



HAL
open science

More surface detail with One-Two-Pixel Matching

Ewelina Rupnik, Marc Pierrot Deseilligny

► **To cite this version:**

Ewelina Rupnik, Marc Pierrot Deseilligny. More surface detail with One-Two-Pixel Matching. [Research Report] IGN - Laboratoire MATIS. 2019. hal-02371337

HAL Id: hal-02371337

<https://hal.science/hal-02371337>

Submitted on 19 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

More surface detail with One-Two-Pixel Matching

Ewelina Rupnik, Marc Pierrot Deseilligny

Abstract—Photogrammetrically derived Digital Surface Models have been widely adopted in geoscientific applications such as 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 enables 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 gold standard: the window-based semi-global matching technique. We evaluate the algorithm on different image datasets spanning various acquisition geometries, radiometric qualities and ground sample distance sizes. In particular, results on Earth satellites (SPOT-7, Pléiades), extraterrestrial (Chang'E3 moon landing), aerial and terrestrial acquisitions are shown. The implementation of the method is available in MicMac – the free open-source software for photogrammetry.

I. INTRODUCTION

Digital surface model (*DSM* or *photogrammetric DSM*) generation using dense image matching is an accepted technique across the geoscience communities. Next to other competitive techniques such as *LiDAR* or radar, image-based reconstruction produces denser 3D information, it is cost-effective and richer as it includes photometric observations that allow, for example, 3D change detection or classification. Photogrammetric DSMs in geoscience applications can be generated from terrestrial images, *unmanned aerial vehicle* (UAV) acquisitions or high-resolution optical satellite imaging.

a) Terrestrial and UAV applications: Modelling of surface roughness parameters [1]; mapping volcanic surfaces [2]; and measuring glaciers' microrelief progression [3] are some of many examples of terrestrial applications carried out with consumer grade cameras and little expert knowledge. UAV-based surveys are increasingly presented as an alternative to terrestrial surveys due to their larger reach, their ease of deployment and reduced operational cost. With respect to resolution, UAV surveys are a compromise between high-resolution close-range and moderate-resolution satellite imaging. The success of the UAV technology is reflected in numerous publications which show that UAV-collected imagery can: enable modelling of forest canopy height [4]; determine the rate and extent of landslide movements [5]; quantify coastal erosion [6], [7] and deposition processes [8] in aeolian research; map ultrafine (i.e. centimetric) tectonic faults in tectonic research [9]; or be employed in repeated surveys of the ice-sheet masses in glaciology [10].

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 older satellites (e.g. CartoSat, ASTER), we can

capture the Earth's surface at ground sampling distances (GSD) in the range of 0.3 metres to some metres, yielding DSMs of a similar resolution [11]. Coupled with daily revisit times, 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 dune migrations [12]; quantify seismic surface displacement [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 planets' topographies [18].

In all of the mentioned applications, the resolution of the surface plays a significant role because it improves the observability of the phenomena. Following the principle that one can more easily spot a small change if one can see more detail, an image matching algorithm that furnishes better resolution surfaces allows possible change detection at finer detail. Put simply, techniques that enable finer resolution could allow 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: Photogrammetrically derived DSMs are typically found by solving an energy function with global, semi-global matching (SGM) or local methods.

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

Whether through a global or SGM approach, the best correspondences are found by minimizing a cost function that is the 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 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: Independent of the algorithm, many different measures exist that describe the similarity (i.e. *photo-consistency*) between pixel correspondences. The very first and most straightforward measure compares the pixel intensity values [28], [29]. However, pixel intensity difference are unreliable for matching in real-world conditions (i.e. non-Lambertian surfaces or, radiometric differences due to illumination change) and in the presence of non-textured zones or repetitive patterns. In practice, photo-consistency measures, e.g. zero-mean normalized cross-correlation (NCC), Census, Rank [30], Mutual Information [31], are defined over a support domain (square windows ordinarily of 3×3 , 5×5 , etc. size) [32]. Including more context within the per-pixel matching is advantageous because 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, this technique neglects the perspective transformations the shape undergoes once projected to 3D and subsequently back-projected to another view (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 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 option carries some 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 to provide good initial surfaces at discontinuities and slopes; and 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 to develop enhanced precision depth reconstruction free from fronto-parallel artefacts and with good performance at surface discontinuities. To realise this goal we:

- formulate a new matching cost defined over a single pixel,
- adopt a second pixel to add context and 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 function (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 the 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) it is less susceptible to noisy data than as if only the single pixel term was used.

Employing the C^C is equivalent to 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 can be associated with arcs (in analogy to C^T), rather than with 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 of our tests, it was also set to 1.0.

$$C^C(d_p, d_q) = Pds_C \cdot \frac{1}{NbV - 1} \cdot \sum_{v=1}^{NbV} \left| \frac{I_{d_p}^v}{I_{d_p}^0} - \frac{I_{d_q}^v}{I_{d_q}^0} \right| \quad (2)$$

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 (1 – intensity ratios) between the master I_0 and the secondary 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 secondary 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 image pair (i.e. master and secondary image couple; see Alg. 1 where δ_v is computed on a 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 of our tests, we set the value to 1.0.

c) *Regularizing term:* The standard regularization function present in *MicMac* library has been adopted and it is not discussed hereafter (see [26] and [37] for more details on the subject).

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

III. RESULTS

A. Datasets

a) *Aerial acquisition – Vaihingen dataset:* A set of four aerial images acquired with the UltraCam-X camera at 0.2m GSD is used in this experiment. The images are part of the EuroSDR Matching benchmark [38]. The camera calibration and orientation parameters were also furnished with the benchmark. The available ground truth (GT) data is a surface that is a median of several DSMs furnished by all of the participants of the benchmark. As a result, many surface details are discarded and the GT is considered unsuitable to evaluate the *One-Two-Pixel* matching algorithm. The evaluation was carried out on an alternative GT (cf. Section III-B) over an area of $\approx 5\text{km} \times 3\text{km}$ (cf. Fig. 7a and Fig. 8a).

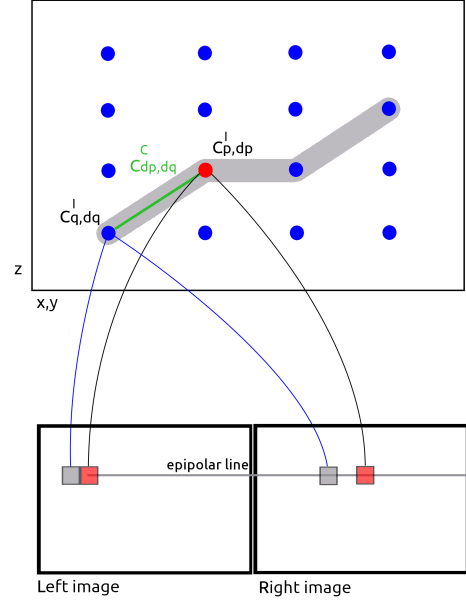


Figure 1: *Top: One-Two-Pixel cost structure corresponding to a stereo pair. In grey is the surface to be inferred from the matching. $C_{p,dp}^I$ and $C_{q,dq}^I$ are the costs of taking disparity d_p at p and d_q at q , respectively. $C_{dp,dp}^C$ is the cost of assigning a given slope. Bottom: A stereo 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.*

b) *Satellite acquisition I – Pléiades dataset:* This scene was acquired by a tri-stereo pair from the Pléiades 1B satellite. It is entirely covered by sand dunes of varying patterns, situated in northern Chile. As a preliminary processing step, the satellite orientation parameters were refined in a RPC-bundle adjustment using methods described in [39]. The evaluation was carried out on an alternative GT (cf. Section III-B) over an area of $\approx 7.5\text{km} \times 5.5\text{km}$ (cf. Fig. 7b and Fig. 8b).

c) *Satellite acquisition II – Spot-7 dataset:* A Spot-7 tri-stereo pair is used in this experiment. The scene represents a rural zone with rich vegetation and varying altitudes. Refined satellite orientation parameters in the form of a replacement grid were furnished with the dataset. Evaluation is carried out within a $\approx 2\text{km} \times 2\text{km}$ region (cf. Fig. 7c and Fig. 9) on a GT described in Section III-B.

d) *Extra-terrestrial acquisition – Chang'E 3 dataset:* In late 2013, the People's Republic of China launched a lunar surface exploration mission known as *Chang'E 3*. To calculate the explorer's trajectory (cf. Fig. 2), over 150 conic images acquired at landing were downloaded¹ and post-processed using *Structure-from-Motion* methods in *MicMac*. The images were available in a heavily compressed *jpeg* format.

