

Image processing of multiphase images obtained via X-ray microtomography: A review

The Faculty of Oregon State University has made this article openly available.
Please share how this access benefits you. Your story matters.

Citation	Schlüter, S., A. Sheppard, K. Brown, and D. Wildenschild (2014), Image processing of multiphase images obtained via X-ray microtomography: A review, <i>Water Resources Research</i> , 50, 3615–3639. doi:10.1002/2014WR015256
DOI	10.1002/2014WR015256
Publisher	American Geophysical Union
Version	Version of Record
Terms of Use	http://cdss.library.oregonstate.edu/sa-termsfuse



REVIEW ARTICLE

10.1002/2014WR015256

Key Points:

- First survey of image processing methods for multiphase fluid images
- A novel protocol is suitable for various types of porous media
- Many routines come with a freely available open-source library

Correspondence to:

S. Schlüter,
schluets@onid.orst.edu

Citation:

Schlüter, S., A. Sheppard, K. Brown, and D. Wildenschild (2014), Image processing of multiphase images obtained via X-ray microtomography: A review, *Water Resour. Res.*, 50, 3615–3639, doi:10.1002/2014WR015256.

Received 4 JAN 2014

Accepted 2 APR 2014

Accepted article online 7 APR 2014

Published online 25 APR 2014

Image processing of multiphase images obtained via X-ray microtomography: A review

Steffen Schlüter^{1,2,3}, Adrian Sheppard², Kendra Brown¹, and Dorthe Wildenschild¹

¹School of Chemical, Biological and Environmental Engineering, Oregon State University, Corvallis, Oregon, USA,

²Department of Applied Mathematics, Research School of Physics and Engineering, Australian National University, Canberra, Australian Capital Territory, Australia, ³Department of Soil Physics, Helmholtz-Centre for Environmental Research—UFZ, Halle, Germany

Abstract Easier access to X-ray microtomography (μ CT) facilities has provided much new insight from high-resolution imaging for various problems in porous media research. Pore space analysis with respect to functional properties usually requires segmentation of the intensity data into different classes. Image segmentation is a nontrivial problem that may have a profound impact on all subsequent image analyses. This review deals with two issues that are neglected in most of the recent studies on image segmentation: (i) focus on multiclass segmentation and (ii) detailed descriptions as to why a specific method may fail together with strategies for preventing the failure by applying suitable image enhancement prior to segmentation. In this way, the presented algorithms become very robust and are less prone to operator bias. Three different test images are examined: a synthetic image with ground-truth information, a synchrotron image of precision beads with three different fluids residing in the pore space, and a μ CT image of a soil sample containing macropores, rocks, organic matter, and the soil matrix. Image blur is identified as the major cause for poor segmentation results. Other impairments of the raw data like noise, ring artifacts, and intensity variation can be removed with current image enhancement methods. Bayesian Markov random field segmentation, watershed segmentation, and converging active contours are well suited for multiclass segmentation, yet with different success to correct for partial volume effects and conserve small image features simultaneously.

1. Introduction

The last decade has seen a tremendous progress in X-ray tomography and imaging techniques providing new means to analyze a multitude of research problems in porous media research. In the scope of water resources research, applications range from soil-water-root interactions and mechanical and hydraulic properties of rocks to pore-scale modeling of multiphase flow and continue to appear in related fields of research (see reviews by *Blunt et al.* [2013], *Cnudde and Boone* [2013], *Wildenschild and Sheppard* [2013], and *Anderson and Hopmans* [2013]). Progress in image processing has kept a comparable pace in terms of new developments in image enhancement, image analysis, and hardware architectures [e.g., *Ketcham and Carlson*, 2001; *Sheppard et al.*, 2004; *Kaestner et al.*, 2008; *Porter and Wildenschild*, 2010; *Tuller et al.*, 2013]. Since X-ray tomography is becoming a standard technique available to an increasing number of research groups in water resources research, more and more scientists have a need for information on how to process their data. Not everyone new to the field has the resources to develop their own image processing toolbox, tailored for the research question at hand, or the budget to take advantage of powerful image processing software that often has a rather comprehensive scope. A relief in this regard are software toolboxes which are freely available to the scientific community like *IMAGEJ* [*Ferreira and Rasband*, 2012], *ITK* [*Ibanez et al.*, 2005], *QUANTIM* [*Vogel et al.*, 2010], *BLOB3D* [*Ketcham*, 2005], or *SCIKIT-IMAGE* [*van der Walt et al.*, 2014], just to name a few. Their multiphase segmentation capabilities are somewhat limited and may require substantial operator input. The software used in this study is described in the Appendix A.

However, comparing the performance of different image processing methods on the same set of test images often leads to very different results. A notorious example is image segmentation of a gray value image into objects and background [*Sezgin and Sankur*, 2004; *Iassonov et al.*, 2009; *Baveye et al.*, 2010]. Yet, these comparative studies often merely list the performance of several segmentation methods with respect

to a certain quality measure or highlight the user-dependency of the segmentation result, but lack in useful information as to why a specific method fails under certain circumstances and how this may be avoided by suitable preprocessing. Another shortcoming is that many recent review papers on image segmentation with a focus on soil images deal with binary segmentation only [Baveye *et al.*, 2010; Wang *et al.*, 2011; Houston *et al.*, 2013a] and do not provide solutions to multiclass segmentation problems.

This review paper has two main objectives. First, we survey various segmentation methods with respect to multiclass segmentation. We focus on methods that operate on a single image, i.e., coupled images scanned at different X-ray energy levels [Rogasik *et al.*, 1999; Costanza-Robinson *et al.*, 2008; Armstrong *et al.*, 2012] or in a wet and dry state [Culligan *et al.*, 2004; Wildenschild *et al.*, 2005] are not discussed here. We refer the reader to Brown *et al.* [2014] where we demonstrate that a single-energy method outperforms a three-energy method and discuss the potential shortcomings of either approach. All of the surveyed methods are locally adaptive, i.e., in addition to global histogram information they consider some neighborhood statistic for class assignment. In particular, we will examine hysteresis segmentation [Vogel and Kretzschmar, 1996], indicator kriging [Oh and Lindquist, 1999], converging active contours [Sheppard *et al.*, 2004], watershed segmentation [Vincent and Soille, 1991], and Bayesian Markov random field segmentation [Kulkarni *et al.*, 2012]. As quality measures, we will use misclassification error, volume fraction, specific interfacial area, and a connectivity measure. Second, we point out that the performance of image segmentation cannot be examined independently of image enhancement prior to classification. To do so, we compare the impact of different denoising methods on the segmentation results. We have chosen standard noise removal methods that were reviewed for application in μ CT data of porous media before [Kaestner *et al.*, 2008; Tuller *et al.*, 2013]. Moreover, we apply efficient algorithms for image artifact removal, such as intensity bias [Iassonov and Tuller, 2010] and ring artifacts [Sijbers and Postnov, 2004]. Finally, we illustrate the impairment of proper threshold detection that is due to low contrast and imbalanced histograms, and present methods to correct it.

The performance of different segmentation methods is evaluated by means of three test images. We start with a synthetic test image of a partially saturated packing of spheres, where the volume fractions and interfacial areas of the wetting, nonwetting, and solid phase are known exactly. The true image is superimposed with ring artifacts, blur, and noise, and the success of different combinations of denoising and segmentation in recovering the morphological properties of the true image is compared. Subsequently, the most suitable combinations are applied to two real images of quite different scopes. The first is a synchrotron image of a three-fluid medium impaired by intensity variation and noise [Brown *et al.*, 2014]; the second is a μ CT image of a soil with macropores, organic matter, and rocks impaired by noise and blur [Houston *et al.*, 2013b].

The paper is organized as follows: in section 2, we provide the details for each image processing method, while image enhancement and segmentation results are compared in terms of visual appearance and morphology measures in section 3. In section 4, we discuss the results and provide recommendations for best practices, and our findings are summarized in section 5.

2. Methods

2.1. Artifact Removal

Due to shortcomings in the image acquisition process, the base signal of an X-ray scan is often superimposed by different kinds of image artifacts [e.g., Ketcham and Carlson, 2001; Wildenschild *et al.*, 2002]. The most frequent impairments are image noise due to a low count of incoming radiation at the detector, and image blur due to movement, hardware constraints, or suboptimal image reconstruction. As discussed below, there are powerful denoising methods that efficiently remove noise in homogeneous locations and at the same time conserve edges between objects. Other image artifacts which are less trivial to remove a posteriori are ring artifacts, due to defective diodes in the detector panel, or beam hardening of polychromatic beams, which manifests itself in the reconstructed image as streakings around high-attenuation objects and intensity variation with distance to the sample center. Note that there are means to avoid some of these artifacts already during image acquisition or image reconstruction like a slightly altering detector panel position during scanning and wedge calibration [Ketcham and Carlson, 2001]. Here we focus on methods that can be directly applied to the reconstructed volume.

Ring artifacts can be removed separately for each image slice z after transforming the image from Cartesian coordinates $\mathbf{x} = (x, y)$ into polar coordinates $\mathbf{x} = (r, \varphi)$, where \mathbf{x} is the location vector, (x, y) is the horizontal and vertical coordinate, and (r, φ) are radius and angle. In this way, the rings appear as vertical lines that can be removed with a moving window W of width $w \ll R$, where R is the radius of the sample. The window detects average variations in gray value median $\tilde{I}(r)$ along r and normalizes them subsequently [Sijbers and Postnov, 2004]. Only homogeneous rows within W contribute to the median $\tilde{I}(r)$ at column r , where the homogeneity threshold H has to be set by the user. This method has some shortcomings since objects aligned to a certain radius may also be removed.

To our knowledge, the removal of streaking artifacts due to beam hardening is an unresolved problem. Intensity bias, on the other hand, can be removed rather easily given that it is not superimposed by changing attenuation coefficients due to variable fluid saturation or matrix porosity. To do so, requires an iterative procedure [Jassonov and Tuller, 2010]: (i) the image is segmented into the class with the highest attenuation and background by simple, histogram-based thresholding. (ii) The mean gray value within the highest density class is stored as a function of radius. (iii) A smooth function is fitted to the data:

$$I(r) = a + b \cos(2\pi r/R) + c \exp(r/R) \tag{1}$$

where a , b , and c are fitting parameters. (iv) The smooth function is used to normalize the data. Steps (i–iv) are repeated until convergence is achieved after two to four iterations.

2.2. Denoising

2.2.1. Median Filter

A good noise removal algorithm should exert significant smoothing in homogeneous regions (i.e., zones with low-intensity gradient $\nabla I(\mathbf{x})$) and minimal modification of edges (i.e., high $\nabla I(\mathbf{x})$ zones), where ∇ is the differential operator with respect to three dimensions ($\nabla = \frac{\partial}{\partial x} + \frac{\partial}{\partial y} + \frac{\partial}{\partial z}$). The simplest method for nonlinear denoising is a median filter (MD) with a cubic kernel of diameter d :

$$\hat{I}_{MD}(\mathbf{x}) = I'(\mathbf{x}) * M_d(\mathbf{x}) \tag{2}$$

where $*$ denotes convolution, I' is the raw image, and \hat{I} is the denoised result. The gray value that divides the set of d^3 sorted gray values within M_d into equal halves is assigned to the current voxel at location \mathbf{x} [Gonzalez and Woods, 2002]. Note that this routine is usually applied in one loop and is rather slow for large kernel sizes, mainly due to sorting. However, a tremendous increase in speed is achieved by applying lookup tables and a moving median, i.e., for a kernel shift of one position only a small amount of d^2 gray values has to be replaced in a table [Huang et al., 1979].

2.2.2. Anisotropic Diffusion Filter

Another popular, nonlinear denoising method is the anisotropic diffusion (AD) filter [Perona and Malik, 1990; Catté et al., 1992]. The rationale of this method is that the Gaussian distribution is the solution to the diffusion equation with a constant diffusion coefficient D . In the same way, applying the diffusion equation with a nonlinear diffusion coefficient amounts to smoothing with a Gaussian kernel of strongly varying standard deviation. Obviously, D should depend on the local intensity gradient $\nabla I(\mathbf{x})$. Hence, anisotropic diffusion calls for a numerical solution of the following partial differential equation (PDE):

$$\hat{I}^0 = I' \tag{3}$$

$$\frac{\partial \hat{I}_{AD}}{\partial t} = \nabla [D(|\nabla(G_\sigma * \hat{I}_{AD})|)] \nabla \hat{I}_{AD} \tag{4}$$

where t is numerical time, \hat{I}_{AD} is short for $\hat{I}_{AD}(\mathbf{x}, t)$, and the gradient of smoothed intensity values, convolved by a Gaussian G_σ of standard deviation σ , serve as an edge detector. The simplest implementation is:

$$D(|\nabla(G_\sigma * \hat{I})|) = \begin{cases} 1, & |\nabla(G_\sigma * \hat{I})| \leq \kappa \\ 0, & |\nabla(G_\sigma * \hat{I})| > \kappa \end{cases} \quad (5)$$

where κ is a diffusion stop criterion. The number of iterations is another important parameter that has to be set manually, because the solution would eventually converge to uniform intensity.

2.2.3. Total Variation Filter

Another PDE-based approach is total variation (TV) denoising [Rudin et al., 1992]. The rationale behind this method is to minimize the intensity variation in the image by means of the following cost function:

$$\hat{I}_{TV} = \operatorname{argmin}_I \left\{ \underbrace{\int |\nabla \hat{I}(\mathbf{x})| d\mathbf{x}}_{\text{regularization}} + \lambda \underbrace{\int |I'(\mathbf{x}) - \hat{I}(\mathbf{x})|^2 d\mathbf{x}}_{\text{fidelity}} \right\} \quad (6)$$

where λ is a scale parameter that controls the trade-off between regularization, i.e., smoothing, and fidelity to the raw data I' . The solution is achieved with the following set of coupled PDE's:

$$\hat{I}^0 = I' \quad (7)$$

$$\frac{\partial \hat{I}_{TV}}{\partial t} = \nabla \cdot \left(\frac{\nabla \hat{I}_{TV}}{|\nabla \hat{I}_{TV}|} \right) + \lambda (I' - \hat{I}_{TV} + K) \quad (8)$$

$$\frac{\partial K}{\partial t} = \alpha (I' - \hat{I}_{TV}) \quad (9)$$

where \hat{I}_{TV} is short for $\hat{I}_{TV}(\mathbf{x}, t)$. The time step control α can be made adaptive to $\partial K / \partial t$. The number of iterations, used as a stopping criterion, is less crucial as compared to \hat{I}_{AD} , because the solution does not converge to uniform intensity due to the fidelity term used in equation (6).

