Automatic Stockpile Volume Monitoring using Multi-view Stereo from SkySat Imagery

Roger Marí, Carlo de Franchis, Enric Meinhardt-Llopis, Gabriele Facciolo

AUTOMATIC STOCKPILE VOLUME MONITORING USING MULTI-VIEW STEREO FROM SKYSAT IMAGERY

Roger Marí [1] *Carlo de Franchis*[1,2] *Enric Meinhardt-Llopis*[1] *Gabriele Facciolo*[1]

[1]Université Paris-Saclay, CNRS, ENS Paris-Saclay, Centre Borelli, France
[2]Kayrros SAS

ABSTRACT

This paper proposes a system for automatic surface volume monitoring from time series of SkySat pushframe imagery. A specific challenge of building and comparing large 3D models from SkySat data is to correct inconsistencies between the camera models associated with the multiple views that are necessary to cover the area at a given time, where these camera models are represented as Rational Polynomial Cameras (RPCs). We address the problem by proposing a date-wise RPC refinement, able to handle dynamic areas covered by sets of partially overlapping views. The cameras are refined by means of a rotation that compensates for errors due to inaccurate knowledge of the satellite attitude. The refined RPCs are then used to reconstruct multiple consistent Digital Surface Models (DSMs) from different stereo pairs at each date. RPC refinement strengthens the consistency between the DSMs of each date, which is extremely beneficial to accurately measure volumes in the 3D surface models. The system is tested in a real case scenario, to monitor large coal stockpiles. Our volume estimates are validated with measurements collected on site over the same period of time.

***Index Terms*—** change detection, stockpiles, RPC refinement, bundle adjustment, SkySat, surface monitoring

1. INTRODUCTION

Stockpile measurement is a task of major importance in a wide variety of industrial activities involving the storage, treatment and transport of bulk materials (e.g. mining [1], landfill management [2]).

Large scale areas represent a particularly adverse scenario to measure vast piles of possibly irregular shapes. Primary methods, such as human-led topographic inspections, can pose a safety risk and are ineffective in terms of time and accuracy. This has raised the interest on technology-aided surveys employing ranging sensors and photogrammetry. Lidar scans and imagery from Unmanned Aerial Vehicles (UAVs), planes or robots, have proven to be highly accurate and have gained great popularity in recent years [1, 2, 3, 4].

However, aerial or ground based solutions are often costly or not viable in restricted areas, which may prevent frequent updates. Recurrent satellite imagery is emerging as an alternative, as it allows to classify, segment and reconstruct large areas without on-site actions [5, 6, 7, 8]. Today's abundance of Earth observation satellite images, notably due to the breakthrough of *SmallSats*, explains this growing trend. SmallSat constellations have many low-cost satellites flying at low-altitude orbits, that produce high-resolution images with short revisit times, e.g. PlanetScope, SkySat [9].

This work was supported by a grant from Région Île-de-France. It was also partly financed by IDEX Paris-Saclay IDI 2016, ANR-11-IDEX-0003-02, Office of Naval research grant N00014-17-1-2552, DGA Astrid project « filmer la Terre » nº ANR-17-ASTR-0013-01, MENRT.

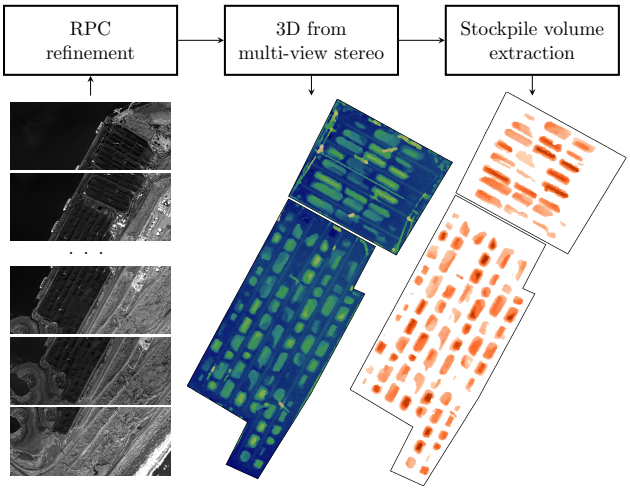

Fig. 1: Diagram of our surface volume monitoring approach. The RPCs of a time series of SkySat acquisitions are refined and used to compute a high-quality surface model for each date, where volume is measured.

