doi:10.1017/S1431927619001752
Microsc. Microanal. 25 (Suppl 2), 2019

A Combination of Un-supervised and Supervised Machine Learning Method for Auto-thresholding Scanning Electron Microscopy Images

Bianzhu Fu[1], Thomas Parson[1], Atanu Das[1], Michael Deangelo[1], Emmanuel Appiah-Amponsah[1]

[1.] Entegris, Inc., 7 Commerce Dr, Danbury, CT 06810

Image thresholding is one of the simplest but still effective image segmentation methods [1] that can find applications in many areas, especially in microscopy. Most microscopy images are grey scale images, from which thresholding produces binary images so that additional information such as phase volume, number of particles amongst others can be calculated. While there are several commercial software and freeware available for image segmentation, auto-thresholding and auto-analysis is preferred due to the advantages of faster throughput of analysis and consistency. This is especially useful in manufacturing environments where higher numbers of images are generated.

While there are several auto-thresholding methods that are currently available, namely, histogram shape-based method [2], clustering-based method [3], entropy-based method [4], and local method [5], the work discussed utilizes a combination of un-supervised and supervised machine learning method to auto-threshold SEM images of chemical-mechanical-planarization (CMP) silica slurry particles on a substrate. The method works by calculating the area fraction of slurry particles from the SEM images generated. This measured parameter is then used as a baseline to evaluate the effectiveness of the post-CMP solutions in removing residue slurry particles. Figure 1 shows the auto-thresholding results using K-means clustering algorithm [6]. Herein, two clusters are selected: one corresponding to the slurry particles and the other corresponding to the background of the SEM images. Figures 1(a) and 1(b), indicate the effectiveness of the algorithm for SEM images with particle area fraction larger than 0.1%. However, in figures 1(c) and 1(d), the approach does not seem to work well due to the smaller area fractions < 0.1%, figure 1(e) and 1(f). Similar problems occur with some other auto-thresholding method as well.

To circumvent the challenges encountered, a novel approach is proposed which uniquely combines the un-supervised and supervised learning method. Since the un-supervised K-means clustering algorithm works well with particles area fraction > 0.1%, these images along with the K-means clustering thresholding results can be used as training data for supervised learning algorithms and the trained model can be applied to auto-threshold images with much less slurry particles. This method is implemented in Python with Sklearn [6] and Skimage [7] libraries. The results are shown in Figure 2. Figure 2(a) shows good linear relationship between area fractions measured from manual and automatic thresholding. Figure 2(b) and 2(c) are example original and segmented binary images with particle area fraction < 0.1% using this method. This method can be applied in atomic force microscopy (AFM) images as well where the surface particles need to be subtracted from images for surface roughness calculations and the results will be discussed as well.

References:

[1] Shapiro, Linda G. & Stockman, George C. in "Computer Vision" (Prentice Hall, Pearson) p. 83
[2] Shapiro, Linda G. & Stockman, George C. in "Computer Vision" (Prentice Hall, Pearson) p. 305-355
[3] B. Sathya, R. Manavalan, International Journal of Computer Applications **29** (2011), p. 27-32

[4] S. Khattak *et al*, International Journal of Computer Control **9** (2015), p. 667-671
[5] Z. Wang, Expert Systems with Applications **87** (2017), p. 30-40
[6] Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011
[7] S. Walt *et al* and the scikit-image contributors. scikit-image: Image processing in Python. PeerJ 2:e453 (2014) http://dx.doi.org/10.7717/peerj.453

