

3DBGrowth: volumetric vertebrae segmentation and reconstruction in magnetic resonance imaging

Jonathan S. Ramos*, Mirela T. Cazzolato*, Bruno S. Faiçal*,

Marcello H. Nogueira-Barbosa†, Caetano Traina Jr.* and Agma J. M. Traina*

*Institute of Mathematics and Computer Science (ICMC), University of São Paulo (USP). São Carlos, SP, BR 13566-590

†Ribeirão Preto Medical School (FMRP), University of São Paulo (USP). Ribeirão Preto, SP, BR 14048-900

Abstract—Segmentation of medical images is critical for making several processes of analysis and classification more reliable. With the growing number of people presenting back pain and related problems, the semi-automatic segmentation and 3D reconstruction of vertebral bodies became even more important to support decision making. A 3D reconstruction allows a fast and objective analysis of each vertebrae condition, which may play a major role in surgical planning and evaluation of suitable treatments. In this paper, we propose 3DBGrowth, which develops a 3D reconstruction over the efficient Balanced Growth method for 2D images. We also take advantage of the slope coefficient from the annotation time to reduce the total number of annotated slices, reducing the time spent on manual annotation. We show experimental results on a representative dataset with 17 MRI exams demonstrating that our approach significantly outperforms the competitors and, on average, only 37% of the total slices with vertebral body content must be annotated without losing performance/accuracy. Compared to the state-of-the-art methods, we have achieved a Dice Score gain of over 5% with comparable processing time. Moreover, 3DBGrowth works well with imprecise seed points, which reduces the time spent on manual annotation by the specialist.

Keywords-3D vertebrae reconstruction; magnetic resonance imaging; Balanced Growth

I. INTRODUCTION

Spinal diseases are increasing worldwide and can cause significant loss of function and compromise quality of life. Surgical spinal treatments have been growing with the aging population, which requires accurate diagnosis to avoid complications [1]. Many spine pathologies can be detected and diagnosed using Magnetic Resonance Imaging (MRI) exams [2], [3]. In a Computer-Aided Diagnosis (CAD) context, the segmentation of each vertebra allows a faster and more objective analysis of the vertebrae condition, aiding in the characterization and quantification of abnormalities [4]. Moreover, an accurate segmentation plays a major role and may assist the medical specialist in surgical planning and evaluation of suitable treatments [5].

The manual segmentation of a vertebral body in a slice-by-slice manner may be time-consuming and prone to errors, due to inter and intra-subject variability. Besides, the subjective judgment that is employed may aggregate even more inaccuracy [6]. Elseways, the knowledge gained over several years of expertise are incorporated. Thus, the semi-

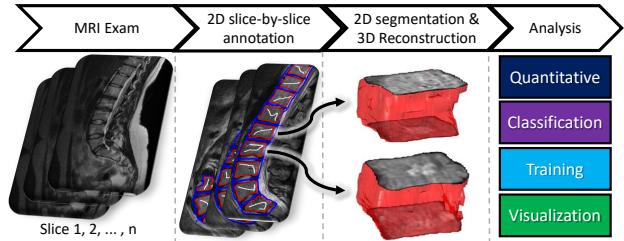


Figure 1: Steps in a semi-automatic segmentation schema.

automatic segmentation assists the specialist, leads to time savings and reduces interpretation errors [7].

The semi-automatic segmentation can be used in several analysis (Figure 1). **Quantitative** measures can be extracted, such as semantic and agnostic features [8], consequently, machine learning techniques can be applied for the **classification** of a given anomaly [9], [10], [11] or for Content-Based Image Retrieval (CBIR) [12], [13]. Interactive segmentation tools can be meaningful during the **training** and education of new radiologists [14]. Students can learn how to correctly segment each vertebra and to detect spine pathologies [15]. This kind of training may avoid potential medical failures, which reduces further complications. In general, the **visualization** of 3D human structures can be used for simulation of medical and surgical procedures [16].

The GrowCut [17] method and its faster version, named as Fast GrowCut [18], which presents slightly lower segmentation accuracy, have been widely used in many medical MRI exams (especially in oncology) [8]. The GrowCut method is based on cellular automata (analogous to a bacteria growth in biology) and works as a region-growing approach with an interactive labeling procedure [17].

