Robust Confidence Intervals for Digital Surface Models Using Satellite Photogrammetry

Roman Malinowski, Emmanuelle Sarrazin, Emmanuel Dubois, Loïc Dumas, Sébastien Destercke

ROBUST CONFIDENCE INTERVALS FOR DIGITAL SURFACE MODELS USING SATELLITE PHOTOGRAMMETRY

Roman Malinowski[1,2,3], *Emmanuelle Sarrazin*[1], *Emmanuel Dubois*[1], *Loïc Dumas*[2], *Sébastien Destercke*[3]

[1] CNES, Toulouse, France [2] CS, Toulouse, France [3]UTC - Heudiasyc, Compiègne, France

ABSTRACT

CNES has developed the CARS pipeline to massively produce Digital Surface Models (DSM) for remote sensing applications. DSM are computed from pairs of very high resolution satellite imagery using multi-view stereo methods. In this paper, we compute robust elevation confidence intervals with more than 90% accuracy alongside the high resolution DSM. We first estimate disparity confidence intervals using possibility distributions during the dense matching step, where most errors usually occur. Disparity confidence intervals are then processed with caution throughout every step of the pipeline to be transformed into elevation confidence intervals. Intervals accuracy is evaluated on both urban and glacier high resolution images.

***Index Terms*—** Image processing, Photogrammetry, Stereovision, Digital Surface Model, 3D, Stereo, Uncertainty, Confidence intervals, Possibility distributions

1. INTRODUCTION

Digital Surface Models (DSM) are used in many remote sensing applications [1, 2]. The CO3D mission (`https://co3d.cnes.fr/fr`), started by CNES and Airbus Defense & Space, aims to automatically produce DSM from Very High Resolution (VHR) images at a global scale. To process the VHR images from satellites, CNES has developed a multi-view stereo pipeline called CARS [3]. This pipeline allows massively parallel computations to ensure scalability, robustness and good performances. It produces DSM from image pairs. It is composed of multiple classical steps, detailed in figure 1. Alongside the DSM, CARS is expected to assess the quality of the height reconstruction. The pipeline currently produces a confidence measure [4] indicating how trustworthy the elevation value of a given DSM pixel is. Although very useful, confidence measures do not indicate if other probable results are near the predicted values or far away. In this paper, we propose to address this issue by introducing confidence intervals, as in figure 2. Confidence intervals enable to capture the notion of error magnitude missing in classical confidence measures.

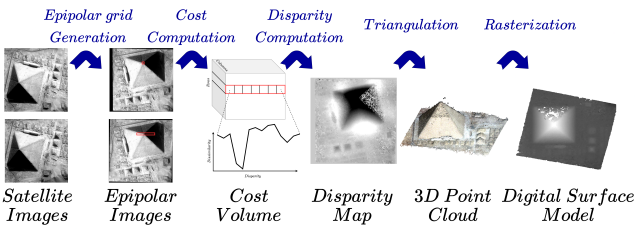

Fig. 1. CARS Pipeline

Classical approaches for computing confidence intervals on DSM are *a posteriori* methods [2, 5, 6], meaning that the intervals are generated using only the output DSM and a reference DSM, regardless of the process used to obtain them. These methods statistically estimate the residuals of squared errors between the computed DSM and a set of reference points, obtained from other sources or by resampling the computed DSM. Thus, it results in a single confidence interval on the mean or median of the squared errors. In this article, we compute elevation confidence intervals alongside the DSM, by estimating the errors potentially made by algorithms in the pipeline. We differ in this regard from classical *a posteriori* methods, as we do not use a reference DSM. Furthermore, we produce confidence intervals on the elevation of every pixel of the output raster with a 90% accuracy, instead of a single confidence interval to estimate the mean or median of squared errors. One of this paper main contribution is also to use robust uncertainty models, named possibility distribution [7], to model the uncertainty of the dense stereo matching step of the pipeline. Possibility distributions are well suited to model epistemic uncertainty (i.e. due to a lack of knowledge),

