Machine Vision and Applications manuscript No.
(will be inserted by the editor)

Fast Parameter-Free Region Growing Segmentation with Application to Surgical Planning

Carlos S. Mendoza · Begoña Acha · Carmen Serrano · Tomás Gómez-Cía

Received: date / Accepted: date

Abstract In this paper we propose a self-assessed adaptive region growing segmentation algorithm. In the context of an experimental virtual-reality surgical planning software platform, our method successfully delineates main tissues relevant for reconstructive surgery, such as fat, muscle, and bone. We rely on a self-tuning approach to deal with a great variety of imaging conditions requiring limited user intervention (one seed). The detection of the optimal parameters is managed internally using a measure of the varying contrast of the growing region, and the stopping criterion is adapted to the noise level in the dataset thanks to the sampling strategy used for the assessment function. Sampling is referred to the statistics of a neighborhood around the seed(s), so that the sampling period becomes greater when images are noisier, resulting in the acquisition of a lower frequency version of the contrast function. Validation is provided for synthetic images, as well as real CT datasets. For the CT test images, validation is referred to manual delineations for 10 cases and to subjective assessment for another 35. High values of sensitivity and specificity, as well as Dice's coefficient and Jaccard's index on one hand, and satisfactory subjective evaluation on the other hand, prove the robustness of our contrast-based measure, even suggesting suitability for calibration of other region-based segmentation algorithms.

Keywords CT · Segmentation · Region Growing · Surgical Planning · Virtual Reality

Carlos S. Mendoza
Dpto. Teoría de la Señal, Escuela Superior de Ingenieros, Av. de los Descubrimientos s/n 41092 Sevilla, Spain
E-mail: csanchez1@us.es

Begoña Acha
Dpto. Teoría de la Señal, Escuela Superior de Ingenieros, Av. de los Descubrimientos s/n 41092 Sevilla, Spain
E-mail: bacha@us.es

Carmen Serrano
Dpto. Teoría de la Señal, Escuela Superior de Ingenieros, Av. de los Descubrimientos s/n 41092 Sevilla, Spain
E-mail: cserrano@us.es

Tomás Gómez-Cía
Unidad de Cirugía Plástica y Grandes Quemados, Hospitales Universitarios Virgen del Rocío, Av Manuel Siurot s/n 41013 Sevilla, Spain
E-mail: tgomezc@gmail.com

1 Introduction

One of the most promising applications of medical image computerized visualization is virtual reality surgical planning. Traditional surgical planning uses volumetric information stored in a stack of intensity-based images, usually from computerized tomography (CT) scanners. Based on a number of these image slices, surgeons build their own mental 3D model of the relevant tissues. This task is difficult even for experienced surgeons. As a consequence, they can miss important information or draw incorrect conclusions due to anatomical variability, either of which can lead to suboptimal treatment strategy decisions [1]. Using volumetric renderings of anatomical structures, and the appropriate virtual tools for basic surgical operations, the complexity of many plastic surgery interventions can be addressed ahead of the actual physical procedure.

A main bottleneck for these computer environments is the delineation of the tissues involved, to such an extent that automated approaches become mandatory. If we can provide a technique that is able to automatically detect the tissues of surgical interest (fat, bone, muscle), then these tissues can be rendered, excisioned directly in 3D, and manipulated or repositioned using the virtual reality platform.

Automatic segmentation is a problem exhaustively addressed in the literature. Any inaccuracies in the process can distort the simulated measures and surgical operations. In this paper we propose a novel segmentation strategy in the

context of the development of a virtual surgical planning environment.

The environment under study was conceived for the simulation of different kinds of reconstructive surgery, providing virtual tools for tissue excision and repositioning, tissue quantification and stereo-lithographic prototyping. In such a framework the need for proper delineation of tissues like fat, muscle and bone becomes crucial. In the initial version of the platform, a simple voxel classification scheme was proposed [2]. Beyond classical segmentation methods like thresholding and simple region growing [3–5], which rely on a number of tuning parameters, developments in automatic parameter-free segmentation approaches become necessary.

In order to pay back the cost and burden of this virtual surgical planning platform development, a wide range of situations were to be covered with the proposed technique. Most available resources, i.e. imaging devices, must be compatible with the method, even ensuring backwards compatibility (for images acquired in the past). As a consequence neither resolution, contrast nor SNR specific standards can be expected. Further, no imaging protocol can be presumed, as regarding reconstruction algorithm, patient positioning in the scanner, presence of radioactive contrast, body segment of the patient to be imaged and so on.

A few authors have referred to the issue of classifying this concrete set of tissue types using a common method. In their work, Zhou et al. [6] developed a technique for skin, fat, muscle, organs and bone segmentation. Their approach was mainly articulated by threshold automatic selection, except for bony tissue, for which the authors made use of a self-assessed adaptive region growing algorithm. Their threshold selection method, based on hierarchical discriminant analysis made assumptions on the histogram that turned out to be unaffordable in our less predictable context. Their strategy for bony tissue, that had been earlier proposed for bronchus segmentation by Law and Heng [7], computed the optimal adaptive threshold by detecting sudden increases in the segmented volume. Its main weakness is the need for an empirical range in this increase, in order to distinguish routinary growth from undesired explosion. This range can hardly be established in our more general problem, since we are faced with varying noise levels, which makes it hard to generalize how much growth must be considered explosion.

Apart from manual trial-and-error adaptive threshold selection [8], some self-assessed region growing strategies, outside our context of application, have been proposed in the past. Some general requirements are expected from any region growing algorithm for clinical use. Such methods must operate in a parameter-free manner, the degree of interaction should be minimal, the output must be provided in reasonable time ranges, and also similar results should be produced for alternative seeding schemes in the region of interest [9].

In their work [10], Hojjatoleslami and Kittler proposed a 3D segmentation method based on finding the global maxima for two different contrast measures which they computed iteratively, as intensity-decreasing pixels were added to the segmented region. The success of the assessment was founded on the assumption that maximal contrast occurred on region boundaries, which is a reformulation of approaches assuming that the variation of the gray values within regions is smaller than across regions, an inherent assumption in all region growing techniques [11]. Unfortunately, the exhaustivity of their per-voxel approach entailed very low computational efficiency, aggravated by their multiple complex peripheral measures. Moreover, the fact that they computed contrast for every voxel addition results in low robustness when facing increasing levels of noise. Revol-Muller et al. [12] used morphological measures to assess the multiplier of the adaptive range in 3D region growing. Instead of computing their assessment function for every pixel addition to the region, they sampled the function for an evenly-spaced set of values.

Another popular region growing technique was proposed by Udupa and Samarasekera [13], based on the fuzzy connectivity measure. This technique computes the path of strongest affinity between each point and the seed point. The fuzzy connectivity algorithm has been used in various medical applications [14–17]. All of these approaches exhibit the inconvenience of parameter-tuning stages prior to their execution, which prevents their use by untrained specialists. For the algorithm in [18], the fuzzy connectivity threshold value is not needed, but as a counter-effect the need for user interaction is considerably increased, and the required computational time rises to impractical values.

In our method we propose an assessment function based on a modified version of the evolving contrast for the region growing sequence. Our approach could be considered as an adaptive sampling strategy for the contrast measure, that is usually computed for every voxel addition to the growing region. This adaptive, regularly-spaced sampling can account for variations in the noise level as it can convey the proper low frequency trend of the evolving contrast, filtering off spurious variations. This results in increased stability and lower computational time, as fewer contrast values ought to be computed.

In Sect. 2 we will analyze the proposed technique, discussing the assumptions made on the intensity distribution of the tissues involved. We will describe the algorithm and the key aspects that constitute the novelty in our approach. In Sect. 3 we will present our validation experiments, and will objectively and subjectively analyze the accuracy, precision and efficiency of the algorithm according to trained expert assessments. The improvements obtained in virtual-reality surgical planning after integration of our technique will also be outlined. Finally, in Sect. 4 we will conclude with some remarks and future possibilities for our technique.