Unlike traditional satellite imagery, SkySat products are delivered as mosaics of images, denoted as *scenes*, with a small geographic footprint. To this end, the pushframe acquisition systems ensure a certain overlap between consecutive scenes [10]. Each scene is provided with a Rational Polynomial Camera (RPC) model, but the different RPCs associated with a mosaic typically exhibit non-negligible bias. Inconsistencies between camera models lead to systematic errors in depth estimation from image correspondences, hindering multi-view 3D reconstruction. RPC correction methods are indispensable to exploit such fragmented data, which is in contrast to the case of large-footprint satellite images [11]. In addition, highly dynamic areas like open-air storage facilities pose an extra challenge for RPC refinement: images acquired at different dates should be handled without assuming a static scene over time.

Motivated by the previous observations, we present an automatic system to monitor stockpiles or large irregular volumes, in general, from time series of SkySat scenes. The method can be applied to stereo and tristereo acquisitions [10]. Our contributions are:

- A generic date-wise RPC refinement, independent of satellite specificities and able to handle areas that change over time.
- A volume tracking strategy based on a time series of high-quality surface models obtained by multi-view 3D reconstruction using the refined RPCs.
- A performance validation based on data acquired on site.

2. RELATED WORK

2.1. Stockpile volume monitoring

Most image based approaches to compute stockpile volume perform a 3D reconstruction of the area based on the dense matching of multiple views [1, 2, 3, 5]. Single image methods using site-specific heuristics or shape from shading have also been explored [8].

Once the 3D geometry of the scene is computed, stockpile volume can be measured in different ways. Cross-section methods model piles as big fairly regular solids, while horizontal section methods divide them into layers following contour lines [2]. For finer estimations on irregular shapes, it is common to discretize the scene into small elementary 3D volumes, i.e. voxels, tetrahedrons, trigonal prisms [12, 4]; or into Digital Surface Models (DSMs), i.e. a 2D grid where each cell is assigned an altitude value [1, 13]. The boundaries of the piles are usually obtained by subtraction of a bare terrain model, possibly combined with a segmentation step [3, 2].

2.2. RPC model refinement

The RPC model is a generic camera model, independent from specific physical properties, widely used to describe satellite optical sensors. The RPC of a satellite image relates 3D space coordinates (latitude, longitude, height) to 2D image coordinates (pixel row and column). The 3D to 2D mapping and its inverse are known, respectively, as the *projection* and *localization* functions.

In practice, commercial satellite imagery RPCs contain inaccuracies caused by internal measurement errors of the complex physical system they encode, which need to be refined [14]. RPC refinement strategies typically rely on a set of tie-points, whose projection across the input views has to coincide. Bundle adjustment methods [15] are a well known solution, which seeks to minimize the reprojection error of the tie-points by optimizing a set of parameters that modify the original camera models. [14] proved that composing each RPC projection with a 2D offset suffices to adjust satellite images covering lengths of 50 km or less, as this compensates the main source of errors, i.e. inaccuracies in sensor attitude. An equivalent solution to remove such errors is to apply a correction rotation around each camera center previous to the RPC projection [11]. Functions of polynomial form, defined in object or image space, depending on whether they are applied before or after the input RPC mappings, can be used to additionally address possible secondary sources of error (e.g. time-dependent drift, lens distortion) [14].

3. METHODOLOGY

3.1. Bundle adjustment to correct camera orientation

Since we do not have access to Ground Control Points (GCPs) of the study area, i.e. points whose object and image coordinates are known in advance, we perform a relative RPC correction based on tie-points derived from feature correspondences. The correction is done in a date-wise manner, solving an independent bundle adjustment problem for the group of cameras of each acquisition date. This ensures that the geometry seen by the cameras is coherent, which is not guaranteed if multiple dates are treated at once. Thus, feature mismatches are minimized and the accuracy of tie-points increases.

Note that a relative correction is sufficient to register the RPCs in a common frame of reference and use them for 3D reconstruction, but the absolute location of the scene remains subject to the geolocation accuracy of the input models. To avoid large drifts in object space, we fix a reference camera for each date, which is not refined.

