Supplement of Earth Surf. Dynam., 7, 859–877, 2019
https://doi.org/10.5194/esurf-7-859-2019-supplement

Supplement of

Introducing *PebbleCounts*: a grain-sizing tool for photo surveys of dynamic gravel-bed rivers

Benjamin Purinton and Bodo Bookhagen

Correspondence to: Benjamin Purinton (purinton@uni-potsdam.de)

S1. Additional Data Dimensions from Point Clouds

The results presented in the main manuscript are similar to other studies segmenting grains from 2D imagery (e.g., Detert and Weitbrecht, 2012). This ignores the potential to exploit the third height dimension of the data from irregularly spaced SfM-MVS (or lidar) point clouds and associated DEMs. Many authors have already begun to look at patch-scale variance or roughness (e.g., Rychkov et al., 2012; Brasington et al., 2012) from point clouds on gravel-bed rivers to determine bulk characteristics, but this stops short of object detection and segmentation. Here, we briefly describe some of our own efforts to incorporate this additional information into *PebbleCounts*.

Our simplest approach was including the gridded DEM information, resampled to the same resolution as the orthomosaic. We inverted the elevation raster and flood-filled from the lowest points (tallest grains) using watershed approaches, similar to lidar tree-detection algorithms (e.g., Chen et al., 2006; Alonzo et al., 2015). For large, prominent grains with semi-spherical shapes, the flooded area was found to linearly increase until reaching the grain boundary, at which point the rate of area change jumped. We explored this break point as a potential segmentation tool for larger grains, but found that in the complex natural setting the shape of most grains is far from spherical, and furthermore, overlapping grains led to inconsistent behavior in the area breaks.

In an additional approach, we calculated both roughness and curvature at a variety of scales (5, 10, 50, 100 mm) directly from the point cloud using the open-source *CloudCompare* software (CloudCompare, 2018). This information was then gridded into a raster of the same resolution of the orthomosaic. While roughness could at times identify the smoother sand patches, it was difficult to discern between a sand patch and flat rock, and a color threshold on the orthoimagery was more successful. Curvature showed some spikes at grain boundaries, with the potential to aid in edge detection, however, we found that curvature was also high on intra-granular features.

In general, this analysis was complicated by vertical noise (scattering around a mean value) inherent to the SfM-MVS technique in the generation of dense point cloud data. In the field, for $\sim$9 near-nadir photos taken from a height of $\sim$4.5 m, the vertical standard deviation of points on a detrended flat surface (one of our coded targets) was found to be 1.7 mm for 13,014 points. On the other hand, in the perfect lab setting with 16 nadir+oblique photos from $\sim$1.5 m, the detrended flat carpet around the pebbles achieved a standard deviation of 0.2 mm (33,371 points), similar to other SfM-MVS studies using large numbers of carefully collected images (e.g., Cullen et al., 2018; Verma and Bourke, 2019). These standard deviations from detrended flat surfaces represent a best-case scenario, whereas, in our field setting, the vertical uncertainty on the complex, overlapping pebbles is likely higher. Such vertical noise is absent from the orthomosaics and limits the applicability of point clouds at these scales.

Ultimately, as the point cloud actually has a lower resolution (since it is based only on matched points) and more vertical noise than the orthomosaic (which exploits the full camera resolution), the imagery alone provided more detail. This is particularly important around grain edges needed for segmentation, which are not captured in nadir imagery alone, as shown in Figure S1. The lab setting resulted in point clouds with sufficient density and precision to identify individual grains with point-cloud processing tools. Thus, achieving higher quality SfM-MVS point clouds is possible, but only through more intense data collection during fieldwork.

Alternatively, lidar point clouds with distance measurements based on phase shifts have a lower standard deviation of $\sim$1 mm in multiple settings and distances (up to $\sim$300 m) and could allow more precise delineation using roughness and curvature calculations directly on the point cloud, however, such devices remain costly. Additionally, the development of affordable hyperspectral cameras with additional wavelengths will help in image segmentation in the spectral domain. To conclude, the potential for additional data dimension integration into pebble counting may be possible using higher dimensional object detection schemes, but, for the time-being, the orthoimagery alone provides satisfying results.

