More surface detail with One-Two-Pixel Matching

Ewelina Rupnik and Marc Pierrot Deseilligny,

***Abstract*—Photogrammetrically derived Digital Surface Models have been widely adopted in geoscientific applications such us mapping and change detection across volcanic surfaces, glaciers, areas of seismic activity, forests, river landforms etc. Resolution of the reconstructed surface is crucial as more accurate information allows for more profound understanding of the phenomena. With this objective in mind, the research presented here proposes a new matching cost function that produces surfaces of enhanced resolution with respect to the golden standard, window-based semi-global matching technique. We evaluate the algorithm on different image datasets spanning various acquisition geometries, radiometric qualities and GSD sizes. In particular, results on Earth satellites (SPOT-7, Pléiades), extraterrestrial (Chang'E3 moon landing), aerial and UAV acquisitions are shown. The implementation of the method is available in the free open-source software for photogrammetry – MicMac.**

***Index Terms*—dense image matching; multi-view; forestry, bare soil, urban area;**

I. Introduction

The dense image matching for digital surface model (*DSM* or *photogrammetric DSM*) generation has been an accepted method across the geoscience communities. Next to other competitive techniques such as *LiDAR* or radar, image-based reconstruction produces denser 3D information, is richer as it includes the photometric observations allowing for e.g. 3D change detection or classification, and is cost effective. Photogrammetric DSMs in geoscience applications can be generated from terrestrial images, UAV acquisitions or high-resolution optical satellite imaging.

a) Terrestrial and UAV applications: Modelling of surface roughness parameters [1]; mapping volcanic surfaces [2]; measuring glaciar's microrelief progression [3] are among examples of terrestrial applications carried out with consumer grade camera and little expert knowledge. UAV-based surveys are increasingly present as an alternative to terrestrial surveys due to their larger reach, the ease of employment and little operational cost. Resolution-wise, it is a compromise between high-resolution close-range and moderate-resolution satellite imaging. The success of the UAV technology is reflected in numerous publications where, e.g., UAV-collected imagery enables modelling of the forest canopy height [4]; determines the rate and extent of landslide movements [5]; in aeolian research quantifies coastal erosion [6], [7] and deposition processes [8]; in tectonic research maps ultrafine (i.e. centimetric) tectonic faults [9]; or in glaciology is employed in repeated surveys of the ice-sheet masses [10].

E. Rupnik and M. Pierrot Deseilligny are with LASTIG IGN, ENSG, Universite Paris-Est, 94160 Saint-Mand, France

Manuscript received April 19, 2005; revised August 26, 2015.

b) Earth satellite and extraterrestrial applications: With the available optical satellite data provided by modern (e.g. Pléiades 1A/B, SPOT-satellites, QuickBird, WorldView 2/3/4, CubeSat) or the older generation satellites (e.g. CartoSat, ASTER) we can capture the Earth surface at Ground Sampling Distances (GSD) in the range of a few m – 0.3m, yielding DSMs of similar resolution [11]. Coupled with the daily revisit times, the satellite images are a suitable tool for mapping the surface evolution. To take a few *state-of-the-art* examples, satellite images are used to: identify patterns and rates of sand dunes migrations [12]; quantify post-seismic surface displacements [13], [14]; estimate timber volume in a densely forested area [15]; study the time-varying coastal shelf [16]; employ multi-temporal DSMs for elevation change studies across glaciers, volcanoes and river landforms [17]; or map other planet's topographies [18].

In all of the mentioned applications resolution of the surface plays a significant role because it improves the observability of the phenomena. Following the principle that if one can see more detail, to the same degree one can more easily spot a small detail change, an image matching algorithm that furnishes better resolution surfaces renders possible change detection at finer-level detail. Put differently, it allows for earlier change detection. The research presented in this publication addresses fine resolution surface reconstruction by proposing a novel image matching technique.

A. Related work

a) Generalities: Photogrammtrically derived DSMs are typically found by solving an energy functional with global, semi-global (SGM) or local methods.

Classically, local methods calculate the best correspondences in three steps: (i) matching cost calculation, (ii) support domain cost aggregation, and (iii) the depth selection by the *Winner-Takes-All* [19], or the belief propagation schemes [20], [21]. The local character of the methods makes them computationally efficient, often at the expense of lower 3D precision and completeness.