Several fully automatic vertebrae segmentation methods have been proposed [19], [20]. However, they take too much processing time, which may not suit clinical practice [21]. More recently, a novel approach called Balanced Growth (BGrowth) [22] has been proposed for the segmentation of crushed vertebral bodies in single slices. Briefly, BGrowth balances the weights along the growing path of a region, so that small intensities transitions are better delineated. The results achieved by BGrowth surpasses all methods from

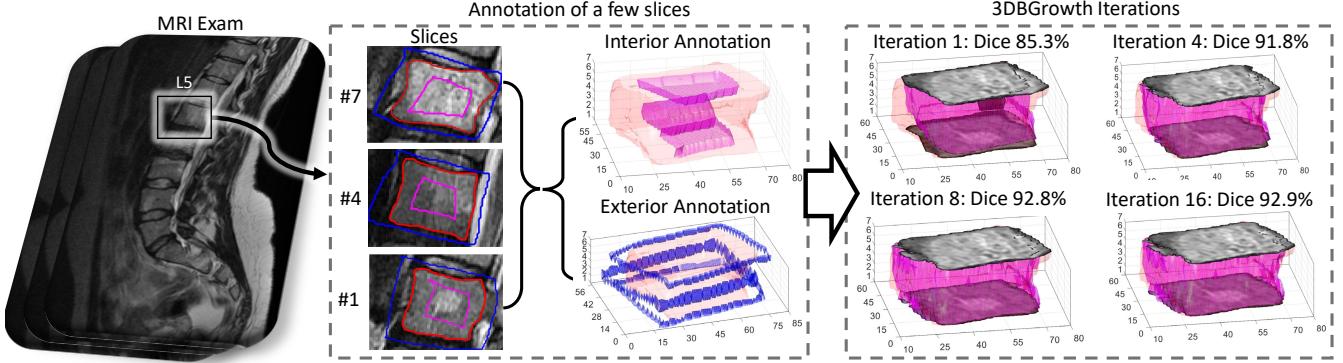


Figure 2: Examples of slices annotation for a single vertebral body (Exam AKA2, L5) and 3DBGrowth's iterations. Ground-truth, interior and exterior annotation in red, magenta and blue, respectively.

the literate, including GrowCut. Moreover, BGrowth is able to achieve promising segmentation results even with very simple/sloppy annotation (seed points).

In this paper, we extrapolate the specialists' annotation up to a fixed limit without losing performance/accuracy, so that the total time spent on manual annotation is reduced. Moreover, we show how to extend BGrowth to deal with the reconstruction of volumetric exams (3D), introducing a novel method called 3DBGrowth. The experimental results show that 3DBGrowth outperforms GrowCut, achieving an average Dice Score of 87% while managing comparable running time. Moreover, the method works well even with rough seed points, which reduces the time spent on manual annotation.

The remainder of the paper is structured as follows. In Section II, we present 3DBGrowth for the segmentation and reconstruction of vertebral bodies in volumetric MRI. In Section III, we detail the experimental design, results and discussion. Finally, Section IV draws the conclusions.

II. 3DBGROWTH: THE PROPOSED METHOD

The usual approach of annotating or stating seeds for segmenting medical images can be cumbersome for large 3D exams. Thus, this work main issue focuses on minimizing the human effort to segment and reconstruct 3D exams built on 2D slices. As illustrated in Figure 2, depending on the MRI exam, not all slices have to be manually annotated by the user to process the 3D reconstruction. If the exams present a small spacing between slices (considering annotations on the sagittal plane), several slices do not need to be annotated, once they are similar. This can be assessed by analyzing the negative slope coefficient [23], which gives the best trade-off between annotation time and performance measures, such as Dice Score Coefficient or Jaccard Coefficient (better explored in the next Section).

The slope between two points, $\{x_1, y_1\}$ and $\{x_2, y_2\}$, can be calculated by $\frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}$, which is the rate of change between the two points. When the slope between two

Input: Image I and labels matrix L .

