Preliminary Report
Automated Analysis of Thin Section

Richard Boyne
CID 01057503
Imperial College London
MSc Applied Computational Science & Engineering
rmb115@imperial.ac.uk

5$^{\text{th}}$ July 2019

Supervisors

Prof. Olivier Dubrule
Imperial College London
Dep. of Earth Science & Engineering
o.dubrule@imperial.ac.uk

MSc Lukas Mosser
Imperial College London
Dep. of Earth Science & Engineering
lukas.mosser15@imperial.ac.uk

Dr Meindert Dillen
Wintershall Dea GmbH
EOR Projects
meindert.dillen@wintershalldea.com

MSc Tobias Thiel
Wintershall Dea GmbH
Exploration & Production
tobias.thiel@wintershalldea.com

Project Motivation

In geology thin sections are samples of a rock or mineral sliced, encased in epoxy and polished (to $\sim 30\mu m$ thick) for inspection under a petrographic or electron microscope. These images can be interpreted to give information about the sample, such as mineral composition, porosity and grain-size distribution (GSD). These give an insight into the geological environment, hence thin sections are an important tool for Exploration and Production in the petroleum industry.

Traditionally this analysis is done by modal analysis, where an experienced geologist categorises several hundred random points in the image to obtain a distribution over the whole sample. This is laborious process and highly subjective, making comparison between samples over time very unreliable. Hence, a consistent automation of this process is highly desirable. One potential method of automation is to segment the image into segments of different material, then by discerning between rock grain, matrix material (e.g. clay) and porous space both the porosity and grain geometries are found. There has been significant research into such automation, specifically with regard to segmentation of grains[1].

This projects goals are split into two sections. Firstly, to implement, and potentially improve upon, pre-existing automatic segmentation algorithms. Secondly, to identify the type of each segmented region as grain, matrix or pore, and also potentially the grain material.

Literature Review

Image Segmentation

Automatic image segmentation has been implemented since 1980s, originally by performing edge detection and attempting to form enclosed regions[2]. More recently superpixel algorithms have been developed[1,3] which instead work on a principle of grouping pixels together based on proximity in some higher dimensional feature space. Such methods are often implemented in computer vision for the identification of a predefined object within an image[4], however in the case of thin sections the desired segments are very similar making the problem somewhat harder.

Within the topic grain separation a commonly used methodology is the watershed transformation[5]. This is illustrated by taking a greyscale image's gradient, interpreting this result as a two dimensional surface and flooding each valley separately to form segments. The main drawbacks here is the need for initial "flooding" markers to feed the water in, commonly chosen as local minima in the image, and that only a grey scale is used removing the information contained in the color channels (though some algorithms combine transformations on each channel).

In the last decade the Simple Linear Iterative Clustering (SLIC[3]) algorithm is being used more. This method uses the unsupervised method k-means clustering in a scaled 5D feature space of position and color. This takes into account all information in a single algorithm and only requires the number of segments to be specified. Significant hyperparameters for this are the relative scaling of the position and color space, determining the priority of each in the algorithm. If multiple images of the same sample are available then these can be used in conjunction to perform a better segmentation. The method MSLIC[1] does this by taking the maximum distance between a pixel and a centroid across all images.

For an unknown number of segments one can initially over segment an image, then combine similar pixels[6]. The recombination involves identifying pixels to merge which can be by similarity in, for example texture (see below), or by somehow evaluating their adherence to boundaries, such as evaluating the color gradients across a boundary. This SLIC segmentation followed by recombination method has been applied to thin sections and shown[1] more accurate segmentation of rock grains than watershed as well as several other segmentation algorithms including Minimum Spanning Trees[7] and Linear Spectral Clustering[8].

Unsupervised Texture Analysis

The grouping of similar textures in an unsupervised manner is another common problem in computer vision, with numerous studies into different methods. Use of hierarchical clustering algorithms is relatively common due to there simplicity and often high performance in spite of this. A drawback is the relatively large computation time cost, often at $\mathcal{O}(n^2)$ where n is the number of data points.

