2862
doi:10.1017/S143192762002303X

Microsc. Microanal. 26 (Suppl 2), 2020

Quantitative Image Analysis of Source Rocks Using Machine Learning Segmentation

Shannon Eichmann[1], Poorna Srinivasan[2], Shawn Zhang[3] and James Howard[3]

[1]Aramco Services Company: Aramco Research Center - Houston, Houston, Texas, United States, [2]Aramco Services Company, Houston, Texas, United States, [3]DigiM Solution LLC, Burlington, Massachusetts, United States

Source rocks are composed of inorganic minerals, clay particles, and organic matter that are compacted to create a composite with texture and pore sizes that vary at the nano- and micro-scales [1-2]. These rocks exhibit chemical variations due to the dominant framework phases (*e.g.* calcite, dolomite, or quartz) and the clay minerals (*e.g.* illite, smectite, or kaolinite), while exhibiting variations in the organic components (kerogen, bitumen, and/or pyrobitumen) where the relative amount and molecular composition varies with thermal maturity [2-4]. Scanning Electron Microscopy (SEM) is commonly used for source rock characterization to quantify the fraction of various components, porosity, and pore size distribution, which are used to evaluate reservoir quality during hydrocarbon exploration. This variation in structural and chemical heterogeneity as well as imaging artifacts such as charging, surface contamination, and surface damage from sample preparation, create images with a broad multi-modal intensity histogram that is challenging to accurately segment [5].

Traditionally, image segmentation applies user-defined or mathematically selected gray-scale threshold cutoffs to separate the image components after pre-processing and has been applied to, and compared for, rock samples [5-9]. For images with few phases, a consistent baseline intensity, and limited surface artifacts, these methods can be effectively applied, and mislabeled regions can be corrected manually or accepted as part of the quantification error. However, for highly heterogeneous samples with more than three components and those with several artifacts including a varying baseline intensity, the mislabeled regions can be numerous, making manual correction impractical and creating significant error in the overall quantification. In these cases, an advanced segmentation method is required that can effectively separate the components using more than intensity variation alone. Recent advances in machine learning has led to the development of image segmentation tools that have been previously demonstrated in natural and manufactured materials at both the SEM and X-Ray micro-Computed Tomography scales [10-12].

Here we demonstrate the use of a supervised machine learning image segmentation technique on high resolution large field-of-view (LgFOV) SEM images of source rocks that were generated by stitching >75 secondary electron tiles (15 nm/pixel). Figure 1A shows an example source rock image after stitching, cropping, and application of a mild non-local means filter. The components shown are the pores (black), organics (dark gray), matrix minerals (light gray), and high density minerals (white). This image also shows common imaging artifacts such as charging (dashed red boxes), uncorrected speckle, and ion mill induced scars (solid red boxes), which are challenging to segment and can negatively affect quantification.

For comparison, LgFOV images were segmented using both intensity threshold ranges (Avizo®, ThermoFisher) and machine learning (DigiM™ I2S, DigiM Solution) methods to separate pores (blue), organics (green), high density minerals (yellow), and matrix (gray). In the intensity threshold range method, a range of intensities in the image histogram are assigned to each component and adjusted to best capture the domains. Figure 1B shows the result from using this method. The image shows sporadically mislabeled speckle and regions with scars and charging that are similarly mislabeled. The charging is commonly mislabeled as high density components (yellow) while the speckle and scars are mislabeled as a mixture of pores (blue), organics (green), and high density components (yellow). In total, these mislabeled regions affect the compositional analysis and the pore size distribution measurements. While it is possible to manually correct these mislabeled issues, this level of manual correction would take considerable time and is impractical as a common practice.

In the machine learning segmentation method, using the web-browser interface, five to ten regions are cropped from the LgFOV SEM image for training purposes using a cloud computing infrastructure that is computationally efficient and scalable. During training, each phase is represented by a trace of pixels that are subjected to a broad range of image processing steps, e.g filtering, feature extraction, etc. This process creates 100+ values for each pixel that are evaluated with a random forest classifier. In each cropped region, the components are labeled and used to train a classifier which is then applied to the whole LgFOV SEM image. Figure 1C shows the resulting segmented image using the machine learning segmentation method. The image shows a significant improvement in the correct assignment of artifacts including both ion mill scars (red solid boxes) and charging (red dashed boxes). In addition, the machine learning segmentation correctly labeled the background speckle. The mislabeled regions are fewer in this case and may be corrected through manual relabeling or possibly with additional classifier training.

