# Review of "Size, shape and orientation matter: fast and automatic measurement of grain geometries from 3D point clouds" by Steer et al.

In their study, Steer et al. present a novel method for point cloud segmentation to retrieve individual pebbles from lidar and SfM point clouds gathered in fluvial environments. The method uses a modified watershed approach, which is common in computer vision segmentation applications, to get grain boundaries. There are several cleaning steps to avoid issues of over-segmentation (single grains being split into many grains) and these must be controlled by a number of parameters and trial-and-error operator tests, with an end result of a qualitatively (to the operator) good segmentation. The choice of model for retrieving the pebble axis dimensions and orientation is explored, and the authors highlight some deficiencies of these models. The method is applied to a number of scenarios in the lab (what they call "synthetic") and field, for which validation data is available, and seems to provide reasonable results for pebble size measurement. Overall, I find the study makes some good progress on pebble segmentation from point clouds, but it needs some significant expansion of the methods to be more explicit and clear, and more discussion of the caveats of the methods (e.g., qualitative segmentation, use of a highly imperfect model for angular grains). I would test the G3P algorithm on some pebble data (the algorithm is published on GitHub, great!), but I am hesitant to do so without a clearer understanding of how the parameters should be set based on point cloud density and other dataset characteristics.

Below, I provide some major comments, followed by a number of line comments (which admittedly sometimes go beyond grammar and citations to verge on major comments), and finish with the important references that I found were missing from the manuscript. I recommend the paper for major revisions before further steps, and would enjoy re-reviewing the paper and testing the algorithm at a later stage.

**Major Comments**

In general, there is some lack of specificity in the methods and the need for some more schematic figures early in the paper showing how grains are segmented and cleaned. Additionally, the authors do not satisfactorily explain how parameters are chosen or how they impact results, especially with regards to more and less angular pebbles, which I imagine is one of the biggest issues with this segmentation approach – not to mention how point cloud density impacts the segmentation. I expand on these concerns below.

In the presentation of the watershed method the authors cite the steepest decent algorithm. Although they make the point that this is usually applied to grids, I did not follow how this algorithm (based on the 8-neighboring grid cells) is the same as the point cloud approach here. If the authors want to talk about flow routing on point clouds, then they should carefully read and cite the FFN method of Rheinwalt et al. (2019). I see the temptation to introduce this in the flow-routing context (especially because of the term "watershed") but I think the connection needs to be elucidated in more detail. Again, this is not the D8 algorithm that is cited, rather it is more of a network-based approach to point cloud segmentation.

Throughout the manuscript the use of “trial-and-error” and “satisfying segmentation” are rather vague. I see that a trained operator who has been to the field site and knows the characteristics of the point cloud data (i.e., collected it themselves) will be required for using this method, and that should be stated explicitly in the abstract and early in the paper. The algorithm needs careful control and likely a lot of tweaking to get “decent” results in a new area. And if this area is large and heterogeneous (in pebble size / shape and point cloud density / quality) it could represent significant effort to get results, which I feel is glossed over in the presentation of the tool. The authors mention the need for validation data at the field cite (i.e., manual Wolman) and that is a significant caveat of the method. If validation data is always needed to “trust” the results, then this needs to be presented up front in the manuscript.

Somewhere in the methods (possibly with reference to the supplement) a schematic figure is needed to show in detail how the “flow” is routed in a neighborhood of pebbles and how the “k” parameter affects this routing. The schematic could also include a cartoon of the three criteria for over-segmentation cleaning, which was not clear based on the description provided. As this is a new method, schematic diagrams (i.e., cartoons, or zoom-in on real data with the points clearly visible) are highly recommended – I would say necessary. This diagram (or diagrams if you put more in the supplement, but I suggest at least one in the main text) should also highlight how the parameters affect the segmentation and cleaning. Alternatively, if this is difficult to show in the diagram, then the parameters should be explicitly treated in a new methods section where the parameters are walked through, with a paragraph or a few sentences for each parameter in which the authors discuss “rules-of-thumb” or quantitative reasons (e.g., point cloud density vs. size of pebbles) for the selection of certain parameters and some trial-and-error scenarios.