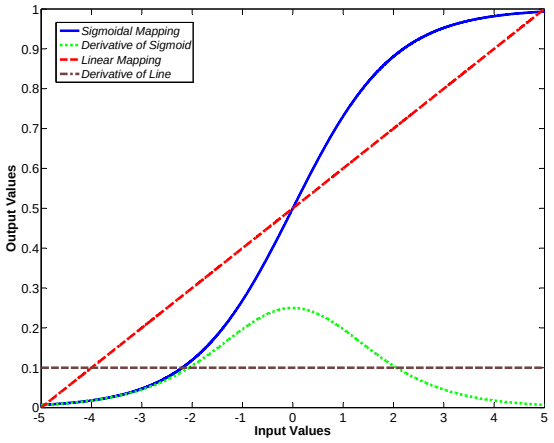


Fig. 1 Comparison of dynamic range extension performed by sigmoidal and linear mappings. When the derivative of the sigmoid overpasses the constant derivative of the linear mapping for a given range of values, then the range gets expanded. The opposite indicates a range contraction.

2 Method

2.1 Normalization and Denoising

From the nature of our regions of interest, we will consider tissue segments that exhibit an inherently constant density, and thus a similar requirement is expected from intensity in the image domain, except for the effects of acquisition noise. We intend to group together all voxels that belong to the same tissue class, and are connected to a provided set of seeds. We model then our object of interest as a connected region whose pixel intensities are sampled from a Gaussian distribution with unknown mean and standard deviation. We may presume that our tissues of interest are surrounded by other tissues derived from other, sometimes adjacent, intensity distributions, like other authors have proposed [19]. Due to the nature of our assessment function as described below, we need intensity values to be mapped in a range of positive values between zero and one. We make use of a sigmoidal mapping with varying slope depending on the statistics of the a region surrounding the provided samples. This can amplify the significance of critical values of the assessment measure in comparison with those obtained with a simpler linear mapping. The reason is that sigmoidal mapping expands the range of values around its center and contracts those away from it. See Fig. 1 for a comparison of the derivatives of linear and sigmoidal functions. Aside from this, any mapping into positive ranges will do the job similarly as concluded from our tests.

If $\mathbf{x}$ is a voxel position, f is a function assigning intensity levels for every position in the image domain, as given by the dataset under study, and $|\cdot|$ denotes cardinality of a set in voxels, we define the following operations on a set A of voxels:

$$\bar{f}_A = \frac{1}{|A|}\sum_{\mathbf{x}\in A} f(\mathbf{x}) \quad , \tag{1}$$

$$\sigma_{f_A} = \sqrt{\frac{1}{|A|}\sum_{\mathbf{x}\in A}\left(f(\mathbf{x})-\bar{f}_A\right)^2} \quad . \tag{2}$$

These expressions are known as the sample mean and sample standard deviation of the population obtained from the elements in A. If N_i is a cubic neighborhood of radius R around the i-th element of a set of M seeds, and $N = \bigcup_i^M N_i$, then $\bar{f}_N$ and σ_{f_N} are the Maximum Likelihood (ML) estimates of the mean and standard deviation of the underlying distribution from the provided population samples if:

$$\bar{f}_N = \frac{1}{M}\sum_{i=1}^{M}\bar{f}_{N_i} \quad , \tag{3}$$

$$\sigma_{f_N} = \frac{1}{\sqrt{M}}\sum_{i=1}^{M}\sigma_{f_{N_i}} \quad . \tag{4}$$

According to these estimates we proceed by application of a non-linear mapping consisting of a sigmoidal window centered at the mean, with width linearly dependent on the standard deviation:

$$f'(\mathbf{x}) = \left(1+\exp\left(-\frac{f(\mathbf{x})-\bar{f}_N}{\left(\frac{K\sigma_{f_N}}{3}\right)}\right)\right)^{-1} \quad . \tag{5}$$

The width of such a sigmoidal window is usually considered to extend $K\sigma$ around the center $\bar{f}_N$ of the mapping. For $K = 3$ the width of the window would be enough to map 99.7% of the samples, of a Gaussian distribution with similar mean and standard deviation. Greater values of K ensure sufficient mapping for the estimated distribution (that of the tissues of interest), avoiding saturation.

Finally, we perform non-linear denoising using an in-slice two-dimensional median filter with kernel radius Γ. Other denoising schemes have been tested without significant impact on the overall performance.

2.2 Self-Assessed Region Growing

Departing from a normalized and filtered version of the image under study, whose intensities lie in the range $[0,1]$, and a manually provided set of seeds, we perform the self-assessed contrast-maximizing algorithm. Basically, the output of the region growing process depends on a confidence interval defined by coefficient k_i, that gets incremented from an initial value k_0 in equal steps of adaptive size Δk. The set of operations for the i-th iteration is described in the following steps:

1. Compute i-th coefficient $k_i = k_0 + i\Delta k$

2. For $j = 0,1,2,...L-1$, and $R_{i0} = N$:
   - Compute tolerance interval I_{ij} using (1), (2) and:
     $$I_{ij} = I(R_{ij}) = \left[\bar{f}'_{R_{ij}} \pm k_i \sigma_{f'_{R_{ij}}} \right] \quad . \tag{6}$$
   - Compute $R_{i(j+1)}$ as the greatest 26-connected set of voxels inside the I_{ij} tolerance interval, that contains R_{ij} and is 26-connected to it.
3. Compute the assessment function $O_i\left(\bar{f}'_{R_{iL}}, \bar{f}'_{P_i}\right)$ according to the intensity average $\bar{f}'_{R_{iL}}$ in R_{iL} and the intensity average $\bar{f}'_{P_i}$ in the external perimeter P_i of R_{iL}:
   $$P_i = (R_{iL} \oplus B) \cap R_{iL} \quad , \tag{7}$$
   where $\oplus$ denotes morphological dilation, and set B is some structuring element, using
   $$O_i\left(\bar{f}'_{R_{iL}}, \bar{f}'_{P_i}\right) = \left| \frac{\bar{f}'_{P_i} - \bar{f}'_{R_{iL}}}{\bar{f}'_{P_i} + \bar{f}'_{R_{iL}}} \right| \quad . \tag{8}$$
4. If O_{i-1} is a local maximum, when compared to O_{i-2} and O_i (only when $i \geq 2$), then the algorithm stops and the output is $R_{(i-1)L}$. Otherwise another iteration takes place

Of all aforementioned parameters only k_0 and Δk are critical for the performance of the algorithm. Higher values of k_0 affects computational efficiency requiring a greater number of iterations before a local maximum of $O\left(\bar{f}_{R_i}, \bar{f}_{P_i}\right)$ is actually found. Therefore, its fine tuning for a specific scanner, could save some computational time. From observation of the region growing sequence, we conclude that these first iterations are typically very fast, so the improvement is frequently negligible. With regard to Δk, the choice must guarantee that the assessment function is being sampled adequately in order to detect its local variations, but also coarsely enough so that contrast variations due to noise do not stop the algorithm on spurious maxima.

Fortunately, the fact that Δk affects the tolerance interval in each iteration through multiplication with the standard deviation, guarantees that the sampling of the assessment function adapts to the noise level in the dataset. As compared to the contrast evolution for regularly increasing intensity values, the resulting sampling will be spaced according to the standard deviation of the grown region (estimated with ever-increasing precision as the region grows). Assuming approximately piece-wise constant regions of interest, this variance can be attributed mostly to noise. As a consequence, for a constant value of Δk, the sampling adapts to the roughness of the assessment function in such a way that the main trend of the evolution is captured, and higher frequency components are ignored.

Beyond these considerations, we have determined the exact value of Δk empirically from the available datasets, and this value has proven to be rather general according to our experimental results.