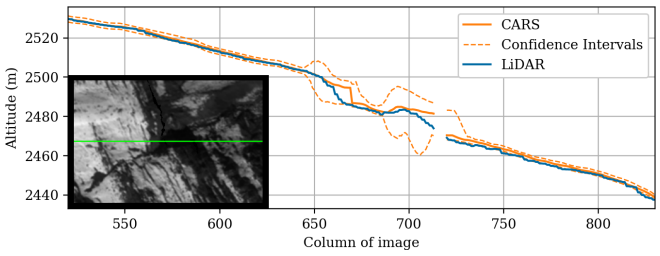

Fig. 2. Elevation intervals along a row of the Peyto scene. Orthoimage with highlighted row appear in the bottom left

and enables to retrieve confidence intervals.

The following section explains the method for creating disparity intervals in the stereo matching step. We then detail the propagation of these intervals to elevation intervals throughout triangulation, filtering and rasterization steps. Finally, we evaluate the accuracy of the confidence intervals on DSM obtained from Pléiades images with a sub-meter resolution. We use LiDAR acquisitions as reference data on the french city of Montpellier, as well as glaciers located in Peyto in Canada, and mountains from the Yosemite Park in the USA.

2. METHOD FOR CREATING DISPARITY INTERVALS

Computation of the disparity map is a crucial step. Its performance drives the quality of the DSM. We therefore decide to determine confidence intervals during the disparity computation step. CARS pipeline uses a stereo matching framework developed by CNES and CS called Pandora [8] for computing the disparity map (see figure 1). To do so, Pandora estimates a cost volume C_V, where $C_V(i, j, d)$ is the matching-cost between pixels (i, j) and $(i, j+d)$ with disparity d from left and right images. We propose to use this information to determine intervals of the most probable disparities for every pixel (i, j). For this purpose, we use *possibility distribution* to model the uncertainty on the considered disparities. Possibility distributions are functions whose values lie in $[0, 1]$, taking the value 1 at least once, strongly related to fuzzy sets [7]. A possibility distribution represents the degree of possibility that d is the correct disparity: 1 meaning that d is fully possible, and 0 meaning that d is impossible. By noting C_V^{min}, C_V^{max} the minimal and maximal values taken by the cost volume, we define for every pixel the possibility $\pi_{i,j}$ as follows:

$$\pi_{i,j}(d) = 1 - \frac{C_V(i, j, d) - \min_\delta C_V(i, j, \delta)}{C_V^{max} - C_V^{min}} \qquad (1)$$

From $\pi_{i,j}$, we compute the set of most probable disparities $\mathcal{D}$ as the disparities who have a degree of possibility at least superior to 0.9. We then define the disparity confidence interval $[\underline{d}, \overline{d}]$ as follows:

$$\mathcal{D} = \{d \mid \pi_{i,j}(d) \geqslant 0.9\},\ \underline{d} = \min \mathcal{D},\ \overline{d} = \max \mathcal{D} \qquad (2)$$

An example of a cost function transformed into a possibility is presented in figure 3, and disparity intervals along the row of an epipolar image is presented in figure 4.

In our experiments, we use the CENSUS similarity measure with a Semi-Global Matching regularization to compute the cost volume. We also use a median filter to correct local outliers and smooth the disparity map. Applying the median filter independently on the disparity map and the interval bounds ensures the consistency of the results. For more details on the method to create disparity intervals, we refer to

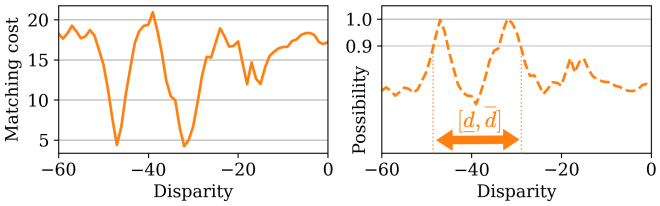

Fig. 3. A cost function (left) transformed into a possibility distribution (right). Arrows represent the disparity interval.