Output: Segmented binary image $L == 1$.

```

1  $W(L \neq 0) \leftarrow 1.0$  // Initial weights
2 for  $\forall(i, j, z)$  do // For every voxel
3   for  $\forall(i_n, j_n, z_n)$  do // and its Neighbors
4      $s \leftarrow W(i, j, z) \times \left[ 1 - \frac{|I(i, j, z) - I(i_n, j_n, z_n)|}{\max_{\forall i, j, z} I(i, j, z)} \right]$ 
5     if  $s > W(i_n, j_n, z_n)$  then
6        $W(i_n, j_n, z_n) \leftarrow (W(i_n, j_n, z_n) + s) / 2$ 
7        $L(i_n, j_n, z_n) \leftarrow L(i, j, z)$ 

```

Algorithm 1: 3DBGrowth method overview.

values of annotation time gets close to a straight horizontal line, there is no gain in annotation time, in other words, the closer the slope gets to 0, the lower the annotation time gain. Moreover, by using a segmentation approach that does not require detailed interior/exterior annotation, such as BGrowth, the total time spent on annotation is greatly diminished. In addition, BGrowth generally requires a simple rectangle-like annotation for the segmentation of individual vertebral bodies. For example, only 3 out of 7 slices were annotated in Figure 2. In average, each slice took 6.5 seconds for annotation and 3DBGrowth took 0.64 seconds to process all 7 slices with only 16 iterations. Summing up, the whole process took $3 \times 6.5 + 0.65 = 20.65$ seconds and achieved a result close to the Ground-Truth.

In our proposed 3DBGrowth method (Algorithm 1), we initially consider the segmentation of foreground and background in gray-scale images. That is, considering a digital image I and its annotations/labels as a matrix L , both with dimension $M \times N \times Z$, representing the number of rows, columns and slices, respectively. Each entry in L has value -1 (background), 0 (unlabelled) or 1 (foreground).

Initially, each entry in a weight matrix W (with the same dimensions as I and L) is set to 1.0 for seeds points and 0 otherwise (line 1). Then, for every voxel (i, j, z)

and each one of its 26 neighbours (i_n, j_n, z_n) , a strength factor s is calculated (line 4). Here, the absolute intensity difference is normalized by the maximum intensity in the image and shifted by 1. Finally, s is multiplied by the current weight $W(i, j, z)$, which produces values within $[0, 1]$. If the strength s is greater than the neighbour's strength ($W(i_n, j_n, z_n)$, line 5), then the neighbour's strength is averaged with the new strength s (line 6) and its label receives the label of the voxel (i, j, z) (line 7).

This process repeats until the algorithm converges or for a fixed number of iterations defined by the user.

III. EXPERIMENTS, RESULTS AND DISCUSSION

The methods and measures used for comparison, as well as the computational set-up, image dataset and experimental design are described as follows.

Segmentation methods: In order to evaluate the performance of 3DBGrowth (BG) in a 3D scenario, we compared it with GrowCut (GC), which has been widely used for the task of vertebrae segmentation [7]. Since Fast GrowCut is an approximation of the original GrowCut, presenting a lower accuracy [18], we consider only GrowCut on the experiments. Due to the limited number of samples (exams) no deep-learning approach was applied.

Comparison measures: The Jaccard Coefficient (JAC), Dice Score Coefficient (DSC) and Hausdorff's Distance (HD) in voxels [24], [25] were considered. The Jaccard (JAC) calculates the intersection of the manual and semi-automatic segmentation, and divides it by the union of them. This indicates the similarity between the segmentations, in which 0 indicates no similarity and, the closer JAC is to 1, the more alike the segmentations [26]. The Dice (DSC) measures the spatial overlap of several segmentations of the same object, i.e., quantifies the overlap degree between two segmented objects. A DSC close to 0 indicates very low overlap, while a DSC closer to 1 indicates a higher overlap. In contrast, the Hausdorff's Distance (HD) indicates how far away (in voxels) the manual and semi-automatic segmentations are. A HD of 0 indicates comparable segmentations. Table I shows a summary of the segmentation methods and comparison measures used in this work.

Table I: Summary of measures/methods used in this work.

Symbol/Acronym	Description
DSC	Dice Score Coefficient
JAC	Jaccard Coefficient
HD	Hausdorff's distance
GC	GrowCut
BG	3D Balanced Growth

Computational set-up: The experiments were performed on a 2.40GHz Intel(R) Core(TM) i7 CPU and 8GB RAM machine, using Matlab(R) version 2018a. The maximum

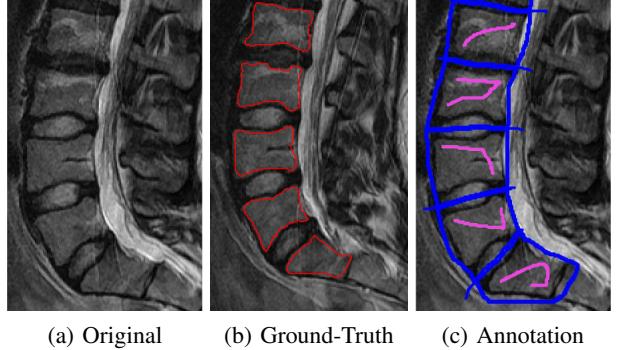


Figure 3: Example of sloppy annotation for a few vertebral bodies in one slice (Aka2, slice 8): ground-truth, interior and exterior annotations in red, magenta and blue, respectively.

number of iterations was set to 50 for GrowCut and 3DBGrowth. No pre or post-processing technique were applied to assure the same conditions for all segmentation methods.

