Earth Surf. Dynam. Discuss.,
https://doi.org/10.5194/esurf-2018-53-RC2, 2018

# *Interactive comment on* "A Structure from Motion photogrammetry-based method to generate sub-millimetre resolution Digital Elevation Models for investigating rock breakdown features" *by* Ankit K. Verma and Mary C. Bourke

**P. Sapirstein (Referee)**

orientalizing@gmail.com

Received and published: 14 August 2018

The paper presents a method to create photogrammetric 3D models of rock surfaces, in support of the quantitative analysis of erosion at a micro-topographic scale (areas less than 10 mˆ2). The objective to capture details down to a resolution of about 0.5 mm, and controlling error within a similar scale, is met with a multistage system of field photography and software processing that involves placing a set of three coded targets in the scene. The reusable target field serves both as a scale bar and as the local coordinate system. Most of the paper is concerned with testing the accuracy of the

proposed photogrammetric method, rather than the geomorphological analysis which motivated its development.

I have waived my anonymity as a referee in part because I have published about methods to improve and assess the accuracy of photogrammetric 3D models in archaeological research. I recommend the authors consider these papers along with the additional literature cited in the bibliographies: P. Sapirstein (2018) "A high-precision photogrammetric recording system for small artifacts" J. of Cultural Heritage 31: 33–45 with 10pp suppl. P. Sapirstein, S. Murray (2017) "Establishing best practices for photogrammetry in archaeology" J. of Field Archaeology 42: 337–50 P. Sapirstein (2016) "Accurate measurement with photogrammetry at large sites," J. of Archaeological Science 66: 137–45 I enjoyed reading this paper, which has reminded me how geologists have developed interests in photogrammetry parallel to those of archaeologists, both working in similar directions. It would be good for the two areas to interact with one another more directly, such as through interdisciplinary citation and conversation.

On the positive side, the authors are to be commended for their thorough citation of geological studies involving photogrammetric modeling. Their triangular target field seems like a good, simple approach to establishing scale at remote sites where the local coordinates and north bearing need not be precisely established. The paper also includes workflows that will be useful for those wishing to learn photogrammetric recording, with many recommendations gained through practical experience.

Still, the contribution of the triangular target field is a relatively small one to the broader field of photogrammetric recording–which has witnessed an explosion in publication since the beginning of this decade about its potential and methods for its application in various contexts. More problematic is the core of the paper, a testing field that the authors use to assess accuracy of their photogrammetric method. As discussed below, the data suggest that the reference measurements on the testing field were distorted, thus invalidating any conclusions about the photogrammetric accuracy. Still, I do not doubt that their proposed method meets the requirements of the geological study, since sub-mm resolution and accuracy is not difficult to attain with photogrammetric modeling at this scale, using the techniques they describe. I believe the paper has potential as a useful publication, but only if it is substantially reworked, beginning with fixing what must be erroneous measurements in the testing chart. Furthermore, given that the literature about photogrammetric methods is well saturated at this point, the paper also needs to make more of an effort to present an actual case study of how geological processes will be analyzed using this 3D data, which seems like the most significant potential contribution of this research.

Specific comments: A) The paper should be framed as a geological case study. It begins this way, but Section 5.3 near the end really should be near the beginning, since it is more of a proposal and justification for generating microscale topographic data in the first place. After the introduction, the paper should justify why a 0.5 or 1-mm resolution/error is needed for doing this sort of analysis, and any other attributes of the 3D models that would serve these objectives. As it stands, the choices of resolution and other processing parameters (e.g., why generate the texture at all?) come off as somewhat arbitrary, raising concerns that the workflow and processing might be needlessly complex and slow–consuming hours of human and computer time rather than minutes to generate potentially viable data. An important omission, required before the final conclusions, is an attempt to show what can be done with the 3D data, specifically the quantitative study of surface roughness, and how this was / will be carried out with the DEMs illustrated toward the end of the paper. The authors do mention a forthcoming paper about this subject, but the readers of the current paper deserve to be given a summary of the results here, and some description of the methods used to assess the 3D models / DEM data.

B) The discussion of other methods (laser scanning, MRMs) could be developed further; as it stands, there is not much basis for comparison provided (such as by laser-scanning and photographing the same subject). The authors might include more explicit estimates for the times required for these methods, at least, so the reader gets

a better notion of how photogrammetry compares practically. The software processing times with photogrammetry can be formidable, and that should be made more explicit in the comparative discussions. The paper calls the other methods “time-consuming” (eg. on page 3), but this seems rather vague, and one could easily characterize photogrammetry in the same way. I agree with the authors’ assertion (page 23, section 5.2) about photogrammetry being cheap and portable, but there are important qualifications to that statement, and it has not been justified well in this paper.

More specifically, how does the roughness analysis of the photogrammetric model compare with that measured by an MRM? Even if there is not a side-by-side test, a little discussion on the resolution, accuracy, etc. of the MRM is warranted.