As an aside, was under-segmentation never (or rarely) an issue with this method? I suggest the authors read and cite Purinton and Bookhagen (2019, 2021) when discussing over- and under-segmentation, as we go into great detail on these affects for 2D image segmentation (e.g., Section 4.2.1 in the 2021 paper). In the 2019 PebbleCounts paper we discuss specifically watershed segmentation techniques and their tendency to over-segment grains in 2D images (cf. Figure 1, Figure 2, Section 1.2). This is based on 2D segmentation, but I can see how there might be some important parallels to the 3D case (i.e., grain angularity is key!).

I had a look at the Otira_1cm_grains.ply file on GitHub. Is this the exact cloud that was used to develop the watershed model? I wonder, because it looks like this is from a single TLS view and thus has significant occlusions on the back-side of grains facing away from the sensor. I imagine that is really problematic for the model development (based on the issues highlighted in Appendix A). Could the authors please comment on this point cloud? Maybe it is just a subset / sample data provided on GitHub and not the full dataset, which is fine.

Regarding the model choice in Section 2.4, I find myself asking: are pebbles really ellipsoids? I think that is an okay assumption and it is the one that is commonly made (well Domokos et al., 2014 make the case for superellipsoids https://en.wikipedia.org/wiki/Superellipsoid, which I’m not sure I agree with either), but it should be noted that this is a very imperfect model done for the sake of simplicity. The labeled grain in the point cloud could also be directly measured on, but historically we have treated grains as ellipsoids with a/b/c axes. The benefit of the proposed method is that you get an entire grain surface not just three axes. The grain surface isn't totally taken advantage of here (a model is still fit and used to extract those three axes) but this fact should be noted: you get the entire (or sampled at the point cloud density) grain surface. One path forward could be in new measurements of grain angularity from point clouds and comparing the volume of pebbles (e.g., convex hull approach that the authors use). Maybe mention that with the ability to measure the entire pebble surface in point clouds, going forward we may be able to measure additional (more accurate) grain characteristics, particularly for grains that are far from ellipsoidal shapes, for instance in mountain headwaters or hillslope deposits. But these measurements will certainly depend on the point cloud density and the size of the pebbles we are interested in. I don't think we are at the point of putting numbers on the appropriate density for a given pebble size, and this is highly dependent on pebble shape, but making some statements (in a new discussion section covering all these points?) along these lines would be some great food for thought and place this paper in the context of advancing our understanding of the true shape characteristics of sedimentary deposits. I think a bunch of the citations that wound up in the introduction and conclusion (Domokos, Szabos, Miller) could be nicely woven into these statements. Have a look at the interesting shape characteristics and two phase abrasion model used by Miller et al. (2014) specifically (the authors cite this paper) – how could the G3P method aid in applying this and measuring surface curvature of the pebbles? Very interesting to think about!

Appendix A is really important. I think this needs to be worked up into the main text, probably at the beginning of the results section. See my comments about that later.

**<u>Line Comments</u>**

Beyond the expansion of scientific content, I suggest another careful read-through of the manuscript for additional and/or similar grammatical errors as I point out in my line edits below. I have highlighted a fair number of grammatical errors that caught my eye, but there are still more in there. As a native English speaker, I found these distracting from the scientific content, but they are all easy fixes. I hope the authors take them in good faith!

P1L7: The first sentence of the abstract is a bit awkward with all the "and" statements. I suggest re-writing this to make it more clear. Consider using an oxford comma here and elsewhere, that could make such lists clearer.

P1L11: Comma after "measurements"

P1L14: "into" should be "of"

P1L15: "individualized" should be "segmented"

P1L16-17: "If different…" sentence is awkward and should be rephrased.

