Optimizing and Evaluating a Graph-Based Segmentation of MRI Wrist Bones

Sonia Nardotto, Roberta Ferretti, Laura Gemme$^{(\boxtimes)}$, and Silvana Dellepiane

DITEN, Università degli Studi di Genova, Genova, Italy
{sonia.nardotto,roberta.ferretti,laura.gemme}@edu.unige.it,
silvana.dellepiane@unige.it

Abstract. In this paper, a quantitative evaluation of the graph-based segmentation method presented in a previous work is performed. The algorithm, starting from a single source element belonging to a region of interest, aims at finding the optimal path minimizing a new cost function for all elements of a digital volume. The method is an adaptive, unsupervised, and semi-automatic approach.

For the assessment, a training phase and a testing phase are considered. The system is able to learn and adapt to the ground truth. The performance of the method is estimated by computing classical indices from the confusion matrix, similarity measures, and distance measures.

Our work is based on the segmentation and 3D reconstructions of carpal bones derived from Magnetic Resonance Imaging (MRI) volumetric data of patients affected by rheumatic diseases.

Keywords: Quantitative evaluation · 3D graph-based segmentation · Carpal bones · MRI volumes

1   Introduction

Rheumatic musculoskeletal diseases are among the most common chronic conditions affecting the European population. Some of these diseases, such as osteoarthritis, are the result of wear and tear; others, such as Rheumatoid Arthritis (RA), are autoimmune diseases. In some cases, chronic inflammation leads to the destruction of cartilage, bone, and ligaments, causing deformity of the joints.

Recent research works have proven that through the analysis of Magnetic Resonance Imaging (MRI) erosive change due to rheumatoid arthritis can be detected with greater sensitivity than by using conventional radiography, particularly in the early stages of the disease [2]. In addition, MRI allows the visualization of the three main signs of RA: synovitis, bone oedema and bone erosion [3].

For these reasons, it is important to perform a quantitative or semi-quantitative analysis, and novel scoring systems have recently been proposed such as the OMERACT rheumatoid arthritis scoring system (RAMRIS), which is based on MRI acquisitions [4].

In this context, volume segmentation plays an important role because it makes it possible to extract every single bone and analyze it to evaluate the disease progression.

The aim of this work is to optimize and evaluate the graph-based segmentation method presented in [1] against the available ground truth, which is representative of the user needs and requirements. A training phase is executed to allow the method to learn and adapt itself to the given data set characteristics and to the user desiderata. Due to the noise, extremely low contrast, and bone conformation, the preliminary segmentation results achieved with no a priori knowledge show some differences from the proposed training results. For instance, in the ground-truth volumes, internal holes are often considered as belonging to the bone tissue.

The training phase evaluates some simple post-processing steps that allow improving the segmented volumes and achieving the best result. To this end, a few Mathematical Morphology (MM) operators are proposed to regularize the shape, fill holes, and make the result more robust to noise. In this step, the Receiving Operating Characteristic curve (ROC) and some similarity measures are taken into account.

In a subsequent phase, the testing data set is processed, and all of the obtained results are tested by applying similarity measures and distance metrics between the real and achieved volume surfaces.

The paper is organized as follows. In Section 2, the proposed segmentation approach is briefly introduced, together with a brief description of the database and its use in the training and testing phases. In Section 3, the parameters used for quantitative evaluation are described and discussed. The application of the method to MRI volumes of the hand district and the performance evaluation are presented in Section 4.

2   Proposed Method

This work focuses on applying the segmentation approach presented in [1] to MRI wrist volumes, a training phase for its optimization, and a subsequent quantitative evaluation. The employed database consists of hand and wrist district MRI volumes whose ground truth is based on a segmentation process accomplished by expert physicians in this medical field.

After giving the main idea of the method and after introducing the reference database (DB), the two phases of training and testing are described.

Despite the very low contrast of the wrist bones with respect to the surrounding tissues, the presence of noise, and the lack of shape and smoothness information in the proposed segmentation method, the extracted Region of Interest (ROI) volumes are quite accurate and well separated from nearby similar structures.

The segmentation method is very sensitive to inhomogeneities, so that fine details and internal holes are tracked even when they are of no interest to the medical expert.