In the specific case of comparing segments the Agglomerate Nesting (AGNES) is shown to be effective at pairing similar segments[9]. This is done in a bottom up manner, where by each data point is considered a cluster, then at each iteration nearest neighbours are merged to form a larger cluster. The exact new centre calculation gives rise to many variants of this algorithm. This is ended when the clusters are all above some distance apart or reach a desired number. One particular advantage of the latter is the ability to form an unknown number of segments, the scale of which is controlled by a single parameter. The algorithm is also robust against outliers, sorting each one into a cluster which only contains themselves.

Porosity from Thin Sections

Often experimental data or more complex imaging methods (e.g. CT-scan) are used to obtain accurate porosity values. The value found by model analysis is less reliable, as a 2D slice is a sample for the overall rock which might not be representative[10]. However, correction factors and judgement can still find meaning from point counted porosity.

Algorithms which automatically discern the area ratio of rock grain to porous space by color spectrum analysis have yielded reasonable accuracy[11]. This relies on the epoxy filling pores so they have a distinct and consistent color. Previous work has compared different techniques and color spaces for this analysis[10]. A multi-threshold method in HSV space was identified as the most accurate with $< 0.1\%$ error in the porosity estimate compared to model analysis, though this is very sensitive to the choice threshold parameters which may require changing for different thin section images.

Project Plan

The Dataset

Whenever Wintershall-DEA explores a site for well production one or two thin slice samples from different depths are selected to be studied. These are photographed in high resolution $\sim 10^4 \times 10^4$ pixels from hundreds of images stitched together and are taken in both white light and polarised light (only one angle). model analysis of at least 100 points is available for each image.

Carbonates and shale composites typically contain a wider range of rock types, so are harder to identify. Here we will focus on sandstone samples. There are a several hundred wells on record with sandstone, from which a small number of images (< 10) with easily determinable grains will be used to develop

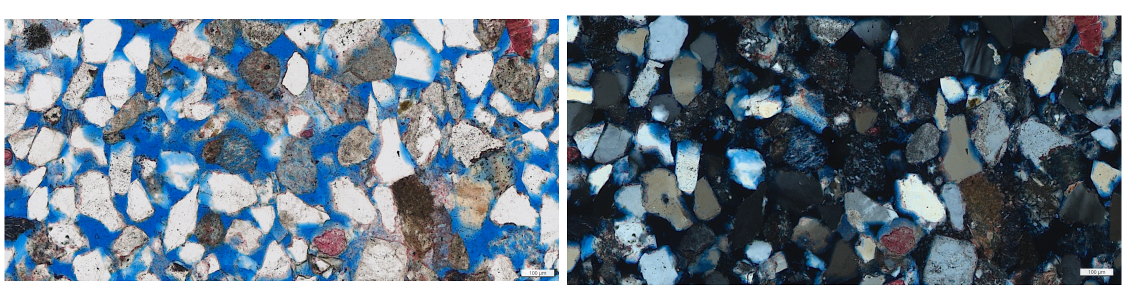

Figure 1: Thin section images of Lithic Sandstone from offshore rigs in Brazos, Texas, USA. Left is the region under white light and right is under polarised light. These images are $\sim 1.5 \times 0.5mm$, with the whole image being $\sim 20 \times 20mm$. Due to the round shape of the whole thin section approximately 350 of these size images are available from this one thin section.

algorithms. Since these images are of high resolution, they can be subdivided to make a larger number of easier to process samples, provided the image size is still large compared to the image features (i.e. grains and pores).

Proposed Method

The planned image analysis workflow is shown in figure 2 and the timeline for developing this is shown in figure 3..

An initial color space transformation is to recognise that image segmentation by most methods are sensitive to this choice. Hence, experimentation with different color spaces might improve performance. The segmentation method will follow the proven method of over segmentation with the SLIC algorithm followed by recombination, since the exact number of grains in any one image is unknown. Though implementations of SLIC exist in the scikit-learn package, a key difficulty will be incorporating both polarised and white light images into this as no existing implementation of MSLIC has been found. Recombination will use a combination of multiple merging criteria, the exact ones yet to be determined. It will follow previous work [1,6] and likely include color similarity via spectrum comparison and texture similarity via Gabor filters [9].

