Icarus 229 (2014) 408–417

Contents lists available at ScienceDirect

Icarus

journal homepage: www.elsevier.com/locate/icarus

A martian case study of segmenting images automatically for granulometry and sedimentology, Part 2: Assessment

Suniti Karunatillake [a,*], Scott M. McLennan [b], Kenneth E. Herkenhoff [c], Jonathan M. Husch [d], Craig Hardgrove [e], J.R. Skok [a]

[a] Louisiana State University, E235 Howe-Russell, Baton Rouge, LA 70803, United States
[b] Department of Geosciences, State University of New York at Stony Brook, Stony Brook, NY 11794, United States
[c] U.S. Geological Survey, Astrogeology Science Center, Flagstaff, AZ 86001, United States
[d] Department of Geological, Environmental, and Marine Sciences (GEMS), Rider University, 2083 Lawrenceville Road, Lawrenceville, NJ 08648, United States
[e] Arizona State University, School of Earth and Space Exploration, Tempe, AZ 85287-1404, United States

ARTICLE INFO

Article history:
Available online 9 October 2013

Keyword:
Mars, surface
Data reduction techniques
Image processing

ABSTRACT

In a companion work, we bridge the gap between mature segmentation software used in terrestrial sedimentology and emergent planetary segmentation with an original algorithm optimized to segment whole images from the Microscopic Imager (MI) of the Mars Exploration Rovers (MER). In this work, we compare its semi-automated outcome with manual photoanalyses using unconsolidated sediment at Gusev and Meridiani Planum sites for geologic context. On average, our code and manual segmentation converge to within ~10% in the number and total area of identified grains in a pseudo-random, single blind comparison of 50 samples. Unlike manual segmentation, it also locates finer grains in an image with internal consistency, enabling robust comparisons across geologic contexts. When implemented in Mathematica-8, the algorithm segments an entire MI image within minutes, surpassing the extent and speed possible with manual segmentation by about a factor of ten. These results indicate that our algorithm enables not only new sedimentological insight from the MER MI data, but also detailed sedimentology with the Mars Science Laboratory's Mars Hand Lens Instrument.

1. Introduction

In a companion work (Karunatillake et al., 2014) we developed an algorithm for semi-automated photoanalyses of unconsolidated sediment. This served an outstanding need (e.g., Cabrol et al., 2008; Karunatillake et al., 2010a; McGlynn et al., 2011) in the planetary community to sedimentologically analyze an extensive database of high resolution images of martian sediment. Developed primarily for the Microscopic Imager (MI) (Herkenhoff et al., 2003, 2008) of the Spirit and Opportunity rovers, the algorithm segments images containing a diverse size range of unconsolidated sediment by operating separately on the fine grains that typically occupy the background (Karunatillake et al., 2014).

Gusev and Meridiani landing sites provide the geological context to assess the effectiveness of our algorithm. Despite broad compositional and physical similarities of martian sediment across landing sites (Yen et al., 2005), variation of secondary sediment phases within Gusev contrasts with that at Meridiani (McGlynn et al., 2012; McSween et al., 2010). Examples of distinctness include well-sorted sand discovered within the Columbia Hills of Gusev Crater (Sullivan et al., 2008) and lag deposits of hematite concretions at Meridiani (Calvin et al., 2008). Such differences reflect the contrasting geology of the two sites, with Gusev's evolution as an impact crater (e.g., Squyres et al., 2004) and Meridiani plains potentially representing a site of ancient ground water upwelling (Andrews-Hanna et al., 2010). Possible pyroclastic constructs such as Home Plate in the Columbia Hills (e.g., Lewis et al., 2008) at Gusev and the sulfate-bearing Burns sedimentary strata at Meridiani (McLennan et al., 2005) exemplify the geologic contrast.

In this manuscript, we compare the semi-automated results of our algorithm with the manual counterpart using MI images of unconsolidated sediment at Gusev and Meridiani. Given deficiencies we experienced (Karunatillake et al., 2010a) in manual segmentation, this assessment serves to reveal relative strengths and weaknesses of both methods. Furthermore, in the planetary context where physical samples of sediment remain unavailable, comparing manual and semi-automated segmentation remains an optimal validation method. We advance the analysis by evaluating the internal consistency of manual segmentation, which provides additional insight into the issues experienced in our earlier work (Karunatillake et al., 2010a).

* Corresponding author. Fax: +1 (225) 578 2302.
E-mail address: wk43@cornell.edu (S. Karunatillake).

0019-1035/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.icarus.2013.09.021

2. Method and data

The spatial resolution of photoanalysis – automated or manual – depends solely on the imaging device; for example, 2 pixels would correspond nominally to 62 μm for MI images. Throughout the text, we list such examples with linear size and Wentworth classification in parenthesis for MI images, such as 2 pixels (MI: 62 μm, coarse silt), to provide context. However, the effective resolution may often differ, depending on the desired accuracy of size for smaller grains and of distribution for larger grains. We express such constraints in image pixels to facilitate scale estimates for any imager, as the product of pixel size and the linear dimension of the image in pixels.

2.1. Analytical and statistical evaluation methods

As the goal of this work, we compare the accuracy and precision of our segmentation algorithm with manual segmentation. Our analyses report primarily phi values (Φ) expressed as $-\log_2$ (mean diameter in mm). We identify relative strengths and weaknesses in two ways: (1) we compare the automated outcome with manual results; (2) we address the absence of physical samples by comparing sequential manual segmentations pairwise to establish context for any disparities between manual and automated segmentation.

The first comparison employed 50 sample images, each at least a 50×50 pixel size pseudo-random excerpt from a martian MI image. The size of the excerpts was limited by manual segmentation; even this small image takes approximately 20 min for a sedimentologist to analyze. Automated segmentation was applied to the entire MI image from which the samples were extracted. We ensured sufficient statistical validity by using 10 excerpts each from five MI images of martian sediment. Two of the original images were used to develop the algorithm, while the remaining three were chosen pseudo-randomly from both Meridiani and Gusev. While the MIs on the two rovers are practically identical (Herkenhoff et al., 2008), significant textural differences – including an abundance of hematitic spherules at Meridiani (Klingelhöfer et al., 2004) – at the two nearly antipodal sites on Mars helped to assess the broad applicability of the code. A single-blind approach with samples segmented manually without seeing the automated output ensured additional rigor; we viewed the whole MI image only where manual segmentation proved particularly difficult.

We segmented 11 pseudo-random excerpts from MI images to assess the internal consistency of manual segmentation. This was achieved by segmenting the images manually once, waiting a few days to lapse to prevent residual memory effects, and segmenting anew. Comparing the re-segmented to the originals revealed a degree of consistency attainable by a human operator.

2.2. Bulk variables

Comparisons consisted of two components, bulk and parametric comparisons. The bulk comparisons of automated and manual segmentation were threefold: (1) difference in the total number, *dN*, (2) difference in the total area, *dA*, and (3) areal mismatch, *mA*, of segmented grains for each of the 50 samples. We minimized bias from varying grain numbers across samples by normalizing *dN* of each sample to the total number of grains from automated segmentation; similarly, *dA* and *mA* were normalized to the automatically segmented total grain area. For succinctness, subsequent references to these variables denotes normalized values even where unstated. Areal mismatch in particular can be revelatory of consistency between one segmentation method and another, which we computed as the total area of segmented components (i.e., both grains and unclassified space) that do not overlap between two independent segmentations of an image.

Grains in imaged sediment samples may distribute in unique, non-Gaussian modes, but the *dN* and *dA* distributions quantify the *difference* between manual and automated segmentation. If the two methods yield fundamentally similar results, the differences would follow Gaussian distributions about a zero mean (e.g., Karunatillake et al., 2010b). Consequently, automated and manual segmentation can be thought to converge if *dN* and *dA* each approximates a random normal distribution (i.e., Gaussian), perhaps to the extent of zero mean and unit variance (i.e., standard normal). The narrowness of the distribution represented by kurtosis would constrain the relative strength of convergence between automated and manual segmentation.