Image Dataset: Due to space limitations, only a meaningful dataset is presented herein, which comprises 17 anonymized MRI exams, ranging from the sacrum (S1) to the mid thoracic (T6-T12) with corresponding manual segmentations. The exams present several health conditions, such as scoliosis, spondylolisthesis and crushed vertebra. The exams have 3.24 ± 0.36 mm of slice thickness and 3.87 ± 0.36 mm of spacing between slices. More information and full access to the dataset is available at [5].

Experimental design: four main parts are analyzed: (A) the performance of each segmentation method is assessed using the whole exam; (B) each segmentation method is tested varying the number of slices annotated for each exam; (C) the vertebral bodies are segmented one-by-one by each method; (D) a statistical test is applied to detect any significant difference between the results of the two methods.

A. Exam segmentation analysis

The initial interior and exterior annotation were performed in a “sloppy” way, i.e., no detailed boundary for accentuated curves were drawn. In general, the annotation looks like a rectangle for the background and a simple line for the foreground (Figure 3). For this experiment, this annotation has been performed on each slice on every exam and, to diminish computational processing, each exam is cropped using the convex hull of the exterior annotation.

Table II shows the average Dice Score (DSC), Jaccard (JAC) and Running Time (RT) in seconds for each one of the 17 exams in the dataset. 3DBGrowth (BG) presented on average 81% DSC and 68% JAC while GrowCut (GC) presented 76% and 61%, respectively. Thus, BG presented higher DSC and JAC percentages than GC for all exams, achieving up to 5% and 7% of DSC and JAC gain, respectively. Moreover, considering DSC and JAC , BG's standard deviation is slightly lower. Analyzing the Running

Time (RT), very often, BG presented a lower average processing time than GC.

Table II: Dice Score (DSC), Jaccard (JAC) and Running Time (RT) in seconds for 3DBGrowth (BG) and GrowCut (GC), considering all slices on each exam (volumetric). The best results are highlighted in bold.

Exam (#slices)	DSC (%)		JAC (%)		RT (s)			
	BG	GC	BG	GC	BG	GC		
DzZ_T1 (12)	85	80	4.86	74	67	7.0	18	21
DzZ_T2 (12)	82	77	4.2	69	63	5.8	27	31
AKa2 (15)	82	77	5.17	69	62	7.1	27	27
AKa3 (15)	78	73	4.78	64	58	6.2	27	29
AKa4 (15)	80	73	7.10	67	58	9.4	26	27
AKs5 (15)	84	78	6.54	73	63	9.2	23	24
AKs6 (15)	84	79	5.44	73	65	7.8	21	24
AKs7 (15)	80	73	7.6	67	57	9.9	21	24
AKs8 (15)	81	78	3.39	68	64	4.7	18	21
S01 (16)	85	82	2.91	74	70	4.3	44	50
S02 (16)	83	78	4.97	70	63	6.9	26	32
F02 (18)	78	74	3.63	64	59	4.7	48	55
S11 (20)	83	80	2.73	71	67	3.9	61	67
F04 (23)	78	75	3.42	64	60	4.5	13	14
AKs3 (25)	80	73	6.42	66	58	8.4	40	37
F03 (25)	80	77	3.57	67	62	4.8	08	09
C002 (31)	71	65	5.85	55	48	6.7	16	14
Mean	81	76	4.9	68	61	6.6	27	30
Std. dev.	3.4	3.9	1.5	4.7	5.0	1.9	13.7	15.2

Considering that the manual annotation of every slice in the exam is too time consuming (for this dataset, on average, 11 minutes/exam), we conducted an experiment to validate the performance of 3DBGrowth and GrowCut when not all slices are annotated, as explored in the next section.