To illustrate the pertinence of our assessment function, we provide in Fig.11 an exhaustive analysis of the contrast evolution for our region growing sequence in three different segmentation scenarios. Using the same dataset, we have provided one single seed for bone, muscle and fat segmentations. Instead of stopping the algorithm at the first maximum, the region growing proceeds until the iteration right before full-image flooding. Contrast evolution curves where all maxima have been arrow-marked can be seen in Subfigs. 11(a)-(c). The segmentation output for these marked iterations are shown in Subfigs. 11(e)-(g) for bone segmentation, in Subfigs. 11(h)-(k) for muscle segmentation and in Subfigs. 11(l)-(s) for fat segmentation. Notice both the adequateness of the first-maximum criterion, and the semantic value of contrast maxima. The adaptation of the sampling process to the characteristics of the tissue surrounding the seeds, allows for an interactive scheme, in which the user can decide to provide additional seeds after getting an incomplete solution, in an iterative manner. This approach has proven to be naturally accepted by the clinical practitioners on field, as they seem to prefer this intuitive way of incorporating their knowledge into the segmentation process. This is especially relevant as the variety of acquisition scenarios grows, like is the case for our domain of application.

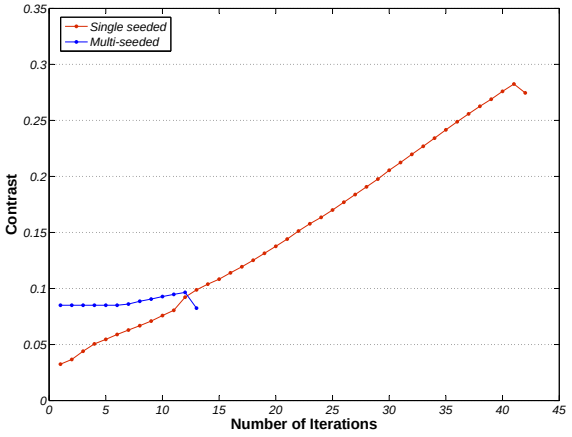


Fig. 2 Contrast evolution comparison for head CT bone segmentation. Red curve depicts single seeding evolution, and blue curve does multiple seeding evolution

To illustrate the effects of multiple seeding, we provide the region growing contrast evolution and growing sequence for a sample bone tissue segmentation in Figs. 2 and 10 until the first maximum is reached. Colors in Fig. 10 indicate iterations, and vary from yellow to violet as the total number is reached. The last iteration occurs earlier when using multiple seeds, as revealed by the smaller number of colors available in Subfig. 10(c). Figure 2 shows how the first maximum is reached earlier, as the assessment function gets more coarsely sampled. This adaptation of the sampling rate simultaneously results in less sensibility to spurious maxima and lower precision in the localization of the maximum of interest. So the greater number of seeds decreases the

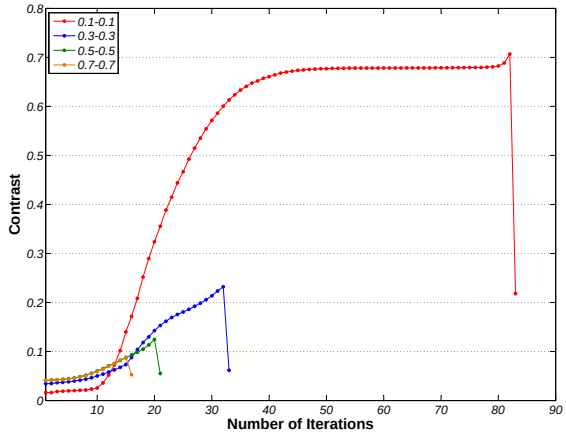
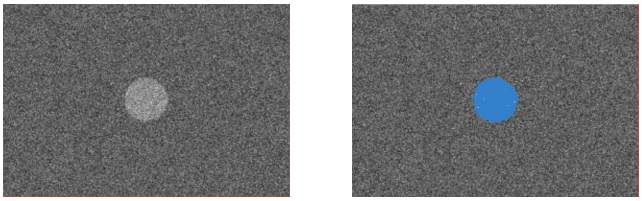


Fig. 3 Contrast evolution for some constant-valued spheres



(a) (b)

Fig. 4 Constant-valued sphere segmentation for $\sigma_b = 0.5$, $\sigma_n = 0.5$. (a) Slice of the test image. (b) Segmentation for the same slice.

precision of the segmentation, but provides a solution for eventual incomplete segmentations that can occur for some of the noisier datasets, where spurious maxima can prematurely stop the algorithm. For color illustrations please refer to the electronic version of this document.

In the following section, experimental validation is provided for both single and multiple seeding schemes.

3 Results

We have implemented our algorithm using open source medical image processing libraries, more precisely the Insight Toolkit [20] for algorithm development, and the command line executable module infrastructure provided by 3DSlicer for fast prototyping, calibration, evaluation, and manual segmentation [21]. The algorithm that we will validate, and that was finally implemented in the virtual reality platform that motivated its development, uses the following parameter values: $R = 2, K = 12, \Gamma = 1, k_0 = 1, \Delta k = 0.1, L = 3$ and B is the 3×3 binary cross structuring element.

To validate our algorithm we have proposed several experiments, based on synthetic and real CT datasets. In their inspiring work, Udupa et al. [22] deployed a methodology for the validation of medical volume segmentation methods. For a segmentation algorithm to be proven useful it

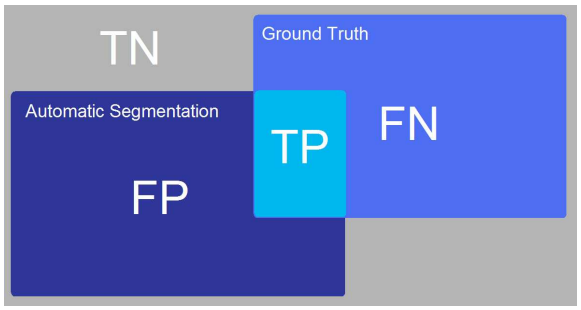


Fig. 5 Sets for accuracy assessment.

has to demonstrate its accuracy (quality of results), precision (parameter independence) and efficiency (human and computational times). *Accuracy* is evaluated in terms of several classical measures, i.e. sensitivity (Sn), specificity (Sp), Dice's coefficient (D) [23], and Jaccard's index (J) [24]. All four measures are dependent on the concepts of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) for a given segmentation and ground truth. To produce such a ground truth a reference delineation of the object must be available. This delineation can be accomplished manually by a trained medical expert. Accuracy can also be assessed by subjective evaluation by an accredited observer in cases where manual delineation is impractical.

Sensitivity and specificity are widely-known, classical assessment rates. They compute the proportion between detected truth and full truth. Furthermore, Dice's coefficient and Jaccard's index are related measures that provide a combination of sensitivity and specificity characteristics.

$$\mathrm{Sn} = \frac{|\mathrm{TP}|}{|\mathrm{TP}| + |\mathrm{FN}|} \quad , \tag{9}$$

$$\mathrm{Sp} = \frac{|\mathrm{TN}|}{|\mathrm{TN}| + |\mathrm{FP}|} \quad , \tag{10}$$

$$\mathrm{D} = \frac{2|\mathrm{TP}|}{|\mathrm{FN}| + |\mathrm{FP}| + 2|\mathrm{TP}|} \quad , \tag{11}$$

$$\mathrm{J} = \frac{|\mathrm{TP}|}{|\mathrm{FN}| + |\mathrm{FP}| + |\mathrm{TP}|} \quad . \tag{12}$$

Precision is also evaluated according to the Jaccard's index applied on two different outputs rather than on a segmentation and a ground truth set. Precision can be computed in terms of inter/intra-operator variability and inter-scanner variability. Since our method relies on only a seed selection procedure, and is supposed to work for a variety of acquisition devices, we decided to compute precision for inter-seed variability. This variability is accounted for by comparing the resulting segmentation from seeds placed randomly inside the tissue of interest. Several seeds can be used so that J is computed for all possible combinations of outputs, and then averaged to obtain a more representative measure.