2.3. Battery of statistical tests

We may not observe sharply-peaked leptokurtic, zero-mean, Gaussian distributions that allow decisive conclusions. But flat and broad platykurtic, many-peaked multimodal, or different-tailed skewed distributions may reveal conditions of convergence. We applied a battery of tests (c.f., Karunatillake et al., 2010b) to each of the three bulk variables (*dN*, *dA*, and *mA*) to assess such deviations from a Gaussian as follows:

1. Shapiro–Wilk test (Mathematica inbuilt) of overlap with estimated normal distribution.
2. Plot of cumulative distribution function (CDF) of sample versus that of the estimated Gaussian; overlain with standard normal for context.
3. Kolmogorov–Smirnov probability (K–S; Mathematica inbuilt) of overlap with estimated normal distribution (Karunatillake et al., 2010b).
4. Estimated normal distribution, excess kurtosis, and skewness.
5. Probability scale plot of the variable's CDF versus value overlain with that of the estimated Gaussian.
6. Frequency histogram of values subject to Kernel density estimate (KDE) with a Gaussian Kernel, overlain with estimated and standard normal histograms for context.

The standard Shapiro–Wilk test revealed goodness-of-fit with the null hypothesis of bulk variables (*dN*, *dA*, and *mA*) drawn from a normal distribution. Consequently, the probability returned by the test represents the likelihood that a distribution sampled from a Gaussian could diverge at least as much as observed due to chance alone. A suitably high probability permits the null hypothesis to be accepted, while a low probability will encourage its rejection and attribute divergence to distinctness of distributions.

The second assessment allowed us to visually compare the distribution of the bulk variable with the Gaussian that best approximates it. The standard normal overlay provided context for divergence, where convergence would be reflected by an overlap between the overlay and the CDF. The K–S test advances the visual comparison quantitatively, albeit with high sensitivity to divergence between the distributional tails.

The fourth test provided a parametric snapshot of the bulk variable's distribution, with the mean and standard deviation of the Gaussian estimate providing important input parameters for the sixth test. Measures of excess kurtosis and skewness in particular provided helpful summary parameterizations to identify the nature of the distribution, and in turn the similarities and differences between manual and automated segmentation. Meanwhile, the fifth test visually evaluated whether the actual distribution of a bulk variable corresponds to a normal distribution, with the

straightness and proximity to a line of reference revealing the nature of convergence.

Lastly, the sixth test allowed an immediate comparison among the estimated Gaussian probability density function (PDF), the observed PDF, and extent of difference from a standard normal distribution. The KDE method of generating the frequency histogram for the bulk variable enabled us to identify the shape and modality of the actual distribution above the level of noise (e.g., Sheather and Jones, 1991). We supplemented the six tests with a visual inspection of sorted *dN*, *dA*, and *mA* values across all 50 samples. Additionally, the plot from test (5) identified samples representing both extreme divergence and maximum convergence between human and code.

Of the three bulk variables, *dN*, *dA*, and *mA*, the last, denoting areal mismatch, is nonnegative. Consequently, the ideal outcome, of convergence between software and manual segmentation, would be a distribution skewed strongly to the right with an asymptotically zero mean and high leptokurty. The convergence or divergence on this basis was determined visually with test (5); quantitatively we used tests (1) and (3).

2.4. Parametric assessment

Bulk comparisons were reinforced with two parametric comparisons. The first compares the mean and standard deviation of the grain diameter distribution of one segmentation method with another for each of the 50 samples. Motivated by Karunatillake et al. (2010a), this employs the difference between the mean values with standard deviation propagated accordingly for manual versus automated or manual versus manual comparisons. A sorted plot of mean differences with propagated standard deviations as error bands provided preliminary qualitative insight. Additional assessments utilized the same battery of tests as for the bulk variables. Since agreement between two methods would yield identical means, the distribution of differences in sample average values between the two methods would ideally converge with a Gaussian distribution where convergence would return a distribution mean approximating zero.

The second parametric assessment employed the K–S distributional comparison (e.g., Karunatillake et al., 2010b) between manual and automated segmentation for each sample image, where a larger probability confers greater confidence that the two distributions are identical. Ideally then, a histogram of K–S probabilities for all 50 samples would be skewed to the left with the peak asymptotically approaching unity. Even if not, peaks of high K–S probability in a multi-modal distribution would indicate that the two methods – manual versus automated – converge satisfactorily. Such tendencies were evaluated by applying our battery of tests to the distribution composed of the K–S probability for each of the 50 samples. A KDE frequency histogram of the K–S probabilities provided additional visual clarity to the analysis.

3. Results and discussion

An unexpected outcome of our assessment was an approximate estimate of the time for a sedimentologist to segment an entire MI image manually, individually identifying most of the $(10–20) \times 10^3$ grains. We estimate no less than 5 days, since each of the roughly 400 50×50 samples constituting one MI image consumes approximately 20 min to segment completely. A more likely time frame would be about sixteen 8-h workdays and perhaps even longer given the tedium of the effort. In addition to internal consistency documented subsequently, automation yields a clear advantage in speed, typically taking no more than 5 min to segment an entire MI image in the Mathematica-8 software environment, including manual adjustment of guide parameters. Without adjustment, the processing time is only about a minute. Similar processing speeds could be expected on a typical business laptop computer for other software environments such as IDL, R, and MATLAB.

3.1. Battery of statistical tests

Upon quantitative evaluation, including the battery of six tests described in Section 2.3, we identified informative differences with manual segmentation and intrinsic limitations of photoanalyses. In bulk comparisons, the striking disparity was in areal mismatch. As Fig. 1 illustrates with KDE frequency histograms corresponding to Test 6 in Section 2.3, areal mismatch frequently exceeds 50% of the total area of grains segmented by the algorithm. This causes the ~0.53 mean and ~0.19 standard deviation observed in the Gaussian (Fig. 1) best fit to the distribution of *Am* values (*Am* normalized to the total area of grains segmented by the algorithm for reasons described in Section 2.2, which we still refer to as *Am* to avoid repetitious use of "normalized." The other bulk variables are also implicitly normalized).

Tests indicate a non-normal distribution of *Am*, enabling additional insight from the actual distribution relative to its Gaussian fit. The Shapiro–Wilk probability of *Am* converging with a normal distribution (Test 1) yields 1.4E−3, while the K–S probability of the same (Test 3) yields 1.2E−2. Consequently, both tests show that areal mismatch, as represented by *Am* values, distributes non-normally at a 0.05 (i.e., 95% statistical confidence) threshold, suggesting that the best-fit Gaussian in Fig. 1 may not represent the actual *Am* distribution. The comparative CDF plot and probability scale plot of the CDF for the *Am* distribution (Fig. 2) support this inference, as does the −0.97 skewness. Consequently, the peaks and shoulders of the KDE distribution (Fig. 1) provides meaningful information on the nature of the actual *Am* distribution.