The classification of segments will be an unsupervised clustering problem, likely involving the shape, texture and color distribution of each segment. Hierarchical clustering approaches, namely AGNES [9], are likely a good choice here as we expect grains of different material to be significantly different that a minimum distance between them in feature space can be defined. Also the ability for outliers, which are common in such sedimentary rocks, to not impede the algorithm is very advantageous. Dimension reduction through PCA or Autoencoders might also be useful here to allow for more features to be considered in the clustering. Once clustered into distinct types a human can label each group giving a semi-automated process. This is not ideal but makes the problem more achievable and robust, as well as making the extension of identifying each grains type a simple matter of user choice.

Once separated grain size and porosity can be compared to the known values to assist in estimating hyperparameters. Testing this routine on unseen images of similar type will then evaluate its performance.

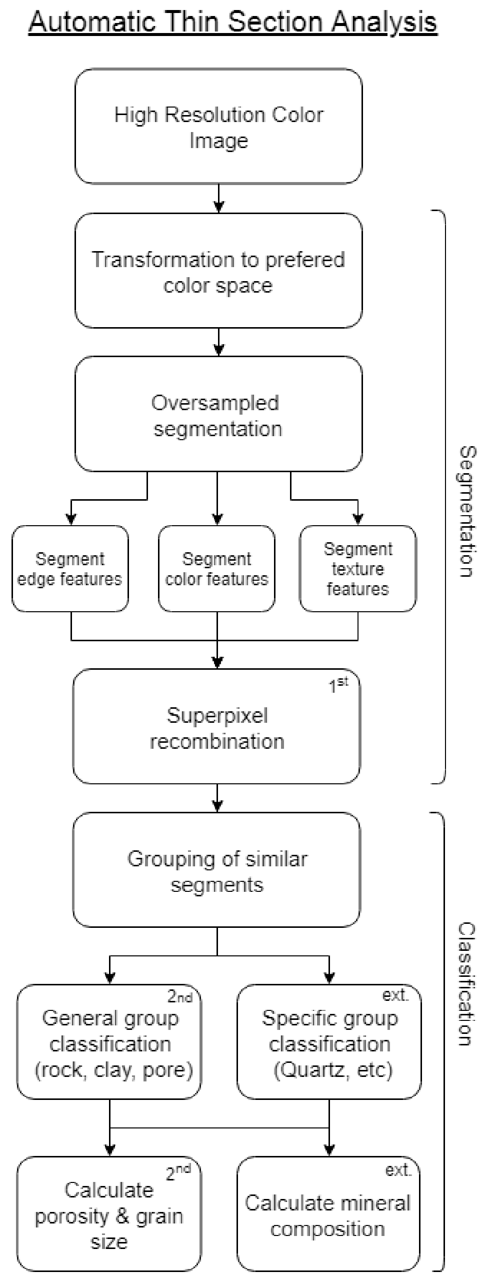

Figure 2: Flow chart for image processing pipe line. The pipeline is broken into two categories, segmentation and classification. The corner text indicates where the project aims is being achieved by this process.