2.2.4. Nonlocal Means Filter

Unlike the previous methods, the nonlocal means filter (NL) is a linear filter, i.e., the gray value at the current location is the average of gray values at other locations, assigned with some suitable weighting factors, w . However, in contrast to standard linear filters (Gaussian filter, mean filter, etc.), it does not use a small-sized kernel, but potentially the entire image as a search window. The rationale is to compare the neighborhoods of all voxels $\mathbf{y} \in I$ with the neighbors of the current voxel at location \mathbf{x} [Buades et al., 2005]:

$$\hat{I}_{NL}(\mathbf{x}) = \sum_{\mathbf{y} \in I} w(\mathbf{x}, \mathbf{y}) I'(\mathbf{y}) \quad (10)$$

Thus, the similarity of a whole neighborhood with fixed size determines the weight $w(\mathbf{x}, \mathbf{y})$ with which a distant voxel will influence the new value of the current voxel. More specifically, the weights are significant only if a Gaussian kernel G_σ with standard deviation σ around \mathbf{y} looks like the corresponding Gaussian kernel around \mathbf{x} :

$$w(\mathbf{x}, \mathbf{y}) = \frac{1}{Z(\mathbf{x})} \exp \left(- \frac{\int G_\sigma(\mathbf{n}) * |I'(\mathbf{x} + \mathbf{n}) - I'(\mathbf{y} + \mathbf{n})|^2 d\mathbf{n}}{h^2} \right) \quad (11)$$

where \mathbf{n} scans the neighborhood, h acts as a filtering parameter that can be adapted to the level of image noise and $Z(\mathbf{x})$ is the normalizing factor. Note that the computational cost for the neighborhood search in the entire image can become excessive, so restricting the search to a certain window size ($\mathbf{y} \in S$ in equation (10)) is required [Buades et al., 2008].

2.3. Edge Enhancement

A notorious problem in image processing is partial volume effects due to image blur. That is, image edges do not manifest themselves as crisp intensity steps, but rather as gradual intensity changes spanning several voxels. A standard method to sharpen the image, i.e., to enhance the intensity gradient locally, is unsharp masking [Sheppard *et al.*, 2004]:

$$\hat{I}_{UM}(\mathbf{x}) = I'(\mathbf{x}) - \frac{w(G_{\sigma} * I'(\mathbf{x}))}{1-w} \quad (12)$$

where σ should roughly match the half width of blurry edges and w defines the degree of edge enhancement, where [0.1, 0.9] is a suitable range. In the context of the last section, this corresponds to the inverse diffusion equation. Evidently, unsharp masking will also enhance noise, so the image should be denoised first. Alternative edge enhancement methods like a difference of Gaussians or a Laplacian of Gaussian filter [Gonzalez and Woods, 2002; Russ, 2006] have a very similar concept and are not further discussed here.

2.4. Image Segmentation

2.4.1. Histogram Bias Correction

The frequency distribution of gray values can have an unfavorable shape for threshold detection. Typical examples are low contrast, imbalanced class proportions, class skewnesses, or class variances. Here we list three methods that can partially remove these histogram traits and thus facilitate more reliable threshold estimates:

1. Gradient mask:

Partial volume effects due to blurred phase edges can cause long tailings in the histogram, which lead to skewed distributions for the lowest and highest intensity class. Partial volume voxels can be identified through high-gradient intensities and treated by different strategies [Panda and Rosenfeld, 1978]. One is to calculate the average gray value of partial volume voxels and use it as an optimal threshold [Schlüter *et al.*, 2010], another is to mask them out and only calculate the histogram for low-gradient regions. The mask is generated by unimodal thresholding [Rosin, 2001] of the histogram of intensity gradients.

2. Histogram equalization:

Image contrast is often enhanced by linear or nonlinear intensity rescaling, sometimes also denoted as histogram stretching [Gonzalez and Woods, 2002; Russ, 2006]. An alternative approach to contrast enhancement is contrast-limited adaptive histogram equalization (CLAHE) [Pizer *et al.*, 1987]. Image contrast can be defined as the slope of the cumulative density function of gray values. Limiting this contrast corresponds to clipping the histogram at a certain cutoff. The histogram area thus removed is uniformly distributed over the range of gray values that occur in the image. In principle, this algorithm operates on the histogram of a certain search window to obtain a locally adaptive contrast enhancement. The method can be generalized such that any transition between global and local contrast enhancement is achieved [Stark, 2000]. In this study, histogram clipping is only applied to the global histogram to improve threshold detection and is not mapped to the corresponding image.

3. ROI dilations:

Some images exhibit unimodal histograms due to very imbalanced class proportions, i.e., a very small volume fraction of a certain phase and a very large volume fraction of the background. A balanced frequency distribution can be obtained with a new, semiautomatic algorithm: (i) pick a threshold manually that detects the class with lowest volume fraction as the region of interest (ROI) and binarize the image. (ii) Dilate the thus obtained mask. (iii) Multiply the original image with the mask in order to compute the ROI histogram. (iv) If it is not yet clearly bimodal (multimodal) return to step (ii).

2.4.2. Global Thresholding

Image segmentation is a crucial step in image processing and affects all subsequent image analyses. In this context, it is common to refer to global thresholding as approaches where classes are assigned to voxels by histogram evaluation only, without considering how the gray values are spatially arranged in the corresponding image. A multitude of different thresholding methods exist today which have been reviewed by various authors [Sahoo *et al.*, 1988; Pal and Pal, 1993; Trier and Jain, 1995; Sezgin and Sankur, 2004]. The

general conclusion, if any, is that none of the reviewed methods excel at all segmentation problems. A comprehensive survey by *Sezgin and Sankur* [2004] compared 40 different thresholding methods, most of them global, and classified them according to fundamental principles. Five out of those methods were chosen for implementation in this study as well as an additional one:

1. G1—Maximum Variance:

This is a classic method based on discriminant analysis [*Otsu*, 1975]. Consider the histogram as an estimator of the probability of a certain gray value and $n - 1$ thresholds to divide the histogram into n classes $\{C_0, C_1, \dots, C_n\}$. The total variance σ_T^2 of the population of gray values can be divided into the sum of within-class variances σ_W^2 and the between-class variance σ_B^2 of class means. The objective is to find the set of $n - 1$ thresholds $\{t_0, t_1, \dots, t_{n-1}\}$ that maximize this between-class variance.

2. G2—Minimum Error:

Minimum Error Thresholding assumes the histogram to be composed of normal distributions for each class [*Kittler and Illingworth*, 1986]. The Gaussian modes that are fitted to the histogram usually overlap at certain gray values. As a consequence, assigning those voxels to only one class will deliberately lead to a certain misclassification error for the other class. The objective is to set the thresholds $\{t_0, t_1, \dots, t_{n-1}\}$ such that the misclassification error is minimal.

3. G3—Maximum Entropy:

This method exists in various modifications. The classic approach, which is implemented here, relies on Shannon Entropy as a measure of the information content of a signal [*Kapur et al.*, 1985]. Assume a threshold to be close to the maximum or minimum gray value. Objects will barely appear in the output image and will be almost completely surrounded by background. Hence, the information content of the resulting image is low. A set of thresholds $\{t_0, t_1, \dots, t_{n-1}\}$ can be adjusted such that the image has the richest detail, i.e., the information transfer is optimal. This is achieved by maximizing the sum of class entropies.

4. G4—Fuzzy C-means:

This method combines the classic k -means algorithm [*Ridler and Calvard*, 1978] with fuzzy set theory [*Jawahar et al.*, 1997]. Membership functions M_0, M_1, \dots, M_n are assigned to each gray value depending on the distance to each class mean $\mu_0, \mu_1, \dots, \mu_n$ and a fuzziness index τ . The optimal set of thresholds $\{t_0, t_1, \dots, t_{n-1}\}$ is detected at the intersections of adjacent M .

5. G5—Shape:

This method detects the thresholds $\{t_0, t_1, \dots, t_{n-1}\}$ at the local minimum between two adjacent histogram peaks [*Tsai*, 1995]. If the number of peaks n_p exceeds the predefined number of classes n , iterative Gaussian smoothing is first applied to the histogram, until $n_p = n$. If $n_p < n$, the missing thresholds are set at the location of maximum histogram curvature instead.

6. G6—Average:

It can be shown that some methods are optimal under certain conditions [*Kurita et al.*, 1992], e.g., equal class probabilities, equal class variances, etc. However, these conditions are hardly ever met in a real image and all methods will be biased to some degree. Assuming that the bias of different methods may partly cancel out due to the different criteria that they optimize, an averaged threshold over all methods may lie closer to the true, unknown optimum. Since some methods may fail completely, outliers have to be removed, i.e., only thresholds within $(\bar{t}_k - \sigma_{t_k}, \bar{t}_k + \sigma_{t_k})$ contribute to the final value, where $k = 0, \dots, n - 1$.

Note that in the first three methods the computational cost for an exhaustive search over all sets of thresholds in a n -dimensional search space increases exponentially with increasing number of classes n . Therefore, the methods are implemented in an efficient way by using lookup tables to store each term of a specific objective function for every possible pair of class boundaries, and employ these tables to calculate the objective function for an arbitrary number of classes [*Liao et al.*, 2001]. For the fuzzy c -means, the exhaustive search is replaced by an iterative search [*Jawahar et al.*, 1997]. The search space for the shape method is one-dimensional irrespective of the number of classes.

2.4.3. Local Segmentation

In contrast to global, histogram-based thresholding, locally adaptive segmentation methods also account for some kind of neighborhood statistic for class assignment in order to smooth object boundaries, avoid noise objects, or compensate for local intensity changes. Due to the added flexibility, local segmentation methods often result in more satisfying segmentation results [Iassonov et al., 2009; Wang et al., 2011]. Five different local segmentation methods, which have all been successfully applied to porous media images in the past, will be used in this study:

1. L1—Hysteresis:

Hysteresis segmentation, which is sometimes also denoted as bilevel segmentation or region growing, was introduced as a method to improve edge continuity in gradient images [Canny, 1986]. In the same way, it improves the class assignment of partial volume voxels in soil images [Vogel and Kretzschmar, 1996; Schlüter et al., 2010]. Two thresholds have to be set by the user, a lower threshold that identifies voxels, which definitely belong to a low-intensity class, and an upper threshold for which the uncertainty of class assignment is highest. A voxel in the intermediate gray value range is only assigned to the low-intensity class, if a neighbor voxel already belongs to the low-intensity class. In other words, low-intensity voxels serve as seed regions for iterations of conditional dilations. Unassigned voxels which cannot be accessed by this region growing process are assigned to the high-intensity class. Hysteresis thresholding is not a multiclass segmentation method in its strictest sense as the procedure has to be repeated $n - 1$ times for n classes.

2. L2—Indicator kriging:

In this geostatistical method, spatial correlation is used as a local assignment criterion [Oh and Lindquist, 1999; Houston et al., 2013b]. Again, two thresholds are specified by the user to define two a priori classes. The upper threshold of the unclassified range is extended toward gray values that definitely belong to a high-intensity class. The class assignment at an unclassified location depends on the weighted average of indicator values in its neighborhood, where the weights are obtained by kriging. Indicator kriging has to be repeated $n - 1$ times for n classes as well.

3. L3—Bayesian Markov Random Field:

The rationale of this probabilistic segmentation method is to find a spatial arrangement of class labels \hat{C} with minimum boundary surface that at the same time honors the gray value data in the best possible way [Berthod et al., 1996; Kulkarni et al., 2012]. This is a combinatorial optimization problem which is solved in the framework of Markov random fields (MRF), i.e., by only evaluating the interaction between direct neighbors:

$$\hat{C} = \underset{\hat{c}}{\operatorname{argmin}} \left\{ \underbrace{\int_{\Omega} \left(\sqrt{2\pi\sigma_c^2} + \frac{I_{\mathbf{x}} - \mu_c}{2\sigma_c^2} \right)}_{\text{class statistics}} + \beta \underbrace{\int_{\Pi} \gamma(c_{\mathbf{x}}, c_{\mathbf{y}})}_{\text{class boundaries}} \right\} \tag{13}$$

with

$$\gamma(c_{\mathbf{x}}, c_{\mathbf{y}}) = \begin{cases} -1, & c_{\mathbf{x}} = c_{\mathbf{y}} \\ +1, & c_{\mathbf{x}} \neq c_{\mathbf{y}} \end{cases} \tag{14}$$

where Ω is the population of all voxels, Π is the population of all pairs of neighboring voxels \mathbf{x} and \mathbf{y} , $c_{\mathbf{x}}$ is the class label at \mathbf{x} , μ_c and σ_c^2 are class mean and variance, and β is a homogeneity parameter that determines the weight of the penalty term for class boundaries. Class updating is achieved in a deterministic order denoted with iterative conditional modes (ICM) [Besag, 1986], where sufficient convergence is usually achieved after three to five loops.

4. L4—Watershed:

The watershed algorithm uses lines of highest gradient to demarcate class borders locally [Beucher and Lantuejoul, 1979; Vincent and Soille, 1991; Roerdink and Meijster, 2000]. The image is preclassified with simple

thresholding to set markers for the immersion process. High-gradient zones are subsequently set to unclassified again, as they will be assigned by the watershed algorithm. These high-gradient zones are either identified by edge detection (Sobel, Canny, etc.) or by a small cubic kernel that detects neighborhoods with nonuniform class assignment. The latter method produced the better results for the images examined in this study. The unclassified zones are filled from different ends using the intensity gradient as a pseudo-height until the watershed line is reached. Finally, the voxels located directly on the separation line and on plateaus are filled with the most representative class in their neighborhood.

5. L5—Converging Active Contours:

This approach is a combination of the watershed method and active contour methods, which uses gradient and intensity information simultaneously [Sheppard *et al.*, 2004]. The method is initialized by identifying seed regions for each class; as in the methods above, this means that each voxel is either assigned to one of the classes or left unassigned, to be classified during the main segmentation step. Like the watershed method described above, this initial classification is normally done by thresholding in both intensity and gradient space; each class has an upper and a lower threshold along with an upper gradient threshold. The main classification algorithm proceeds by simultaneously growing the boundaries of these seed regions toward each other. The speed at which the boundaries advance varies spatially and temporally, depending on the local gradient and optionally on the distance of the local gray value to its class mean. The algorithm ends when all boundaries have converged. The final separation line assignment problem described above that affects some watershed methods is not an issue because every nonseed voxel is traversed by one class boundary before any others. The advancement of the boundaries is implemented efficiently using the fast marching algorithm; despite this, the method is quite computationally intensive and requires parallel implementations to operate on large 3-D images.

2.5. Postprocessing

Denoising and boundary refinement can also be applied to the class image C as postprocessing instead of smoothing the raw image I' . Among the most popular postprocessing methods for binary images are morphological operators like erosion and dilation [Serra, 1982]. However, in multiclass images they would correspond to a minimum and maximum filter, which are not suited for this purpose at all. A median filter, on the other hand, removes segmentation noise and rugged class boundaries rather well. It does however, have the unwanted feature that the results happen to depend on the order in which the classes are numbered. This is avoided by applying a majority filter \hat{C}_{MA} , which assigns the most representative class among all neighbors in a cubic kernel to the central voxel. It is common to enforce two criteria to prevent unwanted features where more than two phases meet: (i) the most representative class exceeds a certain volume fraction and (ii) the volume fraction of the most representative class exceeds the volume fraction of the old class at the central voxel by a certain percentage.

