Unsupervised Segmentation Evaluation Measures for Parameter Optimization in Indicator-Kriging

Tom Bultreys[1], Loes Brabant[2], Matthieu Boone[2], Veerle Cnudde[1] and Luc Van Hoorebeke[2]

[1]UGCT- Department of Geology and Soil Science, Ghent University, Faculty of Science Krijgslaan 281, S8, 9000 Ghent, Belgium [Tom.Bultreys@UGent.be; Veerle.Cnudde@UGent.be]

[2]UGCT - Department of Physics and Astronomy, Ghent University, Faculty of Science, Proeftuinstraat, 86, 9000 Ghent, Belgium [Loes.Brabant@UGent.be; Matthieu.Boone@UGent.be; Luc.Vanhoorebeke@UGent.be]

Keywords: Segmentation, segmentation evaluation, indicator-Kriging

ABSTRACT

This work investigates the performance of four unsupervised evaluation measures for the optimization of the user-defined parameter in the indicator-Kriging segmentation algorithm (Oh and Lindquist 1999). We focus on the application of this algorithm to micro-computed tomography (µCT) scans of porous media. Because ground truth segmentations were required for the set of test images, simulated 3D images were created based on the image acquisition in µCT, starting from segmentations of real µCT-scans.

The tested unsupervised evaluation measures were the intra-class variance, Otsu's parameter, Zeboudj's parameter and the grey value contrast. The intra-class variance proved to be the most efficient at selecting an optimal segmentation parameter.

1. INTRODUCTION

In the literature, a lot of attention has been given to the development of new segmentation algorithms. Comparatively, the methods for validating new segmentation algorithms have gathered less attention. In general, three types of validation can be discerned: visual validation, the use of supervised evaluation measures and the use of unsupervised measures (Zhang 1996, Zhang et al. 2007, Chabrier et al. 2006). Visual validation is intuitive but subjective and non-quantitative. Supervised evaluation measures on the other hand score evaluations by quantifying their likeness to a theoretically correct ("ground truth") segmentation. These methods are objective and quantitative, but ground truth is not always available. Unsupervised measures score a segmentation by quantifying desired properties of the segmented image itself. These measures do not require ground truth, and can therefore be used for online segmentation evaluation and the optimization of free parameters in segmentation algorithms.

This work investigates the efficiency of four unsupervised evaluation measures for the optimization of the free parameter in the indicator-Kriging (IK) segmentation algorithm (Oh and Lindquist 1999). The goal is bifold: increasing the performance of the algorithm and reducing subjectivity caused by parameter selection. The overall performance of the IK algorithm with automatic parameter selection was also investigated. As test images we used simulated µCT scans of porous media, because ground truth segmentations of real µCT images are hard to come by.

1.1. The indicator-Kriging algorithm

The IK algorithm is a locally adaptive segmentation algorithm. It takes into account the grey values in the neighbourhood of each voxel and the spatial correlations in the image. The method requires initialization with two histogram thresholds. To reduce the number of user-defined input parameters, Oh and Lindquist (1999) propose fitting a binormal distribution to the histogram using the expectation-maximization algorithm. The two histogram thresholds can then be reduced to one input parameter. For this work the IK algorithm with expectation-maximization was implemented in the framework of Morpho+ (Brabant et al. 2011).

1.2. Unsupervised evaluation measures

Unsupervised evaluation measures are usually based on one or more of the four criteria proposed by Haralick and Shapiro (1985):

1. Regions should be internally homogeneous.
2. Neighbouring regions should be significantly different in the properties in which they are homogeneous.
3. Regions should be simple, without many small holes.
4. Regions' borders should be smooth and accurate.

The four unsupervised measures were chosen based on Wang et al. (2011), Zhang et al. (2007) and Chabrier et al. (2006). Wang et al. (2011) already used the intra-class variance (ICV) of the pore space, a measure based on the first criterion.

The ICV of a phase in the image is defined as the variance of the pore space grey values normalised by the variance of all the grey values in the image and divided by the relative abundance of that phase in the image. We always used the ICV of the phase which is least abundant in the image.

Otsu's parameter (Otsu, 1979) can be calculated as the sum of the intra-class variances of the pore and the matrix space. It can be proven to be related to both the first and the second criterion of Haralick and Shapiro.

The grey value contrast (GVC) is defined as the absolute value of the difference between the average grey values of the pore and the matrix space over the sum of these quantities. It is therefore based on Haralick and Shapiro's second criterion.