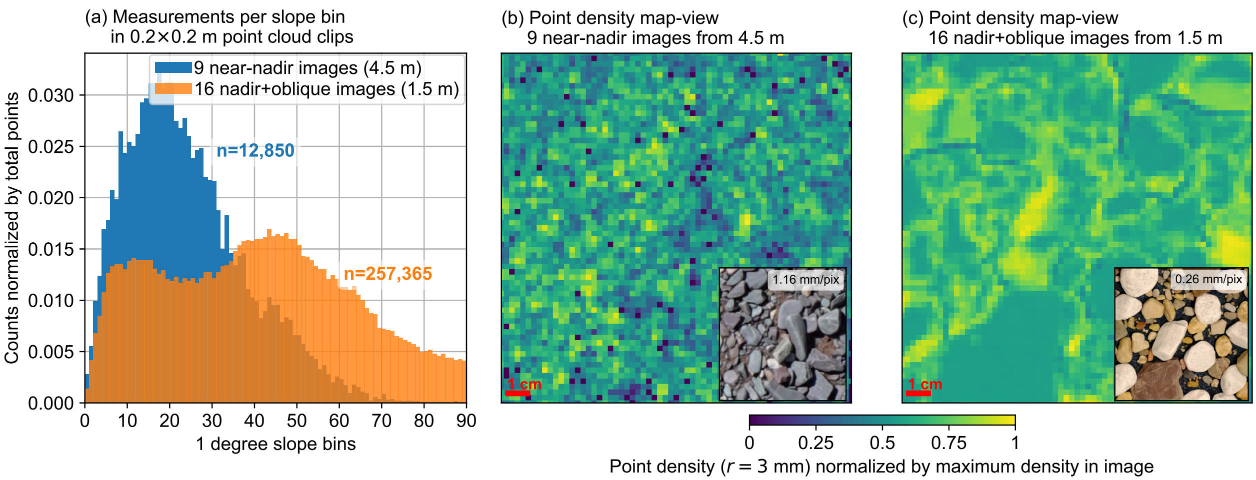

Figure S1. (a) Slope distribution in field (near-nadir) and experimental (nadir+oblique) point cloud clips. The point cloud slope was calculated in *CloudCompare* (CloudCompare, 2018) by first calculating the normals at each point using the 6 nearest neighbors and then extracting the dip of each normal. (b) Map-view of point density normalized by the maximum for the 9 near-nadir field images and (c) the same for the 16 nadir+oblique experimental images. Point density was calculated as the number of points in a radius of 3 mm. The clips were from a 0.2×0.2 m area, visually selected to have similar grain sizes and numbers of grains, shown in the inset images in (b) and (c). The average point density for the 16 nadir+oblique photo setting was 59 points/cm^2, whereas, in the field using 9 near-nadir photos the density was 17 points/cm^2. Note the higher point density on grain edges in (c) compared to (b), which are important for segmenting grains directly on the point cloud.

S2. Command-line Variables and Example Screenshots for *PebbleCounts*

Table S1 shows the command-line variables for *PebbleCounts* (KMS approach) and Table S2 shows the command-line variables for *PebbleCountsAuto* (AIF approach). Examples of the command-line interface and manual clicking steps are shown in Figure S2 and Figure S3, respectively.

Table S1. Command-line variable flags in *PebbleCounts* and their meaning. The default values are effective for most images.

