Introduction

This paper analyzes the accuracy of Structure for Motion (SfM) snow depth data products derived from photography acquired with a small unmanned aerial vehicle (UAV). The authors apply a UAV / SfM technique to three distinct environments throughout the spring ablation period. Results from each environment provide quantitative information about snow depth, but based on error analysis (and low signal to noise ratios (SNR)), the authors claim only snowcovers deeper than 30 cm can be reliably estimated with this technique. Despite this limitation, future UAV / SfM technological and operational advances hold promise for observing snowcovers at unprecedented spatial and temporal resolutions.

Developing reliable techniques for collecting snow depth observations at high spatial and temporal resolutions is a line of inquiry directly relevant to The Cryosphere. The authors succeed in presenting novel tools and data. Applications of UAV/ SfM techniques to snow are in their infancy, and the existing literature is sparse and relies on a small number of survey flights. In this case, a large number of survey flights are considered, which is a useful addition to the literature. Using the Rose criterion ($\mathrm{SNR} \geq 4$) to assess the quality of snow depth estimates is novel to this study and has not been applied in other recent UAV / SfM studies (Bühler et al., 2016, De Michele et al., 2015, Vander Jagt et al., 2015). The results presented in this study are a valuable contribution to our understanding of what challenges future UAV snow surveys may encounter, especially regarding the influence of vegetation on digital surface models (DSMs) during the ablation period. But the authors are not careful enough in differentiating the limitations of their system from those of other methods of applying SfM to snow, and leave some erroneous impressions as a result. With careful attention to correcting that deficiency, as well as an effort to tighten and shorten the paper, it could make a nice contribution once revised.

General Comments

More information, context, and discussion regarding the UAV system would help frame the results and conclusions presented in the study. The world of UAVs and their payloads is broad and quickly expanding. Given the diversity of aircraft, cameras, and processing techniques available, the authors should refrain from representing the results from one UAV system (theirs) as indicative of UAV / SfM snow estimation techniques as a whole. Quantitative results may be particular to the UAV system of choice. More discussion of how the choice of aircraft, payload, and processing software may have influenced results is needed.

For example, the Sensefly Ebee Real Time Kinematic aircraft was shown to be sensitive to wind speeds greater than 6 ms^{-1}. While this conclusion may be useful to future surveyors (i.e. it may not be worth their time to collect data on windy days), other platforms, such as rotary aircraft or even delta wings with more sophisticated autopilots, may be able to compensate and collect consistent data at higher wind speeds. Do the authors recommend future surveys use rotary platforms? Or does the decreased flight range / endurance of rotary aircraft compared to fixed wings outweigh the increased stability? How much of an operational concern is wind sensitivity, given that snow precipitation events and wind events frequently coincide?

A similar discussion of the camera payload would be useful to readers as well. What is the specific model of the Canon IXUS used in this study? A quick Google search yields at least a dozen different models. What are the specifications of the camera? In particular, what is the bit depth? The point about the camera automatically adjusting exposure based on center-weighted values and overexposing some scenes, causing erroneous points, is important. More discussion of this type is useful – for example, that those planning a UAV snow survey should avoid cameras with automatic light metering. Also, the authors mention their system is not equipped with a stabilizing gimbal, which clearly increased wind sensitivity and decreased vertical accuracy. A 3-axis gimbal capable of maintaining an ideal camera orientation is a common feature of many consumer or "prosumer" level UAVs. A gimbal would certainly increase the quality of the SfM inputs, and therefore perhaps the snow depth resolution. Readers interested in snow, but perhaps UAV/SfM novices, would benefit from a more detailed discussion of the camera system used in the study.

The discussion the authors provide about the Postflight Terra 3D data processing is useful and a good model for how the authors could add context to the aircraft and camera components of the UAV system. "Black box" processing algorithms are a frequent frustration for scientific users trying to understand SfM error sources.

Whether or not the erroneous points caused by overexposure are included in the authors' results is unclear upon first reading. For example, section 3.1 (256 - 257) reads "These results exclude areas affected by erroneous points, as described in section 3.3.2, which was small compared to the total snow-covered area." Which results are the authors referencing? Are the authors speaking generally about every single treatment? Or just the alpine-bare? For example, the authors should consider replacing "These results" with "The alpine-bare results" or "All results." In general, an instance of the word "this" or "these" which lacks a referent can be confusing to the reader because they are unsure as to what precisely the writer is referring. After reading section 3.1 it seems the authors did not include the erroneous points for some or all of the results - but upon referencing section 3.3.2 (322 - 324) the reader finds conflicting information: “Erroneous points could be eliminated with the removal of overexposed images. However, reducing the number of images in such a large amount caused a larger bias and gaps in the point cloud, which made this method inappropriate.” Are the overexposed erroneous points included in the results or not? If the erroneous points are included, specify which results are impacted.

Although the literature is sparse regarding SfM estimates of snow, the authors must be wary of comparing results derived from much different methods. For example, in the discussion section (286 - 292) the results are contrasted against the findings of Nolan et al. 2015, despite their methods using a manned aircraft. Similarly, Buhler et al. 2015 is a reference to a manned aircraft experiment. Given the topic sentence of this section begins “Differencing of *UAV* derived DSMs…” (emphasis mine) some readers may find the contrast of the authors’ results with that of a manned aircraft campaign misleading. Also, the 30 cm mean error reported by Nolan et al. is a geolocation error rather than a snow depth error. Snow depth errors were reported as 10 cm, and rigorously documented. Mean snow depths are not reported by Nolan et al. and thus as readers we cannot calculate or assess the SNR of his results, but it does seem like this study is suggesting a higher snow depth threshold for measurements than Nolan et al. That needs to be addressed.

Technical Corrections

Line 222 - 223: “RSME” is twice given as the acronym for root mean square error, rather than RMSE.

Line 244: The mean RMSE for the alpine-bare treatment is 8.1 cm, but in Table 1 the value is 8.7 cm.

Recommendation

I recommend the authors make revisions to the paper based on the comments above. In general, the authors need to discuss the results appropriately with respect to the referenced work and use more precise language. Also, given the limited scope of this study, readers will prefer a considerably shorter paper. Striving for concision may improve the clarity of the paper as well. The paper could easily be shortened by up about 30%.