The shoulders of the *Am* distribution (Fig. 1), corresponding to areal mismatches at 20% and 40%, normalized to the total area of grains segmented automatically, suggest better areal convergence between manual and automated segmentation than evident from the best fit Gaussian peak at 50% mismatch. For example, on a probability scale plot of areal mismatch, areal mismatch decreases

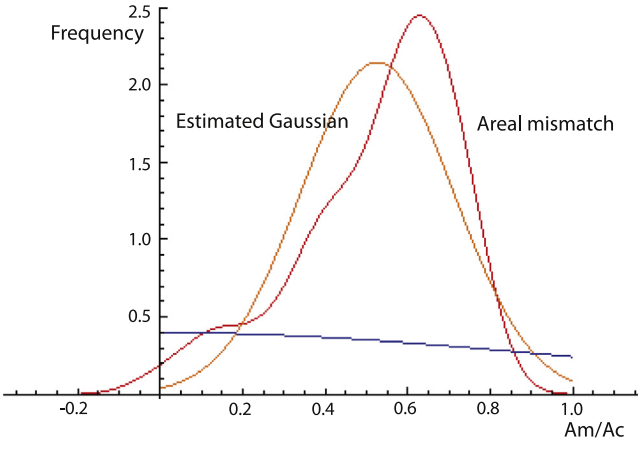

Fig. 1. KDE histogram of total area of mismatch between grains segmented manually and those segmented automatically as implemented in Mathematica-8 (*Am*). Normalized to the total area of grains segmented by the algorithm (*Ac*). The histogram of this 50 sample dataset has been smoothed with a Gaussian Kernel for visual clarity. Best fit Gaussian to the distribution is shown for reference, while the standard normal distribution (blue) provides shape context. Possible multimodality, and a notable shoulder at approximately 10% mismatch, are evident in the histogram. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

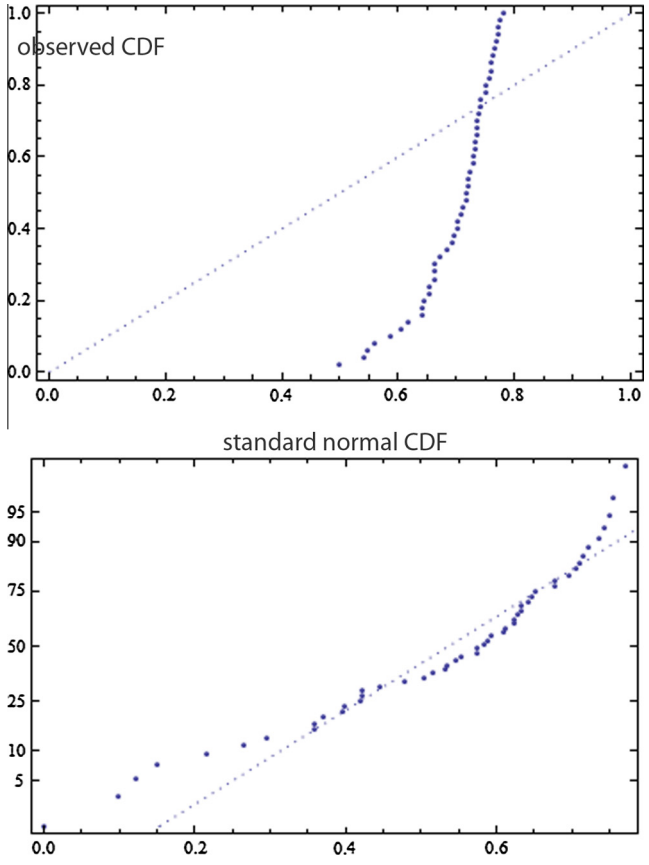

Fig. 2. Top graph showing probability plot of the cumulative distribution function (CDF) of the relative areal mismatch (*mA*) between the two methods relative to those segmented automatically. This normalization to the total area of grains derived from the algorithm minimizes bias from the variability of total grain area across different MI image excerpts. Probability plot reveals extent of correspondence between the actual CDF on *y*-axis and the standard normal distribution on the *x*-axis. Difference between the actual data (blue dots) and the line of perfect correspondence (dotted line) reflects deviation of the *mA* CDF from the standard normal distribution. Section 3.1 presents the insight from this probability plot. Bottom graph showing the scaled (logarithmic) probability plot identifies the strongest deviations from normality. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

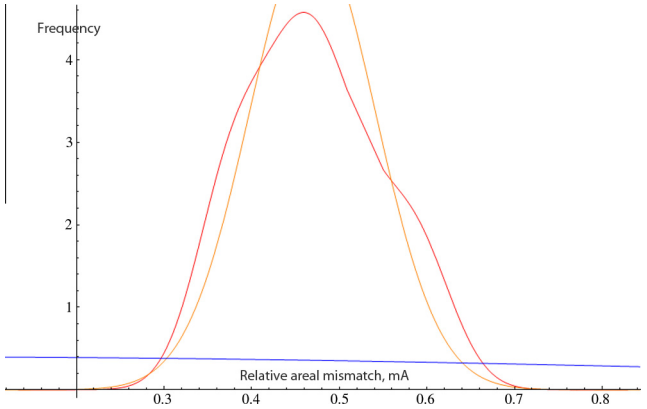

Fig. 3. KDE histogram of relative areal mismatch (*mA*, *x*-axis) between temporally spaced manual segmentations of identical image excerpts. We compute *mA* relative to the second manual segmentation primarily to preventing bias from varying area of grains across different MI images, and secondarily to ensure a consistent denominator. KDE applied with a Gaussian Kernel yields the actual distribution of *mA* (red curve) based on 10 sample excerpts, a subset of the 50 excerpts from 10 MI images. The size of each excerpt, at no smaller than 50×50 pixels reflects the feasibility limits of segmenting manually. The orange curve identifies the best-fit Gaussian to the *mA* distribution. For reference, we also plot the KDE histogram of a standard normal distribution in blue. Areal overlap in grains from sequential manual segmentation would yield a sharply leptokurtic Gaussian skewed sharply to the right. In Section 3.1 and related text, we discuss what deviations from this ideal, as evident from both the best fit and actual KDE histograms, may imply. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

to 10% when grains tens-of-pixels across (MI: ~300 μm, medium sand) are present. This may relate fundamentally to the smallest pixel size of grains that may be segmented correctly (Graham et al., 2005b), which we extrapolate for the planetary context in our companion work (Karunatillake et al., 2014). This may also reflect inferior internal consistency of manual segmentation.

Areal mismatch values of manual segmentation compared to itself demonstrates a major weakness: substantial inconsistency ranging 35% to 50%. The peak of the resulting *Am* distribution suggests approximately 45% internal inconsistency typically, with inconsistency as high as 60% evident from the distributional shoulder (Fig. 3). This suggests that instances of higher areal mismatch between automated and manual segmentation reflects mostly human, not algorithmic, limitations. Furthermore, the internal inconsistency of segmenting manually, as evident in Fig. 3, contrasts starkly with the internal consistency of automated segmentation.

The susceptibility of manual segmentation to internal inconsistency aside, bulk comparisons (*dN*, *dA*) between manual and automated segmentation were favorable. The *dN* distribution indicates that the two methods may produce differences in the number of grains as high as 60% (normalized to the automated number of grains) for individual samples (Fig. 4). KDE histogram of the samples (Fig. 5 red curve) shows that manual segmentation may differ from the automated counterpart between under-estimating the number of grains by about 18%, corresponding to the KDE histogram peak, and over-estimating the number by about 35%, corresponding to the shoulder. Nevertheless, the *dN* distribution itself remains approximately Gaussian based on tests 1 (Shapiro–Wilk) and 4 (K–S). Accordingly, the best fit Gaussian (Fig. 5) peak suggests that manual segmentation typically underestimates the number of grains by no worse than 7% (c.f., Graham et al., 2010) with a distributional standard deviation of ~35%.

Differences in the total area of grains, *dA*, normalized to the total computed area suggest a non-Gaussian distribution as evident in the KDE histogram (Fig. 6) corresponding to Test 6. Specifically, the normalized *dA* distribution diverges from a Gaussian given a Shapiro–Wilk probability of ~1E−3 (Test 1) and K–S probability of ~7E−3 (Test 3), as also evident in the CDF plot (Test 2) and probability scale plot (Test 5) shown in Fig. 7. Consequently, the KDE histogram of the sample generated as a component of Test 6 provides the most insight into the nature of the *dA* distribution (Fig. 6). This shows the primary *dA* peak at approximately −4% (normalized to area of automatically segmented grains), with two secondary peaks at roughly −35% and 48%, along with a diminished peak at ~38%. The relative heights of the peaks suggest that manual segmentation typically underestimates the total area of grains by about 4% relative to the total area of computed grains. Even the less informative Gaussian best fit suggests an underestimate less than 13%. Smaller differences may be attainable depending on the sizes of grains within a sample.