Variable Flag	**Meaning (units)**	**Default Value(s) and Suggested Range**
im	Image to run, including path to folder	No default
ortho	Georeferenced orthoimagery flag	No default, ‘y’ for orthoimagery, ‘n’ for nadir
input_resolution	Input resolution if not orthoimage (mm)	No default, calculate from eq. (3)
subset	Interactively subset image	Default no (‘n’)
*sand_mask**	Name, including path, to a sand mask if one already exists	No default
*otsu_threshold**	Percentage of Otsu value to threshold shadows by (percentage of 100)	No default, suggested value of 50
*maxGS**	Expected maximum a-axis grain size (m)	Default 0.3
*cutoff**	Minimum b-axis length to be counted (pixels)	Default 20, can be raised
*min_sz_factors**	Factors to multiply *cutoff* at each scale, used to cleanup masks for easier clicking	Default [50, 5, 1] for three scales (large to small) for $\sim$1 mm/pixel imagery, double for < 0.8 mm/pixel
*win_sz_factors**	Factors to multiply *maxGS* by at each scale	Default [10, 3, 2] for three scales (large to small), can be changed $\pm$0.5–1.5 to get more or less windows
*improvement_ths**	Improvement threshold values that tell k-means when to halt (fraction of 1)	Default [0.01, 0.1, 0.1] for three scales (large to small), can be varied from 0.01–0.2
*coordinate_scales**	Fraction to scale *x,y* coordinates (fraction of 1)	Default [0.5, 0.5, 0.5] for three scales (large to small), can be varied from 0.3–0.7
*overlaps**	Fraction of overlap between windows (fraction of 1)	Default [0.5, 0.3, 0.1] for three scales (large to small), can be varied from 0–0.5 at each scale
*first_nl_denoise**	Strength of first non-local means denoising	Default 5, can be varied $\pm$1
*nl_means_chroma_filts**	Strength of windowed non-local means denoising	Default [3, 2, 1] for three scales (large to small), can be varied $\pm$1
*bilat_filt_szs**	Size of bilateral filtering windows (pixels)	Default [9, 5, 3] for three scales (large to small), can be varied from 3–9
*tophat_th**	Upper percentile threshold to take from top-hat filter for edge detection (fraction of 1)	Default 0.9, can be varied from 0.8–0.95
*sobel_th**	Upper percentile threshold to take from sobel filter for edge detection (fraction of 1)	Default 0.9, can be varied from 0.8–0.95
*canny_sig**	Canny filtering sigma value for edge detection	Default 2, can be varied from 1–2
resize	Value to resize windows by (fraction of 1)	Default 0.8, can be varied from 0.5–0.99 if you want a smaller (0.5) or larger (0.99) pop-up window

*Influence on results

Table S2. Command-line variable flags in *PebbleCountsAuto* and their meaning. The default values are effective for most images.

Variable Flag	Meaning (units)	Default Value(s) and Suggested Range
im	Image to run, including path to folder	No default
ortho	Georeferenced orthoimagery flag	No default, 'y' for orthoimagery, 'n' for nadir
input_resolution	Input resolution if not orthoimage (mm)	No default, calculate from eq. (3)
subset	Interactively subset image	Default no ('n')
*sand_mask**	Name, including path, to a sand mask if one already exists	No default
*otsu_threshold**	Percentage of Otsu value to threshold shadows by (percentage of 100)	No default, suggested value of 50
*cutoff**	Minimum b-axis length to be counted (pixels)	Default 20, can be raised
*percent_overlap**	Maximum allowable overalp between neighboring ellipses for filtering suspect grains (percentage of 100)	Default 15, can be varied from 5–30
*misfit_threshold**	Maximum allowable misfit between ellipse and grain mask for filtering suspect grains (percentage of 100)	Default 30, can be varied from 10–50
*min_size_threshold**	Minimum area of grain, used to clean the mask (pixels)	Default 10 for $\sim$1 mm/pixel imagery, 40 for $<$ 0.8 mm/pixel
*first_nl_denoise**	Strength of first non-local means denoising	Default 5, can be varied $\pm$1
*tophat_th**	Upper percentile threshold to take from top-hat filter for edge detection (fraction of 1)	Default 0.9, can be varied from 0.8–0.95
*sobel_th**	Upper percentile threshold to take from sobel filter for edge detection (fraction of 1)	Default 0.9, can be varied from 0.8–0.95
*canny_sig**	Canny filtering sigma value for edge detection	Default 2, can be varied from 1–2
resize	Value to resize windows by (fraction of 1)	Default 0.8, can be varied from 0.5–0.99 if you want a smaller (0.5) or larger (0.99) pop-up window