Efficiency relates to the segmentation performance time, human as well as computational time. According to Udupa

et al. [22], it is extremely hard to provide an efficiency measure that proves useful when comparing different methods. One possible way of summing up all efficiency factors is related to the economic cost of usage for an algorithm. Human time is much more expensive than computer time. In our method, usage is designed to be extremely simple, and reduces to placing a handful of seeds on the desired object. For that reason we have computed here only computational times for the execution of our algorithm. The total usage time comprising loading of dataset, seed placement and algorithm execution, ranges from user to user, but seldom surpasses twice the computational time.

All computational times were recorded in a Windows PC with 2 GHz. Intel Core 2 Duo and 2 GB RAM, although our implementation is not multi-threaded. The average dataset size is $512\times512\times350$.

3.1 Constant-valued spheres

To validate the claim that our contrast-based assessment function accurately detects homogeneous-intensity regions, we have created synthetic volume images and corrupted them with some typical CT acquisition artifacts. We have created a 3D volume composed by two ideal tissues, with constant intensity. The first tissue is shaped as a ball centered in the image domain, with radius $r = 20$ voxels. The second tissue is the background which extends to a final image domain with size $512 \times 512 \times 100$ voxels. We argue that this simple phantom is enough to test the effect of artifacts on boundary detection as long as connectivity is preserved. Our method depends only on connectivity and intensity shifts, so the precise shape of the tissue is irrelevant as long as it is connected.

For simulating CT acquisition, we proceeded by blurring the image (accounting for partial volume effect) and adding Gaussian noise. Blurring was performed using a Gaussian smoothing kernel with width σ_{b}. Zero-mean Gaussian noise was generated with standard deviations σ_{n}. We produced segmentations for all combinations for $\sigma_{\mathrm{b}} \in [0.1, 0.9]$ in increments of 0.2, and $\sigma_{\mathrm{n}} \in [0.1L, 0.7L]$ (with L the absolute intensity difference between the two tissues, $L = 1$ in our experiment) in increments of 0.2.

From all the segmentations we computed the aforementioned measures and show them in Table 1. For a representative case ($\sigma_{\mathrm{b}} = 0.5$, $\sigma_{\mathrm{n}} = 0.5$), we have computed precision measure J for all possible combinations of three different seeds. The average similarity according to J is presented in Table 2. For efficiency evaluation, computational time t_{c} has been computed and averaged across all images. We also display average values for the four measures in Table 2. For a representative case ($\sigma_{\mathrm{b}} = 0.5$, $\sigma_{\mathrm{n}} = 0.5$), we have computed similarity Jaccard's index for all possible combinations of three different seeds, which we present in Table 2. For efficiency evaluation, computational time has been recorded and averaged across all images. Notice in Table 2 how the experimental values for Sn and Sp are close to one, indicating very high segmentation fidelity. Precision is close to 1, indicating weak dependence on seed placement. Times show great variance due to the iterative nature of the algorithm.

Table 1 Computed measures for constant-valued spheres

σ_{b}-σ_{n}	Sensitivity	Specificity	Dice	Jaccard
0.1-0.1	1.0000	1.0000	0.9900	0.9802
0.1-0.3	0.9998	0.9998	0.9565	0.9165
0.1-0.5	0.9961	0.9997	0.9159	0.8448
0.1-0.7	0.9265	0.9999	0.9257	0.8618
0.3-0.1	1.0000	0.9998	0.9365	0.8805
0.3-0.3	0.9999	0.9998	0.9345	0.8770
0.3-0.5	0.9943	0.9997	0.9293	0.8680
0.3-0.7	0.9264	0.9998	0.9182	0.8488
0.5-0.1	1.0000	0.9997	0.9276	0.8650
0.5-0.3	0.9996	0.9998	0.9434	0.8928
0.5-0.5	0.9888	0.9998	0.9445	0.8949
0.5-0.7	0.9326	0.9998	0.9161	0.8452
0.7-0.1	1.0000	0.9998	0.9329	0.8743
0.7-0.3	1.0000	0.9995	0.8679	0.7666
0.7-0.5	0.9905	0.9998	0.9313	0.8714
0.7-0.7	0.9340	0.9998	0.9092	0.8335
0.9-0.1	1.0000	0.9997	0.9320	0.8727
0.9-0.3	1.0000	0.9996	0.8847	0.7933
0.9-0.5	0.9916	0.9997	0.9085	0.8323
0.9-0.7	0.9351	0.9997	0.8955	0.8108

Table 2 Average accuracy, precision and efficiency for constant-valued spheres

Avg. Sensitivity	0.9808 ± 0.0298
Avg. Specificity	0.9998 ± 0.0001
Avg. Dice	0.9250 ± 0.0260
Avg. Jaccard	0.8615 ± 0.0452
Precision	0.9250 ± 0.0413
Efficiency (*s.*)	72.6 ± 40.8

For illustrative purposes we include in Fig. 3 a graphical representation of the evolution of the assessment function as the algorithm iterates. See how the fall after the peak decreases as noise and blurring increase.

Also, in Fig. 4 we show a slice of the generated volume for the case 0.5-0.5 and the segmented region marked in blue over the image. Notice the good results facing noise with standard deviation as high as half the intensity difference between the two tissues.

3.2 Continuous-valued spheres

For this experiment we produced a similar synthetic image, only now the intensity inside the ball smoothly varies from 0 to 1 proportionally to the Euclidean distance to the center of the ball. The intensity value for the background was set to 1. In this scheme, no clear boundary is available, the reason is that we wanted to prove that our method does not

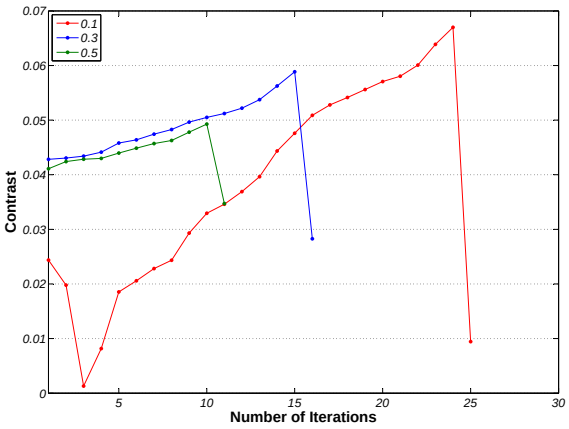
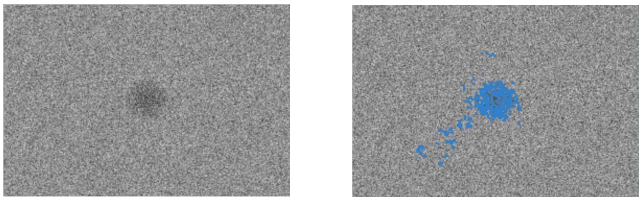


Fig. 6 Contrast evolution for all continuous-valued spheres



(a) (b)

Fig. 7 Continuous-valued sphere segmentation for $\sigma_n = 0.5$. (a) Slice of the test image. (b) Segmentation for the same slice.

Table 3 Computed measures for continuous-valued spheres

σ_n	Sensitivity	Specificity	Dice	Jaccard
0.1	0.9375	1.0000	0.9617	0.9262
0.3	0.8543	0.9998	0.8791	0.7843
0.5	0.7491	0.9990	0.6813	0.5166

Table 4 Average accuracy, precision and efficiency for continuous-valued spheres

Avg. Sensitivity	0.8470 ± 0.0944
Avg. Specificity	0.9996 ± 0.0005
Avg. Dice	0.8407 ± 0.1441
Avg. Jaccard	0.7424 ± 0.2080
Precision	0.9300 ± 0.0176
Efficiency ($s.$)	257 ± 12.9

require abrupt intensity changes, for boundaries to be detected. We corrupted the image only with Gaussian noise of standard deviation $\sigma_n = 0.1, 0.3$ and 0.5. We computed the same accuracy and efficiency measures, as well as precision for the case $\sigma_n = 0.3$. Due to the nature of the values in the regions, seeds must be placed close to the center in order for the condition (which is inherent to any region growing approach) of greater variance across than inside regions, to be met [11]. According to Tables 3 and 4, the results for this extremely subtle boundary are still acceptable. Sn and Sp stay somewhat close to 1, and precision is close enough to 1 as to support the claim of low seed location dependence. Time has increased due to the greater variance of the segmented tissue, which forcer the algorithm to perform more iterations, so that greater frequency maxima can be detected.