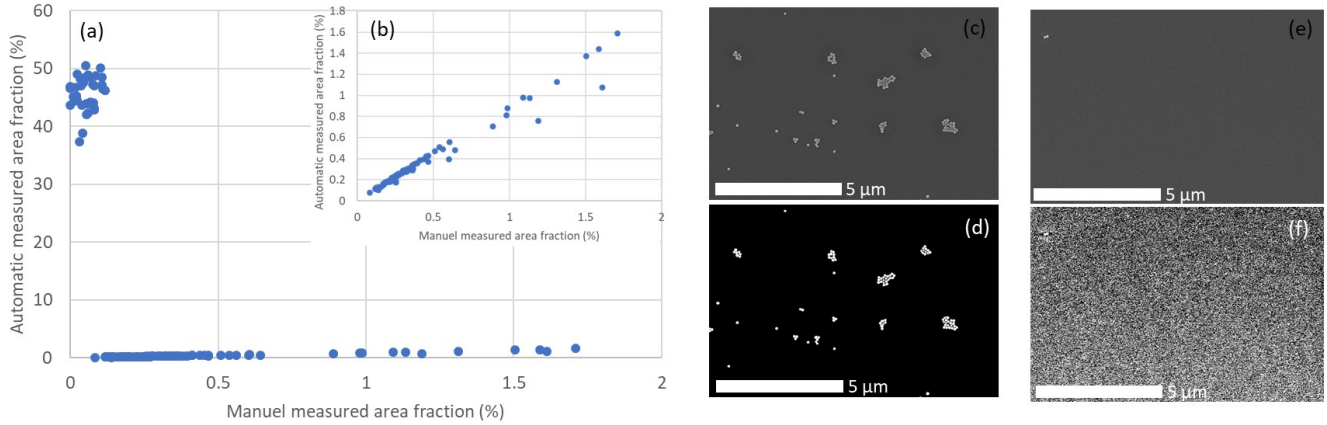

Figure 1. (a) Plot of manual and automatic measured area fraction of slurry particles on the substrate surface, noticing big measurement error occurring at area fraction < 0.1%; (b) Plot of manual and automatic measured area fraction of slurry particles on the substrate surface for area fraction > 0.1%; (c) and (d) are example original images and segmented binary images with area fraction > 0.1%; (e) and (f) are example original images and segmented binary images with area fraction < 0.1%.

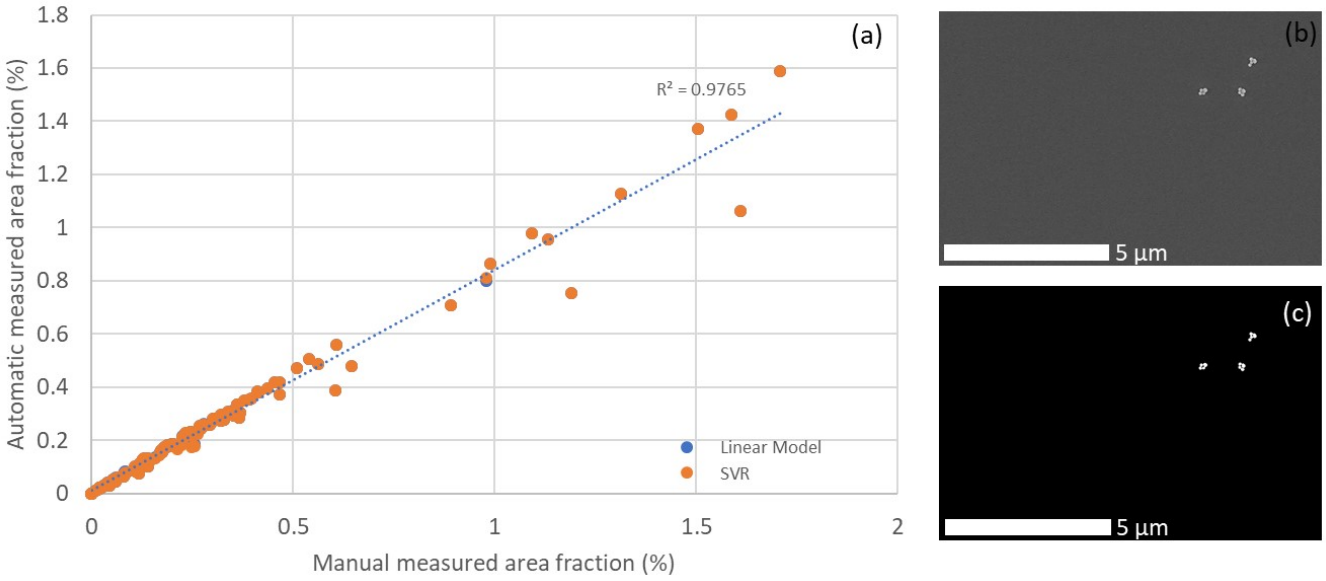

Figure 2. (a) Plot of manual and automatic measured area fraction of slurry particles on the substrate surface using linear and supporting vector regression method; (b) and (c) are example original images and segmented binary images with area fraction < 0.1%.

https://doi.org/10.1017/S1431927619001752 Published online by Cambridge University Press