*Influence on results

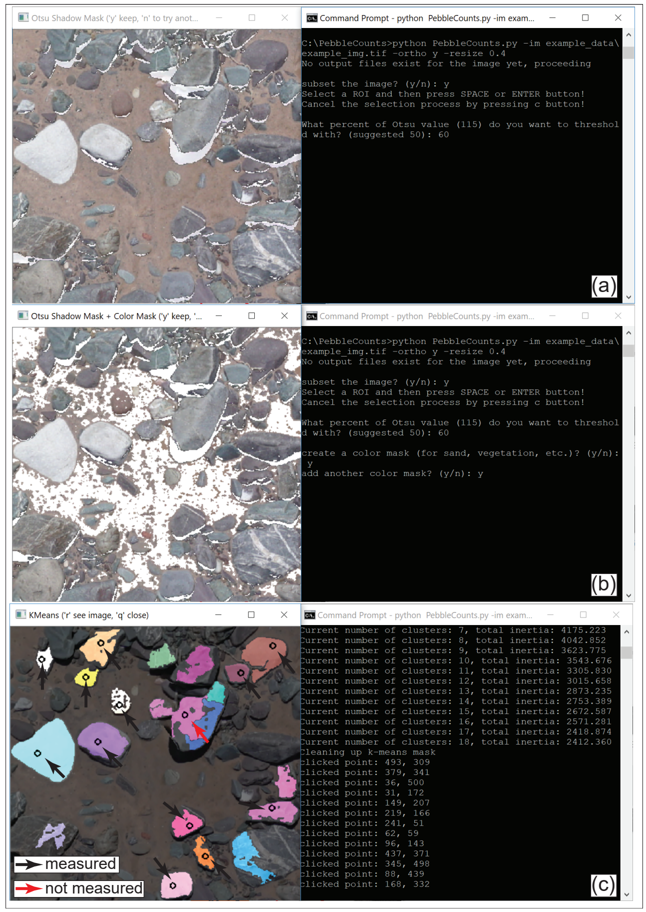

Figure S2. Example of command-line and pop-up interface for *PebbleCounts*. (a) Interactive Otsu thresholding using percentage of Otsu value and yes ('y') or no ('n') confirmation. (b) Interactive color masking by yes ('y') or no ('n') and resulting color mask after selection. (c) K-means clustering and pop-up window for pebble selection by left clicking, with black arrows measured in final output and red arrows ignored after right-click removal (see Fig. S3).

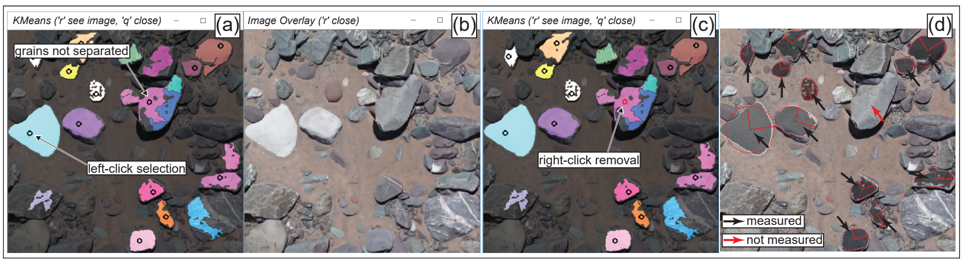

Figure S3. Clicking tutorial continued from Figure S2c. Following k-means clustering at each scale a mask overlaid on the original image is presented (a), and grains are selected by a left click anywhere in the segmented area, resulting in a black circle at the click location. When clicking is finished the mask is closed by pressing ‘q’. To view the original unmasked image the user may press ‘r’ (b). Using this switching the user can see which grains are poorly delineated and remove the last click with a right click on the mouse (c). The original black circle selection turns to red to signify this grain is off and will not be measured in the final output (d).

