection algorithms on medical images." *IEEE Trans. on Medical Imaging* **16(5)**, pp. 642–652, 1997.
5. L. Vincent & P. Soille. "Watersheds in digital spaces: An efficient algorithm based on immersion simulations." *IEEE Trans PAMI* **13(6)**, pp. 583–598, 1991.
6. G. Bueno, O. Musse & F. Heitz. "3-D segmentation of anatomical structures in MR images on large databases." *MRIM* **19(1)**, pp. 73–88, 2001.
7. D. Hagyard & M. Razaz. "Analysis of watershed algorithms for greyscale images." *IEEE Transactions on Image Processing* **1**, pp. 41–44, 1996.
8. S. Beucher. "Watershed, hierarchical segmentation and waterfall algorithm." In J. Serra & P. Soille (editors), *Mathematical Morphology and Its Applications to Image Processing*, pp. 69–76. Kluwer Academic, 1994.
9. E. Grimson, M. Leventon, O. Faugeras et al. "Computer vision methods for image guided surgery." In M. Mirmehdi & B. Thomas (editors), *British Machine Vision Conference*, pp. 1–12. 2000.
10. C. Metz. "Basic principles of ROC analysis." *Semin. Nuclear Medicine* **8**, pp. 283–298, 1978.
11. A. Zijdenbos, B. Dawant, R. Margolin et al. "Morphometric analysis of white matter lesions in MR images: Method and validation." *IEEE Tran Med. Imaging* **13(4)**, pp. 716–724, Dec 1994.
12. A. Kelemen, G. Székely & G. Gerig. "Elastic model-based segmentation of 3-D neuroradiological data sets." *IEEE Trans. Med. Imaging* **18(10)**, pp. 828–839, Oct 1999.