For Fig. 6 we can observe the same effect as in Fig. 3. Notice the increased roughness in these curves as compared to those in Fig. 3, especially in the earlier iterations. This is due to a greater influence of noise on consecutive iterations of the region growing sequence, due to the non-constant intensity of the tissue. This effect has dramatic consequences on the contrast evolution for the case $\sigma_n = 0.1$, where the low value of noise together with the smoothness of the variation produces, after normalization, significant instability for early iterations. Notice that this a limit case, as our tissues of interest display almost constant intensity (apart from noise), rather than linearly varying intensity.

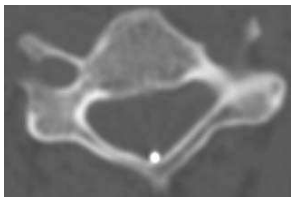
We present in Fig. 7 again a slice of the generated volume for the case 0.5 and its segmentation output. The quality of the segmentation is pretty good even for extremely dim boundaries.

3.3 Ground truth CT images

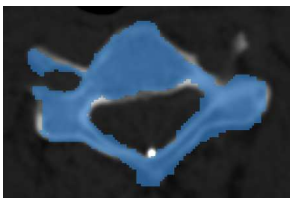
For this experiment we have produced automatic segmentations using a set of real CT images for one of the described tissues (muscle, fat and bone). The testing set proceeds from three different scanners present in our clinical setting (a LightSpeed16 by General Electrics Medical Systems, Milwaukee, Wisconsin USA; an Aquilion by Toshiba Medical Systems Company, Tokyo, Japan; and an AVPS by Philips Medical Systems, Best, the Netherlands). We found almost no coincidences between datasets regarding reconstruction kernel, slice thickness, and intensity windowing. Some of them presented contrast agents in some way.

As explained above, validation of real CT images requires manual segmentations provided by a clinical expert. In our application context this manual segmentation process can be extremely time-consuming, or even intractable, due to considerable tissue delocalization, e.g. manual segmentation of muscular tissue implies manually avoiding all blood vessels and fat traces, for up to 500 slices. For this reason we have computed our accuracy and precision metrics from an average of 20 evenly-spaced slices, in 10 cases.

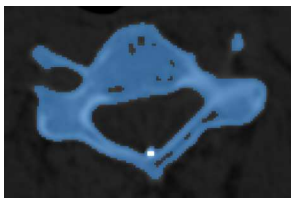
Moreover, the manual accuracy of the segmentation is biased by human high-level knowledge, presenting considerable inter-subject variability, and we (the authors and the clinical practitioners) have observed that in many cases poorer results in accuracy are related to incomplete manual segmentations, rather than incomplete automatic segmentations. Let us say then, that the quality of the segmentation is *at least* as good as the presented results. See in Fig. 8 a cropped comparison sample showing discrepancies between our segmentation output and the manual delineation provided. For



(a)



(b)



(c)

Fig. 8 Close-up comparison of manual segmentation and automatic segmentation. (a) Original caption. (b) Manual segmentation. (c) Automatic segmentation.

Table 5 Accuracy metrics for all CT datasets

	Sensitivity	Specificity	Dice	Jaccard
Case 1	0.884	0.998	0.903	0.824
Case 2	0.967	0.996	0.949	0.902
Case 3	0.783	0.999	0.853	0.743
Case 4	0.828	0.951	0.807	0.676
Case 5	0.650	1.000	0.787	0.649
Case 6	0.869	1.000	0.928	0.866
Case 7	0.744	0.999	0.731	0.576
Case 8	0.988	1.000	0.929	0.867
Case 9	0.702	0.997	0.777	0.635
Case 10	0.805	0.987	0.788	0.651

the results presented in Tables 5 and 6, all segmentations have been used for accuracy and efficiency assessment, and one particular segmentation and several seeds for precision. While specificity stays very high, ensuring self-contained segmented regions, sensitivity is still reasonably close to 1. This prevalence of specificity satisfies our domain of application, since surgical planning depends strongly on boundaries, and incompleteness of our segmentations usually remain in the inner part of the regions. Due to the connected nature of our output regions, over-segmentation would necessarily entail boundary alteration. Computational time ranges between 2 or 3 minutes, which implies a great reduction as compared to previous trial-and-error parameter tuning, according to non-technical users' opinions.

3.4 Ground truth CT samples

In an attempt to circumvent the inherent difficulty of providing manual segmentations for our very extensive regions of interest, we have developed a more tractable experiment. In this evaluation strategy we made use of partial volume samples from a new set of 10 segmented CT scans. As before, we have performed segmentation for bone, fat and muscle tissue indifferently on these 10 sets, but instead of comparing with manual delineations of full slices, we have extracted small regions of interest with size $20(\text{axial}) \times 40 \times 40$. The reason is that for such smaller regions, manual delineation can be produced by our expert with much greater precision. For a comparison between manual/automatic delineations in small samples see Fig.9.

For quantitative evaluation we have computed the values of sensitivity, specificity, Jaccard's index and Dice's coefficient as described at the beginning of this section. See the computed measures in table 7 for each of the concerned datasets. We display in table 8 the average results for

Table 6 Average accuracy, precision and efficiency for CT datasets

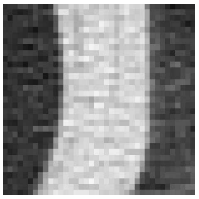
Avg. Sensitivity	0.822 ± 0.109
Avg. Specificity	0.993 ± 0.015
Avg. Dice	0.845 ± 0.077
Avg. Jaccard	0.739 ± 0.117
Precision	0.733 ± 0.171
Efficiency (*s.*)	156.0 ± 36.2

Table 7 Accuracy metrics for all CT samples

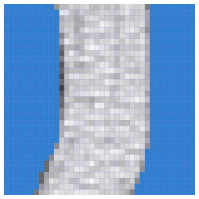
	Sensitivity	Specificity	Dice	Jaccard
Sample 1	0.891	0.986	0.936	0.881
Sample 2	0.928	0.985	0.955	0.914
Sample 3	0.982	0.927	0.845	0.732
Sample 4	0.996	0.867	0.960	0.923
Sample 5	0.985	0.974	0.907	0.830
Sample 6	0.986	0.983	0.961	0.924
Sample 7	0.964	0.947	0.973	0.948
Sample 8	0.988	0.958	0.941	0.888
Sample 9	0.943	0.958	0.933	0.875
Sample 10	0.948	0.983	0.970	0.942

Table 8 Average accuracy, precision and efficiency for CT samples

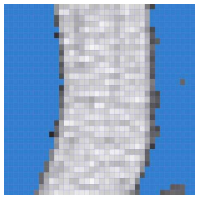
Avg. Sensitivity	0.961 ± 0.0334
Avg. Specificity	0.957 ± 0.0370
Avg. Dice	0.938 ± 0.0382
Avg. Jaccard	0.886 ± 0.0646



(a)



(b)



(c)

Fig. 9 Comparison of manual segmentation and automatic segmentation for the axial view of a sample. (a) Original sample. (b) Manually delineated sample. (c) Segmentation output sample.

this validation experiment. Notice how the use of manageable regions of interest results in greater performance of the algorithm, since the obtention of the ground truth is now trivial.

3.5 Subjective evaluation of real CT cases (Multi-seed scheme)

For complementary accuracy assessment we provide also a subjective evaluation performed by a clinical expert in surgery (a user of the surgical planning platform under development) on 35 datasets. These images proceed from the scanners enumerated in Subsect. 3.3, and also present differences in most imaging acquisition parameters previously described.

Thirty-five segmentations of muscle, fat and bone were performed randomly on the datasets. In this case the multiple seeding scheme was used in a maximum of two seed placement iterations, as described in Subsect. 2.2. The computational time referred below includes both iterations when applicable.