An alternative postprocessing method is size-dependent object removal. For this method, each object of each class has to be labeled. Volumes are then determined by voxel counting, and objects smaller than a user-dependent threshold are filled with the class that completely surrounds the object. Consequently, rugged surfaces of big objects are conserved. Small deleted objects at the boundary between two materials are filled by simultaneous dilation of both materials.

2.6. Structural Analysis

The list of tools with which complex structures like porous media can be analyzed is virtually endless. This study is constrained to only four very simple, but meaningful metrics. If ground-truth information is available from a true image I , a misclassification error ME can be determined:

$$ME = \frac{1}{N_x} \sum_{i=1}^{N_x} \delta(c_i, \hat{c}_i), \quad \delta(c_i, \hat{c}_i) = \begin{cases} 1, & c_i \neq \hat{c}_i \\ 0, & c_i = \hat{c}_i \end{cases} \quad (15)$$

where c_i and \hat{c}_i are the true and current class at location i and N_x is the number of all voxels. Bulk volumes V and surface areas a are examined for each individual phase. Specific interfacial areas between two phases (a and b) are obtained by the following relation:

$$a_{ab} = \frac{1}{2}(a_a + a_b - a_c) \tag{16}$$

where c is the union of all other classes. Both V and a are determined with Minkowski functionals [Vogel et al., 2010]. In this way, surface area estimates are directly obtained from the segmented voxel image and problems associated with aligning isosurfaces of each individual phase are avoided. Finally, a dimensionless connectivity indicator for a specific phase is calculated [Renard and Allard, 2013]:

$$\Gamma_a = \frac{1}{N_a^2} \sum_{i=1}^{N_i} n_i^2 \tag{17}$$

where each cluster of phase a has a label l_i and a size n_i , N_i is the number of clusters and N_a is the number of all phase a voxels. Cluster labeling is typically achieved with a fast method by Hoshen and Kopelman [1976]. Γ_a has the edge over the popular Euler number that it is bounded by (0, 1] and less sensitive to noise.

3. Results

3.1. Synthetic Test Image

3.1.1. Image Generation

Synthetic images provide ground-truth data against which different image enhancement and segmentation methods can be compared. The test image in this study is generated to resemble a partially saturated sand packing. To this end, a volume of $512 \times 512 \times 128$ voxels is filled by nonoverlapping spheres with a radius of $r = 50$ voxels placed at random locations. If a new sphere does not fit at a given location, the radius decreases in steps until the overlap vanishes. The location is completely abandoned when $r = 20$ is reached. The procedure stops after 250 spheres have successfully been placed resulting in a porosity of 0.412. Subsequently, an opening transform [Serra, 1982] is applied to the pore space with a spherical structure element of $r = 15$. The pore space which has been removed by this opening is considered to be filled with the wetting phase (w) and the remaining, larger pore bodies are assigned to the nonwetting phase (n). This does not necessarily produce a physically realistic fluid distribution, but evokes fluid configurations that are similar enough to serve as a test scenario. The resulting volume fractions are $V_w = 0.279$ and $V_n = 0.133$. Gray values of 50, 125, and 205 are assigned to nonwetting, wetting and solid phase, respectively, resulting in the true image I . An axial image slice of I is depicted in Figure 1a.

Subsequently, the image is superimposed with ring artifacts, blur, and noise in order to obtain a more realistic, raw image I' (Figure 1b). The ring artifacts are additive with random height and random radius. The rings manifest themselves as a random gray value offset in the range $[-100,100]$, where the resulting gray value is limited at $[0,255]$. Subsequently, blurring is achieved with a cubic mean kernel of five voxels on a side. Finally, uncorrelated Gaussian noise is added with a signal-to-noise ratio of $\sigma_b/\sigma_n = 2$, where σ_b and σ_n are the standard deviations of gray values in the blurred image and the noise model, respectively.

3.1.2. Denoising and Edge Enhancement

Four different denoising methods (Figures 1d–1g) are applied to the raw image I' in Figure 1b in order to remove noise followed by an edge enhancement step to mitigate partial volume effects caused by blur. The median filter was applied one time with a cubic convolution kernel with a side length of $d = 7$ voxels. The anisotropic diffusion filter was applied with $\kappa = 15$, $\sigma = 1.0$, and $t = 1.5$ (nine iterations). The total variation filter was applied with $\lambda = 2$ stopping after $t = 0.12$ (30 iterations). For the nonlocal means filter, a cubic kernel of $d = 9$ was used for neighborhood search with a Gaussian convolution kernel of $\sigma = 2$ in a reduced search window S of $d = 23$ voxels. The two criteria for the majority filter are set to an absolute majority of 0.33 and a higher occurrence than the class of the central voxel of 0.05.

The settings of each denoising method have been carefully chosen to achieve strong denoising, deliberately accepting some degree of edge smoothing. Subsequently, an unsharp mask filter for edge enhancement with a Gaussian kernel of $\sigma = 1.5$ and a weighting factor of $w = 0.7$ was applied to each denoising result. Finally, the results of the unsharp mask after anisotropic diffusion \hat{I}_{AD+UM} and total variation

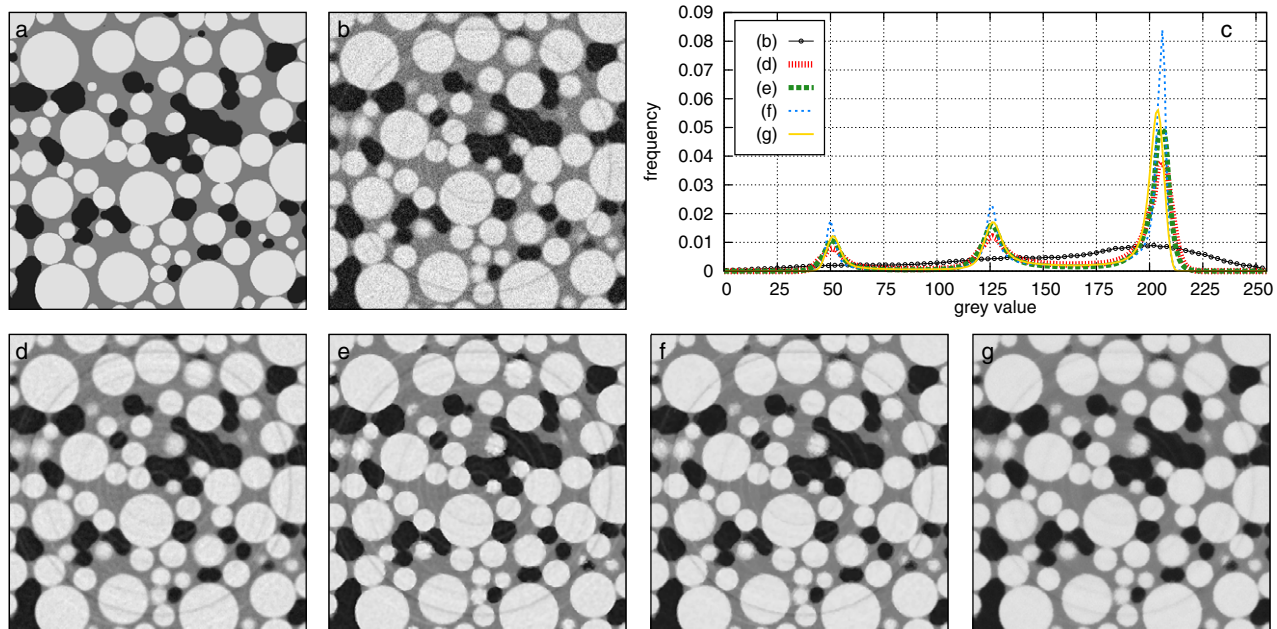


Figure 1. (a) True image I with spheres and size-dependent distribution of wetting and nonwetting phase, (b) raw image I' exhibiting ring artifacts, blur and noise, (c) histograms before and after image enhancement. (bottom row) Image enhancement results with different denoising methods in combination with unsharp masking: (d) \hat{I}_{MD+UM} , (e) \hat{I}_{AD+UM} , (f) \hat{I}_{TV+UM} , and (g) \hat{I}_{NL+UM} .

denoising \hat{I}_{TV+UM} had to be cleaned by a median filter with a $d = 5$ kernel to smooth rough surfaces that were enhanced during unsharp masking.

The histograms in Figure 1c and visual appearances in Figures 1d–1g indicate different success in noise removal and edge preservation. The edges in the median image \hat{I}_{ME+UM} appear slightly more blurry than in the images for the other three denoising methods. Moreover, we observe that the nonlocal means filter \hat{I}_{NL+UM} is superior in removing ring artifacts during the denoising step alone and does not require the additional cleanup of rough edges by a median filter mentioned above.

3.1.3. Ring Artifact Removal

The ring artifact routine of *Sijbers and Postnov* [2004] is applied slice wise with a window size of $w = 27$ pixels and a homogeneity threshold H adapted to the noise level in Figures 1d–1g. The outcome is depicted for the enhanced image \hat{I}_{TV+UM} after TV denoising and unsharp mask edge enhancement (Figure 2c). Obviously, the ring artifact removal works well for some rings, but is incomplete for others even after carefully testing different w and H . For instance, the white ring within the green window exhibits a varying magnitude after denoising (Figure 2a). Since the high artifact magnitude in the pore space contributes less than

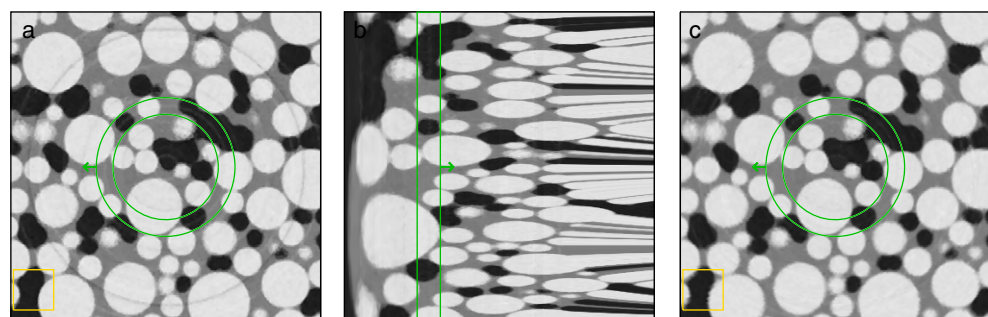


Figure 2. (a) Test image after a combination of TV denoising and unsharp masking \hat{I}_{TV+UM} (corresponds to Figure 1f). Yellow frame marks sharp boundaries before ring artifact removal and the green ring corresponds to the window in polar coordinates. (b) Same image in polar coordinates after line removal. Green rectangle depicts the moving window W . (c) Same image after ring artifact removal. The yellow frame highlights blur due to the back transform into Cartesian coordinates.

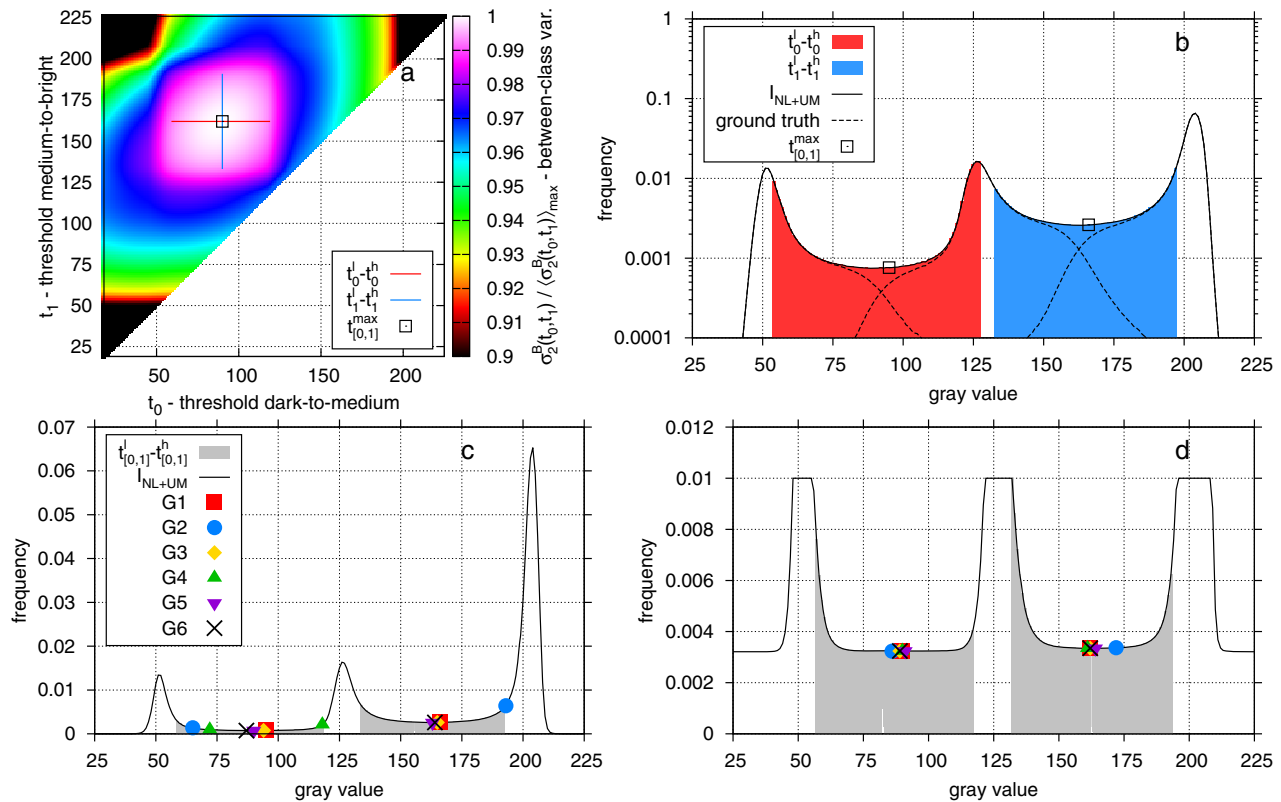


Figure 3. (a) normalized objective function of Otsu's method (G1) with the thresholds detected at the maximum between-class variance and the transition regions stopping at the 0.992 percentile. (b) Histogram of I_{NL+UM} after ring artifact removal with thresholds t_0^{max} and t_1^{max} and transition ranges $([t_0^l, t_0^h])$ and $([t_1^l, t_1^h])$ detected with Otsu's method (G1). (c) Threshold pairs for five different methods (G1–G5) and the corresponding averages after outlier removal (G6). The transition regions $[t_{[0,1]}^l, t_{[0,1]}^h]$ are obtained accordingly. (d) Same averaging method after edge masking and histogram clipping.

half to the window height in Figure 2b, it remains undetected by the median $\tilde{l}(r)$. Stretching the window over more than one slice as suggested by *Ketcham* [2006] does not improve the results. In addition, the back transform from polar into Cartesian coordinates introduces additional blur for high radii (yellow frame).