B. Variation on the number of annotated slices

We used the previous experiment's annotations and left a few slices without annotation: we defined a slice distance, which manages the number of non-annotated slices between two annotated slices. For example, a slice distance of 0 implicates no slice is left without annotation. The slice distance started at 0, increased by 1, up to 7.

As the slice distance increases (Figure 4), the average annotation time decreases and the processing time keeps almost steady for both methods. Also, DSC and JAC drops slowly for both methods. However, BG presented best results than GC for both measures. Considering the negative slope coefficient (as discussed in Section II), highlighted over the magenta line, by using a threshold of -1, the best slice distance would be 3, which presents the best trade-off between annotation time and DSC/JAC .

In the next Section, we conduct experiments using annotations for individual vertebral bodies.

C. Individual vertebrae segmentation

To speed-up the annotation process, we have considered a slice distance of three for this experiment. Each vertebral

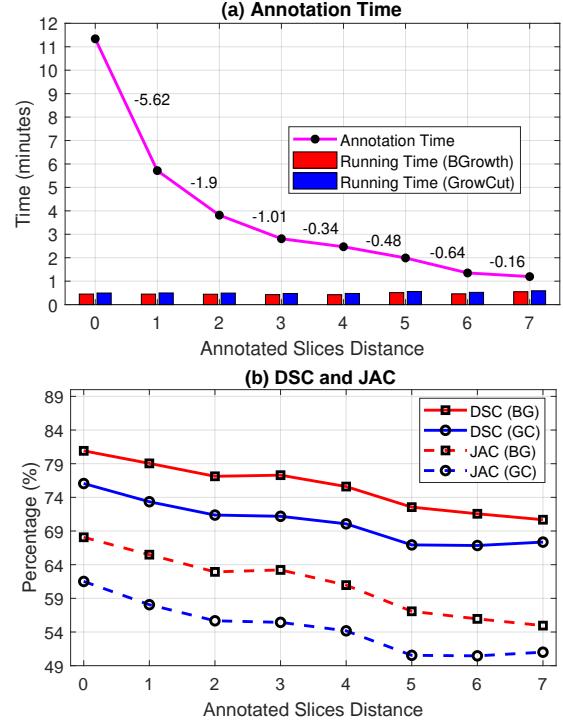


Figure 4: Quality comparison between 3DBGrowth and GrowCut over variations on the number of slices manually annotated: (a) annotation time and running time results; (b) Dice (DSC) and Jaccard (JAC).

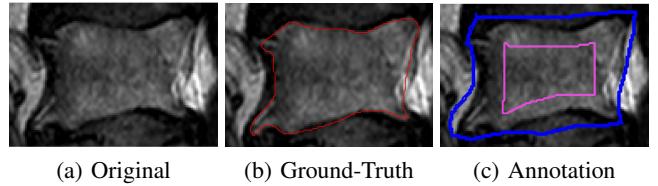


Figure 5: Example of seed points for a single vertebrae (St1, slice 10, L2): ground-truth (GT), interior and exterior annotations in red, magenta and blue, respectively.

body was annotated separately, as exemplified in Figure 5. In general, both the interior and the exterior annotation looks like a rectangle and no detailed borders were drawn.

As reported in Table III, GC and BG presented equal mean Running Time (RT) and BG presented better mean DSC , JAC and HD than GrowCut. Figure 6 depicts the results for a single vertebral body: BG achieved the highest DSC and the lowest HD . GC presented spiculated borders, while BG presented smooth borders (closer to the ground-truth).

Analyzing the average number of annotated slices per vertebra (Table IV), for this dataset, in average, only 37% of the total slices with vertebral content were annotated, which speeded-up the annotation process and took, in average, 36 seconds to annotate each vertebral body.

Table III: Comparison between 3DBGrowth (BG) and GrowCut (GC) for the Dice Score (*DSC*), Jaccard (*JAC*), Hausdorff (*HD*) in voxels and Running Time (*RT*) in seconds. The best values are highlighted in bold.