Whether global or SGM approach, the best correspondences are found by minimizing a cost function that is a sum of the data and the regularity term. Global methods minimize the energy defined over all pixels in a MRF graph with techniques such as α-expansion moves [22], [23] or message passing [24]. Alternatively, definition over a continuous space and sub-pixelar reconstruction precision is is possible with variational methods [25]. Such minimization is NP-hard, and does not allow for concave regularity terms. The first poses memory

issues for large scale reconstructions, while the latter inhibits faithful reconstruction at discontinuities (e.g. buildings).

The SGM methods are solved with multi-directional dynamic programming techniques, impose no constraint on the regularity term and the optimization is resolved along independent lines of pixels [26], [27]. Therefore, the computational times remain reasonable as the image size grows. Since cost function is cumulative along a number of directions, the found optimal solution is close to the global solution.

b) Matching cost calculation: Independently of the algorithm, there exist many different measures to describe the similarity (i.e. *photo-consistency*) between pixel correspondences. The very first and most straightforward measure compares the pixel intensity values [28], [29]. In real-world conditions (i.e. non-Lambertian surfaces, radiometric differences due to illumination change), in the presence of non-textures zones or repetitive patterns, however, pixel intensity differences turn to be unreliable for matching. In practice, photo-consistency measures defined over a support domain (square windows ordinarily of $3x3$, $5x5$, etc. size) are preferred [30], e.g. zero-mean normalized cross-correlation (NCC), Census, Rank [31], Mutual Information [32]. Including more context within the per-pixel matching is advantageous for it increases the uniqueness and the robustness of the matched points, as well as attenuates the noise (caused by the low signal-to-noise ratio of the sensor). Note, however, that (i) it interpolates the image signal within the support and possibly misses the surface fine structures, as well as (ii) a pixel-wise cost aggregated over a support window assumes constant disparity within the aggregated neighbourhood. The latter results in smoothness artefacts or mismatches if it happens to overlap a discontinuity, e.g. a building's edge. Furthermore, if the support window is chosen to be a square region, it takes for granted that the square shape in one view will project to an exact square in another view. Therefore, the perspective transformation the shape undergoes once projected to 3D and subsequently back-projected to another view, are neglected (the so-called *window foreshortening* problem).

Numerous approaches have been proposed to overcome the above shortcomings. To mitigate the mismatches at discontinuities adaptive support windows avoiding discontinuities [33], adaptive windows minimising a cost, or intensity gradient cost aggregation [27] were proposed. Sweeping-planes [34], 3D slanted windows [20], or a fusion of a slanted and adaptive windows (i.e. elongated windows with different orientations) [35] were adopted to account for different surface slopes. Initialisation of the surface planes is carried out by analysing the sparse reconstruction [34], dense reconstruction computed with square windows [33] or by random assignment [20]. Each choice comes with a certain risk of missing the correct surface slope or providing incomplete initialisation. For instance, sparse reconstruction can initialize only in places where 3D points exist hence omits poorly textured areas, classical dense reconstruction may fail at providing good initial surfaces at discontinuities and slopes, whereas random selection gives no guarantee of assigning the suitable plane slope to either of the pixels belonging to a region underlying that plane.

B. Contributions

In this work we aim at enhanced precision depth reconstruction free from fronto-parallel artefacts and with good performance at surface discontinuities. To realise it we:

- formulate a new matching cost defined over a single pixel,
- adopt a second pixel to allow for slope testing with minimum information (a slope is defined by two points in 3D space and it is tested with exactly two pixels),
- embed the image dense matching in a semi-global matching framework which favours slope coherence in the neighbourhood of a point and ensures surface continuity in poorly textured areas.

To increase the robustness of the reconstruction the matching cost is defined in a multi-view constraint.