We refer to the list of image coordinates containing the location of a tie-point across multiple images as a *feature track*. SIFT keypoints [16] are extracted for each image and matched to the keypoints of the overlapping views with a sufficient baseline, using a ratio test of 0.6 to reject outliers. The union-find algorithm from [17] is used to obtain a set of feature tracks of arbitrary length from the unordered collection of pairwise correspondences.

The 3D coordinates of the tie-point associated with each track are initialized by triangulating all the pairwise matches of the track with the input RPC models, as in [7], and taking the mean of the 3D locations. All tie-point coordinates are optimized afterwards, simultaneously with the correction parameters of each camera.

Our method refines the RPC models by composing them with a preceding rotation around an approximation of the camera center. The bundle adjustment problem is therefore expressed as

$$\min_{R_m, \boldsymbol{X}_k} \sum_{k=1}^{K} \sum_{m=1}^{M} \|\mathcal{P}_m(R_m(\boldsymbol{X}_k - \boldsymbol{C}_m) + \boldsymbol{C}_m) - \boldsymbol{x}_{mk}\|^2, \quad (1)$$

where Equation 1 represents the reprojection error of the setting. $\boldsymbol{X}_k$ denotes the k-th tie-point, $\boldsymbol{x}_{mk}$ its observation on the m-th image, and $\mathcal{P}_m(R_m(\boldsymbol{X}_k - \boldsymbol{C}_m) + \boldsymbol{C}_m)$ its reprojection given by the m-th RPC projection $\mathcal{P}_m$ and the correction rotation R_m around the center $\boldsymbol{C}_m$. The camera center $\boldsymbol{C}_m$ is derived by regressing a projective model from each RPC model and it remains fixed during the optimization. This is similar to the bundle adjustment tool of [18].

R_m is initialized as the identity matrix and used in the bundle adjustment procedure using the Euler angles representation, which entails 3 variables to be optimized per camera, i.e.

$$R_m = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & -\sin\phi \\ 0 & \sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (2)$$

where ϕ, θ, α are the 3 Euler angles related to the m-th camera.

3.2. Multi-view stereo reconstruction

After RPC refinement, the open-source satellite stereo pipeline S2P[1] [7] is run with the corrected camera models. For each date S2P is used to reconstruct DSMs of the parts of the study area seen in all pairs of overlapping scenes that meet certain criteria (Fig. 2(a)): the convergence angle of a pair of scenes must be more than 5 degrees and less than 35 degrees; and the bounding box of the triple intersection between the two scenes footprints and the study area must have both its dimensions larger than 200 meters. The reconstructions from different pairs are natively registered in the object space as a result of RPC correction. This allows to easily merge them into a denser and highly accurate model of the entire area by taking the average height value at each DSM cell (Fig. 2(b)). Small holes in the DSMs, due to occlusions between views or lack of texture, are filled by a 5×5 median filter followed by cubic interpolation. The Ground Sample Distance (GSD) is set to 1 m for all DSMs.

Similarly to [19], the DSMs of different dates are aligned by a 3D translation that maximizes the Normalized Cross Correlation between their geometry. Even if the geometry may change between different dates, this alignment serves to minimize the standard deviation of DSM heights over time, which is later exploited to determine a coarse mask of the dynamic parts of the study area (Fig. 3).

Fig. 2(c) and 2(d) highlight the impact of RPC refinement: the point-wise deviation between height values of DSMs from different pairs is of the order of a few meters for the unrefined RPC models, while it is reduced to tens of centimeters using the corrected ones.