Instances of manual-automated convergence are complex outcomes of both grain size distribution and abundance of larger grains that do not necessarily reflect a single factor. Nevertheless, a preponderance of grains sufficiently large for a human to identify without magnification generally enhances convergence with the algorithm. Fig. 8 illustrates an example of convergence in both placement (reflected in a low *mA*) and area (reflected by a low absolute *dA*) of grains.

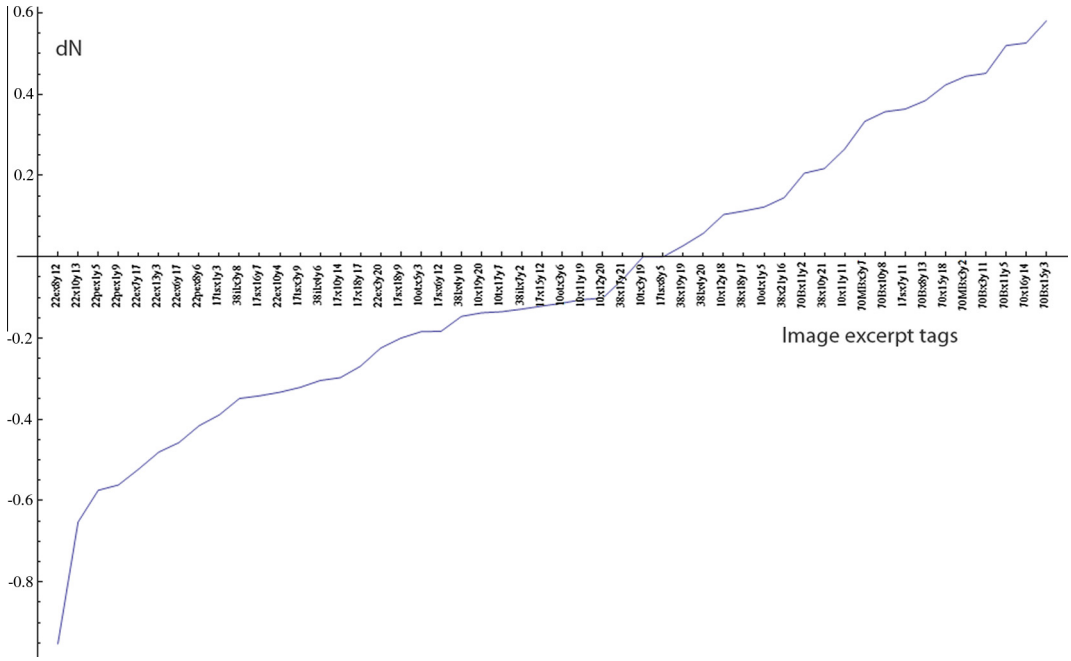

Fig. 4. Relative difference (*dN*, *y*-axis) in total number of grains as segmented automatically and manually as discussed in Section 3.1. For each sample image, we compute *dN* relative to the total number of grains as segmented automatically, preventing bias from varying number of grains across different images. The *x*-axis acts solely as a placeholder for the sorted *dN* values, with the axis tick marks corresponding to the 50 sample excerpts from 10 MI images tagged by pixel location within each image. The size of each excerpt, at approximately 50 × 50 pixels accounts for the feasibility limits of segmenting manually.

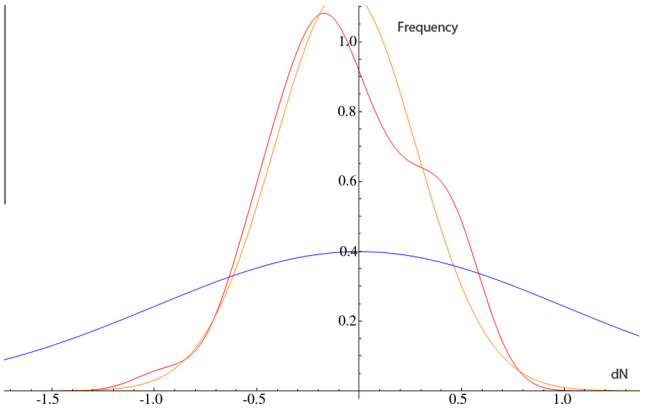

Fig. 5. KDE histogram of relative difference (*dN*, *x*-axis) in total number of grains as segmented automatically and manually as discussed in Section 3.1. We compute *dN* relative to the total number of grains as segmented automatically, preventing bias from varying number of grains across different MI images. KDE applied with a Gaussian Kernel yields the actual distribution of *dN* (red curve) based on 50 sample excerpts from 10 MI images. The size of each excerpt, at no smaller than 50 × 50 pixels accounts for the feasibility limits of segmenting manually. The orange curve identifies the best-fit Gaussian to the *dN* distribution. For reference, we also plot the KDE histogram of a standard normal distribution in blue. Identical results from automated and manual segmentation would yield a Gaussian with zero mean for the *dN* distribution. In Section 3.1 and related text, we discuss what deviations from this ideal may imply. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

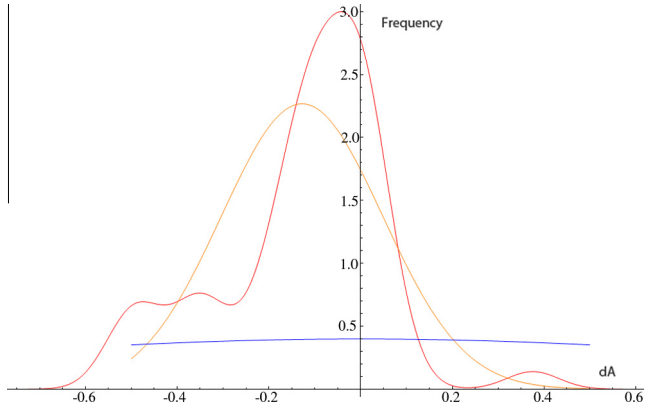

Fig. 6. KDE histogram of relative difference (*dA*, *x*-axis) between total area of grains as segmented automatically and manually as discussed in Section 3.1. We compute *dA* relative to the total area of grains as segmented automatically, preventing bias from varying number of grains across different MI images. KDE applied with a Gaussian Kernel yields the actual distribution of *dA* (red curve) based on 50 sample excerpts from 10 MI images. The size of each excerpt, at no smaller than 50 × 50 pixels, accounts for the feasibility limits of segmenting manually. The orange curve identifies the best-fit Gaussian to the *dA* distribution. For universal shape and peak reference, we also plot the KDE histogram of a standard normal distribution in blue. Identical results from automated and manual segmentation would yield a Gaussian with zero mean for the *dA* distribution. In Section 3.1 and related text, we discuss what the deviations from this ideal may imply. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Overall, our bulk comparisons – revealing mostly minor disparities in total area and number of grains – demonstrate that manual and automated segmentation yield grains of consistent size. However, a human operator is challenged when the image needs to be magnified to identify pixels within a grain, then reduced to locate grain boundaries. In essence, grain boundaries are easier to identify for a human un-magnified since the grayscale transitions between grain edge and background become clearer; but human hand–eye coordination is insufficient to mark individual pixels without magnification. This may cause the relatively chaotic (i.e., internally inconsistent) manual placement of segmented grains that we observed (Fig. 3). In contrast, automation possesses the major strength of positioning grains with internal consistency not only within a single image, but also across different images.