Vertebrae	<i>DSC</i> (%)		<i>JAC</i> (%)		<i>HD</i> (vox.)		<i>RT</i> (s)		
	BG	GC	BG	GC	BG	GC	BG	GC	
Toracic	T6	88	87	79	78	3.16	4.00	0.15	0.16
	T7	85	83	73	69	78.8	79.1	7.95	6.77
	T8	86	85	76	72	79.3	79.2	7.55	8.37
	T9	81	80	50	52	80.0	79.9	7.83	8.38
	T10	87	85	80	76	26.2	26.9	2.95	2.83
	T11	86	84	77	73	6.09	6.99	0.91	0.88
	T12	89	86	79	76	5.63	6.88	1.30	1.36
Lumbar	L1	89	87	78	76	6.56	7.31	1.57	1.56
	L2	88	86	79	76	6.07	7.75	1.52	1.57
	L3	86	85	75	72	6.40	7.49	1.68	1.83
	L4	88	86	76	74	7.16	7.65	1.77	1.88
	L5	87	85	76	74	7.04	8.42	2.33	2.39
Sacral S1	88 86		79	76	6.18	7.57	1.74	1.88	
Mean	87 85		77	74	7.24	7.72	1.52	1.52	
Std. Dev.	.07 .06		.08	.08	4.85	5.00	1.27	1.27	

Table IV: Comparison of the number of annotated slices, considering a slice distance of three.

Vertebrae	Slices annotated	out of (vertebral content)	<i>ANT</i> (seconds)
Toracic	T6	3.0 ± .00	7.0 ± .00
	T7	3.0 ± .00	7.0 ± .00
	T8	3.0 ± .00	7.0 ± .00
	T9	3.0 ± .00	7.0 ± .00
	T10	2.5 ± .52	6.7 ± 3.2
	T11	3.1 ± .53	8.5 ± 2.6
	T12	3.6 ± .50	9.6 ± 1.9
Lumbar	L1	3.9 ± .78	10.2 ± 2.2
	L2	4.2 ± .75	10.9 ± 2.3
	L3	4.3 ± .86	11.6 ± 2.1
	L4	4.5 ± .62	12.5 ± 2.8
	L5	4.5 ± .72	12.5 ± 3.0
Sacral S1	4.1 ± .70	10.9 ± 3.3	35.8 ± 6.3
Mean	4.1 ± .84	10.9 ± 2.9	35.9 ± 8.8
Annotated		37%	

To further investigate the results presented in Table III, we conducted a statistical test, as detailed next.

D. Statistical testing

Considering that the resulting values of each measure had several similar values, the Kolmogorov-Smirnov [27] test was applied to verify the normality of the data. As the null hypothesis that the data follows a normal distribution was rejected for all measures, the Wilcoxon [28] test was used to analyze if there were significant statistical differences. In this test, the null hypothesis is that data from two dependent samples, e.g. the Dice Score (*DSC*) from 3DBGrowth (BG) and GrowCut (GC), were selected from populations having

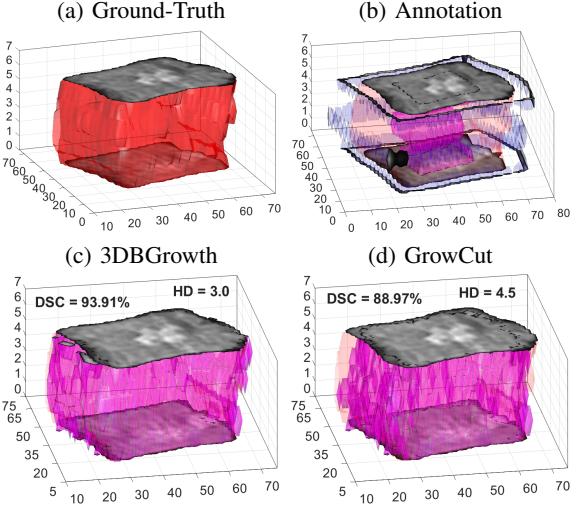


Figure 6: Comparison of results for L2 on exam AKA2: three slices, out of 7, were annotated.

the same distribution, against the opposite alternative.

In the Wilcoxon test results, 3DBGrowth presented significantly better Dice (*DSC*), Jaccard (*JAC*) and Hausdorff's Distance (*HD*) than GrowCut. For the Running Time (*RT*), there was no significant difference, which implicates that both methods presented comparable processing time.

IV. CONCLUSION

The semi-automatic segmentation of vertebral bodies in a volumetric scenario is a challenging task, due to the large number of slices in the exams. To obtain a proper 3D reconstruction of the vertebrae, one has to pay attention on allowing a fast and accurate segmentation of slices. We have investigated this challenge and used the slope coefficient of the annotation time, so that the specialists' annotations were extrapolated from a slice to its neighbours up to a given limit without losing accuracy and, at the same time, reduced the total time spent on manual annotation.