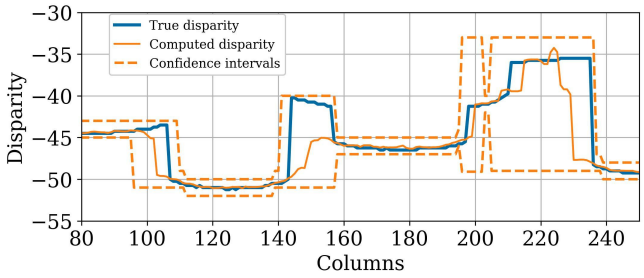

Fig. 4. Disparity intervals along the row of an epipolar image

[9] where the 90% confidence rate of intervals is validated. The following section details the propagation of the resulting disparity intervals into elevation intervals.

3. PROPAGATING DISPARITY INTERVALS INTO ELEVATION INTERVALS

3.1. Triangulation

Knowing the disparity map, as well as the sensor model, it is possible to compute the 3D point coordinates of every pixel of the images by intersecting sensors' lines of sight, also known as triangulation (figure 1). To each pixel $p = (i, j)$ corresponds a 3D point $P_{i,j} = (x, y, z)$. Similarly, we also estimate upper/lower coordinates of every 3D point by intersecting lines of sight using bounds of the lower/upper disparity intervals. To each pixel $p = (i, j)$ thus corresponds a triplet of 3D points $P_{i,j} = (x, y, z)$, $\underline{P}_{i,j} = (\underline{x}, \underline{y}, \underline{z})$, $\overline{P}_{i,j} = (\overline{x}, \overline{y}, \overline{z})$ computed using the disparity prediction and its confidence interval bounds[1]. To facilitate further processing, only elevation changes are computed during the processing of the 3D triplets, meaning we consider that $\underline{x} \approx \overline{x} \approx x$ and $\underline{y} \approx \overline{y} \approx y$. In our case, this approximation can be done as we have a B/H ratio of around 0.3, meaning that the variation of x and y are small compared to the variation of z. For a pixel $p = (i, j)$, we thus get a triplet of points:

$$\underline{P}_{i,j} = (x, y, \underline{z}),\ P_{i,j} = (x, y, z),\ \overline{P}_{i,j} = (x, y, \overline{z}) \qquad (3)$$

[1] In cases where disparities only have negative values, $\overline{P}_{i,j}$ is obtained using $\underline{d}$ and $\underline{P}_{i,j}$ is obtained using $\overline{d}$

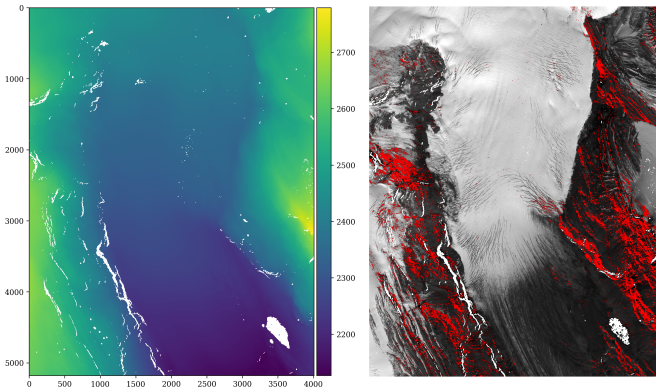

Fig. 5. LiDAR DSM and wrong intervals (red pixels) on the orthoimage for the Peyto region. Accuracy is 92%.

3.2. Filtering

Different filtering steps are applied to ensure the quality of the point cloud. Those steps need to be taken with caution as to maintain coherence between the final prediction and its confidence intervals. During point cloud filtering step, we remove statistical outliers and small clusters of 3D points located far from any other cluster. Applying the filters independently for the predicted point cloud and for its intervals bounds could lead to inconsistencies. Instead, if we filter a point $P_{i,j} = (x, y, z)$ from the main point cloud, then we remove the associated upper and lower points $\underline{P}_{i,j}, \overline{P}_{i,j}$ as well.

3.3. Rasterization

We convert the point cloud into a raster (figure 1) using a weighted mean. Weights are computed using a Gaussian on the distance between each 3D point and the center of the cell. Applying the Gaussian independently to the main point cloud and for the interval bounds maintains their consistency.