[1] https://github.com/cmla/s2p

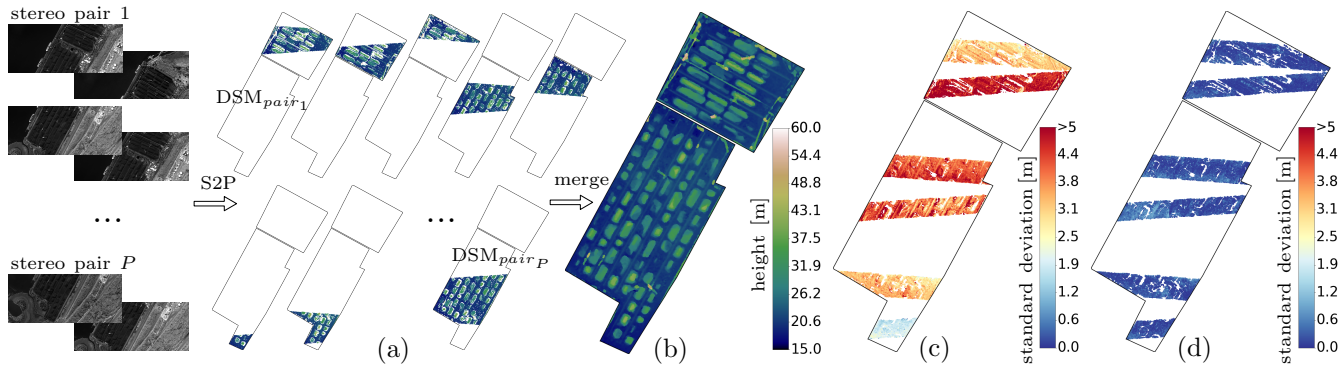

Fig. 2: Illustration of the multi-view stereo process run to reconstruct the study area at each date. (a) Several DSMs of the different parts of the area are computed independently from different stereo pairs. (b) Thanks to RPC correction, the DSMs from different pairs are accurately registered and can be merged directly by taking the average height at each 2D cell. (c) and (d) show, respectively, the standard deviation, in meters, between height values of DSMs from different pairs with and without RPC refinement (the DSMs overlap only on the shown strips).

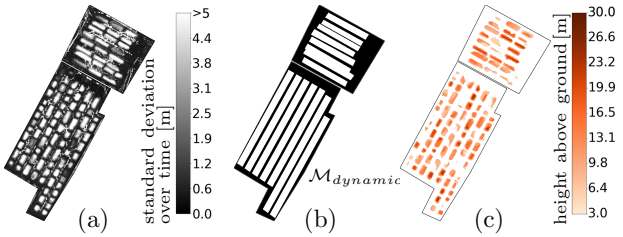

Fig. 3: (a) Point-wise standard deviation of height values along the DSM time series. (b) Mask of dynamic parts. (c) Example of nDSM.

3.3. Volume estimation

Time series of aligned DSMs from remote sensing can be effectively employed to measure and track volumes over an area of interest. As a concrete example of this, we propose an automatic system to monitor stockpiles volume in a real case scenario (Section 4.1).

For each DSM, a Digital Terrain Model (DTM) is subtracted to consider only the heights above ground. Since our study area lies on a flat terrain, we model the DTM as a plane with height equal to the 25th percentile of the DSM heights. Without loss of generality, cloth simulation methods can be used to model non-flat DTMs [20].

Additionally, we determine a site-specific mask $\mathcal{M}_{\text{dynamic}}$ that delimits the dynamic parts of the area (Fig. 3(b)). The labeling of $\mathcal{M}_{\text{dynamic}}$ is based on the point-wise standard deviation of height values over time, across the different DSMs of the area (Fig. 3(a)).

The normalized DSM (nDSM) containing the heights above ground in areas where changes are expected can be expressed as

$$\text{nDSM}(t) = \mathcal{M}_{\text{dynamic}}(\text{DSM}(t) - \text{DTM}(t)) \tag{3}$$

where t represents the acquisition date of the time series.

Furthermore, only values in nDSM(t) between 3 and 30 meters are kept (Fig. 3(c)). Values outside this range are likely to be due to noise, surface roughness or machinery and cranes working in the area. Note that both height thresholding and the use of $\mathcal{M}_{\text{dynamic}}$ are site and task specific post-processing steps aimed at reducing noise and ensuring that only height values related to stockpiles are left in nDSM(t). In the absence of prior knowledge of the facilities or target volumes, these post-processing steps can be omitted at the cost of a small loss of accuracy.

Finally, the volume of the stockpiles left in nDSM(t) is computed in cubic meters as the addition of all individual cell volumes:

$$V(\text{nDSM}(t)) = \sum_i l_i w_i h_i \tag{4}$$

where l_i, w_i, h_i represent the length, width and height of the i-th cell. We use squared cell DSMs, i.e. $l_i = w_i =$ GSD = 1 m.