II. Methods

A. One-Two-Pixel cost function

With the objective of excelling the spatial resolution of reconstructed depths, we propose a reformulated energy functional (1):

$$\mathcal{C}(D) = \sum_{p=1}^{N} C^I(p, d_p) + \sum_{q=Np} C^C(\mathbf{d_p}, \mathbf{d_q}) + \sum_{q=Np} C^T(\mathbf{d_p}, \mathbf{d_q}) \quad (1)$$

where C^I is the single-pixel term, C^C is the two-pixel term, C^T is the smoothing term, N is the number of all pixels, d_p is a disparity at position p and d_q is the disparity at position q belonging to N_p neighbourhood of p. $\mathcal{C}(D)$ is the minimum cost of assigning the optimal depths $\mathcal{D}$ across all pixels. The minimum solution is found in the following three steps: cost assignment (see Alg. 1), cost aggregation (see Alg. 2) and the optimal depth selection. The described methods are implemented in MicMac – the free open-source software for photogrammetry [36].

a) Two-pixel data term: C^C encodes the dissimilarity of a two-pixel window across multiple views (NbV), see Eq. (2). Ideally, the advantages of the minimum window are twofold: (i) it is more apt to reconstruct fine 3D scene structure than the usual square correlation window, and (ii) is less susceptible to noisy data than as if only the single pixel term was used.

Employing the C^C is equivalent of iterating over $\mathbf{d_q}$ and verifying the coherency of a given slope with the image data, the final cost of a slope assignment being an accumulation of the dissimilarities calculated independently for each stereo pair. Speaking in terms of the cost structure, the C^C costs associate with the arcs (in analogy to C^T), rather than with the nodes as is the case of C^I (see Alg. 2 and Fig. 1). Note that the two-pixel term should not be confused with the smoothness term. The latter applies "blind" penalization, without considering the pixel intensity values.

For a depth assignment d_p, the C^C cost will depend on the differences of normalised intensity ratios at position p and q. The choice of ratios was made to guarantee a robust comparison in the presence of additive and multiplicative illumination changes across the views. The role of Pds_C is to control the relative influence of the C^C term in the cost function. In all our tests it was also set to 1.0.

$$C^C(d_p, d_q) = Pds_C \cdot \frac{1}{NbV-1} \cdot \sum_{v=1}^{NbV-1} \left[\frac{I^v_{d_p}}{I^0_{d_p}} - \frac{I^v_{d_q}}{I^0_{d_q}} \right] \quad (2)$$

b) Single-pixel data term: C^I is the cost of assigning a particular depth to the position p. It is implemented as a truly one-pixel measure and conveys the dissimilarity of two pixels in multiple views (NbV), see Eq. (3). To match the dynamic range of the *two-pixel* term, it is expressed as a mean of all intensity ratios between the master I_0 and the slaves I_v. The δ_v parameter is added to model the view-dependent radiometric differences that may disturb the cumulative cost (this is unlike in the C^C expression where the effect of the radiometry change is fully modelled by the difference of ratios). The value of the parameter is a function of intensity ratios between the master image and a slave image. It can be pixel-dependent – if non-constant radiometric corrections across the images are necessary (e.g. specular reflections) – or take a global value per an image (see Alg. 1 where δ_v is computed on the pixel basis). When a pixel-dependent case is chosen, the ratios shall be first smoothed by simple averaging or by 2D polynomial function fit.

Pds_I is a weighting factor and it plays the same role as Pds_C. In all our tests we set the value to 1.0.

c) Regularizing term: ? do we want to talk about it?

$$C^I(p, d_p) = Pds_I \cdot \frac{1}{NbV-1} \cdot \sum_{v=1}^{NbV-1} \cdot \delta_v \cdot \left[1 - \frac{I^v_{d_p}}{I^0_{d_p}} \right] \quad (3)$$

B.