We required our expert to evaluate the percentage of slices in the output segmentation, that required further modifications in order to serve the purpose of surgical planning for the body region comprised in the dataset. We gave him instructions to classify segmentation as Excellent if 100-75% of the slices were good enough, Good for 75-50%, Acceptable for 50-25%, and Useless for 25-0%. We also asked him to provide detailed comments on any non-Excellent segmentations, for future improvements of the algorithm.

See in Fig. 12 some examples of the obtained outputs. First column is for fat segmentation, second for muscle, and third for bone. Three-dimensional surface models obtained by means of the marching cubes algorithm [25], as well as different views are provided. Notice that neither the varying resolution nor the movement artifacts prevented the overall correctness of results.

Notice in Table 9 the very good subjective assessment obtained for most test datasets. Just one segmentation scored below Good for any of the tissues involved, and that should be considered in the context of varying scanners and imaging conditions. Besides these results were obtained in reasonable computational times and with minimal user intervention (two multiple-seeding iterations at most).

Regarding the few cases with non-Excellent qualification the main comments made by the observer concerned issues like poor image quality in the test dataset for muscle and fat segmentation, or abnormal tissue intensity distribution for bone segmentation (Osteoporosis condition).

The interest of this kind of subjective assessment is based in the definition of the evaluation. In this case, the evaluation is made taking in consideration the application for which this technique was developed.

3.6 Performance of our technique once integrated in virtual reality platform VirSSPA

According to our field inquiries, our segmentation technique (that we integrated with the virtual reality surgical planning viewer VirSSPA) has produced dramatic time reduction in surgical planning. Before our algorithm was available for platform VirSSPA, segmentation was usually performed using either rough manual delineation, or iterative threshold selection techniques highly dependent on user interaction. Now, the developed technique is able to produce very precise segmentations just from a couple seeds, which can be introduced quickly and trivially by any operator, thus reducing the cost and burden of the surgical planning cycle.

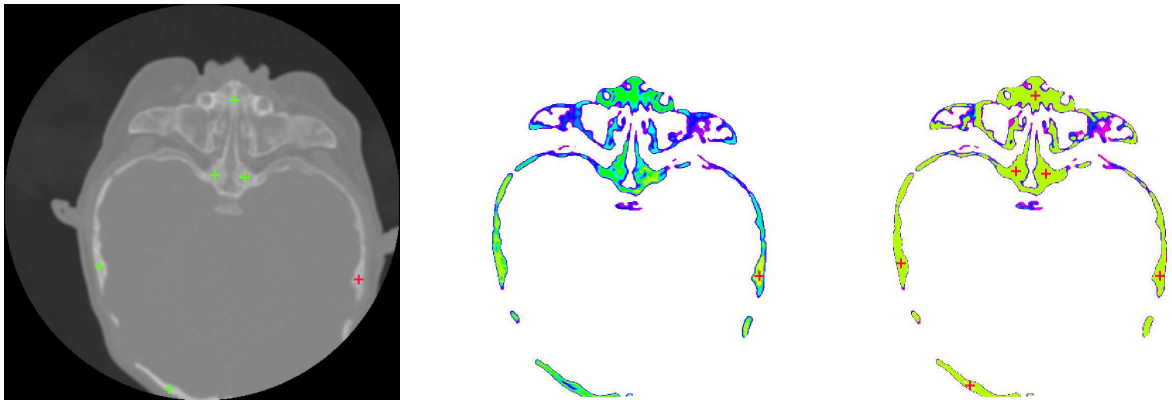
Once the whole tissue of interest has been obtained, a surface rendering is produced through a simple marching cubes technique [25], and then the surgeon in charge of the surgical planning proceeds by 3D manipulation, excision and replacement of different parts of the rendered model. The accuracy of virtual-reality based surgical planning procedures using VirSSPA, exemplified by deep inferior epigastric artery perforator (DIEP) flap surgery, has extensively been demonstrated in [26–28].

Recently, VirSSPA-based surgical planning has been a key factor in successfully performing the first face transplant ever performed in Andalusia (second in Spain, and ninth in the world). The surgical process was simulated using surface models of the skin, muscle, fat, bone and vessels, most of which were segmented using the presented technique.

4 Concluding Remarks, Limitations and Future Work

Our work is basically an attempt to generalize a classical image processing methodology, so it deals properly with real imaging-condition variations. Few efforts have been pursued to achieve this kind of robustness, as most methods are either manually tuned or based on prior unrealistic assumptions that only work for a given set of imaging conditions. Considering the numeric results for the experimental validation, we consider our approach successful in solving the particular needs for which it was conceived. The algorithm has been integrated in the reference platform and prevents from previous trial-and-error segmentation, which was very time-consuming according to its users. Its greatest advantage is thus the absence of tuning parameters and ability to produce nice results for a wide set of acquisition devices, with minimal user intervention, in reasonable time frames. This is founded in the use of our adaptively-sampled contrast measure, and in the normalizing strategy that allows for generalization of the contrast sampling rate.

The most significant limitation to this region growing approach has to do with the fact that connectivity is sometimes too weak as a requirement for some applications (tu-



(a) (b) (c)

Fig. 10 Region growing sequence comparison for head CT bone segmentation. (a) Single and multiple seeding. Red cross-hairs is the single seed, green cross-hairs add to the red one for the multi-seed segmentation. (b) Region growing sequence for single seeding. (c) Region growing sequence for multiple seeding. (Colder colors represent higher iteration number outputs. For colors please refer to the electronic version.)

Table 9 Subjective assessment for segmentation of real CT images

	Excellent (100-75%)	Good (75-50%)	Acceptable (50-25%)	Useless (25-0%)	t_c(s)
Bone Tissue	76.5%	23.5%	0%	0%	52.3±18.3
Fat Tissue	85.8%	7.1%	7.1%	0%	117.5±27.0
Muscle Tissue	62.6%	37.4%	0%	0%	141.6±48.8

mor segmentation, organ segmentation...), as the region of interest may be connected to other similar-intensity regions. For this reason, future improvement for the technique could consist of incorporating some morphological limitations to the growth of the region in order to impose some degree of *stiffness*, and to avoid flooding towards adjacent, slightly connected regions. This improvement could make the algorithm useful in greater variety of surgical planning scenarios.

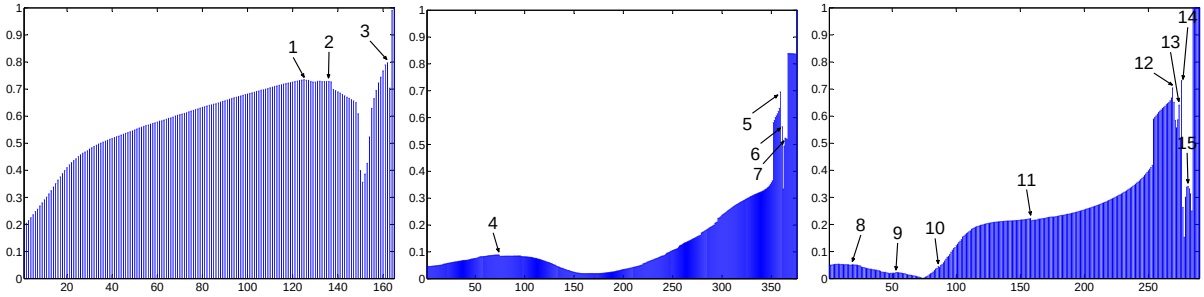
Besides this regularization, another field of exploration consists of transposing our assessment measure into other segmentation paradigms, like watersheds or deformable surfaces, allowing for automatization of these other schemes, which rely on manual tuning and intervention for the obtention of adequate results.

Acknowledgements This work was supported by "Fundación Reina Mercedes" from "Hospital Universitario Virgen del Rocío" (Seville, Spain), and "Consejería de Salud de la Junta de Andalucía". Carlos S. Mendoza was supported by a doctoral scholarship financed by University of Seville, Spain.