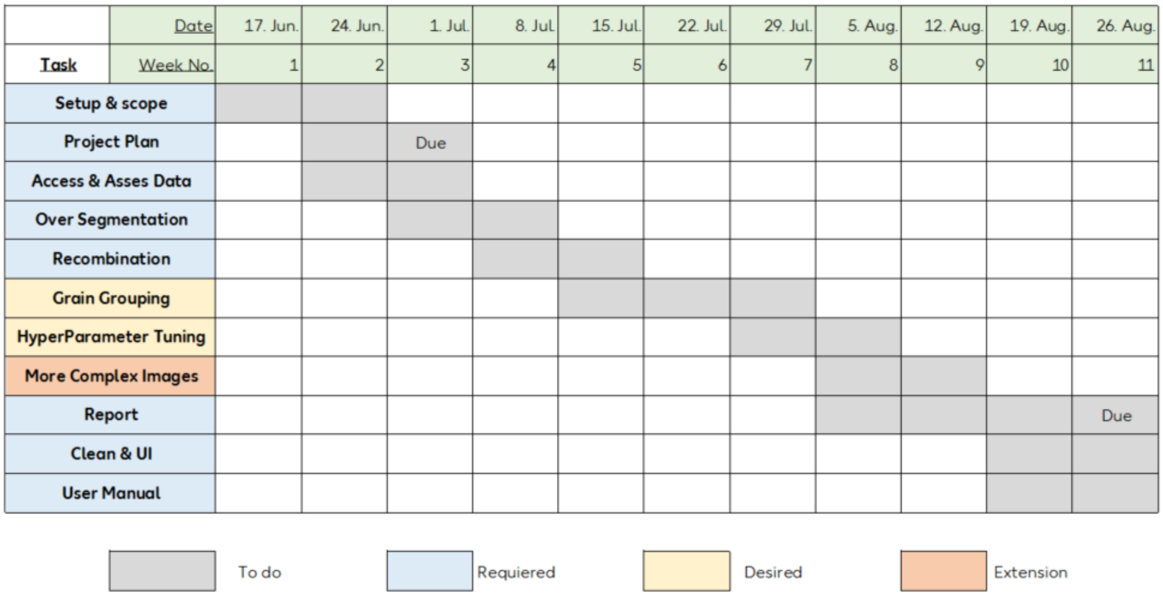

Figure 3: Gantt chart for the timeline developing the framework outlined in figure 2. Time is roughly divided equally between the segmentation and classification aims. The color of a task determines it's priority incase of delays.

Project Aims

The first aim of this project is to achieve a reasonable segmentation of grains, enough to convince an expert that the analysis was done by another expert (i.e. a segmentation Turing test). Following this, the second aim is to obtain the porosity and average grain size. If the grain classification includes grain type this can be extended to include grain composition (i.e. the fraction of different materials). Success on a given image is defined as 80% or greater agreement with values found by model analysis.

Future Work

The natural extension to this project is the expansion to more complex images. By this we mean ones with less distinct grain boundaries and a wider range of grain colors and textures. This usually coincides with a lower porosity for more densely compacted grains and larger range of grain colors and textures. Demonstrating the algorithm to work on such images would be a demonstration of the techniques versatility. Another possible extension to perform the further classifications normally carried out by a geologist from the segmentation. These include rock type (e.g. carbonate, shale), grain sorting (variation of grain sizes) and argilaceous content (amount of non-granular material).

Word Count: 1600 (excluding references and figure captions)

References

[1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, pp. 2274–2282, Nov 2012.

[2] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, pp. 679–698, Nov 1986.

[3] F. Jiang, Q. Gu, H. Hao, N. Li, B. Wang, and X. Hu, "A method for automatic grain segmentation of multi-angle cross-polarized microscopic images of sandstone," *Computers & Geosciences*, vol. 115, pp. 143 – 153, 2018.

[4] S. Yuheng and Y. Hao, "Image segmentation algorithms overview," *CoRR*, vol. abs/1707.02051, 2017.

[5] G. B. Michel Couprie, "Topological gray-scale watershed transformation," in *Topological gray-scale watershed transformation*, vol. 3168, 1997.

[6] R. Kachouri, M. Soua, and M. Akil, "Unsupervised image segmentation based on local pixel clustering and low-level region merging," in *2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, pp. 177–182, March 2016.

[7] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *International Journal of Computer Vision*, vol. 59, pp. 167–181, Sep 2004.

[8] Z. Li and J. Chen, "Superpixel segmentation using linear spectral clustering," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[9] O. Demir and B. Dogan, "Unsupervised image segmentation using textural features," *International Journal of Signal Processing Systems*, vol. 5, pp. 112–115, 09 2017.

[10] R. Richa, T. Mukerji, G. Mavko, and Y. Keehm, "Image analysis and pattern recognition for porosity estimation from thin sections," in *Image analysis and pattern recognition for porosity estimation from thin sections*, vol. 25, 01 2006.

[11] E. Berrezueta, M. J. Domínguez-Cuesta, and Ángel Rodríguez-Rey, "Semi-automated procedure of digitalization and study of rock thin section porosity applying optical image analysis tools," *Computers & Geosciences*, vol. 124, pp. 14 – 26, 2019.