Results will be presented that further demonstrate the differences in these segmentation methods, including a quantitative comparison between the segmented composition and the pore size distributions obtained from each (Figure 2). A demonstration of the ability to segment fractures verses pores using the machine learning method will also be presented. This labeling distinction is not possible when only segmenting with intensity threshold ranges. Finally, benefits and challenges of applying machine learning segmentation for these materials will be discussed with a comparison to data from other measurement techniques when available. Overall, these results demonstrate the application of machine learning segmentation for highly heterogeneous natural materials prior to manual image correction, thus saving time and improving segmentation quality and accuracy [13].

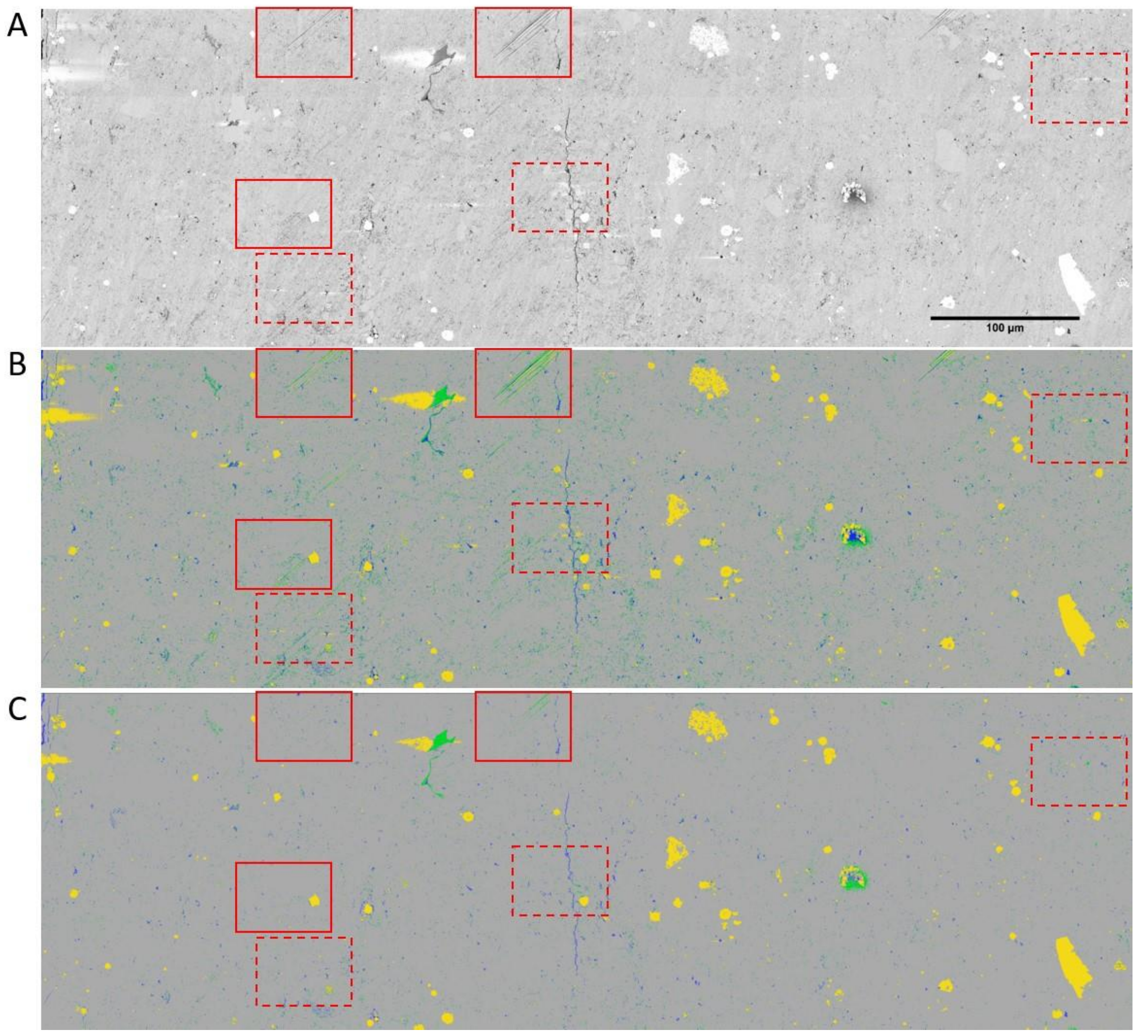

Figure 1. Representative images showing (A) LgFOV SEM, (B) intensity threshold range segmented, and (C) machine learning segmented images. Red solid boxes contain examples of ion mill scars and red dashed boxes contain charging artifacts. In (B) and (C), pores=blue, organics=green, matrix minerals=gray, and high density minerals=yellow. The scale bar is 100 µm.

https://doi.org/10.1017/S143192762002303X Published online by Cambridge University Press

