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. Probably reducing the frame redundancy using the WFS for video data can generate a lack of important photos, needed for an accurate Point Cloud calculation. For that reason, the video results with WFS are generally worse than the RFS. Maybe the WFS method is not appropriate in this case because the photos selected as good were in a non-optimal position. Is clear that the results of a Point Cloud generation are strongly related to the overlap and sidelap percentage between photograms. Maybe with different rules for WFS the results could be even better than using the RFS, even if the latter method is certainly faster.
   Response. There are different video frame selection approaches which are based on weights and thresholds (Hasegawa et al., 2000; Ahmed et al., 2010; Alsadik et al., 2013); however, it is very complex to define an optimal strategy that works in all conditions. WFS method is an approximation, which ensures having a sufficient number of frames for the 3D reconstruction, selected manually based on their IQI and redundancy. Different results are expected using more elaborated methods, but WFS method can work quite well in most of the practical cases.
   Modifications. This is now explained in **page 20 line 5**: "The use of a more elaborated approach might result in more accurate 3DPCs; however, the variability of data and external characteristics limits the development of an optimal frame selection approach that can deal with all conditions".
2. Considering 3D Internal Accuracy assessment. The authors use planar fitting as a method to assess the accuracy of the resulting 3DPCs. It seems that only one planar object was used to perform this analysis. Maybe, considering more planar objects preferably equally distributed, for example, with respect to shaded and well-lighted areas or central or peripheral position could lead to more robust conclusions.
   Response. More objects were analyzed during the research; however, results were statistically similar and due to the paper extent only the most meaningful experiments are shown.

3. Considering 3D External Accuracy: it is not clear how the distance between each 3DPC and the reference one is considered. Is it a mean distance value, calculated over the entire extent? Is it an average value of the distance calculated in correspondence of reference objects or areas? Moreover, how the presence/absence of GPS-measured Ground Control points can affect the final accuracy and usability of the results?
   Response. It refers to the mean distance to the reference 3DPC, calculated from all 3D points distances to the reference 3D points. Besides, the presence/absence of Ground Control Points (GCPs) is a relevant parameter, because video data in most cases do not hold accurate geopositional information and depend either on GPCs or an accurate 3D model. Moreover, the geo-localization of video frames is limited by their oblique perspective and resolution. Low quality GCPs can affect the geometrical accuracy of the video-generated 3DPCs and also the depictability of damage-related geometrical features.
   Modifications. 3DPC external accuracy is now better explained in **page 9 line 31**: "the mean distance of the 3D points to the reference 3D points was computed to determine every 3DPC external accuracy", and also in **page 13 line 6**: "Independent oblique and nadir photos-based 3DPCs registered high accuracies, with mean distances of 2 and 4 cm to the reference model, respectively" and also **Page 15 Figure 6, Y axis**: "Mean distance to the reference model".

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

Ahmed, M. T., Dailey, M. N., Landabaso, J. L., and Herrero, N.: Robust Key Frame Extraction for 3D Reconstruction from Video Streams, in: VISAPP, 2010.

Alsadik, B., Gerke, M., and Vosselman, G.: Automated camera network design for 3D modeling of cultural heritage objects, Journal of Cultural Heritage, 14, 515–526, 2013.

Hasegawa, H., Matsuo, K., Koarai, M., Watanabe, N., Masaharu, H., and Fukushima, Y.: DEM accuracy and the base to height (B/H) ratio of stereo images, International Archives of Photogrammetry and Remote Sensing, 33, 356–359, 2000.