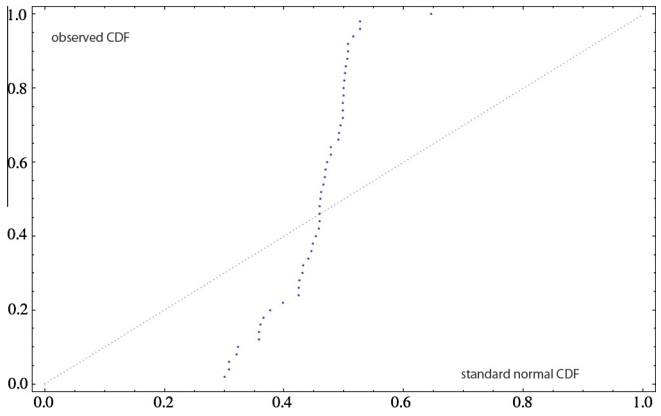

Fig. 7. Probability plot of the cumulative distribution function (CDF) of the relative difference in total area of grains (dA) as segmented manually relative to those segmented automatically. The relative calculation minimizes bias from variability of total area attributable to grains across different MI image excerpts. Probability plot employs correspondence between the actual CDF on y-axis and the standard normal distribution on the x-axis. Deviation of the dA CDF from the standard normal distribution reflected by the difference between the actual data (blue dots) and the line of perfect correspondence (dotted line). Section 3.1 presents the insight from this probability plot. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.2. Parametric assessment

The summary parametric test provided additional insight into the bulk comparisons. First, we visually compared the distribution of areal mean Φ values for the 50 samples resulting from the automatic segmentation with that from manual segmentation (Fig. 9). We employed Gaussian KDE histograms as described for Test 6 in Section 2.2, to focus on distributional features above noise. The comparison highlights convergence both at fine and coarse scales of grain size.

We advanced the visual comparison of areal mean Φ values with the difference of corresponding sample pairs. This involved subtracting mean Φ of each sample from manual segmentation with that from automated segmentation. We illustrate this in Fig. 10 with a Gaussian KDE plot of the difference; the Gaussian best fit (orange) and standard normal PDF (blue) provide context for the nature of distributional differences. Had the two methods generated identical distributions of grain size, the 50 mean Φ differences (red) would have yielded a Gaussian distribution with mean zero. Even though the actual distribution manifests the primary peak at zero, a small peak at approximately −0.5 and a shoulder at ∼0.5 suggest slight deviations from the ideal (Fig. 10). The Gaussian estimate for the difference distribution showing a mean of 0.12 with a standard deviation of 0.25 reflects the effect of the small peak and shoulder on magnifying variance. Summarily then, areal mean Φ values of automated and manual segmentation mostly overlap with zero in the context of propagated standard deviation.

3.3. Areal distribution of grain size compared

Beyond parametric and bulk comparisons, we directly compared the areal distribution of grains resulting from manual segmentation with its automated counterpart. This detailed comparison of the two distributions using the K–S test (described in the context of Test 3 for bulk comparisons in Section 2.3) applied to each of the 50 samples affords additional insight, yielding one K–S probability for each sample. As shown in Fig. 11, the K–S probabilities are distributed multimodally, not Gaussian. The probability (i.e., p-value) indicates the likelihood of differences at least as extreme as the observed for two random samples drawn from the same population. We may state this intuitively as the likelihood that two samples could be at least as different as observed due to chance alone. Accordingly, a higher probability supports convergence, while a lower value supports divergence.

The best-fit Gaussian to the distribution of K–S probabilities (Fig. 11, orange) yields an average K–S probability of 25%, suggesting a 25% likelihood that even samples from identical distributions would differ by at least as much as observed. This high probability strengthens the case that our 50-sample assessment is consistent with statistically indistinguishable areal distributions of grain sizes from automated and manual segmentation. Much lower probabilities, usually less than 5%, would have established statistically distinct distributions. Furthermore, the actual distribution's minor peak approximates unity reflecting substantial convergence of the two methods. As such, the distribution of K–S probabilities constrains the degree to which differences between manual and automated areal distributions relate to chance than systematic issues.

The KDE histogram of K–S probabilities (Fig. 11, red) reveals some divergence between manual and automated segmentation. Specifically, the K–S distribution (as opposed to its best fit) manifests the primary peak at ∼5%, indicating that a substantial fraction of samples diverge between manual and automated segmentation. However, the secondary peak approximating unity and high shoulder connecting the secondary peak with the primary peak introduce substantial positive skewness to the distribution, strengthening the case for convergence represented by higher probabilities. In the context of additional analyses with the cumulative distribution of K–S values (Fig. 12), we may conclude convergence generally, albeit with instances of divergence.

The cumulative distribution of K–S probabilities enables a thorough comparison of results from automated and manual segmentation. We divide the cumulative distribution of K–S probabilities shown in Fig. 12 into three sections using thresholds of K–S probability: first the portion less than 0.1, next between 0.1 and 0.8, and finally exceeding 0.8. The first threshold identifies instances where the areal distributions of grain sizes from the two methods are less than 10% likely to originate from the same population. As stated in the introductory paragraph of this section (Section 3.1), this threshold identifies cases of less than 10% likelihood of sample distributions at least as distinct as the observed originating from the same parent population. In contrast, values exceeding the second, 0.8, threshold identify instances of negligible divergence. The intermediate range from 0.1 to 0.8 represent generally convergent distributions. Our choice of probability thresholds is actually more conservative than the general standard in hypothesis testing where the hypothesis of identical distributions – generally the null hypothesis applied to sample and population – is accepted until the probability decreases to 5% or less (Karunatillake et al., 2010b, 2014).

In the three sections of the cumulative distribution of K–S probabilities for the 50 samples, the first at K–S less than 10% also corresponds to 50% of the samples. This indicates divergent areal distributions of grain sizes for half the samples, moderated however, by the higher confidence threshold and the K–S method's high sensitivity to even subtle differences in the distributional tails. Furthermore, the remaining 50% of samples yield K–S values higher than 10%, indicating minimal divergence. Reinforcing this trend, roughly 15% of the 50 samples show compelling convergence at K–S exceeding 80% (Fig. 12). Summarily therefore, the data do not suggest systematic differences between manual and automated segmentation.

In practice, we anticipate compelling convergence between automated and manual segmentation particularly at the distributional middle as shown by the good visual agreement between