On page 21, the criticisms of Total station and dGPS survey seem overstated. 1) dGPS measurements of a dozen or more targets, with scale bars to fix the scale, should generate a fully georeferenced model; this is common practice in archaeology, where position and orientation are as essential as an accurate scale. I would imagine that this information would be useful in geomorphological recording as well. 2) Total station measurements are more reliable, with local errors of just a few mm, and have the option of shooting reflector-less in inaccessible locations. It would seem either piece of equipment would be advantageous in many contexts, and in fact the system proposed here with the triangular scale bar kit could be integrated for a hybrid method (e.g., placing scales and targets in the area, and measuring the coordinates with a TS). The discussion should be reframed in a more positive light to admit that these different recording methods are not mutually exclusive, but can potentially complement one another.

C) The error testing methods are problematic. First, the testing environment is nearly flat: a printed, 1.4m square printed chart with 5-cm wooden blocks set on it. The chart is also a mostly blank white sheet of paper. Not only is it completely unlike natural rock surfaces, which vary in depth and texture, the printed chart is poor for SIFT keypoint generation. The blank white background and straight black lines are not useful for these descriptors; only the edges of the targets, the printed text, and the blocks are likely to generate reliable matches.

Second, this testing field tells us something about the accuracy of the overall scaling of the model, but not more. We are presented with no tests of the surfaces generated by the photogrammetric software (from the MVS / dense cloud stage), which might be done with repeatability tests (such as creating many models of one outcrop), or comparisons to reference data (such as created by a high resolution laser scanner). This is a significant omission, since it is the key product that is needed for assessing roughness and other parameters related to weathering. For example, poorly calibrated and oriented cameras introduce a significant amount of noise, which would make the restored 3D surface appear much rougher than the reality.

Third, while it is a good idea to separate horizontal and vertical errors, the use of two completely different testing methods (lengths of scale bars between pairs of coded targets, vs. heights of wooden blocks set on a printed sheet) means that the two error values are not comparable. How are the block heights being extrapolated in the software?

Fourth, and most troubling: figure 6a (section 3.3.3), as well as additional charts in the supplement, show a curious result that two independent photogrammetric measurements of horizontal scale bar lengths agree with one another very well, yet differ greatly (about 0.2–0.9 mm, correlated to total length) from the dimensions printed on the testing chart. That is, the consensus of 2/3 of the measurements indicate that the printed chart dimensions are incorrect. Furthermore, the discrepancies in the figures present a sawtooth pattern, flipping in the positive and negative directions (less / more than the printed chart) at a similar scale. The authors explain that they generated the testing chart dimensions in design software and printed it, presumably on a plotter, for the test photography. In their results, the dimensions for S1, S3, S6, S8, S10, S12, S14, and S16 from the two photogrammetric measures are less than the expected length on the printed chart, while the others have positive discrepancies. The explanation for

this distinctive pattern begins with the chart itself: the set I just listed are all vertically oriented on the printed sheet, while the others are horizontal.

I encourage the authors to account for this problem. Photogrammetric recording can be very precise: 1:10,000 is easily obtainable with coded targets (so, errors all below 0.1 mm in length at the size of this testing scene), and it would be difficult to conceive of how a warp that would increase scale on one axis at the expense of another could possibly be introduced. However, scaling distortions of one axis relative to another is common with printing. Some printers are in fact designed to insert subtle distortions to foil counterfeiters, but other reasons like curling of the paper (common with plotter paper) might account for these distortions–which are on the order of just 1 mm, after all, and thus would be hard to see.

Due to these problems with the chart, the conclusions about prime vs. zoom lenses, etc., are invalid, since they were tested against faulty reference measurements. If the actual lengths on the testing chart can be determined, then the photogrammetric estimates could be assessed from the same data, and the authors may be able to reproduce known phenomena in previously published research, such as improved accuracy from a fixed lens (including fixed focus settings) relative to an unstable lens.

D) On the image format (Main 2.3, Table 4, and Supplement 2.5), it is claimed that the JPEG format increases error relative to lossless formats, yet the reported increase in error is so high as to raise flags. The procedure with RAW photography converted later to TIFF adds a significant amount of time and raises storage requirements, which would only be justified if JPEG were indeed much less reliable than TIFF. In my own tests, I found a small effect, with JPEG imagery being about 97-99% as metrically consistent as TIFF images. By that, I mean repeatable for length measurements; so, for example, a TIFF-based scene with length errors of 1.00 mm might have errors of 1.02 mm if based on maximum-quality JPEGs, which for most purposes is negligible.

Of course, this could vary with processing settings and the camera. Since the text and supplement do not specify how the JPEGs were created, it is hard to account for the very high JPEG error, but this may be due to using a relatively high compression ratio. JPEG encoding introduces strong artifacts next to high-contrast straight edges as the quality is reduced below the maximum setting; even 95% quality begins to create artifacts that could interfere with SIFT matching. Furthermore, the testing imagery is basically all black and white lines, which is exactly where JPEG performs its worst. JPEG is designed for photographs of natural forms with comparatively smooth textures, much more like the natural features in the study than the testing chart.

---

Interactive comment on Earth Surf. Dynam. Discuss., https://doi.org/10.5194/esurf-2018-53, 2018.