P1L20: Point "2)", isn't the limit also the size of grains and point cloud density? This is subtle and you state this as a limit at the end of the abstract. Although there you use the term "point cloud resolution". I prefer "point cloud density" and suggest you use that throughout and mention the density of the point clouds (in pts/m2 or pts/cm2 might make more sense for the high-density SfM datasets). I might remove point 2 from this list anyway.

P2L29: "Collection of a"

P2L30: Add citation for Purinton and Bookhagen (2021). We go into great detail on the issue of sample size there, I suggest the authors have a close look at e.g., Sections 4.3, 5.1, 5.2

P3L1: I think the term is more commonly “photo-sieving”

P3L11: “spent in”. Here and elsewhere in the manuscript the phrasing should be “in the field”, not “on the field”.

P3L26: The full and correct name of what we often call SfM processing is “structure from motion with multi-view stereo (SfM-MVS)” (Smith et al., 2015). This should be stated, and the authors could say something like “structure from motion with multi-view stereo, herein referred to as SfM”. I think the lowercase “f” is more common.

P4L6: “oriented”

P4L7: “reorient”

P4L10: Here and elsewhere “point density”, not “resolution”

P4L17: This is commonly called “D8”, please mention this. The Facet Flow Network (FFN) algorithm of Rheinwalt et al. (2019) should certainly be carefully read and cited if you are discussing flow routing on point clouds (cf. https://github.com/UP-RS-ESP/FacetFlowNetwork).

P5L3: Why is k=20 here?

P5L10: “approach is as”

P5L18: The first sentence in Section 2.2 should be broken into multiple sentences. The authors should elaborate on attempts made with different clustering approaches. I agree these are tricky, but be explicit with what methods were tried and possibly what the issues / strengths / weaknesses were of the results.

P5L19: I think “DBSCAN” gets capitalized.

P5L29: I don't understand how the "drainage area" relates to the grain radius. Might need more explanation here, or highlight this in the schematic diagram I proposed.

P5L29-30: Cf seems like a tricky parameter and will have a big effect on the segmentation. I really don't get where this is coming from. Basically, it seems like the segmentation needs to be carefully checked and modified by a trained operator. This should be made clear up front.

P5L32: missing period

P6L2: is k=20 here as well?

P6L3: "between": which normals are compared? Is it the nearest normal in the i and j grain? This would help to show diagrammatically

P6L5: “equals” should be “equal”

P6L6: “60 degrees throughout the paper”. But why 60? And this is a free parameter so I guess not “throughout the paper”. But I don’t understand “in the following” here. “Following” where?

P6L6: Curved how? I'm having a lot of trouble understanding this criterion. I think all of these criteria need a schematic figure.

P6L7: Delete “Therefore”

P6L19: the objective here is to merge small and large grains?

P6L20: “This number of …” don't understand. greater than 10? greater than or equal to k? what is k? is k=20?

P6L24: I guess this is another step that needs very close operator control and field knowledge. I see the algorithm works quickly, but I imagine that this requires a lot of going back and tweaking parameters until a "satisfying" (qualitative) segmentation is reached. This probably takes a while even on a small patch with <1,000 grains as here, and would be difficult to upscale to large areas or to finer point cloud spacing (i.e., SfM at hundreds to thousands of points per cm2) where you would be able to get smaller grains and thus exponentially more segmentation would be performed (log-normal GSD typical of river sediments). SfM is used later on, but this point about the extremely high density of SfM point clouds and the exponential increase in smaller grains is never made? Or I may have missed it.

P6L27-28: "respectively" is maybe missing, if you are referring to the beginning of the sentence.

P6L32: Why wasn’t the example in Figure 2 cleaned? It wasn’t required because the segmentation worked so well?

Figure 2 caption: “fit” not “fitted”, here and elsewhere in the main text. “same color as in panel c”. “relative” not “relatively”

P8L6: “consists of”

P8L7: Don’t understand the use of “adequacy”

P8L9-10: "From directly using the point cloud to describe each grain and measure their sizes and orientations."

