Dear Referee,

Firstly, thank you for your valuable time and contribution. The comments given will definitely help improving the quality of this research paper. The responses to your comments are hereunder detailed point-by-point, and are complemented by a marked-up and a corrected manuscript version.

1. Commercial software have been used (Pix4D, 3DFlow). People working with SfM know that various processing packages produces various results. Can you comment on that in the discussion how do you think that results would have been different if other packages would have been used?
   Response. We used two well known packages that are commonly used by many research groups. These software provide state-of-the-art results. The use of conventional instruments was decided on purpose in order to be closer to the "normal" processing of data. Similar results could be generated with other software resources like AgiSoft and Photoscan Pro, or open source packages such as ColMap, MicMac, etc. Papers like Remondino et al. (2012); Gross (2015) have already discussed the different performances of these software, and differences were not significant.
   Modifications. The explained is now added in **page 18 line 34**: "Similar results are expected with the use of other commercial software (e.g. Agisoft, Photoscan, etc.) or open source packages (e.g. ColMap, MicMac, etc.)".
2. WFS, RFS and IQI refers to 3DFlow functionalities Please explain shortly the principles and provides references.
   Response. IQI functionality is explained in **page 8 line 8**: "Image Quality Index (IQI) indicate possible sources of error for image-based 3D scene reconstruction, such as the presence of low-texture areas or motion-blur effects (3D Flow, 2017)". RFS does not refer to 3DFlow functionalities; however, it is also described in **page 5 line 13** as a method "only defined by an empirical number of randomly-selected frames (defined in Pix4D)".
   Modifications. Wise Frame Selection (WFS) functionalities are now shortly described in **page 6 line 2**: "WFS uses as guideline each frame initial 3D position (generated with Pix4D (2017)) and their correspondent Image Quality Index (IQI) value (computed with 3D Zephyr 3DFlow (2017)), with the aim of discarding the most redundant (i.e. more than 80% overlap) and lowest quality (i.e. an IQI lower than 0.5)

frames, respectively.".

3. Please provide some extra information on the CNN model used (to avoid to have read Vetrivel et al 2017)
   Modifications. In **page 9 line 2** more information about the CNN model is now provided: "A deep learning approach presented by Vetrivel et al. (2017) was tested. This is in based on the imagenet-caffe-alex (Krizhevsky et al., 2012) CNN model which is composed of different groups of layers. Convolutional layers represent the first group, and are a set of filter banks composed of image and contextual feature filters. The following group corresponds to data shrinking and normalization layers. Finally, the last group transforms all the information generated and outputs features with high-level reasoning; usually this layer is connected to a loss function for the final classification. This approach uses a large amount of training samples (i.e. labelled images) to tune the weights of the CNN classification layer."

4. How long would be the full process in an operational context? from the raw video footage arrival to a SDA.
   Response. The present research aims at comparing video data and aerial photos for SDA, and not implicitly at implementing a real-time procedure. Nevertheless, from the analysis performed some limitations are evident, mainly for a rapid video-based 3D reconstruction and damage-feature depictability. These are for example: The lack of proper 3D GCPs for and efficient geo-registration, the lack of pre-existing data, a software-integrated operational flow and an approach to aggregate all the 3D damage-related features.

We hope all the comments were sufficiently clarified and corrected. We remain at your disposal for any additional comment or modification you might deem necessary.

Kind regards,
The authors

References

Gross, J. W.: A comparison of orthomosaic software for use with ultra high resolution imagery of a wetland environment, Center for Geographic Information Science and Geography Department, Central Michigan University, Mt. Pleasant, MI, USA. Available from: http://www. imagin.

org/awards/sppc/2015/papers/john_gross_paper. pdf, 2015.

Remondino, F., Del Pizzo, S., Kersten, T. P., and Troisi, S.: Low-cost and open-source solutions for automated image orientation–A critical overview, in: Euro-Mediterranean Conference, pp. 40–54, Springer, 2012.