In contrast to the three previously mentioned measures, Zeboudj's measure takes into account local information. It uses the local contrast at region edges to punish inaccurate borders, as well as local contrast within regions to punish undersegmentation. It relies on the first, second and fourth previously mentioned criteria to score segmentations. For its definition we refer to Zhang (2007).

2. EXPERIMENTAL

2.1. Simulated µCT-scans

To obtain representative µCT images with objective ground truth, eight simulated images were generated. Eight real µCT scans of porous media (2 aluminium foams, 2 soil samples, Bray sandstone, Savonnière limestone and 2 sand pack samples) were manually thresholded and remaining noise voxels were removed. These segmented images were used as input to a projection simulator developed at the UGCT (De Witte 2010). This software tool generates simulated radiographies from a digital sample by tracing X-rays from a simulated monochromatic point source to a simulated detector, through the attenuating sample. The sample is assigned an arbitrary attenuation coefficient and each phase is considered to be uniformly attenuating. For each sample, 361 projections were generated and Poisson noise was added to them (the count rate was set to be between 600 and 1000). The X-ray source had a cone-beam geometry and the magnification was set so that the input and output images would match. Therefore, the input images can serve as objective ground truth images. Octopus software (www.octopusreconstruction.com) was used for the reconstruction.

The procedure generates images with representative structure, noise and partial volume effects. Furthermore, the procedure can be extended to include other image artefacts typical for µCT, such as beam hardening artifacts (by using a polychromatic projection simulator, also developed at UGCT). An example of a simulated µCT can be found in figure 1.

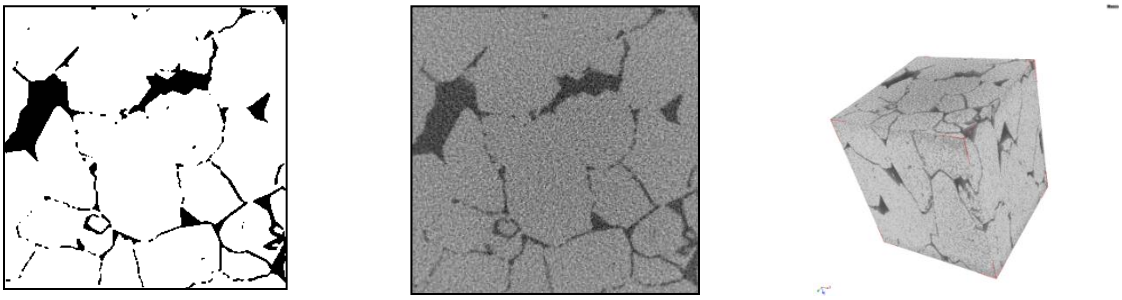

Figure 1: *On the left a slice from a manual segmentation of a real µCT scan of Bray sandstone, in the middle the corresponding slice in the simulated image and on the right the 3D simulated image.*

2.2. Experimental procedure

For each simulated µCT-scan, a set of segmentations was generated by varying the free parameter (for the definition, see Oh and Lindquist, 1999) in the IK algorithm. The parameter was varied from 0.3 to past the value where the two thresholds coincided, in steps of 0.3.

To quantify the efficiency of the unsupervised evaluation criteria, each parameter's similarity rate of correct comparison (SRCC, Chabrier et al. 2006) was calculated over all segmentation sets. The SRCC of an unsupervised parameter indicates the rate with which that parameter is able to select the "best segmentation" (defined as the one with the lowest percentage of wrongly classified voxels) out of a pair of segmentations from the same image. We also indicate whether or not each unsupervised evaluation measure is able to select the overall best segmentation out of a set of images.

Finally, the overall performance of the IK algorithm with automatic parameter selection is determined for the simulated µCT scans.

3. RESULTS AND DISCUSSION

Somewhat surprisingly, the ICV turned out to be the most efficient at selecting the best from a pair of segmentations (see figure 2), while Zeboudj's measure seems to be more theoretically funded.