3.1.4. Global Thresholding

In Figure 3a, the well-known Otsu method (G1) is applied as an example to the histogram of \hat{I}_{NL+UM} after ring artifact removal. Results for the other denoising methods are similar (not shown). The exhaustive search for each pair of thresholds detects the optimum at $t^{max} = \{95; 166\}$. Percentiles of this objective criterion can be used to obtain fuzzy threshold ranges [Oh and Lindquist, 1999], i.e., t_1 is held constant and t_0 decreases (or increases) until in this case $\sigma_b^2 / \langle \sigma_b^2 \rangle_{max} = 0.992$ is reached at t_0^l (or t_0^h). The procedure is repeated for t_1 where t_0 kept is constant. Applying these thresholds and associated ranges to the underlying histogram (Figure 3b) reveals a bias of Otsu's method toward the class with highest volume fraction. Note that the true optimum, which is estimated by the intersection points of the individual gray value frequencies within in each class of the true image, is at $t^{true} = \{91; 162\}$. Some other thresholding methods are also afflicted by bias, yet in different directions (Figure 3c). Fuzzy *c*-means (G3) show the same bias due to imbalanced class probabilities. The shape analysis (G5) performs well since the local histogram minima happen to coincide with the intersection points in Figure 3b. The minimum error method (G2) is biased due to imbalanced skewness, i.e., only the wetting phase exhibits two long tailings due to partial volume effects in both directions, which increases the probability of being assigned to that class. The maximum entropy method (G4) fails completely in identifying meaningful thresholds. Figure 3c also depicts the arithmetic mean of thresholds and associated ranges after outlier removal (G6). This average set of thresholds at $t = \{88; 165\}$ is closer to the true optimum. Note that the specific percentiles for each thresholding method have to be set individually, as each objective criterion exhibits very different ranges, i.e., some are bounded and yet others are logarithmic. A large part of the bias in Figure 3c can be removed easily. First, the impact of imbalanced

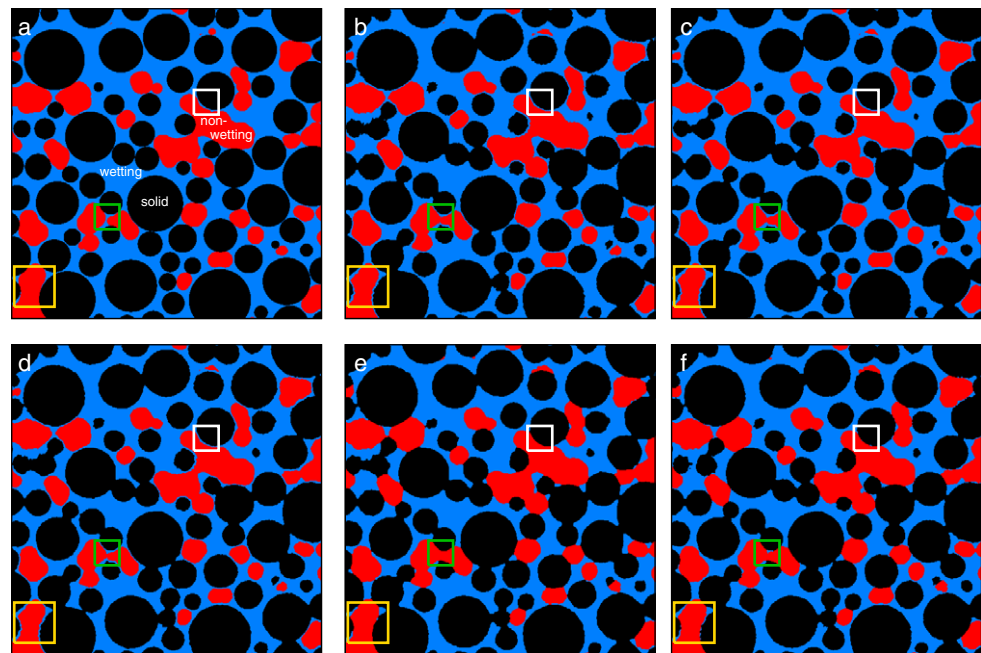


Figure 4. Segmentation results for different local segmentation methods applied on the image after nonlocal means denoising and unsharp masking \hat{I}_{NL+UM} : (a) original image, (b) global thresholding, (c) indicator kriging, (d) Bayesian MRF, (e) watershed, and (f) converging active contours. Differently colored frames highlight failures of various methods.

skewness is mitigated by masking out partial volume voxels at phase edges [Panda and Rosenfeld, 1978]. Second, the impact of imbalanced class probabilities is reduced by histogram clipping [Pizer et al., 1987]. In this way, the skewness of the nonwetting and solid phase is also efficiently removed. The combination of both methods leads to a well-balanced histogram for which each of the surveyed thresholding methods ends up at similar values (Figure 3d). The average after outlier removal at $t = \{90; 162\}$ matches the true optimum almost perfectly.

3.1.5. Local Segmentation

The average set of thresholds (Figure 3d) are determined for each denoising method individually and then used for global thresholding as an initial step in all local segmentation methods. Note that a priori segmentation with Bayesian Markov random field (L3) and watershed segmentation (L4) requires single thresholds t_k^{\max} only, hysteresis segmentation (L1) uses the range $[t_k^l, t_k^{\max}]$, and indicator kriging (L2) and converging active contours (L5) use $[t_k^l, t_k^h]$, with $k = \{0; 1\}$. In addition, converging active contours requires a gradient threshold, which is determined automatically by unimodal thresholding of the gradient histogram.

The results of each local segmentation method on \hat{I}_{NL+UM} are depicted in Figure 4. The segmentation result for hysteresis segmentation (L1) is left out, because it is indistinguishable from the outcome of global thresholding (G6). All segmentation results are free of ring artifacts and noise and exhibit smooth object boundaries which can be mainly credited to successful preprocessing. However, some segmentation results perform better than others in the way they recover image features impaired by blur. Incorrectly identified, apparent wetting films that are due to partial volume voxels at class boundaries between the nonwetting and solid phases (yellow frame) can be mainly attributed to image blur. Indicator kriging (L2) and hysteresis segmentation (L1) cannot cope with this problem, since these are iterative methods, i.e., the segmentation between wetting and nonwetting phase is independent of the segmentation between wetting and solid phase.

In principle, Bayesian MRF segmentation (L3) has a mechanism to remove these films. To do so, the homogeneity factor β has to be set high, thus also removing true features of similar size. A moderate value of $\beta = 0.5$ is used here. Yet, no matter how high β is set only direct neighbors are evaluated in the MRF paradigm and since the films are thicker than two voxels they cannot be penalized. Another fundamental shortcoming of Bayesian MRF segmentation that has previously gone unnoticed is its tendency toward bias in the class statistics (equation (13)). The histogram of the wetting phase exhibits long tailing toward both

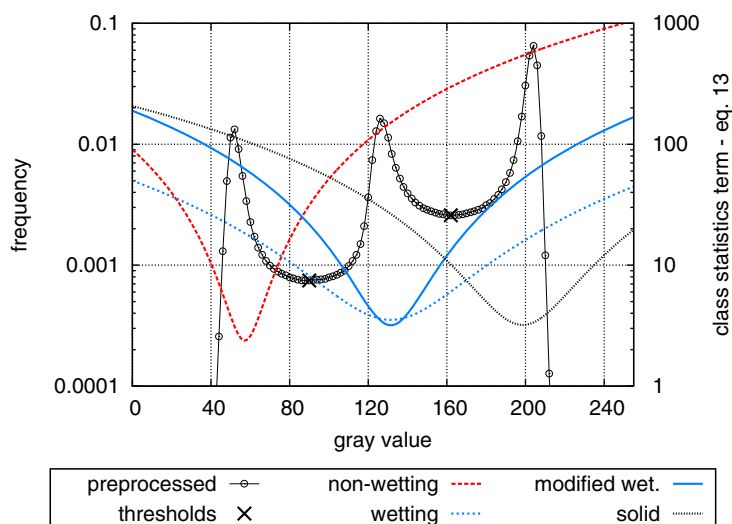


Figure 5. First term of the objective function for MRF segmentation (equation (13)) as a function of gray value for each class. In the modified class statistics, σ_w is replaced by $\sigma_w/2$ for the wetting phase.

directions in contrast to only one tail for the other classes (Figure 3c). Consequently, σ_w is much higher than σ_n and the probability of a voxel in the intermediate intensity range to be assigned to the wetting phase increases accordingly, resulting in even thicker apparent wetting films. This bias is mitigated here by a quick and rather inelegant fix to use $\sigma_w/2$ for the class statistics instead. In this way, the gray values at which the penalty functions for adjacent classes cross each other coincide much better with the previously detected thresholds (Figure 5).

The watershed method (L4) successfully removes apparent wetting films in the yellow frame (Figure 4e). However, true wetting films are also being removed (white frame). Segmentation with converging active contours (L5—Figure 4f) exhibit excessive wetting film removal as well. Finally, no method is capable of restoring the small wetting phase bridge (green frame) in Figure 4, because its intensity is highly smoothed due to image blur.

3.1.6. Structural Analysis

The shortcomings of the different segmentation methods are corroborated by structural properties of the nonwetting phase listed in Table 1. The misclassification error ME is roughly 3–4% for all methods including global thresholding. However, the lowest ME does not guarantee the best recovery of morphological properties. The bulk volume V_n of the nonwetting phase is slightly underestimated by all segmentation methods, except for watershed segmentation and converging active contours which show a tendency for overestimation, due to the way partial volume voxels are treated. Surprisingly, simple thresholding and hysteresis thresholding match the true a_n value best, whereas indicator kriging and Bayesian MRF underestimate it and watershed segmentation and converging active contours overestimate it as a consequence of partial volume voxel treatment. More importantly, the specific interface between fluids a_{wn} only diverges from the total nonwetting surface a_n if false wetting films are successfully suppressed. Simple thresholding, hysteresis thresholding, indicator kriging, and Bayesian MRF segmentation all fail in this respect. The only exception is Bayesian MRF segmentation on the raw data if combined with postprocessing \hat{C}_{MA} , simply because the majority filter removes partial volume voxels. Watershed segmentation and converging active contours underestimate a_{wn} because a lot of true wetting films are removed as well. The connectivity index of the true image Γ_n is matched well by most segmentation methods, because the bubbles remain rather isolated no matter how wetting films are treated. Oversegmentation of the nonwetting phase with watershed segmentation and converging active contours may lead to lower Γ_n , however, because (i) the denominator N_a in equation (17) increases and (ii) the increase in bubble volume is evenly distributed among all clusters in the numerator. The differences in structural properties due to different denoising methods are also listed in Table 1. Obviously, no denoising at all produced the worst results in all respects. Moreover, the structural properties do not vary much among the different denoising methods. Surprisingly, a simple majority filter \hat{C}_{MA} , applied on the segmented raw data without any preprocessing results in the best agreement with the

Table 1. Structural Properties of Segmented Images for Different Combinations of Denoising and Segmentation Methods: Unsharp Mask (UM), Median (MD), Anisotropic Diffusion (AD), Total Variation (TV), Nonlocal Means (NL), Majority (MA)^a

Denoising	ME (-)	V_n (-)	a_n (pix ⁻¹)	a_{nw} (pix ⁻¹)	Γ_n (-)
<i>True Image</i>	0.000	0.133	0.0169	0.0140	0.355
<i>G6—Global Thresholding</i>					
I'	0.140	0.141	0.1005	0.0954	0.312
\hat{I}_{MD+UM}	0.028	0.128	0.0168	0.0167	0.355
\hat{I}_{AD+UM}	0.029	0.124	0.0165	0.0165	0.355
\hat{I}_{TV+UM}	0.027	0.127	0.0169	0.0169	0.355
\hat{I}_{NL+UM}	0.027	0.130	0.0172	0.0170	0.354
\hat{C}_{MA}	0.024	0.129	0.0169	0.0168	0.355
<i>L1—Hysteresis</i>					
I'	0.107	0.131	0.0722	0.0695	0.346
\hat{I}_{MD+UM}	0.027	0.127	0.0167	0.0167	0.355
\hat{I}_{AD+UM}	0.029	0.123	0.0164	0.0164	0.355
\hat{I}_{TV+UM}	0.027	0.127	0.0169	0.0169	0.355
\hat{I}_{NL+UM}	0.027	0.129	0.0171	0.0170	0.354
\hat{C}_{MA}	0.024	0.128	0.0167	0.0167	0.355
<i>L2—Indicator Kriging</i>					
I'	0.037	0.131	0.0202	0.0201	0.386
\hat{I}_{MD+UM}	0.033	0.128	0.0166	0.0166	0.351
\hat{I}_{AD+UM}	0.037	0.122	0.0162	0.0162	0.353
\hat{I}_{TV+UM}	0.032	0.125	0.0165	0.0165	0.352
\hat{I}_{NL+UM}	0.033	0.127	0.0166	0.0166	0.352
\hat{C}_{MA}	0.029	0.131	0.0167	0.0167	0.352
<i>L3—Bayesian MRF ($\beta = 0.5$)</i>					
I'	0.049	0.133	0.0282	0.0272	0.385
\hat{I}_{MD+UM}	0.027	0.124	0.0163	0.0163	0.356
\hat{I}_{AD+UM}	0.033	0.119	0.0160	0.0160	0.357
\hat{I}_{TV+UM}	0.028	0.124	0.0165	0.0165	0.356
\hat{I}_{NL+UM}	0.028	0.127	0.0166	0.0165	0.356
\hat{C}_{MA}	0.038	0.134	0.0170	0.0140	0.356
<i>L4—Watershed</i>					
I'	0.041	0.136	0.0184	0.0112	0.358
\hat{I}_{MD+UM}	0.029	0.135	0.0172	0.0110	0.355
\hat{I}_{AD+UM}	0.025	0.134	0.0170	0.0117	0.355
\hat{I}_{TV+UM}	0.029	0.135	0.0174	0.0111	0.355
\hat{I}_{NL+UM}	0.032	0.137	0.0175	0.0102	0.322
\hat{C}_{MA}	0.037	0.136	0.0174	0.0101	0.359
<i>L5—Converging Active Contours</i>					
I'	0.046	0.136	0.0241	0.0178	0.390
\hat{I}_{MD+UM}	0.042	0.136	0.0173	0.0108	0.324
\hat{I}_{AD+UM}	0.042	0.137	0.0172	0.0107	0.324
\hat{I}_{TV+UM}	0.042	0.136	0.0174	0.0110	0.324
\hat{I}_{NL+UM}	0.043	0.138	0.0175	0.0108	0.324
\hat{C}_{MA}	0.043	0.136	0.0175	0.0105	0.360

^aME is misclassification error, V_n is volume fraction of nonwetting phase, A_n is surface area density of nonwetting phase, A_{nw} is surface area density between nonwetting and wetting phase, and Γ_n is connectivity index of nonwetting phase.

true bulk properties. This is because any kind of preprocessing, i.e., denoising and edge enhancement, removes structural information to some degree.