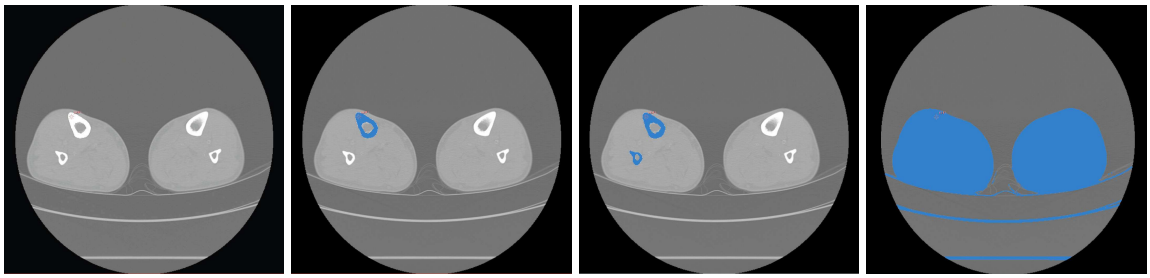
References

1. B. Reitinger, A. Bornik, R. Beichel, and D. Schmalstieg, "Liver surgery planning using virtual reality," *IEEE Comput. Graph. Appl.*, vol. 26, no. 6, pp. 36–47, 2006.
2. C. Suárez, B. Acha, C. Serrano, C. Parra, and T. Gómez, "VirSSPA- a virtual reality tool for surgical planning workflow," *International Journal of Computer Assisted Radiology and Surgery*, pp. 1–7, 2009, article in Press.
3. S. W. Zucker, "Region growing: Childhood and adolescence," *Computer Graphics and Image Processing*, vol. 5, no. 3, pp. 382–399, 1976.
4. G. J. Sivewright and P. J. Elliott, "Interactive region and volume growing for segmenting volumes in MR and CT images," *Medical Informatics*, vol. 19, no. 1, pp. 71–80, 1994.
5. H. Sekiguchi, K. Sano, and T. Yokoyama, "Interactive 3-dimensional segmentation method based on region growing method," *Systems and Computers in Japan*, vol. 25, no. 1, pp. 88–97, 1994.
6. X. Zhou, N. Kamiya, T. Kara, H. Fujita, R. Yokoyama, T. Kiryu, and H. Hoshi, "Automated recognition of human strucure from torso CT images," in *Proceedings of the Fourth IASTED International Conference on Visualization, Imaging, and Image Processing*, 2004, pp. 584–589.
7. T. Y. Law and P. A. Heng, "Automated extraction of bronchus from 3D CT images of lung based on genetic algorithm and 3D region growing," in *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 3979, 2000, pp. I/–.
8. R. Adams and L. Bischof, "Seeded region growing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 6, pp. 641–647, 1994.
9. J. Dehmeshki, H. Amin, M. Valdivieso, and X. Ye, "Segmentation of pulmonary nodules in thoracic CT scans: A region growing approach," *IEEE Trans. Med. Imag.*, vol. 27, no. 4, pp. 467–480, 2008.
10. S. A. Hojjatoleslami and J. Kittler, "Region growing: A new approach," *IEEE Trans. Image Process.*, vol. 7, no. 7, pp. 1079–1084, 1998.
11. R. M. Haralick and L. G. Shapiro, "Image segmentation techniques," *Computer Vision, Graphics, & Image Processing*, vol. 29, no. 1, pp. 100–132, 1985.

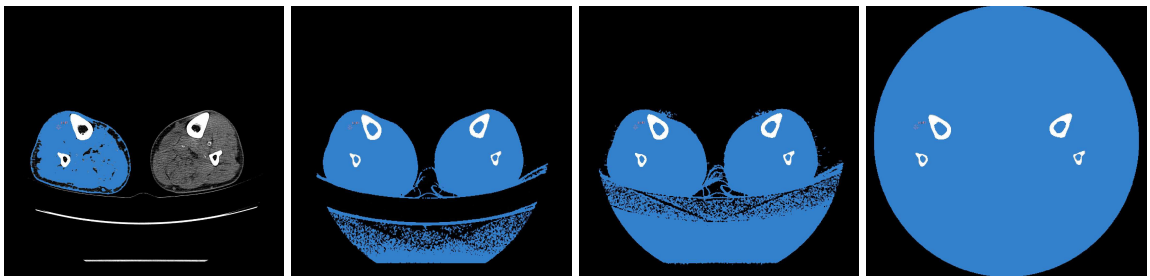
12. C. Revol-Muller, F. Peyrin, Y. Carrillon, and C. Odet, "Automated 3D region growing algorithm based on an assessment function," *Pattern Recognition Letters*, vol. 23, no. 1-3, pp. 137–150, 2002.
13. J. K. Udupa and S. Samarasekera, "Fuzzy connectedness and object definition: Theory, algorithms, and applications in image segmentation," *Graphical Models and Image Processing*, vol. 58, no. 3, pp. 246–261, 1996.
14. J. K. Udupa, "Multiple sclerosis lesion quantification using fuzzy-connectedness principles," *IEEE Trans. Med. Imag.*, vol. 16, no. 5, pp. 598–609, 1997.
15. P. K. Saha, J. K. Udupa, E. F. Conant, D. P. Chakraborty, and D. Sullivan, "Breast tissue density quantification via digitized mammograms," *IEEE Trans. Med. Imag.*, vol. 20, no. 8, pp. 792–803, 2001.
16. S. Luo, X. Li, and G. Zhou, "A simplified fuzzy connectedness method used for segmentation of vessel images," in *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, vol. 1, 2003, pp. 751–753.
17. J. Tschirren, E. A. Huffman, G. McLennan, and M. Sonka, "Intrathoracic airway trees: Segmentation and airway morphology analysis from low-dose CT scans," *IEEE Trans. Med. Imag.*, vol. 24, no. 12, pp. 1529–1539, 2005.
18. J. K. Udupa, P. K. Saha, and R. A. Lotufo, "Relative fuzzy connectedness and object definition: Theory, algorithms, and applications in image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 11, pp. 1485–1500, 2002.
19. W. Jian, Y. Feng, J. L. Ma, X. P. Sun, X. Jing, and Z. M. Cui, "The segmentation and visualization of human organs based on adaptive region growing method," in *Proceedings - 8th IEEE International Conference on Computer and Information Technology Workshops, CIT Workshops 2008*, 2008, pp. 439–443.
20. T. S. Yoo, M. J. Ackerman, W. E. Lorensen, W. Schroeder, V. Chalana, S. Aylward, D. Metaxas, and R. Whitaker, "Engineering and algorithm design for an image processing API: a technical report on ITK–the Insight Toolkit." *Studies in health technology and informatics*, vol. 85, pp. 586–592, 2002.
21. S. Pieper, B. Lorensen, W. Schroeder, and R. Kikinis, "The NA-MIC Kit: ITK, VTK, pipelines, grids and 3D Slicer as an open platform for the medical image computing community," in *2006 3rd IEEE International Symposium on Biomedical Imaging: From Nano to Macro - Proceedings*, vol. 2006, 2006, pp. 698–701.
22. J. K. Udupa, V. R. LeBlanc, Y. Zhuge, C. Imielinska, H. Schmidt, L. M. Currie, B. E. Hirsch, and J. Woodburn, "A framework for evaluating image segmentation algorithms," *Computerized Medical Imaging and Graphics*, vol. 30, no. 2, pp. 75–87, 2006.
23. L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.
24. P. Jaccard, "Étude comparative de la distribution florale dans une portion des Alpes et des Jura," *Bulletin del la Société Vaudoise des Sciences Naturelles*, vol. 37, pp. 547–579, 1901.
25. W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3D surface construction algorithm." *Computer Graphics (ACM)*, vol. 21, no. 4, pp. 163–169, 1987.
26. P. Gacto-Sánchez, D. Sicilia-Castro, T. Gómez-Cía, A. Lagares, T. Collell, C. Suárez, C. Parra, P. Infante-Cossío, and J. M. De La Higuera, "Use of a three-dimensional virtual reality model for preoperative imaging in DIEP flap breast reconstruction," *Journal of Surgical Research*, Article in Press.
27. P. Gacto-Sánchez, D. Sicilia-Castro, T. Gómez-Cía, A. Lagares, T. Collell, C. Suárez, C. Parra, S. Leal, P. Infante-Cossío, and J. M. De La Higuera, "Computed tomographic angiography with VirSSPA three-dimensional software for perforator navigation improves perioperative outcomes in DIEP flap breast reconstruction," *Plastic and Reconstructive Surgery*, vol. 125, no. 1, pp. 24–31, 2010.
28. T. Gómez-Cía, P. Gacto-Sánchez, D. Sicilia, C. Suárez, B. Acha, C. Serrano, C. Parra, and J. Higuera, "The virtual reality tool VirSSPA in planning DIEP microsurgical breast reconstruction," *International Journal of Computer Assisted Radiology and Surgery*, vol. 4, no. 4, pp. 375–382, 2009.