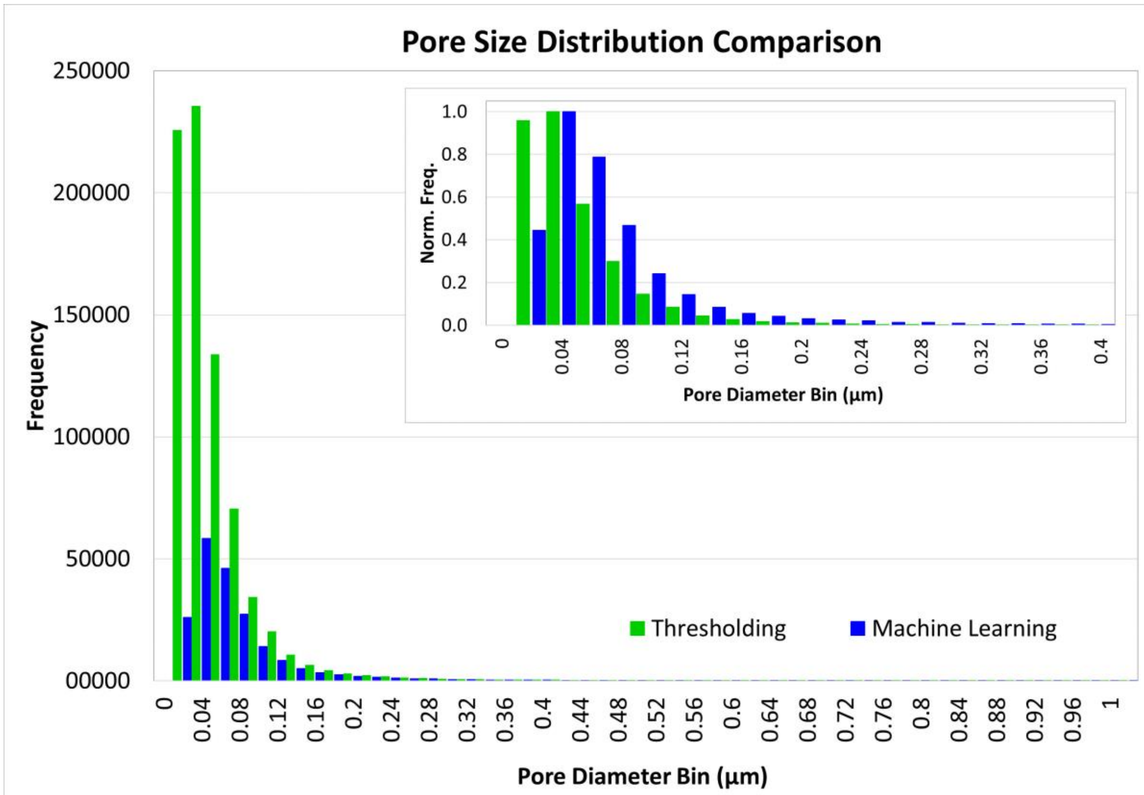

Figure 2. Pore size distribution histograms from the intensity threshold segmented image in Figure 1B (green) and from the machine learning segmented image in Figure 1C (blue) showing the difference in the number of pores identified as well as oversampling of the smallest pores in green. To improve comparison, the inset shows the histograms each normalized to the maximum frequency.

References

[1] KE Peters and MR Cassa, in "The Petroleum System – From Source to Trap" ed. LB Magoon and WG Dow, (AAPG, Tulsa) p. 93.
[2] LT Bryndzia and NR Braunsdof, Elements **10** (2014) p. 271.
[3] M Vandenbroucke, Rev. IFP, Oil & Gas Science and Technology **58** (2003) p. 243.
[4] DW van Krevelen, "Coal: typology, physics, chemistry, constitution" (Elsevier, Amsterdam).
[5] S Schluter *et al*, Water Resources Research **50** (2014) p. 3615.
[6] TJ Collins, Biotechniques **43(S1)** (2007) p. 25.
[7] A Sheppard, R Sok and H Averdunk, Physica A **339(1-2)** (2004) p. 145.
[8] H Andra *et al*, Computers and Geoscience **50** (2013) p. 25.
[9] N Idowu *et al*, SPE Reservoir Evaluation and Engineering **17** (2014) p. 473.
[10] J Howard, S Lin and S Zhang, Petrophysics **60(2)** (2019) p. 240.
[11] S Karimpoulia and P Tahmasebi, Computers and Geoscience **126** (2019) p. 142.
[12] I Varfolomeev, I Yakimchuk and I Safonov, Computers **8** (2019) p. 72.
[13] We would like to thank Josh Brothers for assistance in image collection.

https://doi.org/10.1017/S143192762002303X Published online by Cambridge University Press