3.1.7. Final Workflow

At this point, some preliminary conclusions about a suitable image processing protocol can already be drawn in order to set the workflow for the remaining images:

1. The variability in structural properties among the different denoising methods was rather small. Qualitatively, all filters reduce noise in homogeneous areas well. In addition, only the nonlocal means filter sufficiently denoises along edges while smoothing across the edges is inhibited. In this way, an additional treatment of rugged surfaces after edge enhancement is not necessary. Thus, in the remainder of the paper, only the nonlocal means filter will be used.
2. A combination of methods for histogram bias correction causes a better agreement between different threshold detection methods. The outlier-corrected average of five different threshold detection methods reproduced ground-truth information well and will be used for global thresholding.

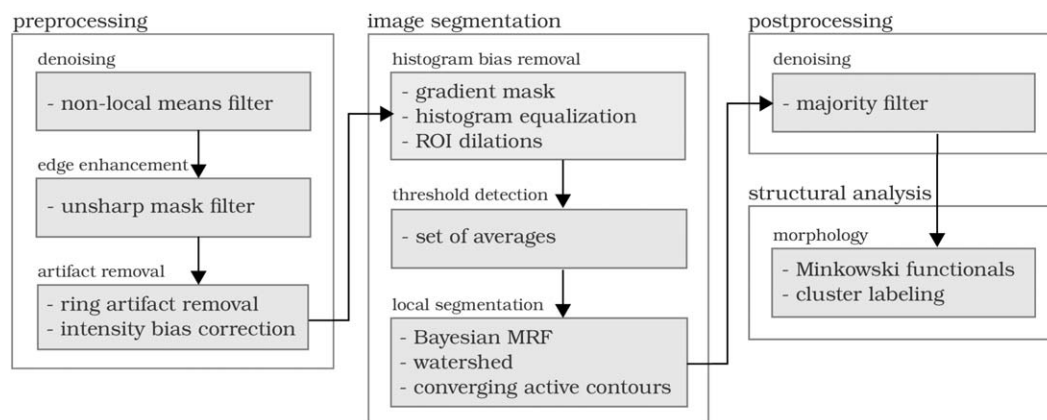


Figure 6. Workflow diagram for multiclass segmentation of the remaining multifluid image and soil image. Headlines denote image processing steps and gray boxes the specific methods.

3. Hysteresis thresholding and indicator kriging are not capable of multiclass segmentation, which leads to misclassification errors where more than two phases meet locally. Therefore, only converging active contours, watershed and Bayesian MRF segmentation will be applied to the remaining images. Simple thresholding will also be applied for sake of comparison.

4. Segmentation of the raw data with a subsequent majority filter performed slightly better in terms of structural properties than denoising the intensity data in advance. Both approaches will be compared with each other.

These findings translate into the workflow diagram shown in Figure 6. Some steps are optional, e.g., the need for edge enhancement depends on the sharpness of the raw image, ROI dilations are only necessary, if the volume fraction of the phase of interest is very low, etc.

3.2. Synchrotron Image of Three-Fluid Phases in a Porous Medium

3.2.1. Image Enhancement

A sample of sintered glass beads with a porosity of roughly 32% has been scanned at a resolution of 9.24 μm . The pore space was partially saturated with air (14%), oil (39%), and water (47%). The cylindrical region of interest has a diameter of 665 voxels and height of 210 voxels. The image is free of ring artifacts and is sharp, without noticeable blur (Figure 7a). The reconstruction method caused a slight decrease in mean intensity for large radii which is evident in the denoised image after nonlocal means denoising (Figure 7b). This radial intensity variation can be removed almost completely (Figure 7c) with intensity bias correction [Jassonov and Tuller, 2010]. This has a considerable impact on the histogram (Figure 7d). The frequency distributions for each class turns from broad and multi-peaked into a narrow band. For the same reasons as discussed for the synthetic test image, edges are masked out and the histogram is clipped into a well-balanced frequency distribution, so that five different global threshold detection methods (G1–G5) yield very similar sets of thresholds. Again, the average after outlier removal (G6) is used for final segmentation using local methods. Note that edge enhancement with unsharp mask filtering has not been applied, as there are hardly any partial volume voxels in this image.

3.2.2. Image Segmentation

Simple thresholding (G6), Bayesian MRF segmentation (L3), watershed segmentation (L4), and converging active contours (L5) are applied to the multifluid image, either on the preprocessed \hat{I}_{NL} or on the raw I followed by a majority filter. A small subset of the segmentation results are depicted in Figure 8. Since the exact arrangement of interfaces is unknown, it is difficult to judge objectively what combination of methods performs best. The edge between beads and air is roughly two voxels thick. In the smooth \hat{I}_{NL} image, the edge is assigned to oil films, water films or both (Figures 8d, 8f, 8h, and 8j), whereas in the noisy image edge voxels are equally assigned to all four classes but subsequently assigned to air or beads since they constitute the most representative class in the neighborhood of an edge (Figures 8c, 8e, 8g, and 8i). It is almost impossible to conclusively determine whether a fluid film thicker than the image resolution really covers the beads entirely. However, small isolated water voxels between solid voxels and oil film are highly unlikely, and should be suppressed. Global thresholding without postprocessing has no mechanism to

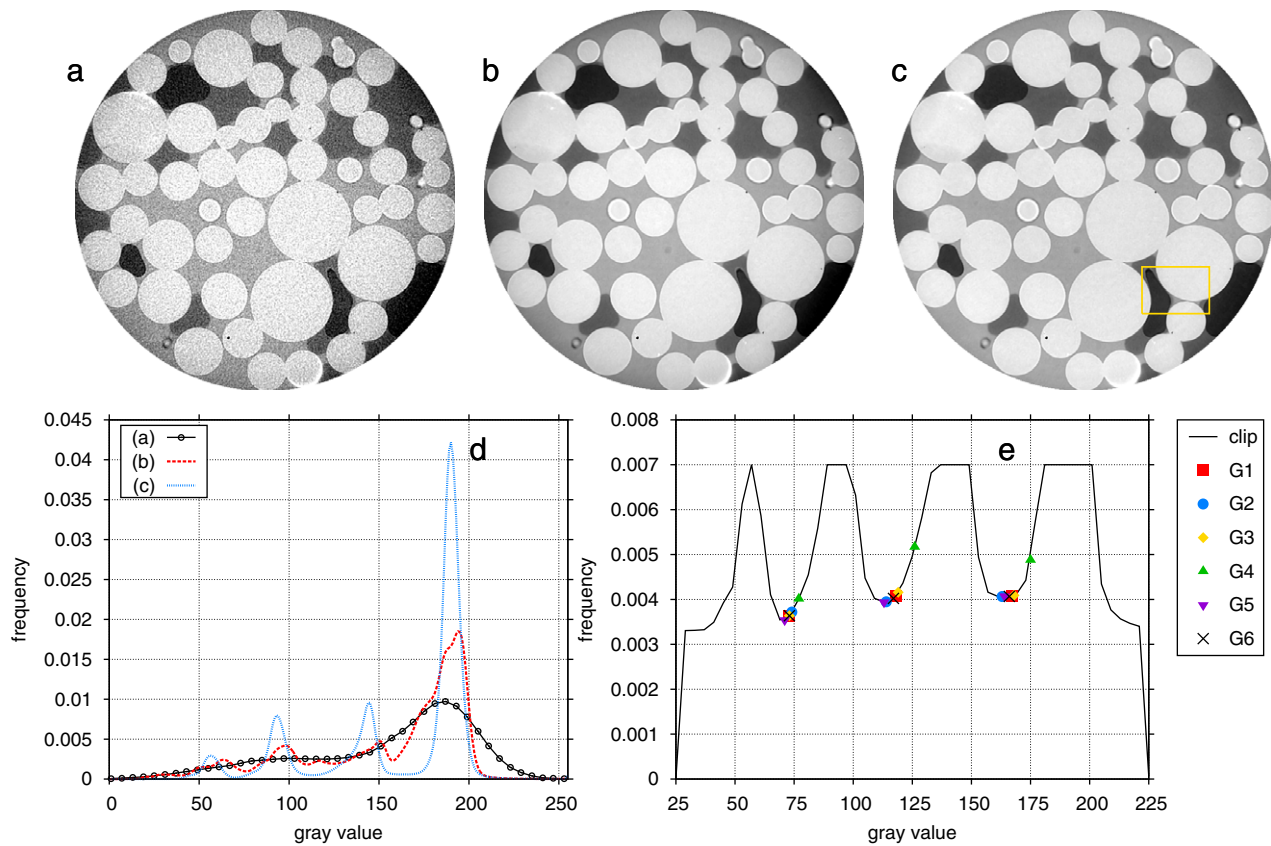


Figure 7. Image enhancement of multifluid image: (a) raw image I' , (b) nonlocal means filter \hat{I}_{NL} on I' , (c) beam hardening removal on \hat{I}_{NL} . The yellow frame outlines the subset in Figure 8. (d) Histograms of Figures 7a–7c. (e) Global threshold detection after histogram clipping.

achieve this. Bayesian MRF segmentation of \hat{I}_{NL} only succeeded in doing so with a relatively high homogeneity parameter ($\beta = 10$). In turn, watershed segmentation fails to detect an oil film at the left side of the pendular water ring, either because of a lacking seed voxel for oil or because the underlying gradient image is not sharp enough to evoke two distinct edges within such a short distance. Converging active contours result in subjectively plausible results for both I' and \hat{I}_{NL} .

3.2.3. Image Analysis

The qualitative analysis illustrated in Figure 8 is corroborated by the results with respect to structural analysis in Table 2. The bulk volumes of air and oil, V_a and V_o , remain largely unaffected by the choice of

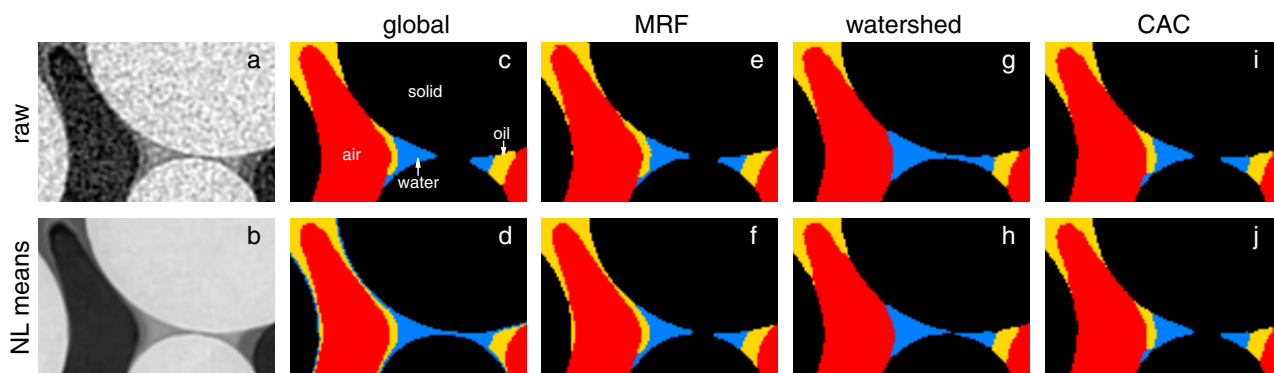


Figure 8. Segmentation results for multifluid image: (a) raw image I' , (b) nonlocal means filter \hat{I}_{NL} , global thresholding on (c) I' with postprocessing or (d) on \hat{I}_{NL} , Bayesian MRF segmentation (e) with $\beta = 0.1$ on I' with postprocessing or (f) with $\beta = 10$ on \hat{I}_{NL} , watershed segmentation (g) on I' with postprocessing or (h) on \hat{I}_{NL} , converging active contours (i) on I' with post-processing or (j) on \hat{I}_{NL} .

Table 2. Volume Fractions V and Surface Area Densities a of Air (a) and Oil (o) for the Segmented Multifluid Image^a

Denoising	V^a (-)	a_a (mm^{-1})	V_o (-)	a_o (mm^{-1})	a_{ao} (mm^{-1})
<i>G6—Global Thresholding</i>					
\hat{C}_{MA}	0.047	0.400	0.125	1.621	0.120
\hat{I}_{NL}	0.046	0.545	0.125	2.281	0.384
<i>L3—Bayesian MRF</i>					
$\hat{C}_{MA} (\beta = 0.1)$	0.047	0.482	0.125	1.840	0.215
$\hat{I}_{NL} (\beta = 0.1)$	0.046	0.505	0.124	2.214	0.369
$\hat{I}_{NL} (\beta = 1)$	0.046	0.459	0.125	2.063	0.327
$\hat{I}_{NL} (\beta = 10)$	0.046	0.397	0.125	1.686	0.149
<i>L4—Watershed</i>					
\hat{C}_{MA}	0.048	0.403	0.127	1.646	0.107
\hat{I}_{NL}	0.048	0.406	0.127	1.631	0.109
<i>L5—Converging Active Contours</i>					
\hat{C}_{MA}	0.047	0.421	0.121	1.642	0.130
\hat{I}_{NL}	0.048	0.407	0.123	1.650	0.119

^aDenoising is either applied prior to segmentation with a nonlocal means filter (\hat{I}_{NL}) or as postprocessing with a majority filter (\hat{C}_{MA}).

segmentation methods. The surface area densities a_o , a_a , and a_{ao} , however, vary considerably among different segmentation methods. This is because small image objects like films, meniscii, and ganglia exhibit large surface-to-volume ratios and at the same time are associated with the highest uncertainty in terms of class assignment. A majority filter \hat{C}_{MA} leads to a general reduction in surface areas and to very similar results for all segmentation methods. The surface areas increase with a decreasing boundary penalty factor β for Bayesian MRF segmentation on \hat{I}_{NL} and are highest for global thresholding on \hat{I}_{NL} , where there is no penalty term at all. The a_o and a_a values for $\beta = 10$ are very similar to all postprocessed class images \hat{C}_{MA} , to the outcome of converging active contours on \hat{I}_{NL} , and to the result after watershed segmentation on \hat{I}_{NL} . Yet, the specific surface area between air and oil ($a_{ao} = 0.149$) is as much as $\approx 27\%$ higher than the average of other methods ($a_{ao} = [0.107 - 0.130]$). Eventually, a decision as to which image is closer to reality needs to be made. But without an analytical solution or other ground-truth information such a decision will suffer from a certain level of subjectivity.