By comparing the results obtained by the proposed segmentation approach with the ground truth, one can appreciate how the detected volumes and surfaces can significantly benefit from simple post-processing steps. Simple mathematical morphology operators greatly improve the result accuracy.

2.1   Segmentation

The proposed method is a 3D seed-based algorithm that applies a graph-based segmentation driven by research into the minimum cost paths for the analysis of digital

volumes. Starting from a single seed point belonging to the ROI chosen by the user, an optimal aggregation algorithm selects the best paths and finds a spanning tree that is minimum with respect to the following cost function [1]:

$$f_{w_s}(v_i) = \min_{\pi(s,v_i)}[\max_{x \in \pi(s,v_i)} w_s(x)] \tag{1}$$

where w is the weight associated with the graph vertices, and $\pi(s, v_i)$ is the path from the seed s to a generic vertex v_i.

The proposed cost criterion is comparable to the single-source shortest path problem for a graph with non-negative vertex weights, producing a Minimum Path Spanning Tree (MPST) in routing problems [1].

It is a non-iterative mechanism adaptive to the image content that takes into account the contextual information. The main advantages are that the segmentation can be obtained in one shot. As opposed to other graph-based methods, this algorithm considers only a single seed point (usually in the literature, two seeds are given: one for the foreground and one for the background), and its growth mechanism produces results that are also optimal from the computational point of view.

To obtain a binary volume from the MPST, it is necessary to apply a graph cut. Because the present paper is devoted to the study of the efficiency of the method in separating single wrist bone tissues, the threshold is chosen in a semi-interactive way corresponding to the maximum graph-cut that does not generate leakage problems. As a consequence, one can prove that the optimal result is encompassed in the MPST, from where it can be extracted.

The elapsed processing time is approximately 30 seconds for each bone when running on an Intel Core i7-4700MQ (2.4 GHz, 6 MB cache, 4 core).

2.2 Database: Training and Testing Phases

The database (DB) used is described in paper [7] and is made of 100 MRI T1-weighted volumes acquired by the 0.2 Tesla ARTOSCAN (Esaote Spa, Genova, Italy). Each volume is made of approximately 120 images of size 256×256 pixels. A sample coronal slice is shown in Fig. 1.

Fig. 1. Coronal MRI slice of the wrist.

A wrist is a complex joint between the distal forearm and the hand, composed of 15 bones: the distal parts of the radius and ulna, eight carpal bones, and the proximal parts of five metacarpal bones.

The bone volumes have been segmented by medical experts with RheumaSCORE Software, implemented by Softeco Sismat srl (Genova, Italy) [8]. The segmentation method takes into account bone and erosion; areas of oedema are considered to belong to the bone tissue.

After the preliminary segmentation step applied without any a priori or model information, the training phase is started where the binary segmented volumes are compared with the ground truth. In this phase, the training of the proposed method is executed to decide which type of post-processing gives the best results. As said before, the approach is an adaptive system that learns from the volumes present in the database.

For all of the binary volumes of the analyzed patients, we considered the following operations of mathematical morphology: a 3D dilation applied one time and two times (with a circular structural element) and a 3D dilation with hole filling. After the application of these processing steps, the results are evaluated by means of the classical parameters (sensitivity and precision) and similarity measure (Dice coefficient). This analysis provides the best post-processing segmentation result, which drives the subsequent testing phase. During the testing phase, all of the test cases are evaluated by the above mentioned measures.

Various measures exist to evaluate the quality of a segmentation; the most common are metrics based on volumetric overlap and surface distance. Depending on the application, one parameter may be favoured over another. In this work, various metrics are employed to allow a global quality evaluation.

3   Quantitative Evaluation

In this section, the already mentioned metrics used for the performance evaluation are described in detail, starting from the similarity measure and followed by distance measures.

The basic indices Precision (PR), Sensitivity (SENS) and DICE are given by these formulae:

$$PR = \frac{TP}{TP+FP}, \quad SENS = \frac{TP}{TP+FN} \tag{2}$$

$$DICE = \frac{2\,TP}{(TP+FP)+(TP+FN)} \tag{3}$$

where TP are the True Positives, FP are the False Positives, and FN are the False Negatives. Precision (or positive predictive value) is the proportion of the predicted positive cases that are correct; Sensitivity tells how many voxels are correctly classified positive of the total positive observations. The values of 1-precision (on the x axis) and sensitivity (on the y axis) are usually plotted in a ROC graph.