P8L10: Delete “the use of”

P8L22: Delete “Anyhow”, too informal

P8L28: It’s not really possible to see anything in Figure 2. You may need another figure with a zoom-in on a couple of grains (the original points, colored by grain label) and the ellipse fit to the points. This may be hard to visualize, but as it stands you cannot see what is being referred to here in Figure 2. cf. Appendix Figure A1. This is going in the direction of a useful figure in panel b, but the points should be increased in size and the model fits made semi-transparent. I realize these figures are screenshots from cloud compare, but I wonder if they could still be somewhat improved by zooming in more and increasing the point size / adjusting the colors?

P8L29: “best solution”. What specifically was tested besides the least squares?

P8L30: Could you provide a reliable citation for “inertia ellipsoids”? I was doing some internet searching and having trouble understanding what these are and how they are computed. I don’t get the relation between the equation of an ellipsoid and this model / mathematical operation. Even an internet link would help me here! I just don't get it, but maybe I need to accept it as mathematical fact.

P8L33: I think singular value decomposition was already defined. Fine to use "SVD"

P9L5: how were the "ellipsoids" and "cuboids" compared? you mean the dimensions of the cuboid (length / width / height) with the axes of the ellipsoids (a /b /c)? Please be explicit here.

P10L3: "axes"

P10L5: "approximates"

P10L11: "metric"

P10L20: "diameters"

P10L27: I like the lab experiment, with the pebbles from the hardware store, but I wouldn't call this "synthetic". Rather "lab experiment" is more appropriate. Or you say "Lab Environment" and "Natural Environment". Please remove the term synthetic from the manuscript. Synthetic implies artificially generated, but these are real pebbles you are laying out and measuring (in a lab setting).

Figure 4: The cuboid volume is definitely way far off from the true volume, but putting it on the x-axis of panel g here implied this is validation data. If you really wanted to get the volume you could do a convex hull of the labelled point cloud and directly measure the volume from this as a control to compare the ellipsoid volume with. Actually, it looks like you do exactly this in Appendix A! Another reason to put Appendix A at the beginning of the results.

Figure 4 caption: "top", "middle", "bottom" should be "left, center, right".

P11L11: This experiment does not test the watershed segmentation at all since the pebbles are not overlapping. Or at least, this is barely a test. That's important. The first experiment is really ONLY for testing the ellipsoid models and axis length measurements as far as I can tell. Be explicit about this point.

P11L12: "captured by pictures" should be "photographed"

P11L13: "Processed" how? Be specific. High resolution? How many interest points? Was any filtering done? What was the resolution of the input images and what camera model was used to take the photos? How many oblique vs. nadir photos and approximately how low of an oblique angle was used?

P11L14: How as the planar surface removed? Manual cropping in CC? If this surface is removed, then the watershed segmentation is really barely necessary, so again this isn't testing the segmentation algorithm -- just the model fits.

P12L10: I wouldn't call these consistent. Maybe it helps to include a metric like RMSE of the two methods compared to the cuboid and plot this number (or two numbers) on each of the subplots in Figure 4e. It actually looks like the bias increases at larger grain sizes, why is that?

P12L15: The discussion of the DLSF and IE models is really interesting. I wonder if this is better highlighted in the Appendix A section, which I recommend becoming a new section at the beginning of the results.

P12L20: I'm not sure about using cuboid volume in any comparisons, this is not a volume of the grain (actually it's really far from the volume of a grain contained in the bounding box).

Figure 5: Struggling with these CDFs. Maybe make the G3P lines dashed in the plot and legend?

Figure 5 caption: “envelopes”

P14L1: So this is equivalent to the cuboids used in the first experiment? May be interesting to note whether the cuboid dimensions did in fact correspond well with manually measured dimensions of the hardware-store bought pebbles. No need for a formal analysis, but was this checked ever? Do the cuboids accurately give the hand-measured lengths?