On the dataset used, on average, only 37% of the slices with vertebral body content had to be annotated, consequently making the process faster (on average, 36 seconds for each vertebral body). We have proposed 3DBGrowth method, which significantly outperforms GrowCut and keeps comparable running time. Moreover, 3DBGrowth presented the best results even with simple/sloppy seed points, which demands less effort on the annotation process.

ACKNOWLEDGMENT

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 and grant No.: 0487/17083480, by the São Paulo Research Foundation (FAPESP, grants No. 2016/17078-0, 2017/23780-2, 2018/06228-7, 2018/24414-2), and the National Council for Scientific and Technological Development (CNPq).

REFERENCES

- [1] M. G. Fehlings, L. Tetreault, A. Nater, T. Choma, J. Harrop, T. Mroz, C. Santaguida, and J. S. Smith, "The aging of the global populationthe changing epidemiology of disease and spinal disorders," *Neurosurgery*, vol. 77, no. 1, pp. 1–5, 2015.
- [2] M. Rak and K. D. Tönnies, "On computerized methods for spine analysis in MRI: a systematic review," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 11, no. 8, pp. 1445–1465, Aug 2016.
- [3] Y. X. J. Wang, F. R. Santiago, M. Deng, and M. H. Nogueira-Barbosa, "Identifying osteoporotic vertebral endplate and cortex fractures," *Quant Imaging Med Surg*, vol. 7, no. 5, pp. 555–591, Oct 2017.
- [4] K. Hammernik, T. Ebner, D. Stern, M. Urschler, and T. Pock, *Vertebrae Segmentation in 3D CT Images Based on a Variational Framework*. Cham: Springer International Publishing, 2015, pp. 227–233.
- [5] Z. Dženan, V. Aleš, E. Jan, H. Daniel, N. Christopher, and K. Andreas, "Robust detection and segmentation for diagnosis of vertebral diseases using routine MR images," *Computer Graphics Forum*, vol. 33, no. 6, pp. 190–204, 2014.
- [6] R. J. Gillies, P. E. Kinahan, and H. Hricak, "Radiomics: images are more than pictures, they are data," *Radiology*, vol. 278, no. 2, pp. 563–577, Feb 2016.
- [7] J. Egger, C. Nimsky, and X. Chen, "Vertebral body segmentation with GrowCut: Initial experience, workflow and practical application," *SAGE Open Med*, vol. 5, pp. 1–5, 2017.
- [8] J. R. F. Junior, M. Koenigkam-Santos, F. E. G. Cipriano, A. T. Fabro, and P. M. de Azevedo-Marques, "Radiomics-based features for pattern recognition of lung cancer histopathology and metastases," *Computer Methods and Programs in Biomedicine*, vol. 159, pp. 23 – 30, 2018.
- [9] P. Casti, A. Mencattini, M. H. Nogueira-Barbosa, L. Frighetto-Pereira, P. M. Azevedo-Marques, E. Martinelli, and C. Di Natale, "Cooperative strategy for a dynamic ensemble of classification models in clinical applications: the case of MRI vertebral compression fractures," *International Journal of Computer Assisted Radiology and Surgery*, vol. 12, no. 11, pp. 1971–1983, Nov 2017.
- [10] L. Frighetto-Pereira, R. M. Rangayyan, G. A. Metzner, P. M. de Azevedo-Marques, and M. H. Nogueira-Barbosa, "Shape, texture and statistical features for classification of benign and malignant vertebral compression fractures in magnetic resonance images," *Computers in Biology and Medicine*, vol. 73, pp. 147 – 156, 2016.
- [11] M. T. Cazzolato, L. C. Scabora, M. R. Nesso, L. F. Milano-Oliveira, A. F. Costa, D. S. Kaster, M. Koenigkam-Santos, P. M. de Azevedo-Marques, C. Traina, and A. J. Traina, "dp-breath: Heat maps and probabilistic classification assisting the analysis of abnormal lung regions," *Computer Methods and Programs in Biomedicine*, vol. 173, pp. 27–34, 2019.
- [12] Z. Xue, L. R. Long, S. Antani, and G. R. Thoma, "Spine X-ray image retrieval using partial vertebral boundaries," in *2011 24th International Symposium on Computer-Based Medical Systems (CBMS)*, June 2011, pp. 1–6.