The Dice coefficient is similarity index that measures the overlap between two sets; a value of 0 indicates no overlap and a value of 1 indicates perfect agreement.

Higher numbers note better agreement between the sets, so when we apply these indices to evaluate the agreement between the segmentation results, the goal is to get as close to 1 as possible.

After some simple operations, the Dice coefficient can be reformulated by these expressions:

$$DICE = \frac{2}{PR^{-1}+SENS^{-1}} = \frac{2}{2+\frac{err}{TP}} \qquad (4)$$

where *err* is the overall segmentation error.

Formula (4) shows the relationship between the Dice coefficient and precision and sensitivity, particularly with the inverse of these two parameters. It may be noted that the Dice coefficient have nonlinear correspondences with the error [5].

The evaluation metrics based on distances [6] perform a comparison of the surface voxels of two segmentations A and B. For each surface voxel of A, the Euclidean distance to the closest surface voxel of B is measured.

Let $S(A)$ be the set of surface voxels of A. The shortest distance of an arbitrary voxel v to $S(A)$ is defined as $d(v, S(A)) = \min_{s_A \in S(A)} \|v - s_A\|$, where $\|\cdot\|$ denotes the Euclidean distance.

The Average Symmetric Surface Distance (ASD) is given by

$$ASD(A,B) = \frac{1}{|S(A)|+|S(B)|} \left(\sum_{s_A \in S(A)} d(s_A, S(B)) + \sum_{s_B \in S(B)} d(s_B, S(A)) \right) \qquad (5)$$

The Root Mean Square Symmetric Surface Distance (RMSD) is

$$RMSD(A,B) = \sqrt{\frac{1}{|S(A)|+|S(B)|} \cdot \left(\sum_{s_A \in S(A)} d^2(s_A, S(B)) + \sum_{s_B \in S(B)} d^2(s_B, S(A)) \right)} \qquad (6)$$

These two distances are strongly correlated, with RMSD giving a large penalty for large deviations from the true contour.

The Maximum Symmetric Surface Distance (Hausdorff distance) is defined as

$$MSD(A,B) = \max\{\max_{s_A \in S(A)} d(s_A, S(B)), \max_{s_B \in S(B)} d(s_B, S(A))\} \qquad (7)$$

This last metric is sensitive to outliers and returns the true maximum error. By the definition of the formulae, the distance is 0 for a perfect segmentation, and greater the value of the distance, the worse the segmentation is.

4   Results and Discussion

For the experiments, 43 patients are analyzed and evaluated; 10 of these cases are non-pathological. In particular, in the training phase, 20 pathological and 4 non-pathological cases are studied. In the testing phase, 13 pathological and 6 non- pathological cases are considered. For the present application, the carpal bones are taken into account: the scaphoid, lunate, triquetrum, pisiform, trapezium, trapezoid, capitate and hamate.

Training Phase. After the segmentation, a binary volume is obtained for each bone. To regularize the results and improve the accuracy with respect to the ground truth, some mathematical morphology steps can be applied to each bone volume. This phase allows the determination of the best MM operator to apply and the corresponding structural element. The considered expansions are 3D dilation, 3D dilation applied twice and 3D dilation with filling holes.

In Fig. 2, it is possible to observe, for the triquetrum bone, the trend of the Dice coefficient in the different patients analyzed for all three dilations. In almost all cases, the best result is the 3D dilation with hole filling.
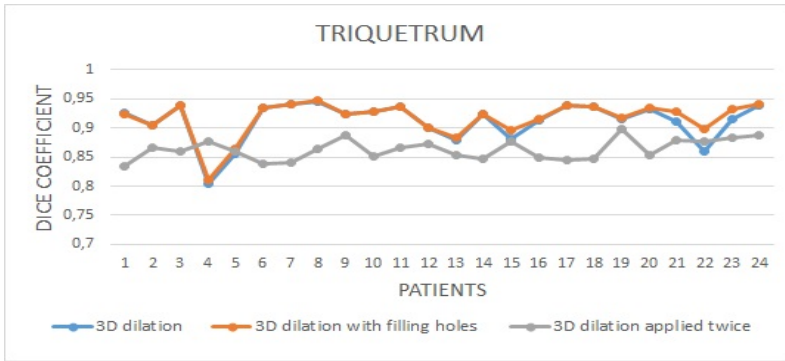
Fig. 2. Trend of DICE coefficient in different patients for all three dilations.