Reference data in the form of a 1.5m orthophoto and a 30m resolution DSM produced from the *Lunar Reconnaissance*

¹<http://moon.bao.ac.cn/>

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 $[-127, 127]$

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

*Orbiter Camera*² (LROC) images served to geo-reference the images. The provided reference data is of much lower resolution than the produced DSMs (i.e. 0.25m). Because of this, an alternative GT was created and is detailed in the following Section III-B.

e) Terrestrial acquisition: The last dataset is a sequence of 5 images taken with a long focal length *DSLR* camera (i.e., $f = 105mm$), see Fig. IV. It does not undergo quantitative evaluation but is shown to demonstrate the performance of the *One-Two-Pixel* matching on large discontinuities (cf. Figs. 12 and 13).

B. Ground truth creation

Due to the lack of GT availability, we adopted a strategy where full resolution images serve to establish the GT surfaces, and are compared against surfaces calculated using lower resolution images. Following this concept, the GTs of the *Vaihingen* / *Pléiades* / *Spot-7* datasets are denoted with the $Z1$

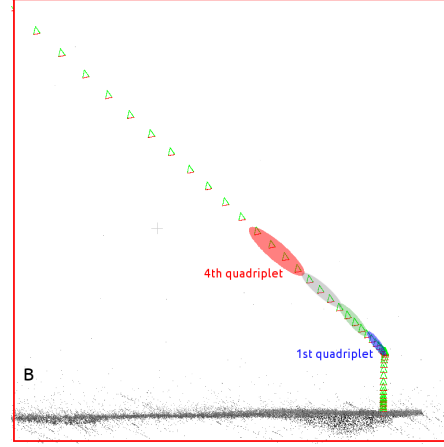
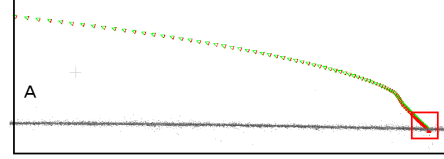


Figure 2: *Chang'E 3* explorer moon landing trajectory. (A) ≈ 150 images acquired before the landing. (B) Close-up of the end trajectory, the highlighted sets of images correspond to images used in the processing. The surface calculated with the 4th quadruplet is compared against a surface produced by a fusion of all the quadruplets.

postfix corresponding to *zoom 1*, or the full resolution images. Analogously, the $Z2$ postfix signifies *zoom 2*, or half the full resolution images. In other words, the matching GSDs are as follows: $GSD_{Z1}^{Vaihingen} = 0.2m$, $GSD_{Z2}^{Vaihingen} = 0.4m$, $GSD_{Z1}^{Pléiades} = 0.5m$, $GSD_{Z2}^{Pléiades} = 1.0m$, $GSD_{Z1}^{Spot-7} = 1.5m$, and $GSD_{Z2}^{Spot-7} = 3.0m$. As there is no obvious choice as to which reconstruction method to employ to create the GT, two sets of GTs are adopted: a GT calculated with cross-correlation similarity measure ($CORR_{Z1}$) and a GT obtained with the *One-Two-Pixel* measure ($12Pix_{Z1}$).

Because the *Chang'E 3* dataset features a large number of images of varying scales, we have chosen to calculate the GT based on the images obtained with larger scales (i.e. those taken closer to the planet surface). Using this GT, we then evaluate surfaces obtained from images taken at higher altitudes. Nonetheless, the footprints of the images closer to the surface are smaller, and consequently, cover a smaller area than the surface under evaluation. To match the surface extent and further enhance the quality of the GT, we perform surface reconstruction from multiple sets of quadruplets, and subsequently fuse them in 3D as presented in [11]. In Fig. 2, the evaluated image set (i.e., 4th quadruplet) is highlighted in red, while the gray-green-blue sets (i.e., 1st-3rd q.) are used to produce the GT. All surfaces were re-sampled to a 0.25m grid which corresponded to the mean GSD of 4th quadruplet. See also Fig. 10 (G)-(H) for the output GT surface and the LROC orthophoto, respectively.