S3. Resampling and Parameter Selection in AIF Approach

Figure S4 demonstrates the percentage of grains with a match found in the AIF approach when increasing resampling from a factor of 0.6–2.6 by 0.1 steps using Lanczos resampling (Lanczos, 1950). As the resampling factor increases, there is progressive reduction in the number of found grains after filtering, therefore we selected the original resolution (resampling factor of one). Figure S5 and Figure S6 demonstrate two cases where the resampling slightly improved the resulting grain-size distribution. Both images were of relatively low quality with significant blurring and the presence of many weak edges between grains of similar color.

We selected a maximum percent misfit between the ellipse and grain of 30% as the 90^{th} percentile of misfits for the KMS approach was 30%. Furthermore, we allowed a maximum overlap between neighboring ellipses of 15%, visually selected to minimize overlapping grain measurement and over-segmentation of discrete grains. For the higher resolution imagery it was necessary to use a lower sobel and top-hat threshold (0.85), since we consider all the edges at once in the AIF approach, rather than in a windowed subset as in the KMS approach, and many edges are not found when using the 0.9 threshold given the increased number of pixels under consideration.

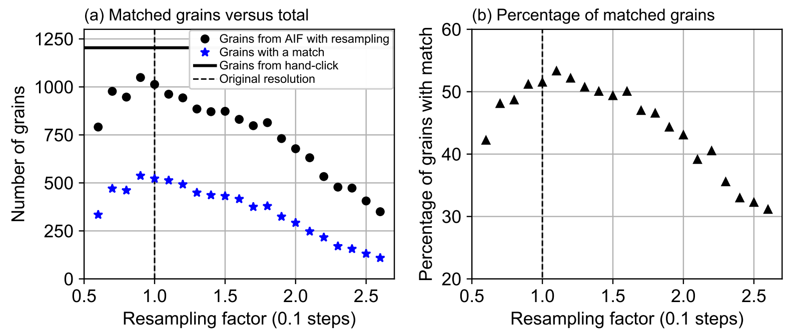

Figure S4. Matching grains found in each filtered mask versus the resampling factor (where 1 is the original image) for the ∼1.16 mm/pixel resolution images. Matches are defined as an AIF grain within 5 pixels of the hand-clicked line or the KMS grain centroid and with a 1 cm maximum b-axis difference between the AIF grain and the match.

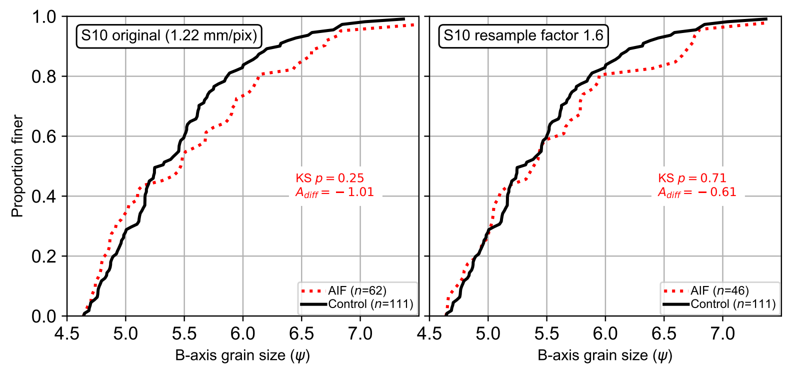

Figure S5. Slight improvement (increase in p and decrease in A_{diff}) in result using a 1.6-times resampling factor prior to running the AIF algorithm for the difficult (somewhat blurry, weak edges) S10 orthoimage.

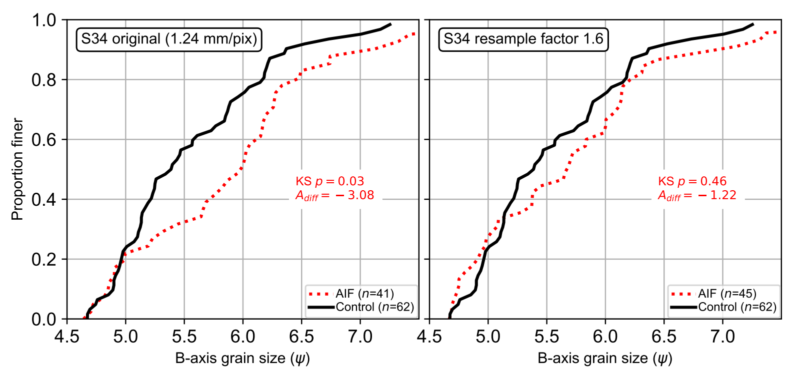

Figure S6. Slight improvement (increase in p and decrease in A_{diff}) in result using a 1.6-times resampling factor prior to running the AIF algorithm for the difficult (very blurry, weak edges) S34 orthoimage.