- [13] A. Gururajan, S. Kamalakannan, H. Sari-Sarraf, M. Shahriar, R. Long, and S. Antani, "On the creation of a segmentation library for digitized cervical and lumbar spine radiographs," *Computerized Medical Imaging and Graphics*, vol. 35, no. 4, pp. 251 – 265, 2011.
- [14] D. Karimi, G. Samei, C. Kesch, G. Nir, and S. E. Salcudean, "Prostate segmentation in MRI using a convolutional neural network architecture and training strategy based on statistical shape models," *Int. J. Computer Assisted Radiology and Surgery*, vol. 13, no. 8, pp. 1211–1219, 2018.
- [15] P. Stefan, S. Habert, A. Winkler, M. Lazarovici, J. Fürmetz, U. Eck, and N. Navab, "A radiation-free mixed-reality training environment and assessment concept for C-arm-based surgery," *International Journal of Computer Assisted Radiology and Surgery*, vol. 13, no. 9, pp. 1335–1344, Sep 2018.
- [16] P. Banerjee, M. Hu, R. Kannan, and S. Krishnaswamy, "A semi-automated approach to improve the efficiency of medical imaging segmentation for haptic rendering," *Journal of Digital Imaging*, vol. 30, no. 4, pp. 519–527, Aug 2017.
- [17] V. Vezhnevets and V. Konouchine, "GrowCut - interactive multi-label N-D image segmentation by cellular automata," *International Conference on Computer Graphics and Vision - GraphiCon*, vol. 1, Nov 2005.
- [18] L. Zhu, I. Kolesov, Y. Gao, R. Kikinis, and A. Tannenbaum, "An effective interactive medical image segmentation method using Fast GrowCut," vol. 17, 2014.
- [19] R. Korez, B. Likar, F. Pernuš, and T. Vrtovec, "Model-based segmentation of vertebral bodies from MR images with 3D CNNs," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. Unal, and W. Wells, Eds. Cham: Springer Int. Publishing, 2016, pp. 433–441.
- [20] B. Gaonkar, Y. Xia, D. S. Villaroman, A. Ko, M. Attiah, J. S. Beckett, and L. Macyszyn, "Multi-parameter ensemble learning for automated vertebral body segmentation in heterogeneously acquired clinical MR images," *J. of Translational Engineering in Health and Medicine*, vol. 5, pp. 1–12, 2017.
- [21] G. Hille, S. Saalfeld, S. Serowy, and K. Tnnies, "Vertebral body segmentation in wide range clinical routine spine MRI data," *Computer Methods and Programs in Biomedicine*, vol. 155, pp. 93 – 99, 2018.
- [22] J. S. Ramos, C. Y. V. Watanabe, M. H. Nogueira-Barbosa, and A. J. M. Traina, "BGrowth: an efficient approach for the segmentation of vertebral compression fractures in magnetic resonance imaging," *Symposium on Applied Computing*, pp. 1–8, April 2019.
- [23] L. Vanneschi, M. Clergue, P. Collard, M. Tomassini, and S. Vérel, "Fitness clouds and problem hardness in genetic programming," in *Genetic and Evolutionary Computation Conference*. Springer, 2004, pp. 690–701.
- [24] P. Jaccard, "The distribution of the flora in the alpine zone," *New Phytologist*, vol. 11, no. 2, pp. 37–50, Feb. 1912.
- [25] T. Sørensen, *A Method of Establishing Groups of Equal Amplitude in Plant Sociology Based on Similarity of Species Content and Its Application to Analyses of the Vegetation on Danish Commons*, ser. Biologiske skrifter. I kommission hos E. Munksgaard, 1948.
- [26] P. D. Barbieri, G. V. Pedrosa, A. J. M. Traina, and M. H. Nogueira-Barbosa, "Vertebral body segmentation of spine MR images using Superpixels," in *28th IEEE International Symposium on Computer-Based Medical Systems*, C. T. Junior, P. P. Rodrigues, B. Kane, P. M. de Azevedo Marques, and A. J. M. Traina, Eds. São Carlos and Ribeirão Preto, Brazil: Conference Publishing Services (CPS), June 2015, pp. 44–49.
- [27] F. J. Massey, "The Kolmogorov-Smirnov test for goodness of fit," *Journal of the American Statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.
- [28] F. Wilcoxon, S. Katti, and R. Wilcox, "Critical values and probability levels for the Wilcoxon rank sum test and the Wilcoxon signed rank test," *Selected Tables in Mathematical Statistics*, vol. 1, pp. 171–259, 1970.