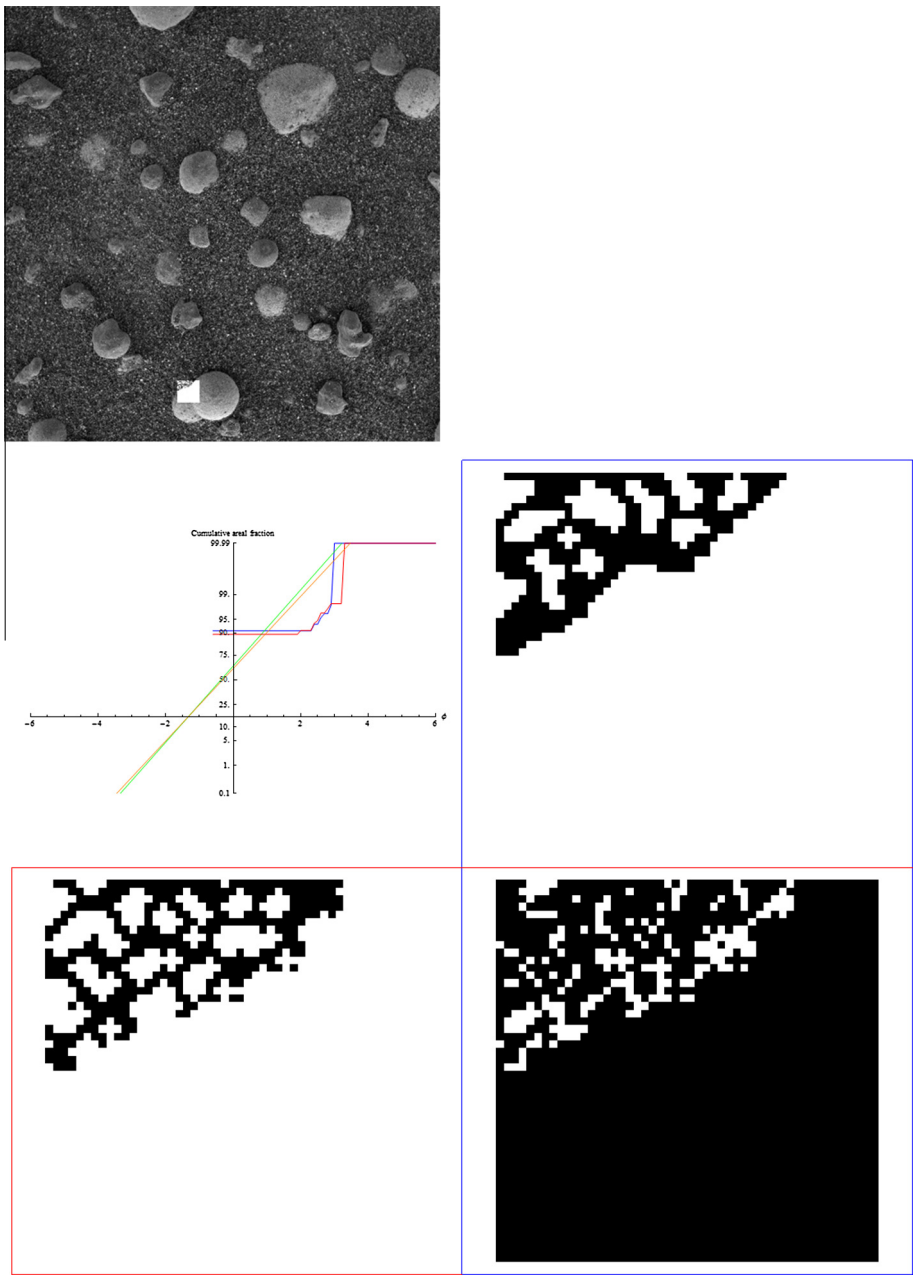

Fig. 8. Example of simultaneously minimal areal mismatch and low absolute *dA* between manual and automated segmentation highlights an instance of better agreement between the two segmentation methods. Upper context image approximately 3 × 3 cm, with 50 × 50 pixel sample excerpt highlighted as brightened square at bottom center. Lower graphs in the bottom analytical figure summarize comparisons graphically (top left) with the cumulative areal fraction distribution (CDF) of grain sizes versus phi. The blue curve corresponds to manual segmentation while the pale green straight line fit represents the best fit Gaussian. The red curve corresponds to the automated outcome with orange best fit. Binary image on the upper right shows the manual segmentation outcome, while the lower left shows the automated result. The broad white region corresponds to the portion of the compounded spherules captured in the sample window (top context image). Lower right binary image identifies pixels of mismatch as 1 (white) and overlap as 0 (black). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

mean grain size histograms (Fig. 9) as well as the secondary peak approximating unity in the histogram of K–S values (Fig. 11). Such convergence is strengthened by the conservative nature of the K–S test with its strong sensitivity to divergence in the distributional tails (Karunatillake et al., 2010a) where areal grain size distributions are most likely to diverge (Fig. 9 by Graham et al., 2005a).

The robustness of our algorithm and its software implementation in Mathematica-8 surmounts key limitations of photoanalyses at least as successfully as a sedimentologist. In the context of two-dimensional imagery for which we developed our algorithm, one such constraint is imbrication, typically yielding smaller apparent size than the actual by the overlap of grains (e.g., Fig. 6 by Graham et al., 2010). It may also cause separate grains to appear fused (e.g., Wang, 2008). Sophisticated solutions exist in terrestrial software that employ stereoscopic information to distinguish individual grains better (e.g., Maerz et al., 1996). Such algorithms can be enhanced further to address foreshortening from the projection onto a plane. However, that comes at the cost of substantially higher complexity and computing burden (e.g., Maerz et al., 1996). A third key limitation results from the iceberg effect (Fig. 6 by Graham et al., 2010) where large grains partially concealed by neighbors may appear smaller than actual.

Some disparities between photoanalysis and manual sieving resolve easily. For example, systematic differences with square-holed

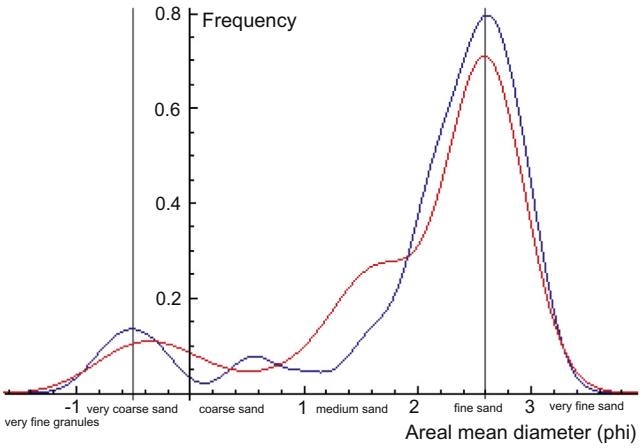

Fig. 9. Histograms of areally-weighted mean diameters of grains in Φ units for all 50 samples; smoothed with a Gaussian Kernel for visual clarity. The histogram for manual segmentation is blue, while the computed histogram is red. The two gray vertical lines highlight peak overlap, indicating convergence both at fine grain sizes and coarse grain sizes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

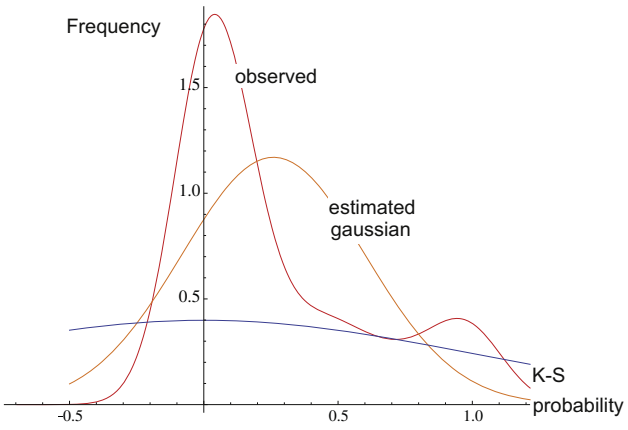

Fig. 11. Distribution of K–S probabilities for manually-derived areal distribution of grains compared to the computed for all 50 samples. Smoothing with a Gaussian Kernel for visual clarity causes the K–S distribution to appear to extend beyond 0 and unity even though it is bounded between. Estimated Gaussian (orange) and standard normal (blue) distributions provide context for the shape of the actual distribution (red) of K–S values. Peaks occur at approximately unity and zero joined by a broad shoulder. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

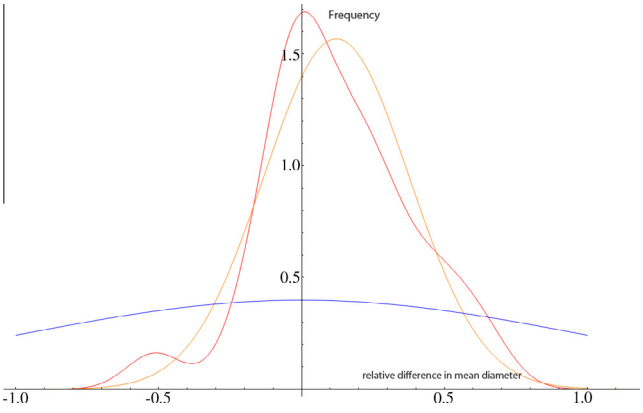

Fig. 10. KDE histogram of relative difference in mean diameter (x-axis) between automated and manual segmentation as discussed in Section 3. We compute the difference relative to the mean diameter of grains as segmented automatically, preventing bias from diverse grain distributions across different MI images. KDE applied with a Gaussian Kernel yields the actual distribution (red curve) based on 50 sample excerpts from 10 MI images. The size of each excerpt, at no smaller than 50×50 pixels, accounts for the feasibility limits of segmenting manually. The orange curve identifies the best-fit Gaussian to the distribution. For universal shape and peak reference, we also plot the KDE histogram of a standard normal distribution in blue. Identical results from automated and manual segmentation would yield a Gaussian with zero mean. In Section 3 and related text, we discuss what the deviations from this ideal may imply. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

physical sieves can be addressed with scaling factors (para. 17 by Graham et al., 2005b). Equally addressable is the bias for larger grains introduced by counting all grains along the image perimeter, mitigated by counting all grains along the top and left edges while excluding all grains on the bottom and right edges (Fig. 2 and para. 15 by Graham et al., 2005b).