S4. Agisoft Orthomosaic Generation

Agisoft (Agisoft, 2018) processing was carried out in the following steps for the indoor handheld imagery (with field-gathered mast imagery differences in parentheses following the step):

1. Image quality detection and the exclusion of photos with quality metric < 0.7. This step analyzes pixel contrast to estimate sharpness with values ranging from 0/blurred to 1/sharp. We found 0.7 to be a sufficient lower cutoff upon visual inspection of results.

2. Detection of 12-bit coded targets in the remaining photos, with two targets placed at each of the four corners of the area and ensuring that the diameter of the printed targets' center circle was limited to 10–30 pixels in image resolution for successful automated detection.

3. Input of scale for the orthomosaic output, provided by the distances between the targets at each corner, resulting in four distance measurements, with 0.5 mm accuracy using a ruler with cm and mm demarcations. (For the field images: The scale was provided by the XYZ coded target locations in UTM zone 19S, WGS84 ellipsoidal datum.)

4. Photo alignment at high quality with a 40,000 key-point and 2000 tie-point limit.

5. Dense cloud generation from the aligned photos at the medium output and with moderate depth filtering. Given the high quality of the photos more aggressive options did not improve results. (For the field images: Given the increased complexity of the setting and imperfect photo collection, the dense point cloud was generated at high quality with aggressive depth filtering.)

6. DEM building from the dense cloud with default settings in a local coordinate system. (For the field images: The DEMs and orthomosaics were also output in UTM zone 19S projections, providing undistorted pixels with resolution in m/pixel.)

7. Generation of an orthomosaic using the DEM for orthorectification at the default settings.

8. Output of the orthomosaic to a GeoTiff file with resolution provided in m/pixel.

S5. KMS and AIF Results Separated by Site

Here we show all of the results (following 20-pixel truncation) for each of the 12 sites in Figure S7. These results are aggregated in curves shown in the main manuscript Figure 11 and a comparison of the individual percentiles of interest is shown in the main manuscript Figure 12.