²<http://lroc.sese.asu.edu>

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

```

C. Evaluation strategy

The results are interpreted in terms of accuracy measures suited to (i) normal distribution observations (i.e., *mean*, *median*, *standard deviation*) calculated on DSMs' signed differences Δh , as well as (ii) non-normal errors, insensitive to outliers and calculated on absolute differences $|\Delta h|$ (i.e. *Normalised Mean Absolute Deviation (NMAD)* and sample quantiles) [40]. Sample quantiles enable us to infer the percentage of errors that fall within a specified interval. The statistics are presented in Tab. I and Fig. 4.

Additionally, theoretical histograms and histograms of Δh are plotted in Fig. 5. In Fig. 6 height profiles are drawn, and in Figs. 7, 8, 9, 10, qualitative results in the form of gray-shaded DSMs are given.

D. Discussion

a) Aerial acquisition – Vaihingen dataset: Where the $CORR_{Z1}$ was considered the GT, all Δh manifest a systematic shift of the DSM and a strongly non-normal distribution (compare the *mean* with the *median* and the σ and *NMAD* in Tab. Ia). Further evidence of the non-normality of the distribution is the mismatch between the theoretical histogram and the histogram calculated on the data, as is shown in Fig. 5a. The $\approx 2m$ standard deviations are probably due to the fact that the $CORR_{Z2}$ DSM is incapable of discerning the small vegetation rows which make up a lot of the area of interest. Furthermore, note the fidelity with which the $12Pix_{Z2}$ and $1Pix_{Z2}$ reconstruct at discontinuities, and in flat, non-textured zones (cf. Figs. 6a, 7a and 8a).

According to the sample quantiles, there is a tendency for DSMs compared to $12Pix_{Z1}$ (cf. dashed lines in Fig. 4) to perform better than those compared to $CORR_{Z1}$. The three best sample quantiles corresponding to the size of the

GSD are 76%, 71%, and 68% for $12Pix_{Z1} - 12Pix_{Z2}$, $12Pix_{Z1} - 1Pix_{Z2}$ and $CORR_{Z1} - CORR_{Z2}$, respectively. In other words, 76%, 71%, and 68% of errors will be found within the interval of $\langle 0, GSD \rangle$.

b) Satellite acquisition I – Pléiades dataset: The dataset's quantitative error metrics and the qualitative interpretation of the results are incoherent. The lowest *mean* and *median* scores are observed for $12Pix_{Z1} - 1Pix_{Z2}$, followed by $CORR_{Z1} - CORR_{Z2}$ and $CORR_{Z1} - 12Pix_{Z2}$. However, the standard deviations and *NMAD* follow the inverse order: the performance of $12Pix_{Z1} - 12Pix_{Z2}$ precedes that of $CORR_{Z1} - CORR_{Z2}$ (cf. Fig. 1b). This could be partially explained by the increased level of noise observed in the images. While $12Pix$ and *CORR* can attenuate the noise due to their support regions and the cost function smoothing term, the $1Pix$ has only the latter. However, this does not necessarily mean that more detailed surfaces are reconstructed with either – observe the surface details reconstructed with $12Pix$ and $1Pix$ yet invisible to *CORR* in Figs. 6b, 7b and 8b. This puts in question the chosen GT and the error metric used.

The spread between the *NMADs* and the standard deviations signals, as previously, a non-normal error distribution.

c) Satellite acquisition II – Spot-7 dataset: The specificity of the images is the changing illumination across the triplet pair (cf. Fig. 3 for image intensity ratios). To demonstrate the efficacy of the radiometric calibration proposed in the algorithm, the $12Pix_{Z2}$ DSM is calculated twice: with and without the calibration. Calibration using a global value per image pair turned sufficient to model the illumination effect. Note in Tab. Ic the significant improvement in accuracy with the correction applied.

The Δh in all cases match the normal distribution errors relatively well (cf. Fig. 5c; except the non-calibrated result). The $12Pix_{Z2}$ performs best of all (cf. Tab. Ic) and obtains the

lowest $NMAD$ and σ measures, while its *mean* and *medium* scores are placed *ex aequo* with $CORR_{Z2}$. All surfaces but the $CORR_{Z2}$ exhibit a systematic shift when compared to the $CORR_{Z1}$ GT.

According to the sample quantile curves, when the $CORR_{Z1}$ is considered GT, 85% of $CORR_{Z2}$, 76% of $1Pix_{Z2}$ and 78% of $12Pix_{Z2}$ DSMs' errors lay within the $\langle 0, GSD \rangle$ interval. When the $12Pix_{Z1}$ was used, the proportions of DSMs' errors are: 84% for $CORR_{Z2}$, 87% for $1Pix_{Z2}$, and 89% for $12Pix_{Z2}$ DSMs.

d) *Extra-terrestrial acquisition – Chang'E 3 dataset:*

$1Pix$ DSM comprises a systematic error revealed by the mean and median measures, while in contrast $CORR$ and $12Pix$ DSMs reveal correspondingly insignificant or no bias (cf. Tab. Id). This behaviour can be explained by the fact that $1Pix$ cost function, defined on a single pixel, becomes too ambiguous to find the correct surface. On the other hand, $CORR$ and $12Pix$ cost functions include more context therefore remain close to the true surface. Until now, this behaviour has not been observed, nonetheless, the quality of the images in other datasets was incomparably better and free from compression artefacts. Notice the different character of noise here and in *Satellite acquisition I*, having per pixel structure and arranged in blocks, respectively.