A major issue with segmentation, manual or otherwise, is the difference between areal distributions and mass distributions. While the areal distributions enabled by our algorithm compare far better with the sieved mass distributions than point counts of previous planetary work (Cabrol et al., 2008; Calvin et al., 2008; Yingst et al., 2008), the scaling function between area and mass distributions is non-linear; differences cannot be corrected with a constant scale factor. Even stereoscopic algorithms (Maerz et al., 1996) would not resolve this, since the imager can only sample the uppermost layer of sediment, not a sample in bulk. It will remain a caveat whenever comparisons are made with planetary granulometry and sieve-based terrestrial counterparts. On the other hand, grid-by-number terrestrial distributions can be converted to the areal distribution or vice versa (para. 18 by Graham et al., 2005b).

3.4. Terrestrial analogs

The major caveat of our work, the physical inaccessibility of martian sediment, will be addressed by using terrestrial analogs in an upcoming project. Entailing sieving for uniformity of samples, this would effectively calibrate our algorithm independent of human vision. While beyond the scope of this study, our preliminary observations of terrestrial sediment indicate that translucent (quartz-rich) grains would challenge the algorithm by making grain edges brighter than interiors. Translucence undermines the segmentation process, since our algorithm utilizes higher reflected radiant emittance from grain interiors. We did not encounter such issues with martian sediment images and do not encounter them with Costa Rican beach sediment, since their basaltic composition minimizes translucence.

The Costa Rican beach sand (Playa Hermosa, Jaco; $\sim$ 0.1–1.0 mm grain size fine-to-coarse sand with a small fraction of finer material) had a large, locally derived basaltic component on which we placed foreground pebbles ($\sim$5–10 mm grain size fine-to-medium pebbles) (Fig. 13). Fig. 13A shows an extracted portion of the original image, taken with a Nikon D60 digital SLR camera with an 18–55 mm auto focus, variable focal length lens (the image was taken at a focal length of 55 mm), and a CCD of 3872×2592 pixels. With an effective resolution of $\sim$17 μm/pixel (this value varies slightly with focus), the image approximates the best resolution of the MAHLI camera on the Curiosity Rover.

Fig. 13B shows the segmentation results overlain on the original, with the algorithm operating in the distinct foreground mode. Note that the three foreground pebbles are resolved satisfactorily, with no background sand grains included in the foreground, except for a few contiguous grains with similar albedo adjacent to the pebbles. Visual inspection (e.g., Fig. 13C) indicates that in general,

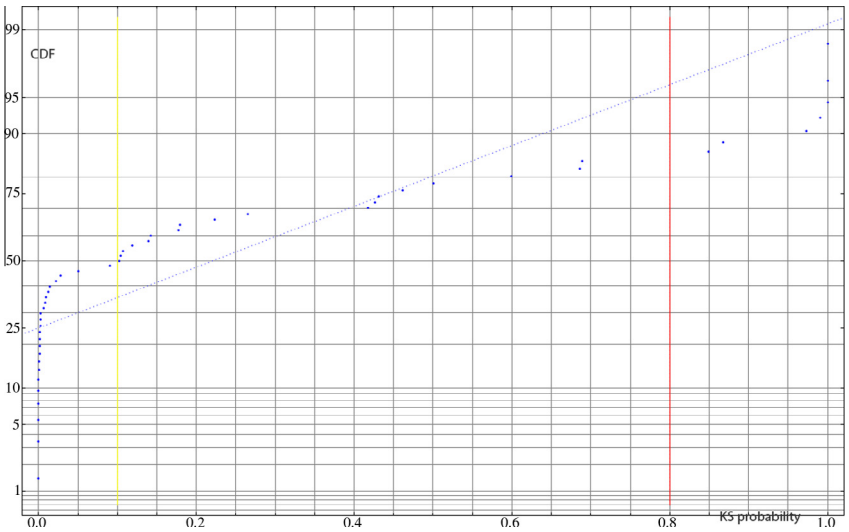

Fig. 12. Cumulative distribution function (CDF, *y*-axis) of K–S probabilities (*x*-axis) for areal distributions of automated segmentation compared to the manual counterparts. Each of the 50 sample MI excerpts yields one K–S probability. We partition the K–S probability cumulative distribution in three with two thresholds at 0.1 (yellow grid line corresponding to 10% likelihood that automated and manual distributions could have differed by at least as much as the observed due to chance alone) and 0.8 (red grid line corresponding to 80% likelihood of chance divergence). 50% of samples yield K–S values higher than 10%, while ~15% of the samples exceed 80% indicating minimal divergence. Detailed inferences from the cumulative distribution are presented in Section 3.1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

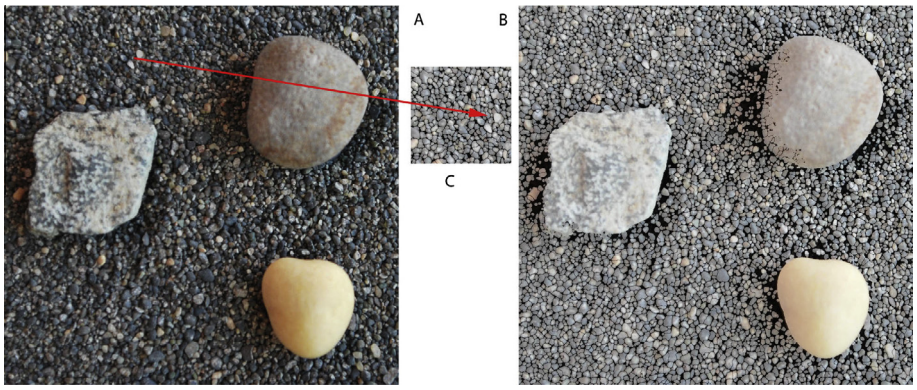

Fig. 13. Image of terrestrial sample, ~27.2 mm to a side, of Costa Rica sediment (Playa Hermosa, Jaco; ~0.1–1.0 mm grain size fine-to-coarse sand with a small fraction of finer material) on the left (A). Foreground pebbles (~5–10 mm grain size fine-to-medium pebbles) were placed on the sediment. (A) This figure shows the original prior to segmentation, while (B) shows the segmented result overlain on the original. Excerpt shown in (C), with location in original identified with an arrow. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

grains were segmented successfully with few fused or fragmented individuals. This initial result with a terrestrial sample enhances our confidence that the algorithm can accurately segment images even when the sediment sample is unavailable for hands-on analysis, provided the algorithm parameters are optimized and the sample characteristics lend themselves to automated segmentation (e.g., relatively non-reflective, opaque grains). Our future study with terrestrial samples will extend, detail, and quantify these initial results, while retaining the robust statistical assessment of the current work.

4. Conclusions and future work

Our algorithm, as described in our companion work (Karunatillake et al., 2014), operates robustly relative to a human operator at the grain size scales prevalent in MER MI images. Also rapid, it processes a single MI image within 1–5 min. Such strengths and consistency across different images, locations, and perhaps missions and planets holds the promise of robust granulometry and inferred sedimentology, of planetary sediment, as an improvement over current ability. As with terrestrial photoanalysis, the focal axis of the imager must be oriented normally to the sediment surface for a constant pixel-to-physical size conversion factor, or the device data must offer computable image distortion; angular distortion must be minimal relative to grain size.