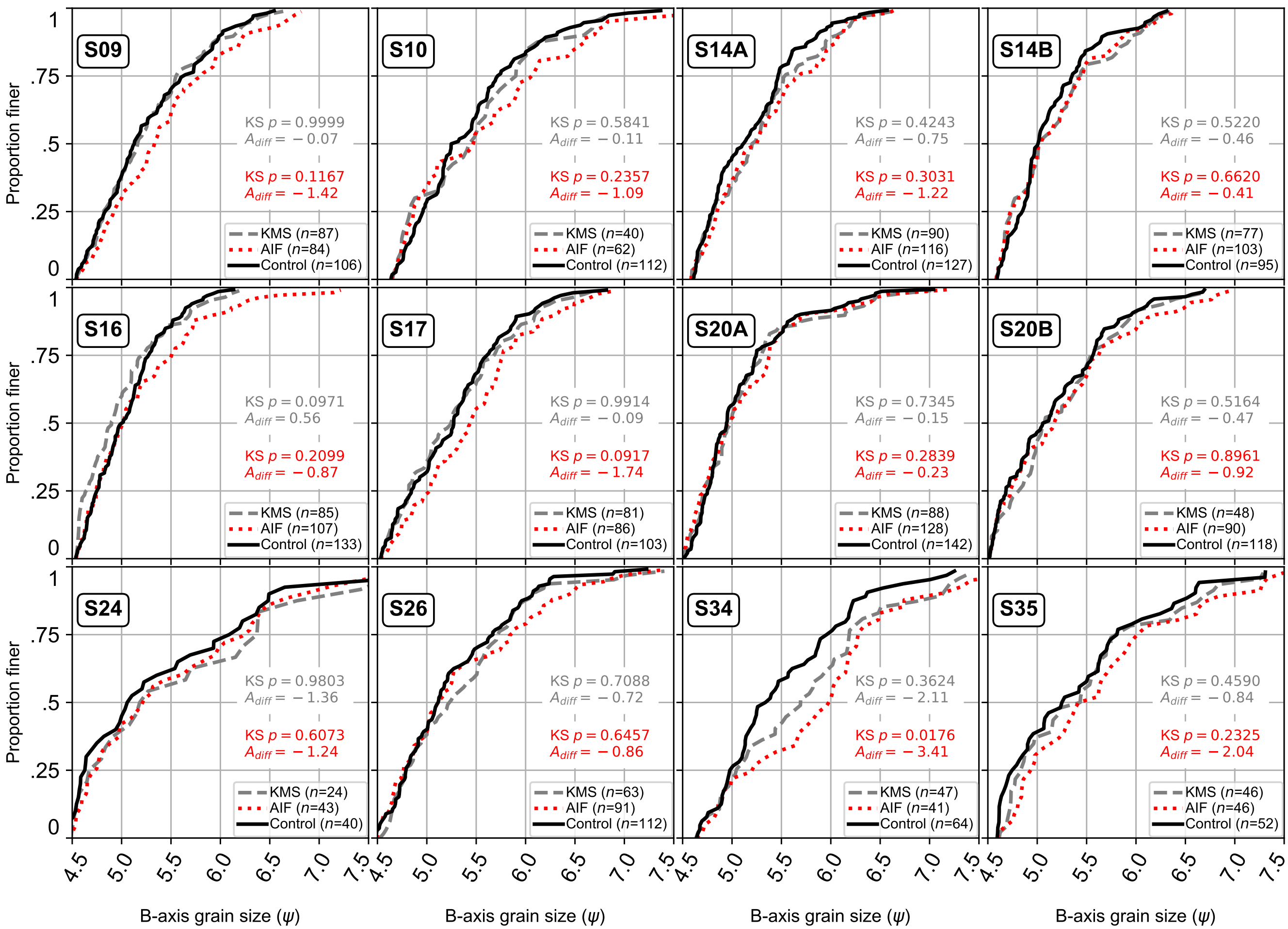

Figure S7. Comparison of 20-pixel truncated grain-size distributions between hand-clicked control (black line), KMS *PebbleCounts* (gray, dashed line), and AIF *PebbleCountsAuto* (red, dotted line) for the 12 × ∼1.16 mm/pixel control sites. In corresponding colors are the p-value results of a KS-test and the A_{diff} approximate integral between the curves for each approach versus the control data. The legend indicates the number of grains (n) making up each curve. See Figure 6b in the main manuscript for sites.

S6. Misidentification in the AIF Approach

Figure S8 demonstrates remaining issues with the AIF approach in a few map-view examples. On a grain-by-grain basis, there are many inaccuracies falling into three main categories: over-segmentation of grains with internal edges and the selection of each segment as a separate grain, under-segmentation and merging of neighboring grains that have weak edges sometimes caused by image blur, and misidentification of non-grain objects or clusters of small grains. It is clear from this analysis that caution must be used when interpreting AIF results, particularly in complex or blurry images.