Fig. 3 shows the 3D visualization of all bones obtained using the 3D dilation with hole filling (on the left) and the ground truth (on the right).
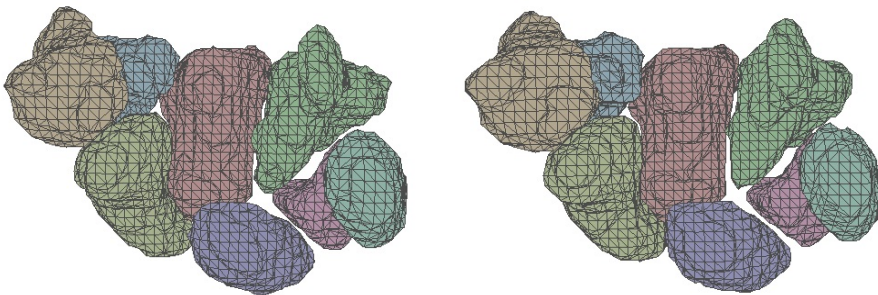
Fig. 3. 3D visualization of results retrieved by our method (on the left) and by the database (on the right)

Testing Phase. Once the 3D dilation with hole filling is defined as the best procedure, 19 patients are evaluated. A first analysis is performed using the Dice coefficient. As shown in Tab.1 (Testing phase), the mean, minimum and maximum values are computed for each bone. The value 1 corresponds to the two coincident volumes. In our cases, the Dice coefficient takes values of approximately 0.91, so the results obtained

are good. By analysing the minimum and maximum values, one can note that they are distributed within a limited range.

By comparing the training and testing phases, it results that the mean values are similar. This underlines the validity of the testing phase.

Table 1. Dice coefficients in the Training and Testing phases for each carpal bone.

TRANING PHASE				*TESTING PHASE*			
DICE COEFFICIENT				**DICE COEFFICIENT**			
	Mean	**Minimum**	**Maximum**		**Mean**	**Minimum**	**Maximum**
CAPITATE	0.9323	0.8128	0.9626	CAPITATE	0.9189	0.8430	0.9620
TRIQUETRUM	0.8967	0.7573	0.9435	TRIQUETRUM	0.8921	0.6818	0.9411
PISIFORM	0.9142	0.8298	0.9515	PISIFORM	0.9214	0.8494	0.9463
SCAPHOID	0.9346	0.8324	0.9603	SCAPHOID	0.9308	0.8571	0.9603
LUNATE	0.8994	0.7098	0.9539	LUNATE	0.8893	0.6203	0.9464
TRAPEZIUM	0.9016	0.7858	0.9456	TRAPEZIUM	0.9036	0.7817	0.9477
TRAPEZOID	0.9165	0.8116	0.9463	TRAPEZOID	0.9002	0.7241	0.9509
HAMATE	0.8938	0.6484	0.9569	HAMATE	0.9014	0.7622	0.9448
MEAN	*0.9111*	*0.7735*	*0.9526*	*MEAN*	*0.9072*	*0.7650*	*0.9499*

A second analysis applied to the volumes is based on the distances ASD, RMSD, and MSD. The results obtained for each bone and the mean values are shown in Tab.2 (Testing phase). The results are also good for these parameters. In fact, for ASD and RMSD, the obtained values are on an average under unity. The values of MSD are larger than the previous ones because they are not normalized over the total surface and they take into account the worst case. However, on an average, the results are close to 4 voxels. Also in this case the mean values in the two phases are similar.

Table 2. Distance measure in the Training and Testing phases for each carpal bone.