40% of the $CORR$ DSM, 56% of the $1Pix$ DSM and 73% of the $12Pix$ DSM errors lay within the interval of $\langle 0, GSD \rangle$ (cf. Fig. 4). The long tails of the theoretical histogram distribution in $12Pix$ and $1Pix$ DSMs suggest that some outliers are present in the data (cf. Fig. 5d). Irrespective of that, the $12Pix$ DSM is the unquestionable winner among other results for this dataset. The height profiles drawn over craters and on flat ground in Fig. 6d and the DSMs shown in Fig. 10 confirm the high fidelity of $12Pix$ DSM.

The evaluation is carried out within a region of interest marked in blue in Fig. 10.

e) *Terrestrial acquisition:*

Finally, in Fig. 13 several areas of interest of the reconstructed statue are shown. The evaluation metrics are (i) performance on slopes and (ii) the level of surface detail. The results are: the $CORR$ surfaces are smoother than the $1Pix$ and $12Pix$; the $1Pix$ produces artefacts at large slopes; the $12Pix$ performs best in terms of the level of detail and at discontinuities. In particular, note the following points of interest:

- in Fig. 13a, the band over the forehead and the cheek close to the ear;
- in Fig. 13b, the repetitive pyramid pattern and upper left image corner for directional matching artefacts;
- in Fig. 13c, the cords of the guitar as well as the level of noise on the flat area; and
- in Fig. 13d, the separation of fingers, as well as the parts of the guitar.

IV. CONCLUSION

We have presented a new matching cost function that is embedded in a semi-global matching framework. The method was tested on 5 different datasets (1 aerial, 2 satellite, 1 extra-terrestrial and 1 terrestrial dataset), 4 of which were evaluated

qualitatively and quantitatively. The quantitative assessment is twofold: based on comparing DSMs' reconstructions calculated with different similarity measures (i.e. cross-correlation, one-pixel and one-two-pixel); and by comparing to reference surfaces of higher resolution. During evaluation, normal and non-normal error distribution metrics are taken into account, as well as histograms, surface profiles and gray shaded surfaces.

The normal surface metrics do not reveal a clear winner among all methods and suggest their inappropriateness for our experiments. The distribution-free error metrics (i.e., $NMAD$ and sample quantiles) unanimously point to our method as the better performer. In all experiments the $12Pix$ sample quantile curves report the best measurement accuracies.