4. EXPERIMENTS

4.1. Data

We tested our method on a time series of SkySat panchromatic L1B scenes covering the Richards Bay Coal Terminal (RBCT) in South Africa, which has an open-air storage area of $\sim$1.6 km^2. The RBCT is one of the world's leading coal export terminals. Tonnes of coal stockpiles are managed 24 hours a day to be shipped overseas.

The time series comprises 43 acquisition dates, distributed non-uniformly between January and July 2020. The distance between consecutive dates oscillates between 1 and 20 days, falling below 1 week in most cases. For each date there are 6 to 10 scenes, captured by the same sensor among the 3 sensors of the SkySats, with a difference of a few seconds. All images are free from clouds.

SkySat L1B scenes have a nadir GSD of $\sim$0.72 m and a total size of 1349 $\times$ 3199 pixels. Each scene is delivered with a RPC camera model. The absolute geolocation accuracy of the provided RPC models is of $\sim$30 m, with SkySats orbiting at an altitude of $\sim$500 km [10].

4.2. Validation using on-site stock weight measurements

The volume V of coal stockpiles for each date t of the sequence was computed using the method described in Section 3 (Fig. 4(b)). Coal stock weight measurements, S, collected by on-site agents during the same time period, are shown in Fig. 4(a) in mega tonnes (Mt). The correlation between both sets is recognizable to the naked eye.

To assess the performance of our system, we propose a simple approach to predict coal weight from stockpiles volume, so that the available measurements can be compared in equivalent units. The conversion is not straight-forward, as multiple date-dependent factors may be involved (e.g. humidity factors, non-uniform coal types or pile densities). To this end, we linearly interpolate the two sets of

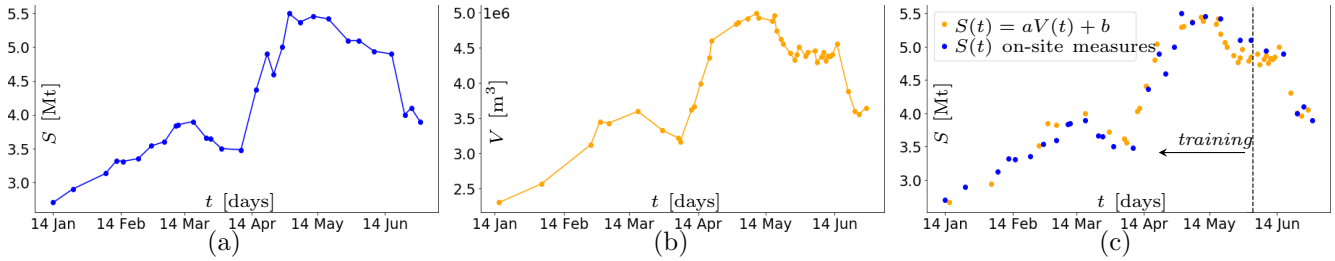

Fig. 4: (a) On-site coal stock measurements, S, in mega tonnes (Mt). Both in this plot and (b) the line linking the set of discrete measures, represented as dots, is linearly interpolated. (b) Stockpile volume, V, as obtained from the time series of photogrammetric DSMs derived from satellite imagery. (c) Linear least-squares regression to predict coal mega tonnes as a function of stockpile volume estimates.

measurements and apply a least-squares regression to fit two coefficients a and b satisfying $S(t) = aV(t) + b$. The result is shown in Fig. 4(c). We obtain $a = 1.02$, $b = 0.3$, where a can be interpreted as the best-fitting bulk density and b as a ground offset. Observe that 85% of the interpolated samples were used to fit a and b (the *training* set), but the strong correlation extends to the rest of dates, stressing the robustness of the method.

It remains difficult to quantify the exact accuracy of the system without measurements that match exactly in time, specially since shipments or stock arrivals can occur within a few hours. In general, Fig. 4(c) seems to indicate that the differences between our weight estimates derived from remote sensing and on-site measurements from neighboring dates are smaller than 0.3 Mt. Occasional larger differences can be due to noisy or interpolated data in the photogrammetric DSMs used for volume estimation, specific stockpile properties or a sudden strong activity in the area.

5. CONCLUSION

An automatic system for surface volume monitoring using recurrent satellite imagery was presented. The system is based on a generic RPC refinement step, independent from satellite specificities, which allows to accurately measure volume from multiple independent DSMs obtained with an open-source satellite stereo pipeline. A time series of SkySat images distributed over ∼6 months was used to validate the system in a real case scenario concerning a highly dynamic area of coal stockpiles.