P14L2: “wo”?

P14L5: Why no a and c axes here? Time constraint? Just curious.

P14L9: This is a fair approach, but it would also be nice here to cite the binomial modeling approach of Eaton et al. (2019) and its application in Purinton and Bookhagen (2021).

Table 1: Thanks for noting the final parameters, but what were the steps in their selection (what other values were tried?). Is this just something the operator needs to "get a feel" for? Or are there good reasons for these values? This goes towards one of my major comments regarding the parameters.

P14L19: What is the MP resolution of the images?

P14L19: Here and elsewhere in the manuscript it should be “square meters” not “squared meters”.

P14L23: "trial-and-error" here and elsewhere is vague. Can you be more specific about how these tests were run? How were initial values selected and how were the modified based on the runs? Was there systematic adjustment of the parameters or was it a bit more "random"?

P15L1: “grains are detected”

P15L5: Again, incorrect use of “synthetic”. You are measuring real grains on the point cloud.

P15L13: “percent”

Figure 6: Why only considering these percentiles? Maybe It would be cool to have a QQ plot where you show several percentiles plotted against each other for the WoI vs. G3P, with one QQ plot for each site and for each axis (so a 3x3 figure with 9 QQ plots, well 7 plots actually because Herault is missing a and c). Then you could plot e.g., the 5, 16, 25, 50, 75, 84, 95 percentiles against each other in each plot (common grain size percentiles). Could also include the common bias and accuracy metrics on the plot (cf. Purinton and Bookhagen 2019, Figure 12). For each percentile the uncertainty could also be visualized with vertical and horizontal error bars. Would be useful and you could drop Table 2.

P15L22: “quartile” should be “quantile” here and elsewhere. Quartile refers to the 25$^{th}$, 50$^{th}$, 75$^{th}$, and 100$^{th}$ percentiles. Alternative to using quantile, consider just using the term “percentile” which is more common in grain-size studies. Quantile is usually the 0.1, 0.2, 0.3, etc., whereas percentile is 10, 20, 30.

Table 2: I suggest to remove this. A Figure with vertical and horizontal error bars as I suggest in place of Figure 6 would be much easier to read and present the results more strongly.

P17L5: Did you mean "point/cm2"? is this supposed to be a point density? or is this a distance between points?

P17L6: “10^6”. But the SfM datasets had way more points right? On the order of 10^7-9 is what I would expect from SfM point clouds of pebbles. I saw you reduced the number to "speed processing", how long would the processing have been at full resolution? Days?

P17L19-20: “Suitable values…” A bit more guidance should be provided. what would be interesting is to show (maybe in the supplement) a zoomed in area of the point cloud with labels resulting from e.g., k = 10, 20, 30, 40, 50, to see how this effects the segmentation.

Section 4.2: By reducing the density of SfM points you are removing one of the key benefits of this method: denser point clouds! That should be noted. Point density could be very useful for measuring smaller grains…

P17L24-26: Yes! we have noted this too on SfM point clouds. This may be a Metashape issue, or an issue with the number of photos and the angle of the photos. It should be noted that the source of these issues may be related to the quality of photos taken (lighting, resolution, blurriness) and/or lack of sufficient coverage / view angles -- it is not necessarily a result of the underlying SfM algorithms, but we need to explore this more.

P17L29: “built-in”

P18L4: Be consistent of your use of commas with "e.g." and "i.e.", in other places you do not use a comma after.

P18L12: “> 10^2”… well that depends on the size of the grains versus size of surveyed areas and the density of the point cloud. It’s tricky to put a number on this.

P18L13: Here and elsewhere I think you mean “representation”, not “representativity” (not a word).

P18L14: Yes! Here you could mention that an entire surface of the grain is retrieved, so we are not limited to fitting ellipsoid models (though this is still useful w.r.t. historical approaches).