Similarly, the visual assessment of the surface quality indicates that both the $12Pix$ and the $1Pix$ DSMs convey more detail than the correlation-based DSMs. The $12Pix$ and $1Pix$ exhibit similar performance except in two cases: when considerable flaws in image quality are present and for large slopes. In these two cases, the $12Pix$ similarity measure outperforms the $1Pix$.

ACKNOWLEDGMENT

We would like to thank Renaud Binet from CNES for providing us with the *Chang'E 3* dataset, Tristan Faure and Antoine Levenant from IGN for giving us access to *Spot7* images and Arthur Delorme from IPGP for the *Pléiades* images.

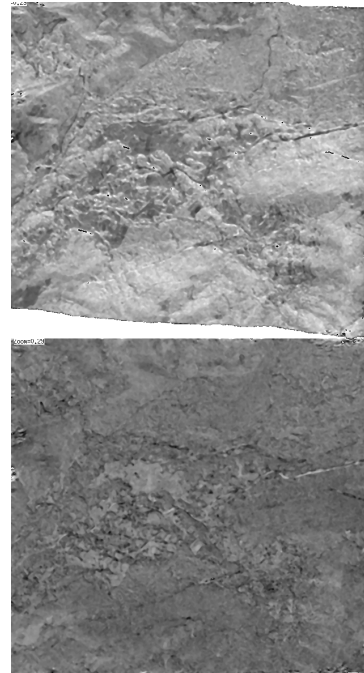


Figure 3: *SPOT-7* dataset. Ratios of intensities between the master image and two secondary images. The images are ortho-rectified to the geometry of the master image. If the ratios had similar values, the radiometric correction would be superfluous.

Table I: Mean, median and standard deviation scores calculated on DSMs signed differences, $|\Delta h|$. Three GT DSMs are present in the table: $CORR_{Z1}$ computed with cross-correlation, $12Pix_{Z1}$ with One-Two-Pixel measures and GT computed as a fusion of multiple quadriplets. The values in read are mean $NMAD$ scores computed on all available GTs. The values are expressed in meters.

(a) *Vahingen* dataset

| Method | mean | median | NMAD | σ |
|---------------------------|-------|--------|------------|----------|
| $CORR_{Z1} - CORR_{Z2}$ | -0.08 | -0.02 | 0.34 /0.38 | 2.32 |
| $12Pix_{Z1} - CORR_{Z2}$ | 0.08 | 0.0 | 0.42 /0.38 | 2.24 |
| $CORR_{Z1} - 1Pix_{Z2}$ | -0.15 | 0.0 | 0.54 /0.42 | 2.75 |
| $12Pix_{Z1} - 1Pix_{Z2}$ | 0.0 | 0.0 | 0.30 /0.42 | 0.68 |
| $CORR_{Z1} - 12Pix_{Z2}$ | -0.19 | -0.02 | 0.49 /0.38 | 2.71 |
| $12Pix_{Z1} - 12Pix_{Z2}$ | -0.03 | -0.02 | 0.27 /0.38 | 0.60 |

(b) *Pléiades* dataset

| Method | mean | median | NMAD | σ |
|---------------------------|-------|--------|------------|----------|
| $CORR_{Z1} - CORR_{Z2}$ | -0.14 | -0.06 | 1.06 /1.23 | 1.76 |
| $12Pix_{Z1} - CORR_{Z2}$ | 0.19 | 0.13 | 1.39 /1.23 | 1.96 |
| $CORR_{Z1} - 1Pix_{Z2}$ | -0.43 | -0.28 | 2.91 /2.30 | 2.91 |
| $12Pix_{Z1} - 1Pix_{Z2}$ | -0.10 | 0.0 | 1.68 /2.30 | 1.65 |
| $CORR_{Z1} - 12Pix_{Z2}$ | -0.52 | -0.36 | 1.66 /1.23 | 2.22 |
| $12Pix_{Z1} - 12Pix_{Z2}$ | -0.19 | -0.17 | 0.79 /1.23 | 1.00 |

(c) *SPOT-7* dataset

| Method | mean | median | NMAD | σ |
|-----------------------------------|-------|--------|------------|----------|