III. Results

+ MUST - show that larger baseline works better and if yes then add to Contributions !!! +A MUST : SHOW PROFILES TO BACKUP THE FRONTO-PARALLEL DISCOURSE +cmpr with s2p + stereo vs triplet

Evaluation :

- qualitative
  - gray shade
  - profiles across trees, buildings, bare soil
  - differences on discontinuities?
- quantitative
  - 1-/3-pix disparity diff between GT

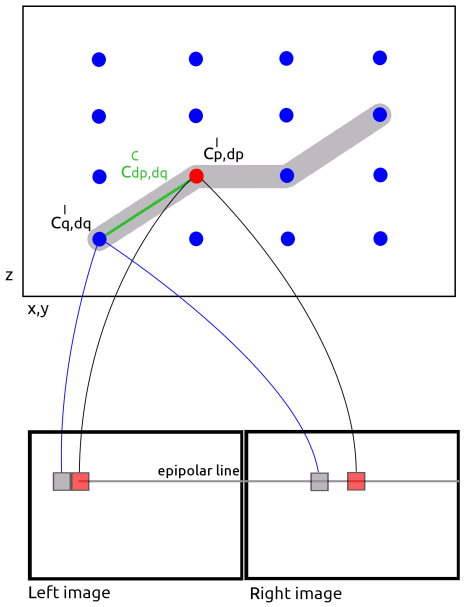

Fig. 1. *Top: One-two-Pixel cost structure corresponding to a stereo pair. In grey is the surface to be inferred from the matching.* $C^I_{p,dp}$ *and* $C^I_{q,dq}$ *are the costs of taking disparity dp at p and dq at q , respectively.* $C^C_{dp,dp}$ *is the cost of assigning a given slope. Bottom: A stereo pair that feeds the cost structure. The cost structure is illustrated on a stereo pair in epipolar geometry, however, the implemented algorithm is not constrained by the number of views and, by default, operates in the original image geometry.*

A. Extra-terrestrial

B. Bare soil - greece Pleiades

C. Forest - Spot7

D. Urban zoned - aerial and/or sattelite?

Satellite/aerial acquisitions
Middlebury benchmark

IV. Conclusion

The conclusion goes here.

Appendix A
Proof of the First Zonklar Equation

Appendix one text goes here.

Appendix B

Appendix two text goes here.

Acknowledgment

The authors would like to thank...

Algorithm 1 Cost assignment

$VIm[v].ImOrth(X,Y)$ - intensity in image v at position (X,Y); all images are orthorectified to the geometry of the master

$ratio(int,int)$ - calculate ratios
$norm(int)$ - normalise to [0-255]
δ - radiometric calibration
$SurfOpt$ - optimiser

▷ Assignment

for $Z = Z_{min}; Z < Z_{max}; Z++$ **do**
  for $X = 0; X < Sz.x; X++$ **do**
    for $Y = 0; Y < Sz.y; Y++$ **do**

$V_0 = VIm[0].ImOrth(X,Y)$
$C^I = 0$
for $v = 0; v < NbV; v++$ **do**

$V_k = VIm[v].ImOrth(X,Y)$
$Val = ratio(V_0, V_k)$
$Val = norm(Val)$
$r.push\_back(Val)$

$V_k^{cor} = \delta_v V_k$
$Val^{cor} = ratio(V_0, V_k^{cor})$
$C^I += Abs(Val^{cor})$
end for

▷ Communicate the vector of ratios and the data term to the optimiser

$SurfOpt(X,Y,Z) = r$
$SurfOpt(X,Y,Z) = C^I$

    end for
  end for
end for

REFERENCES

[1] F. Bretar, M. Arab-Sedze, J. Champion, M. Pierrot-Deseilligny, E. Heggy, and S. Jacquemoud, "An advanced photogrammetric method to measure surface roughness: Application to volcanic terrains in the piton de la fournaise, reunion island," *Remote Sensing of Environment*, vol. 135, pp. 1–11, 2013.

[2] S. Kolzenburg, M. Favalli, A. Fornaciai, I. Isola, A. Harris, L. Nannipieri, and D. Giordano, "Rapid updating and improvement of airborne lidar dems through ground-based sfm 3-d modeling of volcanic features," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 11, pp. 6687–6699, 2016.

[3] A. Kääb, L. M. R. Girod, and I. T. Berthling, "Surface kinematics of periglacial sorted circles using structure-from-motion technology," 2014.

[4] J. Lisein, M. Pierrot-Deseilligny, S. Bonnet, and P. Lejeune, "A photogrammetric workflow for the creation of a forest canopy height model from small unmanned aerial system imagery," *Forests*, vol. 4, no. 4, pp. 922–944, 2013.