P18L16: Well, drone usage in this case is really challenging, and maybe not realistic, see Section 5.4 in Purinton and Bookhagen 2019.

P18L18: “hundred grains”

P18L21: “spent in”

Section 4.5: Good that you make some of these statements, but I think they can be teased earlier. You are not just limited to the a/b/c axes when we start labeling point clouds, but a wealth of other more accurate information about the grain shape (cite Domokos’ work here).

P19L22: Note, orientation is also provided by 2D image segmentation methods (e.g., PebbleCounts), although this is only the "azimuth" and not the "dip".

P20L26: I think "select" not “generate”. Why only 1000 grains? Why not use all grains?

Figure 9 caption (and Figure S7 caption): Check your definition of the boxplot. If 100% of the data was always in the whiskers there would be no red outliers outside the whiskers.

P21L13: "solves the issue" is a little strong here. I would say "makes progress on the issue".

P21L14: I think "methodological advance" is also a little strong. Rather, the authors take the watershed approach concept used by many other algorithms / studies and apply it to point clouds in the framework of a network-based approach. This is an "alternative" and a "unique application to point clouds", but I would refrain from "advance" w.r.t. other studies that use different approaches.

P21L19: Not sure what is meant by this last statement, maybe rephrase.

P21L20: Yes, it detected all grains in the lab experiment, but is that an interesting or notable result? The grains were far from each other and not overlapping (and the area between them was removed manually I think?). The lab experiment was more about testing the ellipse models.

P22L7-8: Shouldn't this statement be in the discussion?

P22L9-13: This is really late to be introducing previous and highly relevant work. A discussion of this alternative approach should be up in the introduction or methods section. I agree with the point about G3P used to generate training data, but that could go in the discussion.

Appendix:

P29L1: This appendix is enormously important! I think it should really be a section at the beginning of the results. It really helps my understanding of the models and the limitations based on grain shape. Grain shape is extremely important if you also make the statement about the ellipsoid model being an imperfect descriptor of natural fluvial sediments

P29L7: “consist of”

P29L9: Be specific, how were all sides of the grains collected? Was the object rotated to get a photo from every side? were they laid on a flat surface (so there is still part of the grain "missing")?

P29L10: I'm not sure if you just want a percentage or a rather what is more interesting is taking a percentage from a certain region, or dropping an area of points to simulate an occlusion. Is that what is done? Not clear from this description.

P29L12: “seeked”?

P29L15: Convex hull is used here! It should also be used in Figure 4 rather than "cuboid volume".

Figure A1: Figure is pretty low-resolution. It would be nice if when I zoom in on the panel b I can see the points and ellipsoids, but the resolution is too low for that right now. Not strictly necessary, but this would be really helpful.

P30L14: But these ellipsoidal models are likely increasingly "wrong" for more angular grains. I don't think you need to modify the analysis but this point should definitely be noted! I think the path forward for grain-size measurement from point clouds does not lie in ellipsoidal models, but rather measurements directly on the labelled points.

P30L15-16: This is important! DLSF overestimates c. We see that in Figure 4. That's why this appendix / figure should come at the beginning of the results.

P30L17: “assuming that”

**References**

Purinton, B., & Bookhagen, B. (2021). Tracking downstream variability in large grain-size distributions in the south-central Andes. Journal of Geophysical Research: Earth Surface, 126, e2021JF006260. https://doi.org/10.1029/2021JF006260

Rheinwalt, A., Goswami, B., & Bookhagen, B. (2019). A network-based flow accumulation algorithm for point clouds: Facet-Flow Networks (FFNs). Journal of Geophysical Research: Earth Surface, 124. 2013–2033 https://doi.org/10.1029/2018JF004827

Smith, M., Carrivick, J., and Quincey, D.: Structure from motion photogrammetry in physical geography, Prog. Phys. Geog., 40, 247–275, https://doi.org/10.1177/0309133315615805, 2015