3.3. μ CT Image of Soil

3.3.1. Image Enhancement

The soil image corresponds to image 2 in *Houston et al.* [2013b] with a size of $256 \times 256 \times 256$ voxels and a resolution of $32 \mu\text{m}$. The sample consists of macropores ($V_p \approx 12\%$), organic matter ($V_o \approx 13\%$), soil matrix ($V_m \approx 72\%$), and dense particles like rocks ($V_r \approx 3\%$). The raw data I' (Figure 9a) is again denoised with a nonlocal means filter \hat{I}_{NL} (Figure 9b). Afterward, edges are enhanced by an unsharp mask \hat{I}_{NL+UM} with $\sigma = 1$ and $w = 0.5$ so that intensity values of partial volume voxels are forced closer to their respective class means (Figure 9c). The gradient mask in Figure 9d detects partial volume voxels at phase edges which are excluded from subsequent threshold detection. Each of these methods has a favorable impact on the intensity histogram (Figure 9e) in that valleys between the class peaks are much more pronounced. The combination of all methods, i.e., edge masking on \hat{I}_{NL+UM} , together with histogram clipping leads to a modified histogram for which all five threshold detection methods (G1–G5) lead to similar results (Figure 9e). The average after outlier removal (G6) is again used for all locally adaptive segmentation methods.

3.3.2. Image Segmentation

Only some of the segmentation methods introduced in this paper are applied to the soil image. Global thresholding of the raw data I' in combination with a majority filter \hat{C}_{MA} (Figure 10a) shall serve as a reference to which the other methods can be compared. Evidently, simple thresholding already leads to rather satisfying results if it is accompanied by suitable postprocessing. However, the segmented image clearly suffers from false organic coatings around macropores which can be attributed to incorrect assignment of partial volume voxels (violet frame). This misclassification of boundary voxels can be avoided with watershed segmentation (Figure 10b) and converging active contours (Figure 10c) both applied to denoised and edge-enhanced image \hat{I}_{NL+UM} . In addition, even thin macropores are correctly detected (green frame). Bayesian MRF segmentation is applied to \hat{I}_{NL+UM} with $\beta = 0.1$, $\beta = 1$, and $\beta = 10$,

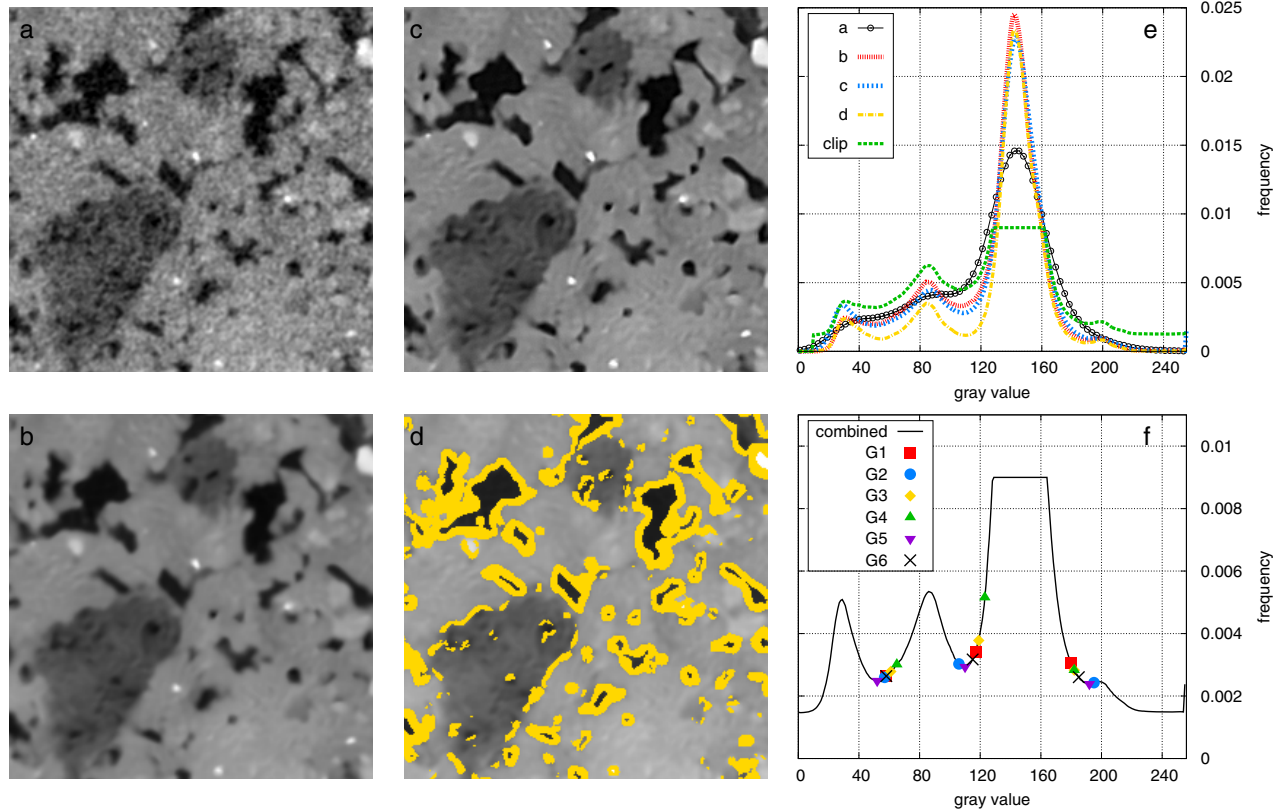


Figure 9. Image enhancement of a soil image: (a) raw image I' [Houston et al., 2013b], (b) \hat{I}_{NL} after nonlocal means denoising, (c) \hat{I}_{NL+UM} after unsharp mask, (d) gradient mask on \hat{I}_{NL+UM} , (e) histogram of Figures 9b–9d and histogram clipping of Figures 9b. (f) Histogram after combined postprocessing (b + c + d + clip) together with the corresponding thresholds obtained by various global thresholding methods.

respectively (Figures 10d–10f). We observe that the segmented images look very different for different β . If the penalty term for class boundaries (second term of equation (13)) has a low weight ($\beta = 0.1$), the image looks very similar to (Figure 10a), i.e., all image objects are well preserved, yet all macropores exhibit false coatings of organic matter. In turn, if the homogeneity parameter is set very high ($\beta = 10$), partial volume effects are suppressed and so are small image objects in general, like the thin macropore in the green frame. A moderate value ($\beta = 1$) leads to an unsatisfactory trade-off between the two problems.

3.3.3. Image Analysis

The qualitative interpretation is again corroborated by structural properties of macropores and organic matter summarized in Table 3. Bulk volume and surface area vary much more among different segmentation methods as compared to the multifluid image. For instance, the surface area density a_{op} between organic matter and macropores decreases by a factor of 2.5 if the homogeneity parameter for Bayesian MRF segmentation is increased from $\beta = 0.1$ (Figure 10d) to $\beta = 10$ (Figure 10f), and is even smaller for watershed segmentation (Figure 10b) and converging active contours (Figure 10c). In addition, the connectivity indicator Γ is also sensitive to the choice of segmentation method, because small objects close to the image resolution have a high impact on the continuity of a phase. Therefore, macropore connectivity increases from $\Gamma_p = (0.72 - 0.80)$ to $\Gamma_p = (0.83 - 0.84)$ if thin macropores that connect larger pore bodies are correctly identified, e.g., the green frame in Figure 10. At the same time, the connectivity of organic matter decreases drastically from $\Gamma_p = (0.94 - 0.96)$ to $\Gamma_p = (0.64 - 0.74)$ if false organic coatings of macropores (violet frame) are successfully suppressed.

4. Discussion

4.1. Image Enhancement

We have demonstrated how essential a suitable combination of image enhancement methods can be for subsequent image segmentation. The correction of intensity bias due to beam hardening was adequately

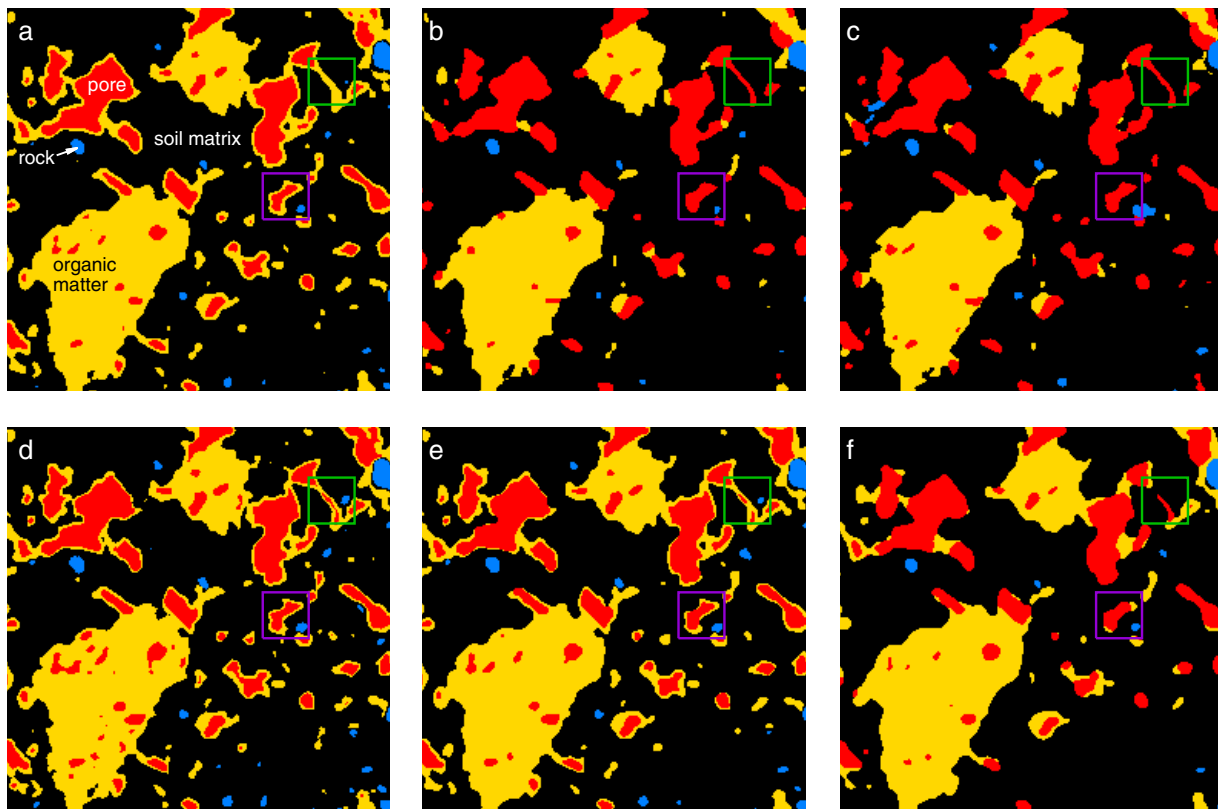


Figure 10. Image segmentation of the soil image: (a) global thresholding on I' with postprocessing, (b) watershed segmentation on \hat{I}_{NL+UM} , (c) converging active contours on \hat{I}_{NL+UM} , (d–f) Bayesian MRF segmentation on \hat{I}_{NL+UM} with $\beta = \{0.1, 1, 10\}$, respectively.

corrected by the method of *Iassonov and Tuller* [2010], since the examined sample was cylindrical and fairly homogeneous. If the intensity bias had a more complex shape, which cannot be readily described as function of radius, the bias model could have been obtained by interpolation instead. For instance, the approach of *Yanowitz and Bruckstein* [1989] to use the Laplace equation to interpolate a threshold surface between slowly varying gray values along phase boundaries can be adapted to interpolate a bias surface between objects of the highest intensity class. Even better results can be expected by using the Poisson equation for this purpose [*Pérez et al.*, 2003]. The ring artifact removal routine [*Sijbers and Postnov*, 2004] succeeded in removing most of the rings. However, the ring artifact removal is incomplete if the artifact magnitude is not constant along rotation angle ϕ . Moreover, the back transform from polar into Cartesian

Table 3. Volume Fractions V , Surface Area Densities a , and Connectivity Indices Γ of Pores p and Organic Residues o for the Soil Image^a

Denoising	$V_p (-)$	$V_o (-)$	$a_o (mm^{-1})$	$a_{op} (mm^{-1})$	$\Gamma_p (-)$	$\Gamma_o (-)$
<i>G6—Global Thresholding</i>						
\hat{C}_{MA}	0.082	0.213	2.74	0.91	0.721	0.953
\hat{I}_{NL}	0.096	0.190	3.30	1.29	0.777	0.940
<i>L3—Bayesian MRF</i>						
$\hat{C}_{MA} (\beta = 0.1)$	0.085	0.208	2.63	0.91	0.734	0.953
$\hat{I}_{NL} (\beta = 0.1)$	0.093	0.197	2.97	1.12	0.787	0.950
$\hat{I}_{NL} (\beta = 1)$	0.086	0.193	2.13	0.71	0.795	0.961
$\hat{I}_{NL} (\beta = 10)$	0.087	0.176	1.46	0.44	0.822	0.963
<i>L4—Watershed</i>						
\hat{C}_{MA}	0.110	0.137	1.05	0.30	0.802	0.652
\hat{I}_{NL}	0.117	0.127	0.84	0.25	0.834	0.644
<i>L5—Converging Active Contours</i>						
\hat{C}_{MA}	0.088	0.135	1.24	0.40	0.730	0.744
\hat{I}_{NL}	0.129	0.116	0.66	0.24	0.825	0.681

^aDenoising is either applied prior to segmentation with a nonlocal means filter (\hat{I}_{NL}) or as postprocessing with a majority filter (\hat{C}_{MA}).

coordinates introduces additional blur that increases with radius r . Thus, the performance is always somewhat worse as compared to line removal directly applied on the sinograms when this is possible [Ketcham, 2006]. Also, some methods operate on rings in Cartesian space directly [Freundlich, 1987]. Finally, Fourier and wavelet filters generally lead to improved ring removal [Raven, 1998; Münch et al., 2009].

The surveyed denoising methods were efficient in removing noise while keeping the blurring of edges at a minimum, given that the associated parameters are set adequately. In fact, image noise today has become a secondary issue for successful image analysis. Instead, image blur has been identified as more of a pitfall for the success of the various segmentation methods in this study. Edge enhancement with unsharp masks partly mitigates image blur, but is not capable of removing partial volume effects completely. Surely, future advances in X-ray tomography hardware and reconstruction software will lead to steady improvements in image quality, so that sharper images can be acquired on a routine basis.

One of our significant findings was that even though image enhancement is often indispensable for robust threshold detection, it does not necessarily imply that the segmentation itself also has to be applied to the enhanced image. Instead, segmenting the raw data with the thus obtained thresholds may lead to fewer misclassification errors if suitable postprocessing is applied to the segmented images. This is because any image enhancement inevitably destroys some structural information in the raw data. For the images examined in this study, a majority filter on the segmented raw data produced good results, mostly because the objects had rather smooth, convex boundaries. In turn, a majority filter can be a less desirable option if true objects are thin, concave, have rough surfaces or acute angles. It is up to the user to always compare and decide which is the most favorable option.