[5] U. Niethammer, S. Rothmund, M. James, J. Travelletti, and M. Joswig, "Uav-based remote sensing of landslides," *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 38, no. Part 5, pp. 496–501, 2010.

[6] S. d'Oleire Oltmanns, I. Marzolff, K. D. Peter, and J. B. Ries, "Unmanned aerial vehicle (uav) for monitoring soil erosion in morocco," *Remote Sensing*, vol. 4, no. 11, pp. 3390–3416, 2012.

[7] M. James and S. Robson, "Straightforward reconstruction of 3d surfaces and topography with a camera: Accuracy and geoscience application," *Journal of Geophysical Research: Earth Surface*, vol. 117, no. F3, 2012.

[8] C. H. Hugenholtz, K. Whitehead, O. W. Brown, T. E. Barchyn, B. J. Moorman, A. LeClair, K. Riddell, and T. Hamilton, "Geomorphological mapping with a small unmanned aircraft system (suas): Feature detection and accuracy assessment of a photogrammetrically-derived digital terrain model," *Geomorphology*, vol. 194, pp. 16–24, 2013.

[9] K. Johnson, E. Nissen, S. Saripalli, J. R. Arrowsmith, P. McGarey, K. Scharer, P. Williams, and K. Blisniuk, "Rapid mapping of ultrafine fault zone topography with structure from motion," *Geosphere*, vol. 10, no. 5, pp. 969–986, 2014.

[10] J. C. Ryan, A. L. Hubbard, J. E. Box, J. Todd, P. Christoffersen, J. R. Carr, T. O. Holt, and N. A. Snooke, "Uav photogrammetry and structure from motion to assess calving dynamics at store glacier, a large outlet draining the greenland ice sheet," 2015.

[11] E. Rupnik, M. Pierrot Deseilligny, and A. Delorme, "Multi-view DSM fusion. Application to satellite images." *Journal of Photogramemtry and Remote Sensing*, 2017, in preparation.

[12] E. Hermas, S. Leprince, and I. A. El-Magd, "Retrieving sand dune movements using sub-pixel correlation of multi-temporal optical remote sensing imagery, northwest sinai peninsula, egypt," *Remote sensing of environment*, vol. 121, pp. 51–60, 2012.

[13] A. Rosu, M. Pierrot-Deseilligny, A. Delorme, R. Binet, and Y. Klinger, "Measurement of ground displacement from optical satellite image correlation using the free open-source software micmac," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 100, pp. 48–59, 2015.

[14] A. Vallage, Y. Klinger, R. Grandin, H. Bhat, and M. Pierrot-Deseilligny, "Inelastic surface deformation during the 2013 mw 7.7 balochistan, pakistan, earthquake," *Geology*, vol. 43, no. 12, pp. 1079–1082, 2015.

[15] C. Straub, J. Tian, R. Seitz, and P. Reinartz, "Assessment of cartosat-1 and worldview-2 stereo imagery in combination with a lidar-dtm for timber volume estimation in a highly structured forest in germany," *Forestry*, vol. 86, no. 4, pp. 463–473, 2013.

[16] K.-H. Tseng, C.-Y. Kuo, T.-H. Lin, Z.-C. Huang, Y.-C. Lin, W.-H. Liao, and C.-F. Chen, "Reconstruction of time-varying tidal flat topography using optical remote sensing imageries," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 131, pp. 92–103, 2017.

[17] L. Girod, C. Nuth, A. Kääb, R. McNabb, and O. Galland, "Mmaster: Improved aster dems for elevation change monitoring," *Remote Sensing*, vol. 9, no. 7, p. 704, 2017.

[18] K. Gwinner, F. Scholten, F. Preusker, S. Elgner, T. Roatsch, M. Spiegel, R. Schmidt, J. Oberst, R. Jaumann, and C. Heipke, "Topography of mars from global mapping by hrsc high-resolution digital terrain models and orthoimages: characteristics and performance," *Earth and Planetary Science Letters*, vol. 294, no. 3-4, pp. 506–519, 2010.

[19] K.-J. Yoon and I.-S. Kweon, "Locally adaptive support-weight approach for visual correspondence search," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2. IEEE, 2005, pp. 924–931.