4. EVALUATION

4.1. Datasets

Using the CARS pipeline, we compute DSM at 0.5m resolution from Pléiades images at 0.7m and then resampled at 0.5m. Validation is performed with reference DSM acquired by LiDAR. The computed DSM and the ground truth are co-registrated using the method in [10]. Studied scenes are located in Montpellier in France (MTP), Peyto in Canada (Peyto) and two scenes are in California, USA (Cal 1 & 2). More details are presented in table 1. We thank the Engineering Research Council of Canada, the Tula Foundation (Hakai Institute), the CESBIO and the Airborne Snow Observatories for kindly providing ground truth data.

Location	Size	Date image	Date LiDAR
MTP	4001×4000	2019-09-12	2021-07-30
Peyto	5178×4009	2016-09-13	2016-09-13
Cal 1	8354×8315	2017-05-01	2017-02-05
Cal 2	8354×8315	2017-05-02	2017-02-05

Table 1. Size and acquisition date of evaluation data

4.2. Evaluation metrics

We use three metrics to evaluate the performance of our method. First, we compute the intervals accuracy I_{acc}, defined as the proportion of intervals that correctly include the ground truth. It is however possible to obtain 100% accuracy by extending the confidence intervals to unnecessarily large sizes. To ensure that the size of the intervals are usable, we calculate their median size I_{size} relatively to the *disparity to altitude ratio* r_{alt}. This ratio, expressed in m/pix, represents the amount of altitude gained by a shift of one pixel during the disparity computation step. The ratio changes from one scene to another, depending on the viewing angles of the satellite. For intervals that do not capture the true elevation Z, we compute the median of the error magnitude to ensure that wrong intervals remain close to the ground truth. I_{err} is defined as the distance between the true elevation and the interval bounds. The error is normalized by the r_{alt} to be expressed in pixels.

$$I_{acc} = \frac{\#correct\ intervals}{\#number\ of\ intervals} \quad (4)$$

$$I_{size} = med\left(\frac{\overline{z} - \underline{z}}{r_{alt}}\right) \quad (5)$$

$$I_{err} = med\left(\frac{1}{r_{alt}} \times \min(|Z - \underline{z}|, |Z - \overline{z}|)\right) \quad (6)$$

4.3. Results

An example of confidence intervals along the row of a DSM can be seen in figures 2 and 6. The intervals have different sizes depending on their position in the image, adapting to areas where the algorithm struggles to correctly estimate the elevation. Some extreme cases in dense urban scenery are presented in figure 6, where intervals get pessimistic near building borders. Results on each scene is detailed in table 2, where metrics are evaluated globally, and also for different slope categories. The main result showing the robustness and validity of the method is the accuracy superior to 90% on every scene when considering every slope. The accuracy drops below 90% only for slopes superior to 20° for the MTP and Peyto scenes. The error is spatially correlated to steep slopes, as seen in figure 5. Intervals sizes are between 2 and 3 pixels in altitude, which is relatively small. When intervals do not

Metric	Slope (°)	MTP	Peyto	Cal 1	Cal 2
I_{acc} (%)	Global	90.7	92.0	96.5	94.1
	$[0, 20[$	93.8	99.3	97.0	95.0
	$[20, 45[$	89.7	86.9	98.1	96.7
	$[45, 90]$	88.1	59.0	93.9	90.0
I_{size} (pix)	Global	2.2	2.1	2.9	2.5
	$[0, 20[$	2.0	2.1	3.0	2.6
	$[20, 45[$	1.9	2.1	3.4	2.7
	$[45, 90]$	2.7	2.4	2.4	2.3
I_{error} (pix)	Global	0.2	0.3	0.3	0.3
	$[0, 20[$	0.2	0.1	0.3	0.3
	$[20, 45[$	0.2	0.2	0.3	0.2
	$[45, 90]$	0.3	0.5	0.3	0.3

Table 2. Metrics evaluated for different slope categories.