| $CORR_{Z1} - CORR_{Z2}$ | 0.06 | 0.0 | 1.44 /1.50 | 2.21 |
| $12Pix_{Z1} - CORR_{Z2}$ | -0.08 | -0.04 | 1.55 /1.50 | 2.16 |
| $CORR_{Z1} - 1Pix_{Z2}$ | 0.22 | 0.0 | 2.05 /1.77 | 2.84 |
| $12Pix_{Z1} - 1Pix_{Z2}$ | 0.08 | 0.0 | 1.48 /1.77 | 1.87 |
| $CORR_{Z1} - 12Pix_{Z2}^{nocal}$ | -0.15 | 0.0 | 3.91 /3.81 | 4.03 |
| $12Pix_{Z1} - 12Pix_{Z2}^{nocal}$ | -0.29 | 0.0 | 3.73 /3.81 | 3.75 |
| $CORR_{Z1} - 12Pix_{Z2}$ | 0.21 | 0.07 | 1.85 /1.57 | 2.65 |
| $12Pix_{Z1} - 12Pix_{Z2}$ | 0.07 | 0.0 | 1.30 /1.57 | 1.73 |

(d) *Chang'E 3* dataset

| Method | mean | median | NMAD | σ |
|------------------|------|--------|------|----------|
| $GT - CORR$ | 0.04 | 0.04 | 0.47 | 0.49 |
| $GT - OnePix$ | 0.19 | 0.15 | 0.29 | 0.34 |
| $GT - OneTwoPix$ | 0.02 | 0.00 | 0.12 | 0.24 |

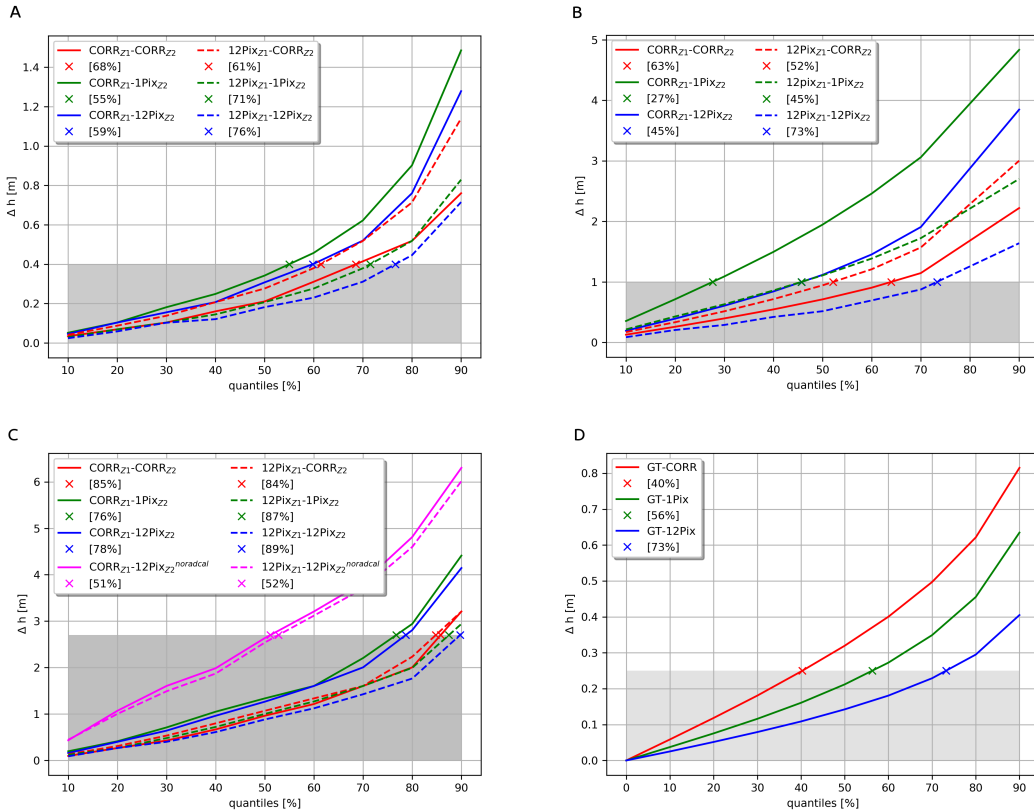


Figure 4: Sample quantiles calculated on: A. *Vahingen*, B. *Pleiades*, C. *Spot-7*, D. *Chang'E 3* datasets. The gray zone indicates $|\Delta h|$ values that are $\leq GSD$ (of the lower resolution images). The \times markers correspond to sample quantiles for respective GSD values, i.e., they signal the percent of all Δh for which $\langle 0 \geq |\Delta h| \leq GSD \rangle$.

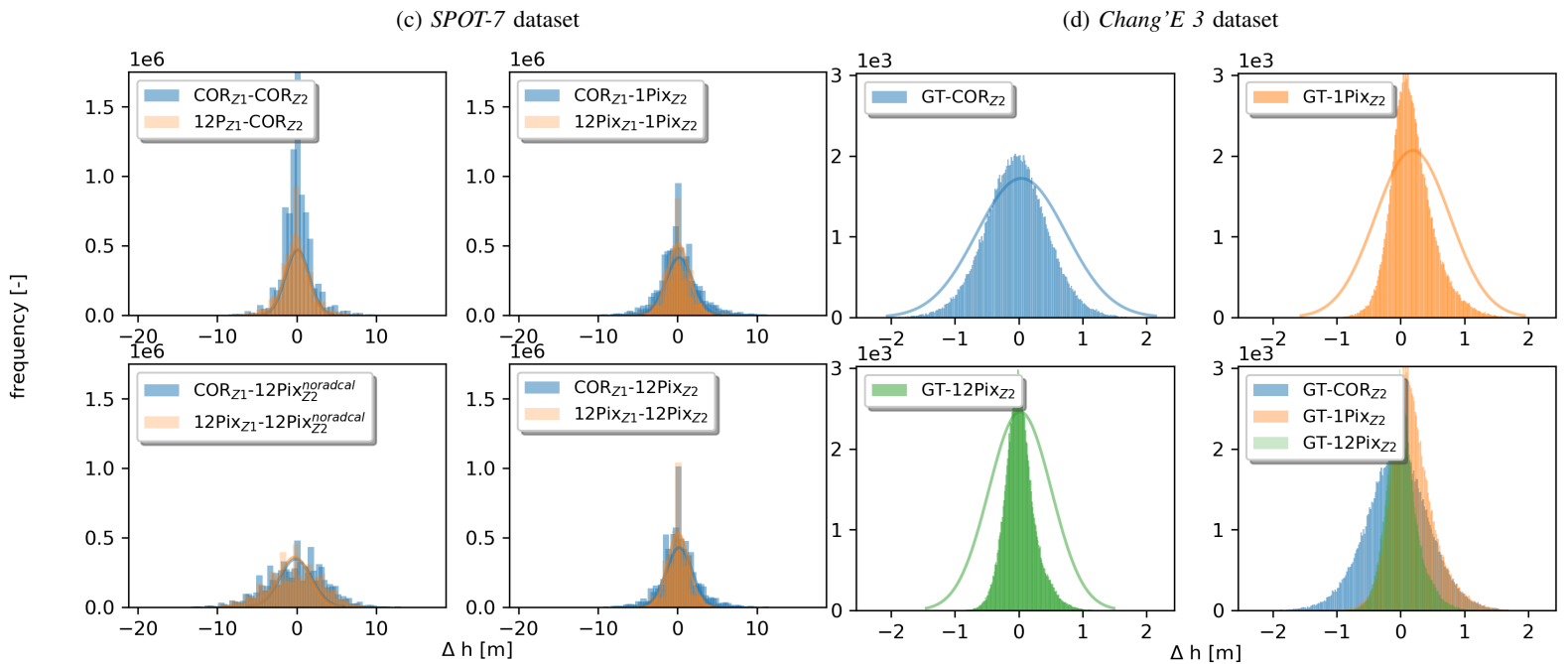
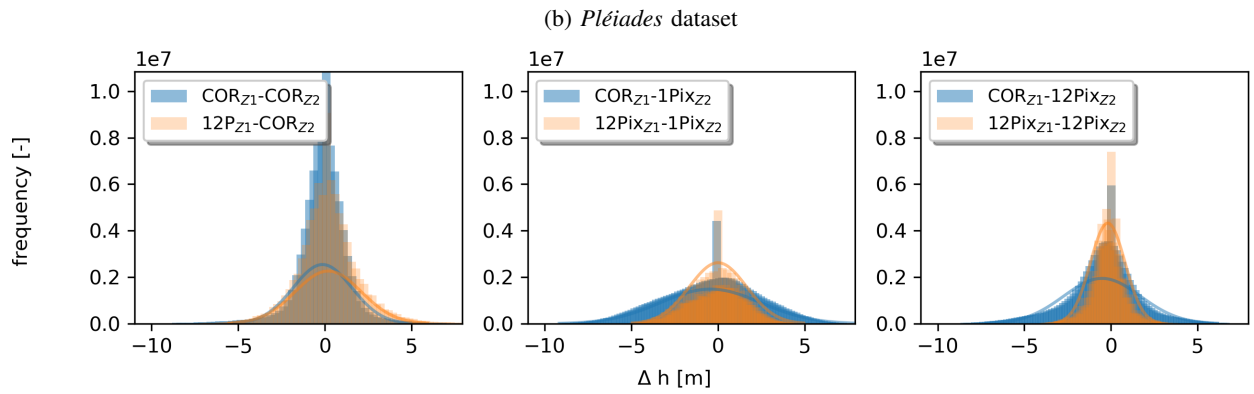
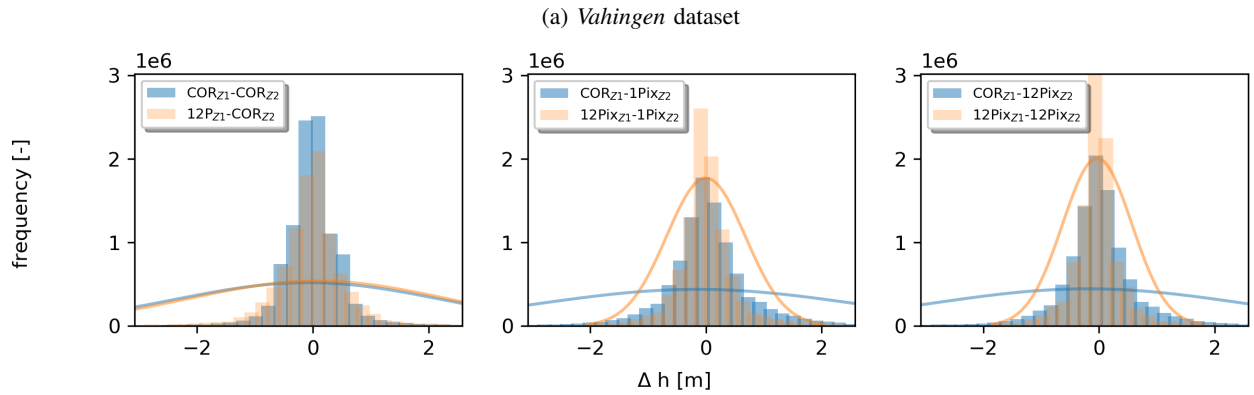


Figure 5: Δh histograms calculated for all datasets and *Ground Truths* (i.e., $CORR_{Z1}$, $12Pix_{Z1}$ and GT). Superimposed curves are theoretical histograms calculated from non-robust error metrics.

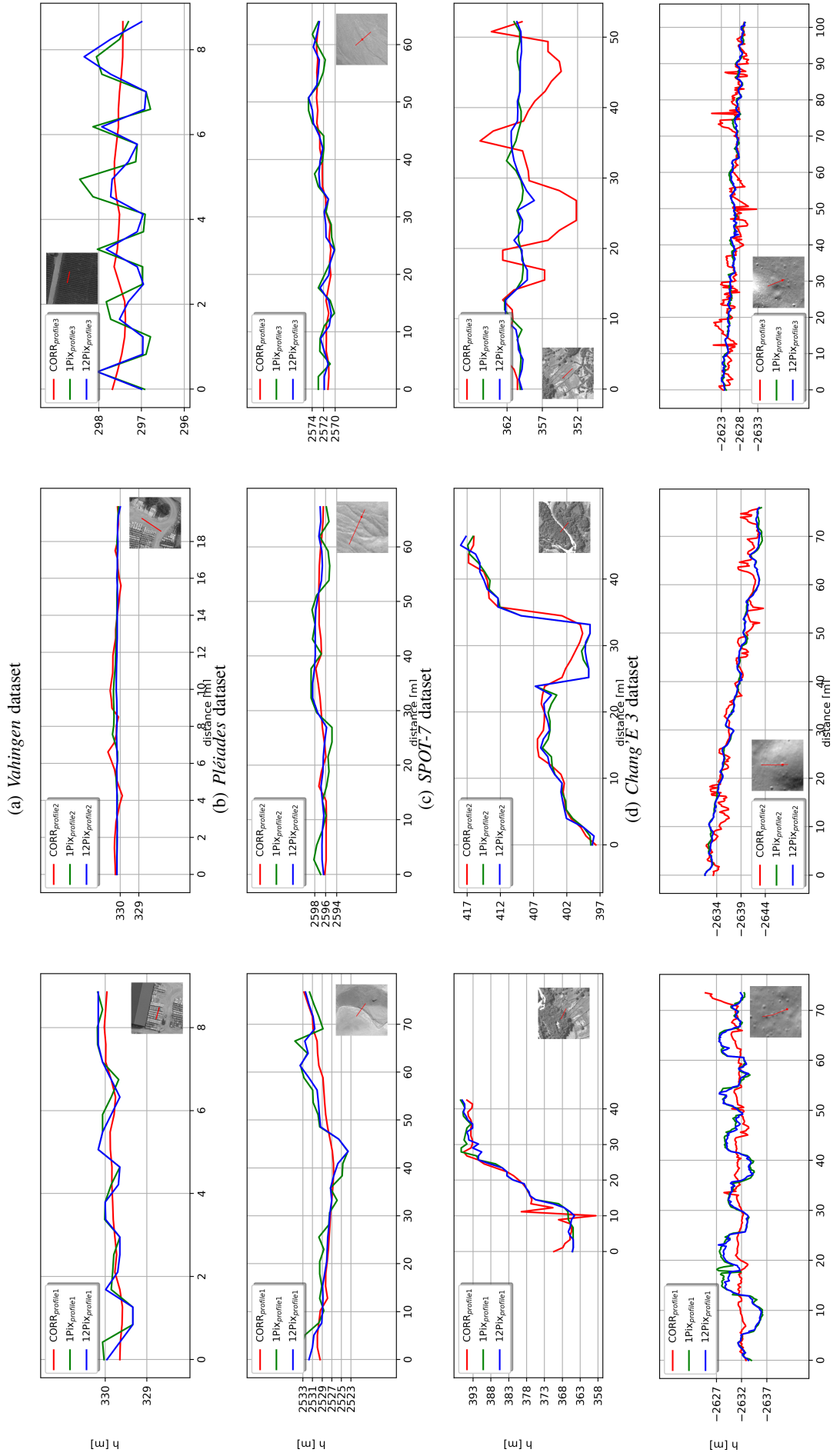


Figure 6: Height profiles drawn over DSMs computed with $CORR_{z1}$ (in red), $1PIx_{z1}$ (in blue) and $12PIx_{z1}$ (in green).

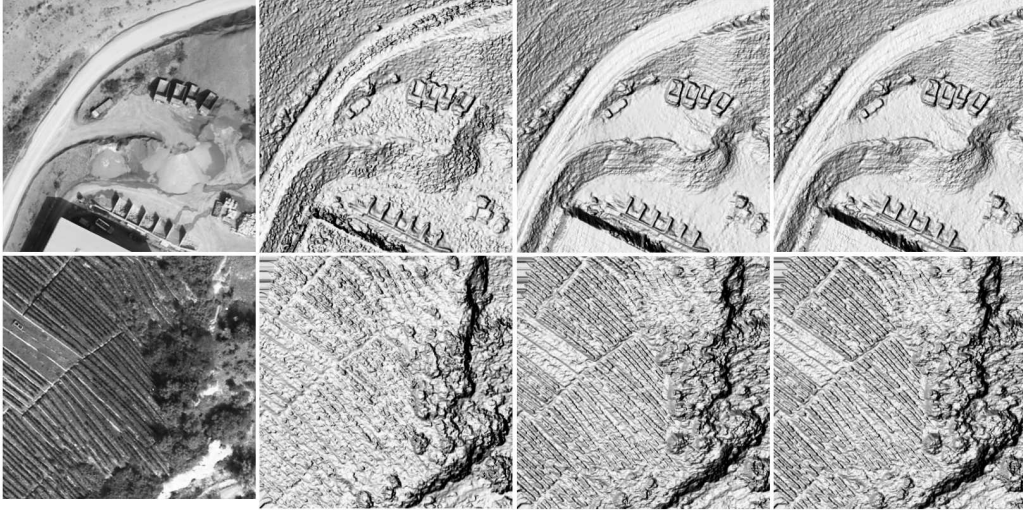
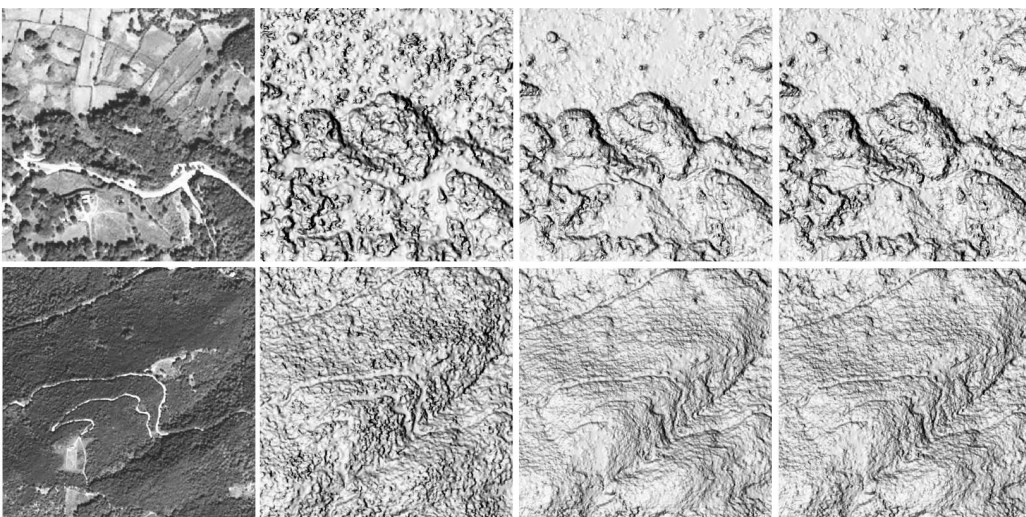
(a) *Vahingen* dataset(b) *Pléiades* dataset(c) *SPOT-7* dataset

Figure 7: Close-ups on selected DSM areas. The columns starting from left: the image, grayshaded DSMs computed with *CORR*, 1Pix and 12Pix , respectively.

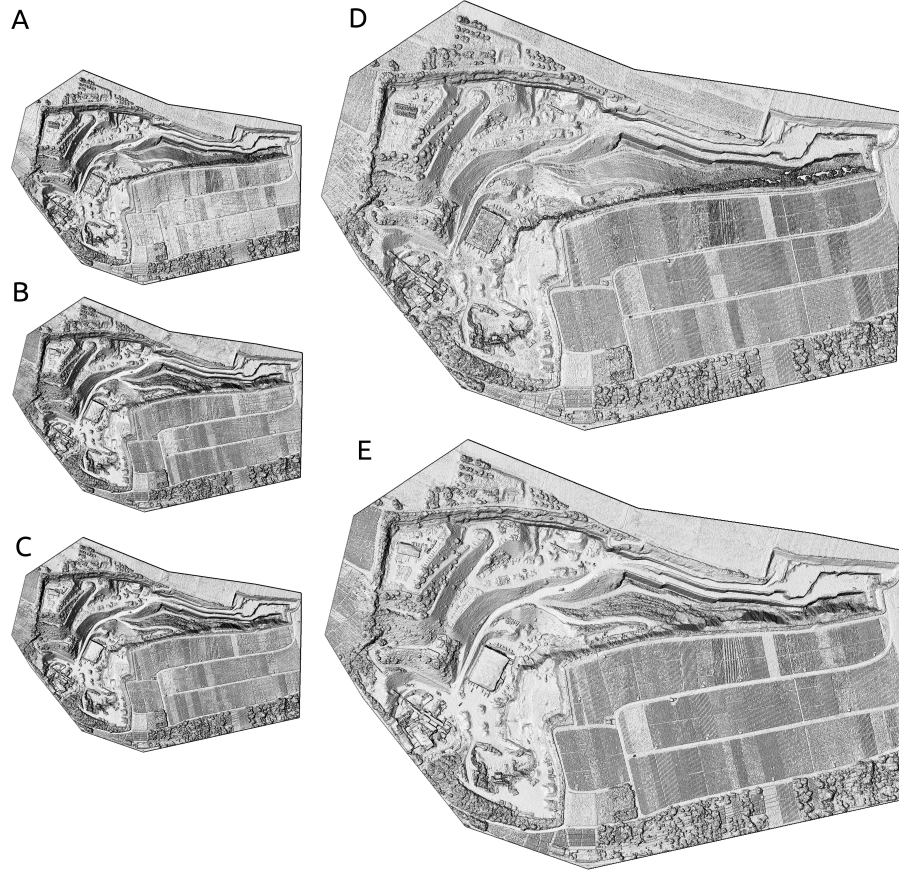
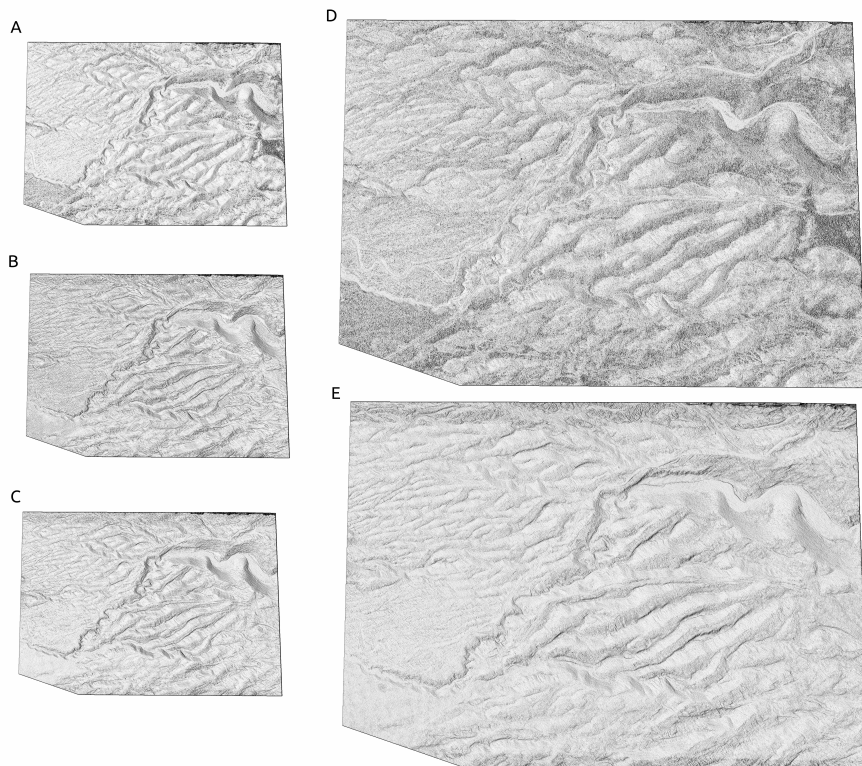
(a) *Vahingen* dataset.(b) *Pléiades* dataset.

Figure 8: Grayshaded representation of surfaces reconstructed with different methods and at different resolutions. A. $CORR_{Z_2}$, B. $1Pix_{Z_2}$, C. $12Pix_{Z_2}$, D. $CORR_{Z_1}$, E. $12Pix_{Z_1}$.



Figure 9: *Spot-7* dataset. Grayshaded representation of surfaces reconstructed with different methods and at different resolutions. A. $CORR_{Z2}$, B. $12Pix_{Z2}^{noRadCal}$, C. $1Pix_{Z2}$, D. $12Pix_{Z2}$, E. $CORR_{Z1}$, F. $12Pix_{Z1}$.