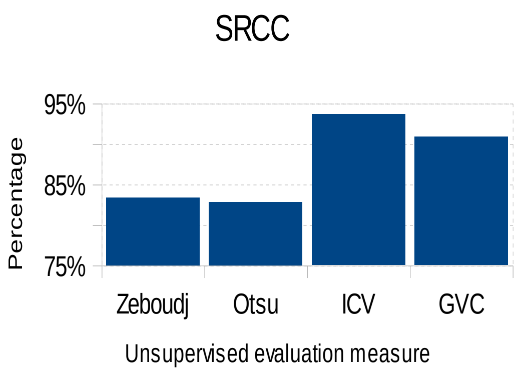

Figure 2: *SRCC of the tested unsupervised evaluation measures*

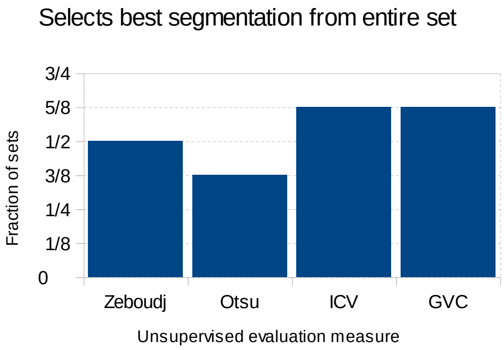

Figure 3: *Fraction of images for which the evaluation measures select the optimal segmentation*

It is also useful to check whether a measure is able to select the optimal segmentation from the whole set (figure 3). The ICV and the GVC measures score best. Zeboudj's measure only selected a different optimal segmentation from these two measures in one image case.

However, the ICV of the phase that is most abundant in the image turned out to follow the opposite trend of the ICV of the least abundant phase, i.e. it would select the worst segmentation from the set instead of the best. Therefore, the ICV was determined to drop in SRCC as the porosity approached 50%, since there is a transition zone between the ICVs of the two phases being good evaluation measures.

The average percentage of misclassified voxels of the IK algorithm with automatic parameter selection was 1,62 % for the tested images if the ICV, the GVC or the Zeboudj measure were used for parameter selection (Zeboudj selected a different parameter only once, and in this case the segmentation it selected was hardly discernible from the one selected by ICV and GVC). With the Otsu measure, the overall percentage of misclassification was 1,74 %, however it failed completely for one out of eight images (the results for this image were left out of the average). For comparison, the k-means clustering algorithm (after application of a median filter) misclassified 2 % of all voxels on average and completely failed to segment one of eight images (result was left out of the average).

4. CONCLUSIONS

In this study, the intra-class variance of the least abundant phase in the image gives the best results out of the tested unsupervised segmentation evaluation measures for automatic parameter selection in the IK algorithm with expectation-maximization. However, if the porosity is close to 50 %, it breaks down and it would be better to use the Grey Value Contrast or Zeboudj's measure. The IK algorithm with automatic parameter selection is determined to be more accurate than k-means clustering after median filtering.

It should be noted that further tests (i.e. more test images) are needed to generalize the results in this abstract.

5. ACKNOWLEDGEMENTS

The Special Research Fund of the Ghent University (BOF) is acknowledged for the doctoral grant to Loes Brabant.

6. REFERENCES

Brabant, L.; Vlassenbroeck, J.; Witte, Y. D.; Cnudde, V.; Boone, M.; Dewanckele, J. & Hoorebeke, L. V.Three-Dimensional Analysis of High-Resolution X-Ray Computed Tomography Data with Morpho+. *Microscopy and microanalysis,* 2011*, 17*, 252-263.

Chabrier, S.; Emile, B.; Rosenberger, C. & Laurent, H. Unsupervised performance evaluation of image segmentation. *EURASIP Journal on Applied Signal Processing,* 2006, 1-12.

De Witte, Y. Improved and practically feasible reconstruction methods for high resolution X-ray tomography. PhD Thesis. *Ghent University,* 2010.

Haralick, R. & Shapiro, L. Survey: image segmentation techniques. *Computer vision, graphics, and image processing,* 1985*, 29*, 100-132.

Oh, W. & Lindquist, B. Image thresholding by indicator Kriging. *IEEE Transactions on Pattern analysis and machine intelligence,* 1999*, 21*, 590-602.

Otsu, N. A threshold selection method from gray-level histograms. *IEEE Transactions on systems, man and cybernetics,* 1979*, SMC-9*, 62-66.

Wang, W.; Kravchenko, A.; Smucker, A. & Rivers, M. Comparison of image segmentation methods in simulated 2D and 3D microtomographic images of soil aggregates. *Geoderma,* 2011*, 162*, 231-241.

Zhang, H.; Fritts, J. & Goldman, S. Image segmentation evaluation: a survey of unsupervised methods *Computers Vision and Image Understanding,* 2007*, 110*, 260-280.

Zhang, Y. A survey on evaluation methods for image segmentation. *Pattern Recognition,* 1996*, 29*, 1335-1346.