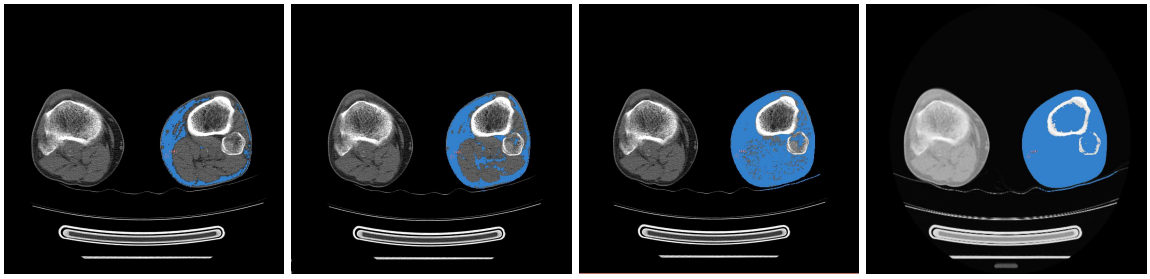
(a) Bone Segmentation Contrast (b) Muscle Segmentation Contrast (c) Fat Segmentation Contrast



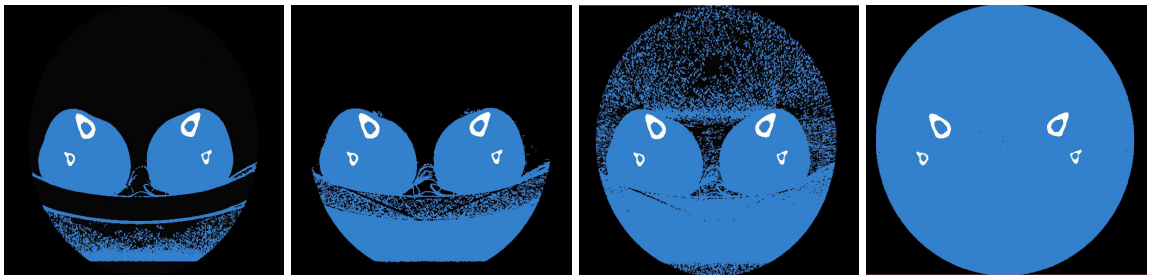
(d) Original dataset (e) Peak #1 (f) Peak #2 (g) Peak #3



(h) Peak #4 (i) Peak #5 (j) Peak #6 (k) Peak #7

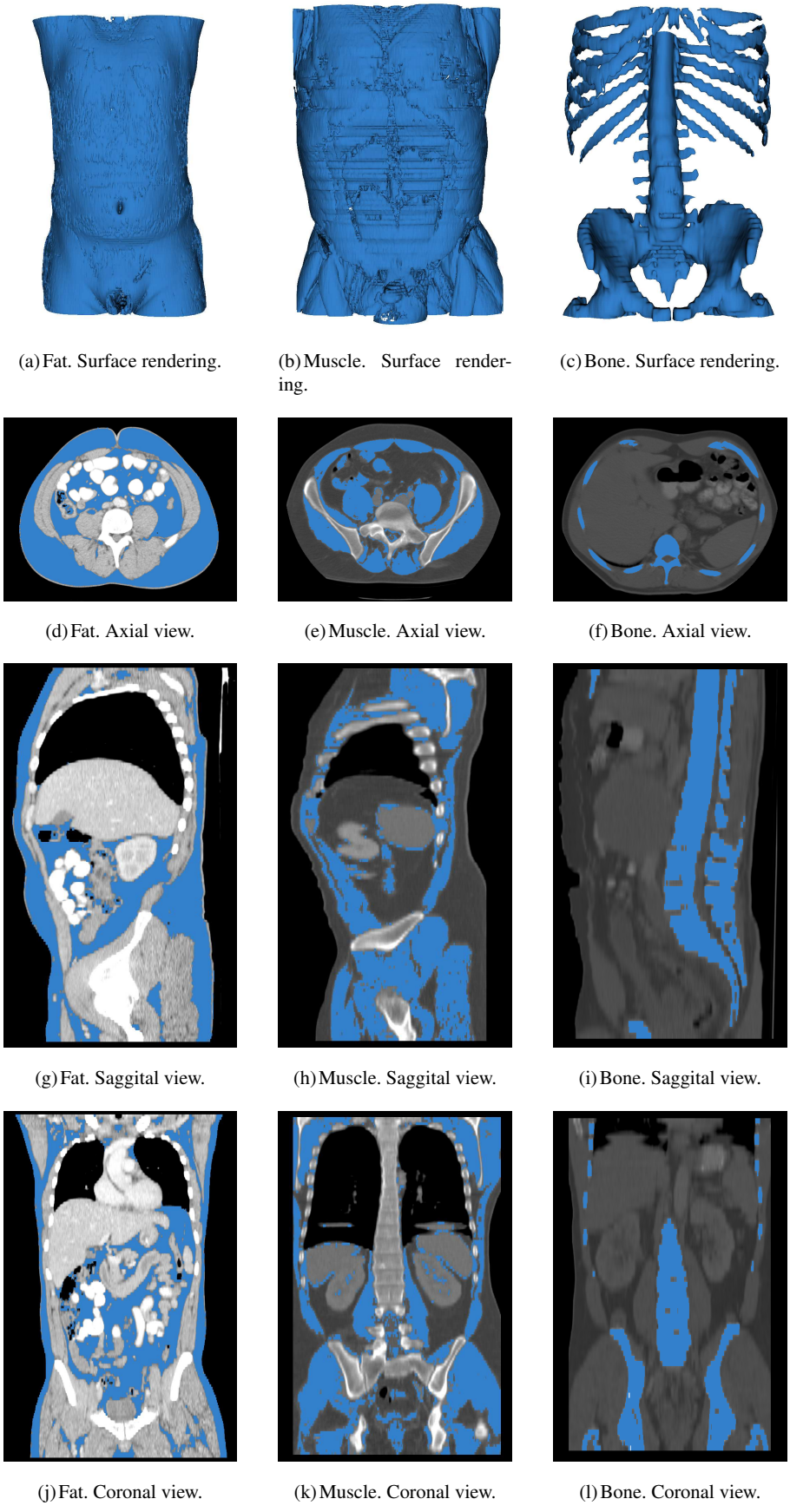


(l) Peak #8 (m) Peak #9 (n) Peak #10 (o) Peak #11



(p) Peak #12 (q) Peak #13 (r) Peak #14 (s) Peak #15

Fig. 11 Fat, muscle and bone contrast maxima. The areas in blue indicate segmentations for the arrow-marked maxima. Contrast curves are shown across growing number of iterations



(a) Fat. Surface rendering. (b) Muscle. Surface rendering. (c) Bone. Surface rendering.

(d) Fat. Axial view. (e) Muscle. Axial view. (f) Bone. Axial view.

(g) Fat. Saggital view. (h) Muscle. Saggital view. (i) Bone. Saggital view.

(j) Fat. Coronal view. (k) Muscle. Coronal view. (l) Bone. Coronal view.

Fig. 12 Fat, muscle and bone segmentation examples. The blue region is the segmentation output.