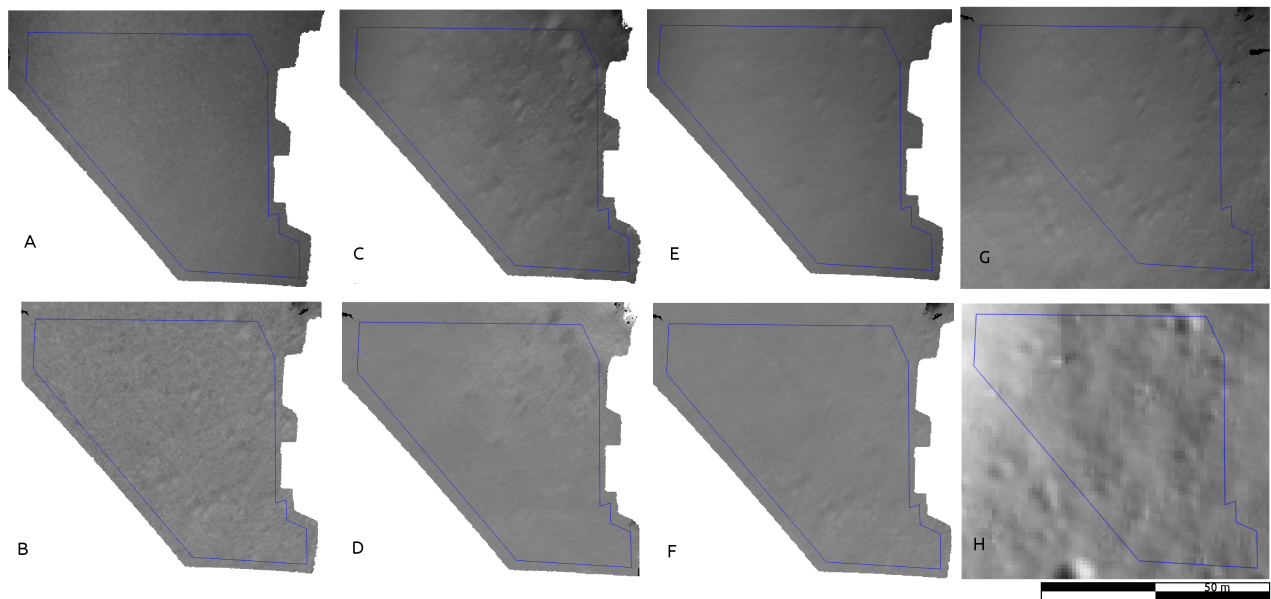


Figure 10: *Chang'E 3* dataset. Surfaces reconstructed with (A) COR , (C) $1Pix$, (E) $12Pix$ similarity measures, as well as (G) the GT surface. The surface differences Δh for (B) $GT - COR$, (D) $GT - 1Pix$, (F) $GT - 12Pix$. (H) is the LROC orthophoto and in blue is the region of interest.



Figure 11: *Terrestrial* dataset. A sequence of 5 images used to generate the surface in Figs. 12 and 13.

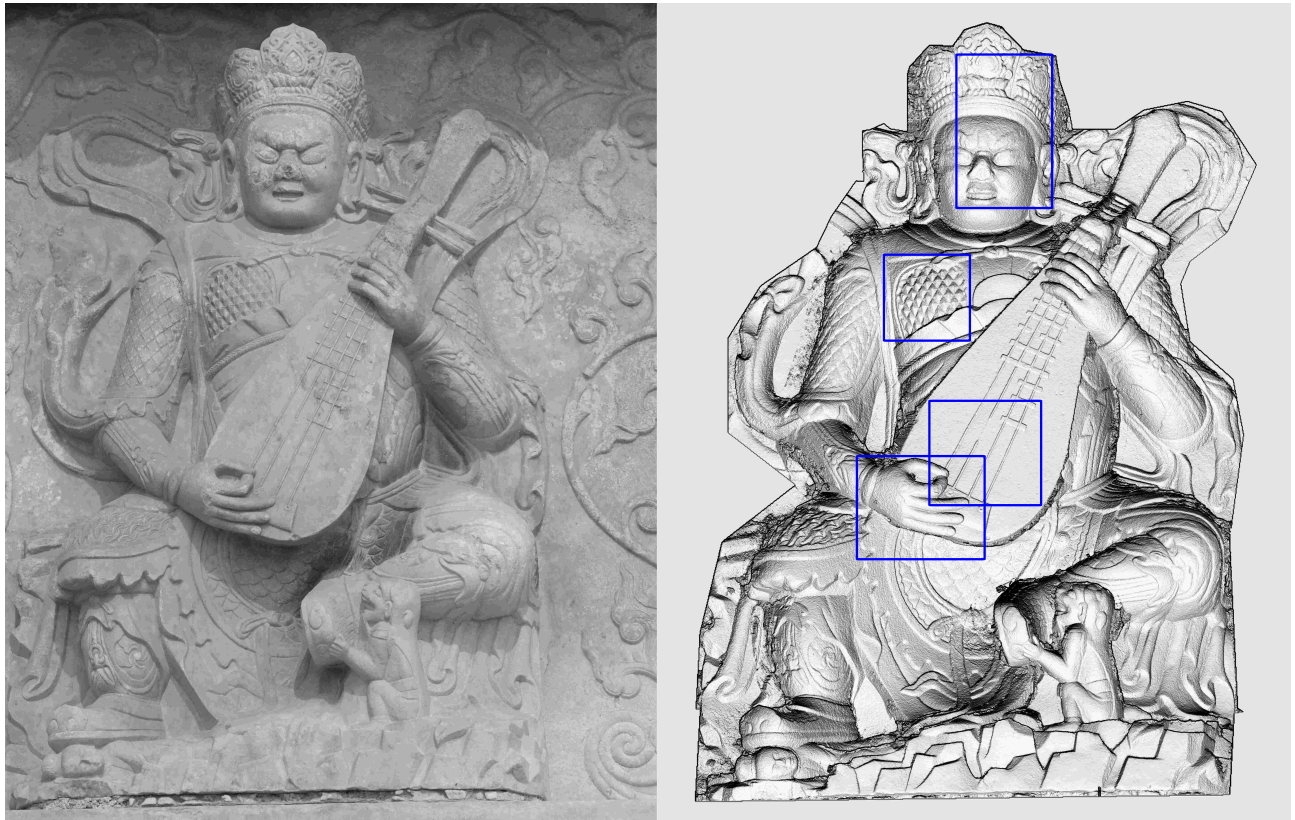


Figure 12: *Terrestrial dataset*. Left: the master image; right: the DSM and in blue the regions of interest.

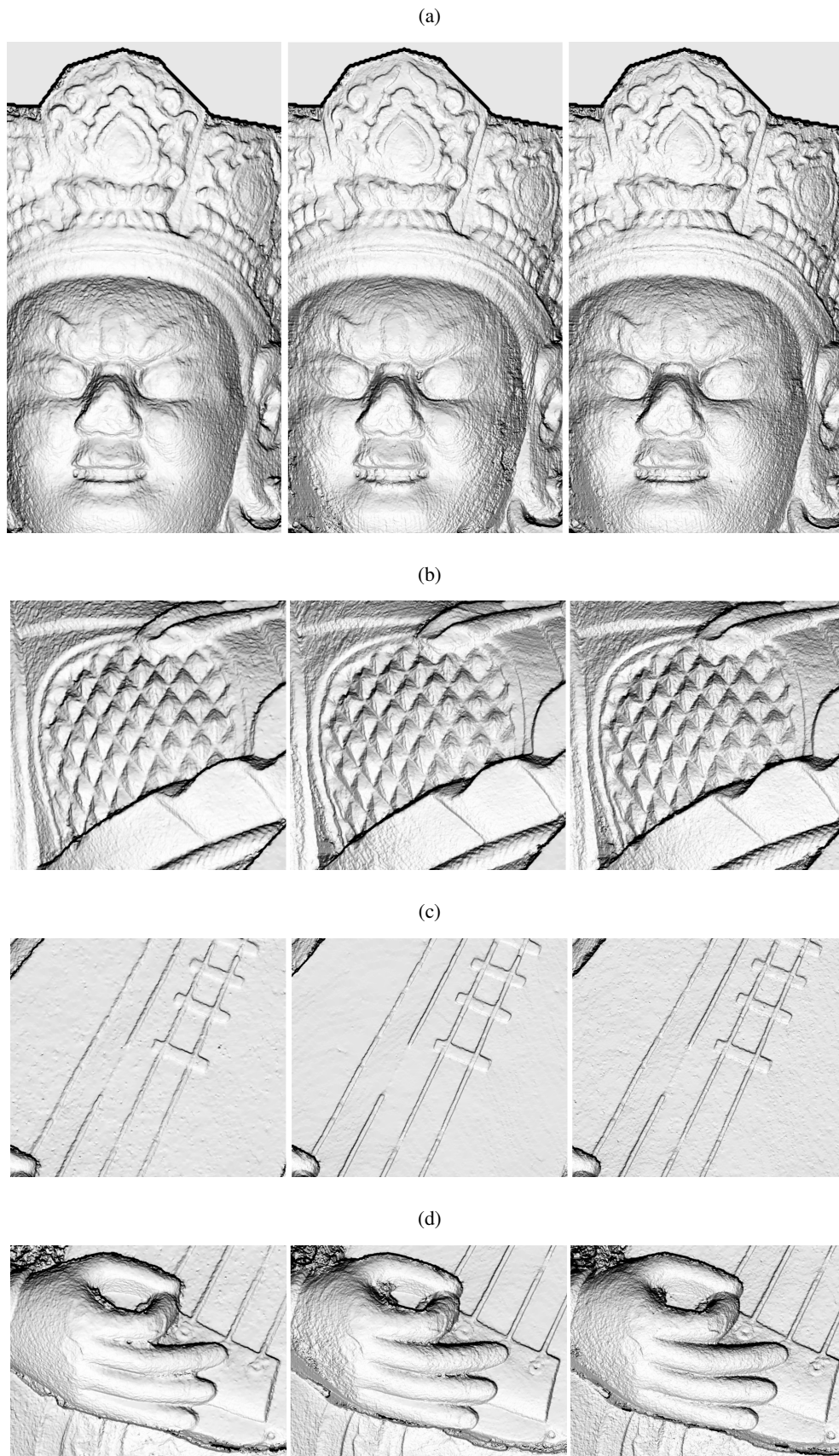


Figure 13: *Terrestrial* dataset. Close-ups on selected DSM areas. The DSMs were computed with *CORR* (left column), *1Pix* (middle column) and *12Pix* (right column) similarity measures.

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. Fornaciari, 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. Käab, 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. Marzolf, 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, "3D reconstruction from multi-view VHR-satellite images in MicMac," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 139, pp. 201–211, 2018.
- [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. Käab, 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.
- [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] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," in *European conference on computer vision*. Springer, 1994, pp. 151–158.
- [31] P. Viola and W. M. Wells III, "Alignment by maximization of mutual information," *International journal of computer vision*, vol. 24, no. 2, pp. 137–154, 1997.
- [32] 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.
- [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. 1, p. 14, 2017.
- [37] M. Pierrot-Deseilligny, E. Rupnik, L. Girod, J. Belvaux, G. Maillet, M. Deveau, and G. Choqueux, "Micmac, apero, pastis and other beverages in a nutshell," 2016.
- [38] N. Haala, "Dense image matching final report," *EuroSDR Publication Series, Official Publication*, vol. 64, pp. 115–145, 2014.
- [39] E. Rupnik, M. Pierrot-Deseilligny, A. Delorme, and Y. Klinger, "Refined satellite image orientation in the free open-source photogrammetric tools Apero/MicMac," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2016.
- [40] J. Höhle and M. Höhle, "Accuracy assessment of digital elevation models by means of robust statistical methods," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 64, no. 4, pp. 398–406, 2009.