4.2. Global Thresholding

We have demonstrated that every histogram-based thresholding method relies on certain assumptions about the histogram shape. This introduces some bias if the class modes have different variance, skewness, or proportion. Many bias correction techniques for the standard methods used here have been suggested. For instance, minimum error thresholding can be corrected for imbalanced overlap [Cho et al., 1989] or can be used with Poisson distributions for each class instead of Gaussian distributions [Pal and Bhandari, 1993]. Shannon entropy can be replaced by Tsallis entropy for maximum entropy thresholding, allowing for an additional degree of freedom that can be used to tune the results [de Albuquerque et al., 2004]. Fuzzy c -means can be corrected for imbalanced class proportions [Jawahar et al., 1997]. The list is virtually endless. However, we found it more useful to (i) put some effort into suitable image enhancement prior to thresholding, (ii) alleviate the impact of bias by histogram clipping, and (iii) use the average after outlier removal to determine thresholds for subsequent segmentation. Even with this preprocessing methodology there might be soil images that still exhibit unimodal histograms due to a lot of unresolved porosity or very imbalanced class proportions [Wang et al., 2011; Baveye et al., 2010]. The second problem can be avoided with a semiautomatic algorithm based on ROI dilations (Figure 11). For instance, small rocks constitute only 3% to volume in the soil image and are hard to distinguish as an individual class in the histogram. The threshold for the region of interest (ROI) is set to the class mean ($t = \mu_s = 203$). Consecutive dilations of the ROI mask lead to a clearly bimodal histogram for which a threshold between the soil and rock class can be easily identified. Note that the thus obtained local histogram minimum at 182 is very close to $t = 185$ in Figure 9f, but much easier to identify. An alternative to deal with unimodal histograms, which is, however, restricted to two-class segmentation, is to estimate a threshold from the mean gray value within edge regions [Panda and Rosenfeld, 1978; Schlüter et al., 2010]. The problem of too much unresolved porosity or too gradual intensity changes is more severe and puts the entire concept of segmentation into question. In this case, some morphology analysis can be deployed to the intensity data directly, including distance transforms [Jang and Hong, 2001], isosurfaces [McClure et al., 2007], skeletonization [Chung and Sapiro, 2000], or tortuosity [Gommes et al., 2009].

4.3. Local Segmentation

We have corroborated the importance of using locally adaptive methods truly capable of multiclass segmentation instead of applying iterations of binarizations [Tuller et al., 2013]. Methods that fulfill this criterion are Bayesian MRF segmentation [Berthod et al., 1996; Kulkarni et al., 2012], watershed segmentation [Beucher and Lantuejoul, 1979; Vincent and Soille, 1991; Roerdink and Meijster, 2000], and converging active contours [Sheppard et al., 2004]. Bayesian MRF segmentation was originally developed for image classification in the

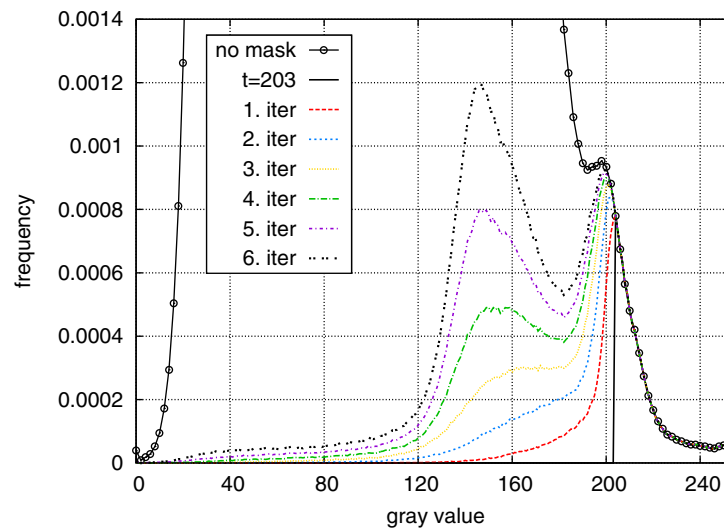


Figure 11. Iterative algorithm to identify a truly bimodal histogram at high gray values in the soil image.

presence of additive noise and is based on the assumption that individual class modes follow a Gaussian distribution. This causes failure of the method for denoised images with pronounced partial volume effects due to image blur, which leads to either one-sided or two-sided tailings in the histogram. As a consequence, special care has to be taken to correct for this histogram bias. In addition, the results depend heavily on the homogeneity parameter β . Erroneous assignment of partial volume voxels could only be suppressed with a high penalty on class boundaries ($\beta = 10$), which at the same time removed true image features of similar size. A promising improvement of the Bayesian MRF segmentation, especially if applied to fluid images, would be to replace the unspecified penalty term (equation (14)) with real surface tensions [Knight *et al.*, 1990; Silverstein and Fort, 2000]. In fact, any noninvasive laboratory technique which provides independent measurements about a structural property of the same sample can potentially help to condition segmentation parameters. This has been recently demonstrated for three-phase segmentation of limestone via hysteresis thresholding, where the region growing parameters were conditioned by independent porosity measurements [Mangane *et al.*, 2013].

For the watershed segmentation in its current implementation, a benefit is that it only requires simple thresholding and a gradient image. However, the lack of additional parameters also hampers flexibility. Therefore, excessive elimination of wetting films in the synthetic image could not be avoided. In contrast, the somewhat similar converging active contours method honors local gradient and intensity information at the same time. A reported drawback of this method is its sensitivity to seed region detection and the adjustment of parameters for the speed function [Iassonov *et al.*, 2009]. In this paper, we demonstrated strategies to define the seed regions automatically. In addition, the parameters for the speed function can be obtained directly by a careful evaluation of the gradient histogram (Figure 12). The frequency of local gradients is almost always unimodal, and therefore, difficult to threshold with standard methods. Yet, every unimodal histogram exhibits some characteristic features such as a point of maximum curvature [Tsai, 1995] or the point of maximum distance from an auxiliary line between the histogram mode and the maximum bin [Rosin, 2001] (Figure 12, inset). In addition, some gradient histograms may exhibit secondary shoulders if the intensity data comprise fairly constant edge heights. Picking one of these points as a gradient threshold would result in a fully objective segmentation method. In this study, we always used the gradient cutoff detection with Rosin's method. In turn, adjusting this cutoff adds some flexibility to the speed function in the same way as β scales the penalty term in the Bayesian MRF method. For instance, we achieved a very good reproduction of the true interfacial area between fluids a_{wn} in the synthetic test image by setting the gradient cutoff to a higher value.

As expected, objects that are close in size to the image resolution were associated with the highest uncertainty in all test images. Five voxels in diameter has been suggested as a rough estimate for an object size limit for which image analysis is reliable [Lehmann *et al.*, 2006; Vogel *et al.*, 2010]. Features like the thin oil

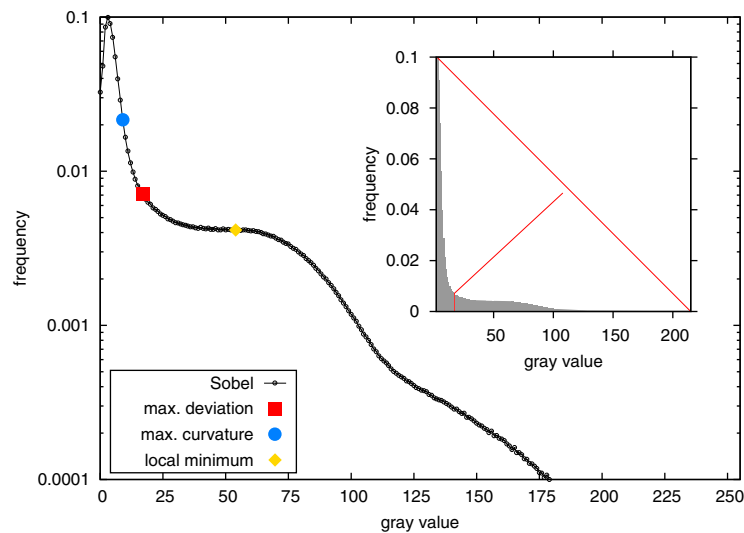


Figure 12. Gradient histogram computed on a Sobel image of \hat{I}_{NL+UM} for the synthetic image. Each unimodal histogram has a unique point of maximum curvature and maximum distance from the auxiliary line depicted in the inset. In addition, the shoulder in the histogram produces a local minimum, which can be detected as well.

films and pendular rings at grain contacts in the synchrotron image clearly do not meet this criterion. *Brown et al.* [2014] compared surface area densities of smooth fluid interfaces for the same kind of three-fluid samples when improving the image resolution from 10.6 to 5.3 μm and observed an increase in the range of 5–16%, depending on the specific fluid pair. An improvement in resolution usually comes at the cost of a smaller field of view as far as industrial scanners are concerned, as well as increased noise unless the scan time is increased as well. For heterogeneous media, especially, there is a trade-off between a representative image size and sufficient resolution to capture important details. However, the scale window can also be extended toward smaller objects by scanning the same sample at different resolutions, even with different imaging techniques, and merging the images with image registration [*Latham et al.*, 2008], statistical fusion of images [*Mohebi et al.*, 2009] or scale fusion applied to structure analysis data only [*Vogel et al.*, 2010; *Schlüter et al.*, 2011]. Finally, for this study, we have merely compared the outcome of different local segmentation methods with each other. Of course, several methods can also be combined. For instance, Bayesian MRF segmentation with surface tensions could be used for postprocessing the outcome of converging active contours. If necessary, postprocessing with a majority filter could also be applied in addition to denoising prior to segmentation. After all, the choice of a suitable protocol always depends on characteristics of the raw data.

5. Conclusions

We have surveyed recent advances in image enhancement and image segmentation of multiphase X-ray microtomography data. Image enhancement methods included ring artifact removal, intensity bias correction, edge enhancement, image denoising, and contrast enhancement. Image segmentation methods comprised six global segmentation and five different locally adaptive segmentation methods. Some general conclusions can be drawn from our findings:

1. Image blur is the major cause of poor segmentation results in this study, since image noise and other image artifacts could be removed with current image processing methods.
2. A lot of uncertainty in threshold detection can be removed by suitable preprocessing and correction of histograms for bias in the class statistics prior to thresholding.
3. Bayesian Markov random field segmentation, watershed segmentation and converging active contours are suited for multiclass segmentation. The converging active contour method has potentially the highest flexibility to correct for partial volume effects and simultaneously conserve small image features.

Appendix A

Most image processing steps presented in this study can be done with freely available software. The only exceptions are the nonlocal means filter for denoising and the watershed method for segmentation, which were carried out with AVIZO® FIRE (<http://www.vsg3d.com/avizo/fire>). Moreover, the converging active contours method for segmentation was performed with MANGO (<http://physics.anu.edu.au/appmaths/capabilities/mango.php>).

The conversion between Cartesian and polar coordinates for ring artifact removal was performed with the freely available Polar Transformer (<http://rsbweb.nih.gov/ij/plugins/polar-transformer.html>) plug-in for IMAGEJ. All other image processing steps are either built on or directly implemented in the QUANTIM (<http://www.quantim.ufz.de>) open-source image processing library.

Acknowledgments

We thank Alasdair Houston (SIMBIOS, Abertay, UK) for providing the μ CT soil image and three anonymous reviewers for their helpful comments. S. Schlüter is grateful to the Alexander-von-Humboldt Foundation for granting a Feodor-Lynen scholarship. A. Sheppard acknowledges the support of an Australian Research Council Future Fellowship.

References

- Anderson, S., and J. Hopmans (Eds.) (2013), *Soil–Water–Root Processes: Advances in Tomography and Imaging*, vol. 61, SSSA Spec. Publ., Soil Sci. Soc. of Am., Madison, WI, doi:10.2136/sssaspecpub61.
- Armstrong, R., M. Porter, and D. Wildenschild (2012), Linking pore-scale interfacial curvature to column-scale capillary pressure, *Adv. Water Resour.*, *46*, 55–61.
- Baveye, P. C., et al. (2010), Observer-dependent variability of the thresholding step in the quantitative analysis of soil images and X-ray microtomography data, *Geoderma*, *157*(1–2), 51–63, doi:10.1016/j.geoderma.2010.03.015.
- Berthod, M., Z. Kato, S. Yu, and J. Zerubia (1996), Bayesian image classification using Markov random fields, *Image Vision Comput.*, *14*(4), 285–295.
- Besag, J. (1986), On the statistical analysis of dirty pictures, *J. R. Stat. Soc., Ser. B*, *48*, 259–302.
- Beucher, S., and C. Lantuejoul (1979), Use of watersheds in contour detection, paper presented at International Workshop Image Processing, Real-Time Edge and Motion Detection/Estimation, CCETT, Rennes, France.
- Blunt, M. J., B. Bijeljic, H. Dong, O. Gharbi, S. Iglauer, P. Mostaghimi, A. Paluszny, and C. Pentland (2013), Pore-scale imaging and modelling, *Adv. Water Resour.*, *51*, 197–216, doi:10.1016/j.advwatres.2012.03.003.
- Brown, K., S. Schlüter, A. Sheppard, and D. Wildenschild (2014), On the challenges of measuring interfacial characteristics of three-phase fluid flow with x-ray microtomography, *J. Microsc.*, *253*(3), 171–182, doi:10.1111/jmi.12106.
- Buades, A., B. Coll, and J. M. Morel (2005), A non-local algorithm for image denoising, in *Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005*, vol. 2, pp. 60–65, IEEE Comput. Soc., San Diego, CA, doi:10.1109/CVPR.2005.38.
- Buades, A., B. Coll, and J.-M. Morel (2008), Nonlocal Image and Movie Denoising, *Int. J. Comput. Vision*, *76*(2), 123–139, doi:10.1007/s11263-007-0052-1.
- Canny, J. (1986), A computational approach to edge detection, *IEEE Trans. Pattern Anal. Mach. Intel.*, *8*(6), 679–698, doi:10.1109/TPAMI.1986.4767851.
- Catté, F., P. Lions, J. Morel, and T. Coll (1992), Image selective smoothing and edge detection by nonlinear diffusion, *SIAM J. Numer. Anal.*, *29*(1), 182–193, doi:10.1137/0729012.
- Cho, S., R. Haralick, and S. Yi (1989), Improvement of kittler and illingworth's minimum error thresholding, *Pattern Recognition*, *22*(5), 609–617, doi:10.1016/0031-3203(89)90029-0.
- Chung, D. H., and G. Sapiro (2000), Segmentation-free skeletonization of gray-scale images via PDEs, in *Proceedings 2000 International Conference on Image Processing*, IEEE International Conference on Image Processing, vol. 2, pp. 927–930, Vancouver, BC, Canada, doi:10.1109/ICIP.2000.899868.
- Cnudde, V., M. N. Boone (2013), High-resolution X-ray computed tomography in geosciences: A review of the current technology and applications, *Earth-Science Reviews*, *123*, 1–17, doi:10.1016/j.earscirev.2013.04.003.
- Costanza-Robinson, M. S., K. H. Harrold, and R. M. Lieb-Lappen (2008), X-ray microtomography determination of air-water interfacial area-water saturation relationships in sandy porous media, *Environ. Sci. Technol.*, *42*(8), 2949–2956, doi:10.1021/es072080d.
- Culligan, K., D. Wildenschild, B. Christensen, W. Gray, M. Rivers, and A. Tompson (2004), Interfacial area measurements for unsaturated flow through a porous medium, *Water Resour. Res.*, *40*, W12413, doi:10.1029/2004WR003278.
- de Albuquerque, M. P., I. Esquef, A. G. Mello, and M. P. de Albuquerque (2004), Image thresholding using Tsallis entropy, *Pattern Recognition Lett.*, *25*(9), 1059–1065, doi:10.1016/j.patrec.2004.03.003.
- Ferreira, T., and W. Rasband (2012), *ImageJ User Guide—J 1.46r*, National Institute of Health, Bethesda, MD.
- Freundlich, D. A. (1987), Ring artifact correction for computerized tomography, Patent [4,670,840], U.S. Patent and Trademark Off., Washington, D. C.
- Gommes, C., A. Bons, S. Blacher, J. Dunsmuir, and A. H. Tsou (2009), Practical methods for measuring the tortuosity of porous materials from binary or gray-tone tomographic reconstructions, *AIChE J.*, *55*(8), 2000–2012, doi:10.1002/aic.11812.
- Gonzalez, R., and R. Woods (2002), *Digital Image Processing*, pp. 793, Prentice Hall, Upper Saddle Creek, N. J.
- Hoshen, J., and R. Kopelman (1976), Percolation and cluster distribution. I. Cluster multiple labeling technique and critical concentration algorithm, *Phys. Rev. B Solid State*, *14*, 3438–3445, doi:10.1103/PhysRevB.14.3438.5.
- Houston, A., S. Schmidt, A. Tarquis, W. Otten, P. Baveye, and S. Hapca (2013a), Effect of scanning and image reconstruction settings in X-ray computed microtomography on quality and segmentation of 3D soil images, *Geoderma*, *207*, 154–165, doi:10.1016/j.geoderma.2013.05.017.
- Houston, A., W. Otten, P. Baveye, and S. Hapca (2013b), Adaptive-window indicator kriging: A thresholding method for computed tomography images of porous media, *Comput. Geosci.*, *54*, 239–248, doi:10.1016/j.cageo.2012.11.016.
- Huang, T., G. Yang, and G. Tang (1979), A fast two-dimensional median filtering algorithm, *IEEE Trans. Acoust. Speech Signal Process.*, *27*(1), 13–18, doi:10.1109/TASSP.1979.1163188.
- Iassonov, P., and M. Tuller (2010), Application of segmentation for correction of intensity bias in X-ray computed tomography images, *Vadose Zone J.*, *9*(1), 187–191, doi:10.2136/vzj2009.0042.
- Iassonov, P., T. Gebrenegus, and M. Tuller (2009), Segmentation of X-ray CT images of porous materials: A crucial step for characterization and quantitative analysis of pore structures, *Water Resour. Res.*, *45*, W09415, doi:10.1029/2009WR008087.