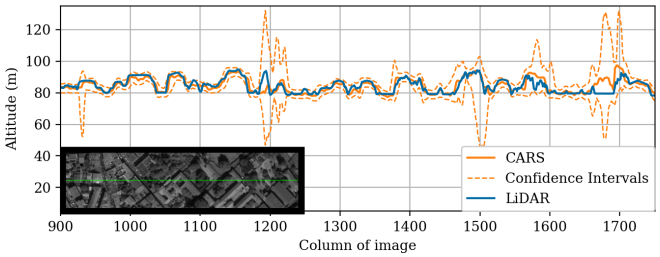

Fig. 6. Elevation intervals along a row of the MTP scene. Orthoimage with highlighted row appear in the bottom left.

contain the ground truth, they usually miss the ground truth by less than half a pixel.

5. CONCLUSION

This paper presents a method for computing confidence elevation intervals based on the uncertainty of the algorithms used to create a DSM. The intervals are propagated with caution and validate the desired accuracy. Contrary to current methods for evaluating the uncertainty of DSM, our method does not need a reference DSM to create the intervals. The intervals automatically adapt their size depending on the terrain and the difficulty experienced by the algorithm to evaluate the elevation along the image. The code is freely available and implemented in the CARS pipeline. We hope that this work will help users handling DSM. Future work will address the performances of the method for very steep slopes. We will also investigate the origin of the sub-pixel error, to understand if it results from our method or from the re-sampling when switching from sensor geometry to epipolar geometry.

6. REFERENCES

[1] Etienne Berthier, Christian Vincent, Eyjólfur Magnússon, Ágúst Þór Gunnlaugsson, Pierre Pitte, Emmanuel Le Meur, Mariano Masiokas, Lucas Ruiz, Finnur Pálsson, Joaquín M.C. Belart, and Patrick Wagnon, "Glacier topography and elevation changes derived from Pléiades sub-meter stereo images," *The Cryosphere*, Dec. 2014.

[2] César Deschamps-Berger, *Apport de la photogrammétrie satellite pour la modélisation du manteau neigeux*, Ph.D. thesis, Université Toulouse 3 - Paul Sabatier, Aug. 2021.

[3] Julien Michel, Emmanuelle Sarrazin, David Youssefi, Myriam Cournet, Fabrice Buffe, Jean-Marc Delvit, Aurelie Emilien, Julien Bosman, Olivier Melet, and Céline L'Helguen, "A New Satellite Imagery Stereo Pipeline Designed for Scalability, Robustness and Performance," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Aug. 2020.

[4] Emmanuelle Sarrazin, Myriam Cournet, Loïc Dumas, Véronique Defonte, Quentin Fardet, Y. Steux, N. Jimenez Diaz, E. Dubois, D. Youssefi, and F. Buffe, "Ambiguity Concept In Stereo Matching Pipeline," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, June 2021.

[5] Romain Hugonnet, Fanny Brun, Etienne Berthier, Amaury Dehecq, Erik Schytt Mannerfelt, Nicolas Eckert, and Daniel Farinotti, "Uncertainty Analysis of Digital Elevation Models by Spatial Inference From Stable Terrain," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2022.

[6] Bin Wang, Wenzhong Shi, and Eryong Liu, "Robust methods for assessing the accuracy of linear interpolated DEM," *International Journal of Applied Earth Observation and Geoinformation*, Feb. 2015.

[7] Didier Dubois and Henri Prade, "When upper probabilities are possibility measures," *Fuzzy Sets and Systems*, July 1992.

[8] Myriam Cournet, Emmanuelle Sarrazin, Loïc Dumas, Julien Michel, Jonathan Guinet, David Youssefi, Véronique Defonte, and Quentin Fardet, "Ground Truth Generation and Disparity Estimation for Optical Satellite Imagery," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Aug. 2020.

[9] Roman Malinowski, Emmanuelle Sarrazin, Loïc Dumas, Emmanuel Dubois, and Sébastien Destercke, "Robust Confidence Intervals in Stereo Matching using Possibility Theory," Apr. 2024, arXiv:2404.06273 [cs].

[10] Christopher Nuth and Andreas Kääb, "Co-registration and bias corrections of satellite elevation data sets for quantifying glacier thickness change," *The Cryosphere*, Mar. 2011.