In future work, we will consider diverse examples analyzing terrestrial sediment as references to compare with sieved results, assess previous planetary granulometry with MER MIs (Cabrol et al., 2008; Karunatillake et al., 2010a; McGlynn et al., 2011), and determine grain size distributions in the context of Thermal Infrared characterization (Fergason et al., 2006; Hardgrove et al., 2009). We expect this to expand the scope of our work further, by integrating thermal characterizations of remotely sensed surfaces with remote imaging, including aerial, campaigns.

Acknowledgments

Peter Overmann, Shadi Ashnai, Theodore Gray, and other members of the Wolfram Research Image Processing Team provided analytical and coding solutions to key hurdles, without which we

would have been unsuccessful in developing our algorithm. Peter Borg at Rider University provided a critical imaging campaign for preliminary terrestrial sediment analyses. We thank Dave Rubin and Aileen Yingst for vital feedback on an earlier version of the manuscript, which also helped to clarify our path for future research. We were supported by NASA Mars Data Analysis Program Grants NNX07AN96G, NNX10AQ23G, NNZ11AI94G, and NNX13AI98G. Louisiana State University's Geology and Geophysics Department in the College of Science provided postdoctoral funding to support J.R. Skok. Undergraduate students Jade Bing, Jacqueline Bleakley, Thomas Vajtay, and Thomas Weindl at Rider University provided helpful suggestions, enhancing the readability of our work.

References

Andrews-Hanna, J.C., Zuber, M.T., Arvidson, R.E., Wiseman, S.M., 2010. Early mars hydrology: Meridiani playa deposits and the sedimentary record of Arabia Terra. J. Geophys. Res. 115 (E6), E06002. http://dx.doi.org/10.1029/2009JE003485.

Cabrol, N.A. et al., 2008. Soil sedimentology at Gusev Crater from Columbia Memorial Station to Winter Haven. J. Geophys. Res. 113 (E6), E06S05. http://dx.doi.org/10.1029/2007JE002953.

Calvin, W. et al., 2008. Hematite spherules at Meridiani: Results from MI, Mini-TES, and Pancam. J. Geophys. Res. Planets 113 (E12). http://dx.doi.org/10.1029/2007JE003048.

Fergason et al., 2006. Physical properties of the Mars Exploration Rover landing sites as inferred from Mini-TES–derived thermal inertia. J. Geophys. Res. E2 (2), E02S21. http://dx.doi.org/10.1029/2005JE002583.

Graham, D.J., Rice, S.P., Reid, I., 2005a. A transferable method for the automated grain sizing of river gravels. Water Resour. Res. 41, 12. http://dx.doi.org/10.1029/2004WR003868.

Graham, D.J., Rice, S.P., Reid, I., 2005b. Comment: Photographic techniques for characterizing streambed particle sizes. Trans. Am. Fish. Soc. 134 (6), 1599–1603. http://dx.doi.org/10.1577/T04-146.1.

Graham, D.J., Rollet, A.-J., Piégay, H., Rice, S.P., 2010. Maximizing the accuracy of image-based surface sediment sampling techniques. Water Resour. Res. 46, 15. http://dx.doi.org/10.1029/2008WR006940.

Hardgrove, C., Moersch, J., Whisner, S., 2009. Thermal imaging of alluvial fans: A new technique for remote classification of sedimentary features. Earth Planet. Sci. Lett. 285 (1–2), 124–130, http://dx.doi.org/j.epsl.2009.06.004.

Herkenhoff, K.E. et al., 2003. Athena Microscopic Imager investigation. J. Geophys. Res. Planets 108 (E12), 8065. http://dx.doi.org/10.1029/2003JE002076.

Herkenhoff, K.E. et al., 2008. In-situ observations of the physical properties of the martian surface. In: The Martian Surface: Composition, Mineralogy and Physical Properties. Cambridge University Press.

Karunatillake, S., McLennan, S., Herkenhoff, K.E., 2010a. Regional and grain size influences on the geochemistry of soil at Gusev Crater, Mars. J. Geophys. Res. 115, E00F04. http://dx.doi.org/10.1029/2010JE003637.

Karunatillake, S., McLennan, S.M., Herkenhoff, K.E., Husch, J.M., Hardgrove C., Skok, J.R., 2014. A martian case study of segmenting images automatically for granulometry and sedimentology, Part 1: Algorithm. Icarus 229, 400–407.

Karunatillake, S. et al., 2010b. Recipes for spatial statistics with global datasets: A martian case study. J. Sci. Comput. 46, 439–451. http://dx.doi.org/10.1007/s10915-010-9412-z.

Klingelhöfer, G. et al., 2004. Jarosite and hematite at Meridiani Planum from Opportunity's Mössbauer Spectrometer. Science 306 (5702), 1740–1745. http://dx.doi.org/10.1126/science.1104653.

Lewis, K.W. et al., 2008. Structure and stratigraphy of Home Plate from the Spirit Mars Exploration Rover. J. Geophys. Res. 113 (E12), E12S36. http://dx.doi.org/10.1029/2007JE003025.

Maerz, N.H., Palangio, T.C., Franklin, J.A., 1996. WipFrag image based granulometry system. In: Franklin, J., Katsabanis, T. (Eds.), Measurement of Blast Fragmentation: Proceedings of the Fragblast-5 Workshop on Measurement of Blast Fragmentation, Montréal, Québec, Canada, 23–24 August 1996. A.A. Balkema, Rotterdam, Netherlands; Brookfield, VT, USA, pp. 91–99.

McGlynn, I.O., Fedo, C.M., McSween Jr., H.Y., 2011. Origin of basaltic soils at Gusev Crater, Mars, by aeolian modification of impact-generated sediment. J. Geophys. Res. 116, 15. http://dx.doi.org/10.1029/2010JE003712.

McGlynn, I.O., Fedo, C.M., McSween, H.Y., 2012. Soil mineralogy at the Mars Exploration Rover landing sites: An assessment of the competing roles of physical sorting and chemical weathering. J. Geophys. Res. 117 (E1), E01006. http://dx.doi.org/10.1029/2011JE003861.

McLennan, S.M. et al., 2005. Provenance and diagenesis of the evaporite-bearing Burns formation, Meridiani Planum, Mars. Mars. Earth. Planet. Sci. Lett. 240 (1), 95–121. http://dx.doi.org/10.1016/j.epsl.2005.09.042.

McSween, H.Y., McGlynn, I.O., Rogers, A.D., 2010. Determining the modal mineralogy of Martian soils. J. Geophys. Res. 115 (E7), E00F12. http://dx.doi.org/10.1029/2010JE003582.

Squyres, S.W. et al., 2004. The Spirit rover's athena science investigation at Gusev Crater, Mars. Science 305 (5685), 794–799. http://dx.doi.org/10.1126/science.3050794.

Sullivan, R. et al., 2008. Wind-driven particle mobility on Mars: Insights from Mars Exploration Rover observations at "El Dorado" and surroundings at Gusev Crater. J. Geophys. Res. 113, E06S07. http://dx.doi.org/10.1029/2008JE003101.

Wang, W., 2008. Rock particle image segmentation and systems. In: Yin, P.-Y. (Ed.), Pattern Recognition Techniques, Technology and Applications. I-Tech, Vienna, Austria, pp. 197–226.

Yingst, R.A., Crumpler, L., Farrand, W.H., Li, R., Cabrol, N.A., Neakrase, L.D., 2008. Morphology and texture of particles along the Spirit rover traverse from sol 450 to sol 745. J. Geophys. Res. Planets 113, E12S41. http://dx.doi.org/10.1029/2008JE003179.