6. REFERENCES

[1] P.L. Raeva, S.L. Filipova, and D.G. Filipov, "Volume computation of a stockpile - A study case comparing GPS and UAV measurements in an open pit quarry," *ISPRS*, vol. 41-B1, 2016.

[2] G. Tucci, A. Gebbia, A. Conti, L. Fiorini, and C. Lubello, "Monitoring and computation of the volumes of stockpiles of bulk material by means of UAV photogrammetric surveying," *Remote Sensing*, vol. 11, no. 12, pp. 1471, 2019.

[3] H. He, T. Chen, H. Zeng, and S. Huang, "Ground control point-free unmanned aerial vehicle-based photogrammetry for volume estimation of stockpiles carried on barges," *Sensors*, vol. 19, no. 16, pp. 3534, 2019.

[4] W. Zhang and D. Yang, "Lidar-based fast 3D stockpile modeling," in *ICICAS*, 2019, pp. 703–707.

[5] B. Schmidt, M. Malgesini, J. Turner, and J. Reinson, "Satellite monitoring of a large tailings storage facility," in *Tailings and Mine Waste*, 2015.

[6] P.M. Salim, M.A. Jais, N. Sahriman, A.M. Samad, M.A. Abbas, I. Maarof, and N. Tarmizi, "Monitoring quarry areas using remote sensing techniques," in *CSPA*, 2017, pp. 323–328.

[7] C. de Franchis, E. Meinhardt-Llopis, J. Michel, J-M. Morel, and G. Facciolo, "An automatic and modular stereo pipeline for pushbroom images," *ISPRS*, vol. 2-3, pp. 49–56, 2014.

[8] M. d'Autume, A. Perry, J-M. Morel, E. Meinhardt-Llopis, and G. Facciolo, "Stockpile monitoring using linear shape-from-shading on PlanetScope imagery," *ISPRS*, vol. 5-2, 2020.

[9] G. Curzi, D. Modenini, and P. Tortora, "Large constellations of small satellites: A survey of near future challenges and missions," *Aerospace*, vol. 7, no. 9, pp. 133, 2020.

[10] Planet Labs, "Planet imagery product specifications," 2019.

[11] R. Marí, C. de Franchis, E. Meinhardt-Llopis, and G. Facciolo, "To bundle adjust or not: A comparison of relative geolocation correction strategies for satellite multi-view stereo," in *ICCV Workshops*, 2019.

[12] J. Zhu, J. Yang, J. Fan, D. Ai, Y. Jiang, H. Song, and Y. Wang, "Accurate measurement of granary stockpile volume based on fast registration of multi-station scans," *Remote Sensing Letters*, vol. 9, no. 6, pp. 569–577, 2018.

[13] C. Arango and C.A. Morales, "Comparison between multicopter UAV and total station for estimating stockpile volumes," *ISPRS*, vol. 40-1, 2015.

[14] J. Grodecki and G. Dial, "Block adjustment of high-resolution satellite images described by rational polynomials," *Photogrammetric Engineering & Remote Sensing*, vol. 69-1, 2003.

[15] B. Triggs, P.F. McLauchlan, R.I. Hartley, and A.W. Fitzgibbon, "Bundle adjustment — A modern synthesis," in *International Workshop on Vision Algorithms*, 1999.

[16] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," *IJCV*, vol. 60, no. 2, pp. 91–110, 2004.

[17] P. Moulon and P. Monasse, "Unordered feature tracking made fast and easy," in *CVMP*, 2012.

[18] R.A. Beyer, O. Alexandrov, and S. McMichael, "The Ames Stereo Pipeline: NASA's open source software for deriving and processing terrain data," *Earth and Space Science*, vol. 5, no. 9, pp. 537–548, 2018.

[19] G. Facciolo, C. de Franchis, and E. Meinhardt-Llopis, "Automatic 3D reconstruction from multi-date satellite images," in *CVPR Workshops*, 2017.

[20] W. Zhang, J. Qi, P. Wan, H. Wang, D. Xie, X. Wang, and G. Yan, "An easy-to-use airborne lidar data filtering method based on cloth simulation," *Remote Sensing*, vol. 8-6, 2016.