- Ibanez, L., W. Schroeder, L. Ng, and J. Cates (2005), *The ITK Software Guide*, 2nd ed., Kitware Inc.
- Jang, J.-H., and K.-S. Hong (2001), A pseudo-distance map for the segmentation-free skeletonization of gray-scale images, in *Proceedings of Eighth IEEE International Conference on Computer Vision, ICCV 2001*, vol. 2, pp. 18–23, Vancouver, BC, Canada, doi:10.1109/ICCV.2001.937586.
- Jawahar, C., P. Biswas, and A. Ray (1997), Investigations on fuzzy thresholding based on fuzzy clustering, *Pattern Recognition*, *30*(10), 1605–1613, doi:10.1016/S0031-3203(97)00004-6.
- Kaestner, A., E. Lehmann, and M. Stapanoni (2008), Imaging and image processing in porous media research, *Adv. Water Resour.*, *31*(9), 1174–1187, doi:10.1016/j.advwatres.2008.01.022.
- Kapur, J., P. Sahoo, and A. Wong (1985), A new method for gray-level picture thresholding using the entropy of the histogram, *Comput. Vision Graphics Image Process.*, *29*(3), 273–285.
- Ketcham, R., and W. Carlson (2001), Acquisition, optimization and interpretation of X-ray computed tomographic imagery: Applications to the geosciences, *Comput. Geosci.*, *27*(4), 381–400.
- Ketcham, R. A. (2005), Computational methods for quantitative analysis of three-dimensional features in geological specimens, *Geosphere*, *1*(1), 32–41.
- Ketcham, R. A. (2006), New algorithms for ring artifact removal, *Proc. SPIE*, *6318*, 6318007, doi:10.1117/12.680939.
- Kittler, J., and J. Illingworth (1986), Minimum error thresholding, *Pattern Recognition*, *19*(1), 41–47, doi:10.1016/0031-3203(86)90030-0.
- Knight, R., A. Chapman, and M. Knoll (1990), Numerical modeling of microscopic fluid distribution in porous media, *J. Appl. Phys.*, *68*(3), 994–1001.
- Kulkarni, R., M. Tuller, W. Fink, and D. Wildenschild (2012), Three-dimensional multiphase segmentation of X-Ray CT data of porous materials using a Bayesian Markov random field framework, *Vadose Zone J.*, *11*(1), pp. 74–85, doi:10.2136/vzj2011.0082.
- Kurita, T., N. Otsu, and N. Abdelmalek (1992), Maximum likelihood thresholding based on population mixture models, *Pattern Recognition*, *25*(10), 1231–1240, doi:10.1016/0031-3203(92)90024-D.
- Latham, S., T. Varslot, and A. Sheppard (2008), Image registration: Enhancing and calibrating X-ray micro-CT imaging, paper presented at Proceedings of the Society of Core Analysis, SCA2008-35, pp. 1–12, Abu Dhabi, UAE.
- Lehmann, P., P. Wyss, A. Flisch, E. Lehmann, P. Vontobel, M. Krafczyk, A. Kaestner, F. Beckmann, A. Gygi, and H. Flüher (2006), Tomographical imaging and mathematical description of porous media used for the prediction of fluid distribution, *Vadose Zone J.*, *5*, 80–97.
- Liao, P.-S., T.-S. Chen, and P.-C. Chung (2001), A fast algorithm for multilevel thresholding, *J. Inf. Sci. Eng.*, *17*(5), 713–728.
- Mangane, P. O., P. Gouze, and L. Luquot (2013), Permeability impairment of a limestone reservoir triggered by heterogeneous dissolution and particles migration during CO₂-rich injection, *Geophys. Res. Lett.*, *40*, 4614–4619, doi:10.1002/grl.50595.
- McClure, J., D. Adalsteinsson, C. Pan, W. Gray, and C. Miller (2007), Approximation of interfacial properties in multiphase porous medium systems, *Adv. Water Resour.*, *30*(3), 354–365, doi:10.1016/j.advwatres.2006.06.010.
- Mohebi, A., P. Fieguth, and M. Ioannidis (2009), Statistical fusion of two-scale images of porous media, *Adv. Water Resour.*, *32*, 1567–1579.
- Münch, B., P. Trtik, F. Marone, and M. Stapanoni (2009), Stripe and ring artifact removal with combined wavelet—Fourier filtering, *Opt. Express*, *17*(10), 8567–8591, doi:10.1364/OE.17.008567.
- Oh, W., and B. Lindquist (1999), Image thresholding by indicator kriging, *IEEE Trans. Pattern Anal. Mach. Intel.*, *21*(7), 590–602.
- Otsu, N. (1975), A threshold selection method from gray-level histograms, *Automatica*, *11*(285-296), 23–27.
- Pal, N. R., and D. Bhandari (1993), Image thresholding: Some new techniques, *Signal Process.*, *33*(2), 139–158, doi:10.1016/0165-1684(93)90107-L.
- Pal, N. R., and S. K. Pal (1993), A review on image segmentation techniques, *Pattern Recognition*, *26*(9), 1277–1294, doi:10.1016/0031-3203(93)90135-J.
- Panda, D., and A. Rosenfeld (1978), Image segmentation by pixel classification in (gray level, edge value) space, *IEEE Trans. Comput.*, *C-27*(9), 875–879, doi:10.1109/TC.1978.1675208.
- Pérez, P., M. Gangnet, and A. Blake (2003), Poisson image editing, *ACM Trans. Graphics*, *22*(3), 313–318.
- Perona, P., and J. Malik (1990), Scale-space and edge detection using anisotropic diffusion, *IEEE Trans. Pattern Anal. Mach. Intel.*, *12*(7), 629–639.
- Pizer, S. M., E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. T. H. Romeny, J. B. Zimmerman, and K. Zuiderveld (1987), Adaptive histogram equalization and its variations, *Comput. Vision Graphics Image Process.*, *39*(3), 355–368, doi:10.1016/S0734-189X(87)80186-X.
- Porter, M., and D. Wildenschild (2010), Image analysis algorithms for estimating porous media multiphase flow variables from computed microtomography data: A validation study, *Comput. Geosci.*, *14*(1), 15–30.
- Raven, C. (1998), Numerical removal of ring artifacts in microtomography, *Rev. Sci. Instrum.*, *69*(8), 2978–2980, doi:10.1063/1.1149043.
- Renard, P., and D. Allard (2013), Connectivity metrics for subsurface flow and transport, *Adv. Water Resour.*, *51*, 168–196, doi:10.1016/j.advwatres.2011.12.001.
- Ridler, T., and S. Calvard (1978), Picture thresholding using an iterative selection method, *IEEE Trans. Syst. Man Cybern.*, *8*(8), 630–632, doi:10.1109/TSMC.1978.4310039.
- Roerdink, J. B., and A. Meijster (2000), The watershed transform: Definitions, algorithms and parallelization strategies, *Fundam. Inf.*, *41*(1), 187–228.
- Rogasik, H., J. Crawford, O. Wendroth, I. Young, M. Joschko, and K. Ritz (1999), Discrimination of soil phases by dual energy X-ray tomography, *Soil Sci. Soc. Am. J.*, *63*, 741–751.
- Rosin, P. (2001), Unimodal thresholding, *Pattern Recognition*, *34*(11), 2083–2096.
- Rudin, L., S. Osher, and E. Fatemi (1992), Nonlinear total variation based noise removal algorithms, *Physica D*, *60*, 259–268, doi:10.1016/0167-2789(92)90242-F.
- Russ, J. C. (2006), *The Image Processing Handbook*, CRC Press, Boca Raton, FL.
- Sahoo, P., S. Soltani, and A. Wong (1988), A survey of thresholding techniques, *Comput. Vision Graphics Image Process.*, *41*(11), 233–260.
- Schlüter, S., U. Weller, and H.-J. Vogel (2010), Segmentation of X-ray microtomography images of soil using gradient masks, *Comput. Geosci.*, *36*(10), 1246–1251, doi:10.1016/j.cageo.2010.02.007.
- Schlüter, S., U. Weller, and H.-J. Vogel (2011), Soil-structure development including seasonal dynamics in a long-term fertilization experiment, *J. Plant Nutr. Soil Sci.*, *174*(3), 395–403, doi:10.1002/jpln.201000103.
- Serra, J. (1982), *Mathematical Morphology and Image Analysis*, Academic, San Diego, Calif.
- Sezgin, M., and B. Sankur (2004), Survey over image thresholding techniques and quantitative performance evaluation, *J. Electron. Imaging*, *13*(1), 146–165.
- Sheppard, A., R. Sok, and H. Averdunk (2004), Techniques for image enhancement and segmentation of tomographic images of porous materials, *Physica A*, *339*(1), 145–151.

- Sijbers, J., and A. Postnov (2004), Reduction of ring artefacts in high resolution micro-CT reconstructions, *Phys. Medicine Biol.*, *49*(14), N247–N253.
- Silverstein, D., and T. Fort (2000), Prediction of air-water interfacial area in wet unsaturated porous media, *Langmuir*, *16*(2), 829–834.
- Stark, J. (2000), Adaptive image contrast enhancement using generalizations of histogram equalization, *IEEE Trans. Image Process.*, *9*(5), 889–896, doi:10.1109/83.841534.
- Trier, O., and A. Jain (1995), Goal-directed evaluation of binarization methods, *IEEE Trans. Pattern Anal. Mach. Intel.*, *17*(12), 1191–1201, doi:10.1109/34.476511.
- Tsai, D.-M. (1995), A fast thresholding selection procedure for multimodal and unimodal histograms, *Pattern Recognition Lett.*, *16*(6), 653–666, doi:10.1016/0167-8655(95)80011-H.
- Tuller, M., R. Kulkarni, and W. Fink (2013), Segmentation of X-ray CT data of porous materials: A review of global and locally adaptive algorithms, in *Soil–Water–Root Processes: Advances in Tomography and Imaging*, *SSSA Special Publications*, vol. 61, edited by S. Anderson and J. Hopmans, pp. 157–182, Soil Sci. Soc. of Am., Madison, WI.
- van der Walt, S., J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, T. Yu (2014), scikit-image: Image processing in Python, *PeerJ PrePrints*, *2*, e336v2, doi:10.7287/peerj.preprints.336v2.
- Vincent, L., and P. Soille (1991), Watersheds in digital spaces: An efficient algorithm based on immersion simulations, *IEEE Trans. Pattern Anal. Mach. Intel.*, *13*(6), 583–598, doi:10.1109/34.87344.
- Vogel, H.-J., and A. Kretzschmar (1996), Topological characterization of pore space in soil sample preparation and digital image-processing, *Geoderma*, *73*(1), 23–38.
- Vogel, H.-J., U. Weller, and S. Schlüter (2010), Quantification of soil structure based on Minkowski functions, *Comput. Geosci.*, *36*(10), 1236–1245, doi:10.1016/j.cageo.2010.03.007.
- Wang, W., A. Kravchenko, A. Smucker, and M. Rivers (2011), Comparison of image segmentation methods in simulated 2D and 3D microtomographic images of soil aggregates, *Geoderma*, *162*(3–4), 231–241, doi:10.1016/j.geoderma.2011.01.006.
- Wildenschild, D., and A. P. Sheppard (2013), X-ray imaging and analysis techniques for quantifying pore-scale structure and processes in subsurface porous medium systems, *Adv. Water Resour.*, *51*, 217–246, doi:10.1016/j.advwatres.2012.07.018.
- Wildenschild, D., J. Hopmans, C. Vaz, M. Rivers, D. Rikard, and B. Christensen (2002), Using X-ray computed tomography in hydrology: Systems, resolutions, and limitations, *J. Hydrol.*, *267*(3), 285–297.
- Wildenschild, D., J. Hopmans, M. Rivers, and A. Kent (2005), Quantitative analysis of flow processes in a sand using synchrotron-based X-ray microtomography, *Vadose Zone J.*, *4*(1), 112–126.
- Yanowitz, S., and A. Bruckstein (1989), A New Method for Image Segmentation, *Comput. Vision Graphics Image Process.*, *46*(1), 82–95.