TRAINING PHASE				*TESTING PHASE*			
	ASD	**RMSD**	**MSD**		**ASD**	**RMSD**	**MSD**
CAPITATE	0.6038	0.9283	4.9703	CAPITATE	0.6468	0.9566	4.8575
TRIQUETRUM	0.6810	1.0198	4.6848	TRIQUETRUM	0.8259	1.2548	5.1991
PISIFORM	0.5321	0.7981	3.2313	PISIFORM	0.5325	0.8062	3.1545
SCAPHOID	0.4920	0.7677	3.5450	SCAPHOID	0.5471	0.8289	3.4033
LUNATE	0.6645	0.9913	4.2234	LUNATE	0.6945	1.0275	4.3575
TRAPEZIUM	0.6699	0.9854	4.1040	TRAPEZIUM	0.7062	1.0335	4.1369
TRAPEZOID	0.5288	0.7994	3.3465	TRAPEZOID	0.6020	0.9163	3.7777
HAMATE	0.7309	1.0788	4.9428	HAMATE	0.6665	0.9798	4.7162
MEAN	*0,6129*	*0,9211*	*4.1310*	*MEAN*	*0.6527*	*0.9754*	*4.2003*

5   Conclusions

In this work, a quantitative evaluation of the graph-based segmentation method has been performed. The proposed approach essentially consists of three steps: the segmentation, the training phase, and the testing phase. The segmentation, starting from a

single seed and using graph theory, allows the extraction of each bone separately. The training phase has been crucial to defining the best result of our system with respect to the ground truth by computing some common parameters (sensitivity and precision) and similarity measure (Dice coefficient). The testing phase proves the robustness of our method by computing similarity and distance measures (Average Symmetric Surface Distance, Root Mean Square Symmetric Surface Distance, Hausdorff Distance).

The application consists of the extraction of carpal bones from real MRI volumes. It is possible to apply this method to different anatomical districts and different pathologies using appropriate qualitative validation.

The results show the good accuracy and precision of the method with respect to the ground truth. As future work, we intend to automate the optimization process and validate all the data sets.

References

1. Gemme, L., Dellepiane, S.: A new graph-based method for automatic segmentation. In: 18th International Conference on Image Analysis and Processing, ICIAP, Genova (in press, 2015)
2. Włodarczyk, J., Czaplicka, K., Tabor, Z., Wojciechowski, W., Urbanik, A.: Segmentation of bones in magnetic resonance images of the wrist. International Journal of Computer Assisted Radiology and Surgery, 1–13 (2014)
3. Cimmino, M.A., Innocenti, S., Livrone, F., Magnaguagno, F., Silvestri, E., Garlaschi, G.: Dynamic gadolinium-enhanced magnetic resonance imaging of the wrist in patients with rheumatoid arthritis can discriminate active from inactive disease. Arthritis & Rheumatism 48(5), 1207–1213 (2003)
4. Boesen, M., Østergaard, M., Cimmino, M.A., Kubassova, O., Jensen, K.E., Bliddal, H.: MRI quantification of rheumatoid arthritis: current knowledge and future perspectives. European Journal of Radiology 71(2), 189–196 (2009)
5. Chang, H.H., Zhuang, A.H., Valentino, D.J., Chu, W.C.: Performance measure characterization for evaluating neuroimage segmentation algorithms. Neuroimage 47(1), 122–135 (2009)
6. Heimann, T., Van Ginneken, B., Styner, M.A., Arzhaeva, Y., Aurich, V., Bauer, C., Wolf, I.: Comparison and evaluation of methods for liver segmentation from CT datasets. IEEE Transactions on Medical Imaging 28(8), 1251–1265 (2009)
7. Tomatis, V., et al.: A database of segmented MRI images of the wrist and the hand in patients with rheumatic diseases. In: Murino, V., Puppo, E., Sona, D., Cristani, M., Sansone, C. (eds.) ICIAP 2015 Workshops. LNCS, vol. 9281, pp. 143–150. Springer, Heidelberg (2015)
8. Parascandolo, P., Cesario, L., Vosilla, L., Viano, G.: Computer aided diagnosis: state-of-the-art and application to musculoskeletal diseases. In: 3D Multiscale Physiological Human, pp. 277–296. Springer, London (2014)