[20] M. Bleyer, C. Rhemann, and C. Rother, "Patchmatch stereo-stereo matching with slanted support windows." in *Bmvc*, vol. 11, 2011, pp. 1–11.

[21] S. Galliani, K. Lasinger, and K. Schindler, "Massively parallel multiview stereopsis by surface normal diffusion," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 873–881.

[22] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 23, no. 11, pp. 1222–1239, 2001.

[23] T. Taniai, Y. Matsushita, Y. Sato, and T. Naemura, "Continuous 3d label stereo matching using local expansion moves," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.

[24] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient belief propagation for early vision," *International journal of computer vision*, vol. 70, no. 1, pp. 41–54, 2006.

[25] R. Ranftl, S. Gehrig, T. Pock, and H. Bischof, "Pushing the limits of stereo using variational stereo estimation," in *Intelligent Vehicles Symposium (IV), 2012 IEEE*. IEEE, 2012, pp. 401–407.

[26] M. Pierrot-Deseilligny and N. Paparoditis, "A multiresolution and optimization-based image matching approach: An application to surface reconstruction from spot5-hrs stereo imagery." in *ISPRS Workshop On Topographic Mapping From Space*, vol. 36(1), Ankara, Turkey, February 2006.

[27] H. Hirschmüller, "Stereo processing by semiglobal matching and mutual information," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 30, no. 2, pp. 328–341, 2008.

Algorithm 2 Cost aggregation

for each direction + **do** ▷ forward
  for each direction - **do** ▷ backward

    for $X = 0; X < Sz.x; X++$ **do**
      for $Y = 0; Y < Sz.y; Y++$ **do**
        for $DZ = Z_{min}; DZ < Z_{max}; DZ++$ **do**

          $Cost = CostOfDZ(DZ)$ ▷ get the value of concave cost function at DZ
          $\mathbf{r}_k = SurfOpt.rOfk(k)$ ▷ recover the vectors of ratios
          $\mathbf{r}_{k+1} = SurfOpt.rOfk(k+1)$
          ▷ calculate the two-pixel cost and add to cost
          $Cost+ = TwoPixCost(\mathbf{r}_k, \mathbf{r}_{k+1})$

          ▷ communicate the two-pixel cost to the optimiser
          $SurfOpt.UpdateCostOneArc(\mathbf{r}_{k+1}, Dir, Cost)$

        end for
      end for
    end for
  end for
end for

[28] T. Kanade, a. K. S. Kano, H., A. Yoshida, and K. Oda, "Development of a video-rate stereo machine," in *Intelligent Robots and Systems 95.'Human Robot Interaction and Cooperative Robots', Proceedings. 1995 IEEE/RSJ International Conference on*, vol. 3. IEEE, 1995, pp. 95–100.

[29] S. Birchfield and C. Tomasi, "A pixel dissimilarity measure that is insensitive to image sampling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 4, pp. 401–406, 1998.

[30] Y. Furukawa and C. Hernández, "Multi-view stereo: A tutorial," *Foundations and Trends® in Computer Graphics and Vision*, vol. 9, no. 1-2, pp. 1–148, 2015.

[31] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," in *European conference on computer vision*. Springer, 1994, pp. 151–158.

[32] P. Viola and W. I.W.M., "Alignment by maximization of mutual information," *International journal of computer vision*, vol. 24, no. 2, pp. 137–154, 1997.

[33] K.-J. Yoon and I. S. Kweon, "Adaptive support-weight approach for correspondence search," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 4, pp. 650–656, 2006.

[34] D. Gallup, J.-M. Frahm, P. Mordohai, Q. Yang, and M. Pollefeys, "Real-time plane-sweeping stereo with multiple sweeping directions," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. IEEE, 2007, pp. 1–8.

[35] A. Buades and G. Facciolo, "Reliable multiscale and multiwindow stereo matching," *SIAM Journal on Imaging Sciences*, vol. 8, no. 2, pp. 888–915, 2015.

[36] E. Rupnik, M. Daakir, and M. Pierrot Deseilligny, "Micmac - a free, open-source solution for photogrammetry," *Open Geospatial Data, Software and Standards*, vol. 2, no. 14, pp. 1–9, 2017.