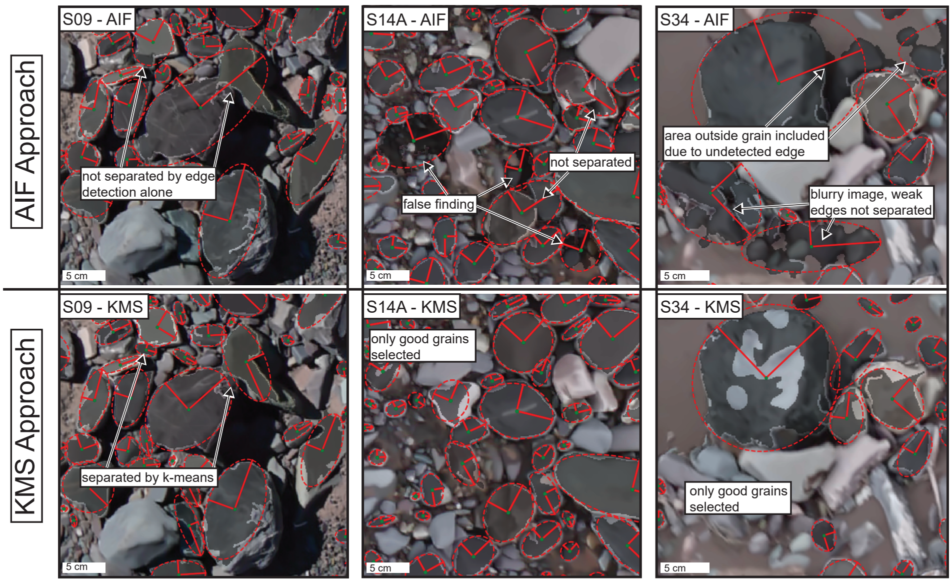

Figure S8. Resulting delineated grains using the AIF *PebbleCountsAuto* function (top row) versus the same area from the KMS *PebbleCounts* function (bottom row). Labels indicate the issues with the AIF results and improvement in KMS results. Note the poor results for the blurry image on the right (S34).

References

Agisoft: AgiSoft PhotoScan Professional, http://www.agisoft.com/downloads/installer/, 2018.

Alonzo, M., Bookhagen, B., McFadden, J. P., Sun, A., and Roberts, D. A.: Mapping urban forest leaf area index with airborne lidar using penetration metrics and allometry, Remote Sensing of Environment, 162, 141–153, https://doi.org/10.1016/j.rse.2015.02.025, 2015.

Brasington, J., Vericat, D., and Rychkov, I.: Modeling river bed morphology, roughness, and surface sedimentology using high resolution terrestrial laser scanning, Water Resources Research, 48, W11 519, https://doi.org/10.1029/2012WR012223, 2012.

Chen, Q., Baldocchi, D., Gong, P., and Kelly, M.: Isolating Individual Trees in a Savanna Woodland Using Small Footprint Lidar Data, Photogrammetric Engineering & Remote Sensing, 72, 923–932, https://doi.org/10.14358/PERS.72.8.923, 2006.

CloudCompare: CloudCompare Software, http://www.cloudcompare.org/, 2018.

Cullen, N. D., Verma, A. K., and Bourke, M. C.: A comparison of structure from motion photogrammetry and the traversing micro-erosion meter for measuring erosion on shore platforms, Earth Surface Dynamics, 6, 1023–1039, https://doi.org/10.5194/esurf-6-1023-2018, https://www.earth-surf-dynam.net/6/1023/2018/, 2018.

Detert, M. and Weitbrecht, V.: Automatic object detection to analyze the geometry of gravel grains–a free stand-alone tool, in: River flow 2012 : Proceedings of the international conference on fluvial hydraulics, San José, Costa Rica, September 5-7, 2012, pp. 595–600, Taylor & Francis Group, London, 2012.

Lanczos, C.: An iteration method for the solution of the eigenvalue problem of linear differential and integral operators, J. Res. Natl. Bur. Stand. B, 45, 255–282, 1950.

Rychkov, I., Brasington, J., and Vericat, D.: Computational and methodological aspects of terrestrial surface analysis based on point clouds, Computers & Geosciences, 42, 64–70, https://doi.org/10.1016/j.cageo.2012.02.011, 2012.

Verma, A. K. and Bourke, M. C.: A method based on structure-from-motion photogrammetry to generate sub-millimetre-resolution digital elevation models for investigating rock breakdown features, Earth Surface Dynamics, 7, 45–66, https://doi.org/10.5194/esurf-7-45-2019, https://www.earth-surf-dynam.net/7/45/2019/, 2019.