AUTOMATIC DIGITAL SURFACE MODEL GENERATION FROM PLÉIADES STEREO IMAGES

Carlo de Franchis[1], Enric Meinhardt-Llopis[1], Julien Michel[2], Jean-Michel Morel[1], Gabriele Facciolo[1]

1: Ecole Normale Supérieure de Cachan, 61 Avenue du Président Wilson, 94230 Cachan, France

2: CNES - DCT/SI/AP, 18 Avenue Edouard Belin, 31401 Toulouse, France

Résumé

Nous proposons une chaîne stéréo complètement automatique pour produire des modèles numériques d'élévation à partir d'images Pléiades. L'agilité des satellites Pléiades leur permet d'acquérir plusieurs vues d'une scène au cours d'un même passage, ce qui ouvre la voie à de nouvelles applications tirant parti de ces séquences quasiment simultanées d'images à haute résolution. En pratique le mode d'acquisition tri-stéréo permet de réduire les occlusions et d'effectuer une validation croisée des points reconstruits. Cet article présente notre chaîne, baptisée *s2p*, et étudie des modèles numériques d'élévation et des nuages de points obtenus à partir de triplets stéréo fournis par Airbus DS et le CNES via le programme RTU. L'algorithme s2p a la particularité de permettre l'utilisation des outils habituels de mise en correspondance stéréo. Pour cela il effectue une stéréo-rectification très précise de chaque paire d'images. Bien que le système d'acquisition ne corresponde pas au modèle pinhole, qui est nécessaire pour stéréorectifier, les erreurs commises en approximant le système par un modèle pinhole sont négligeables pour des portions d'images suffisamment petites. Une image complète peut ainsi être traitée par petites tuiles qui sont traitées indépendamment.

Mots clés : reconstruction 3D, stéréo, modèle numérique d'élévation, télédetection, capteur barrette

Abstract

We propose a fully automated stereo pipeline for producing digital elevation models from Pléiades satellite images. The agility of the Pléiades satellites allows them to capture multiple views of the same target in a single pass, enabling new applications that exploit these quasi-simultaneous high-resolution images. Concretely the tri-stereo acquisition modality permits to reduce the occlusions and to cross-validate the observed points. This paper gives an overview of our pipeline, named s2p*, and presents some digital elevation models and 3D point clouds built from Pléiades tri-stereo datasets. The data was provided by Airbus DS and the CNES through the RTU program. The particularity of the s2p algorithm is that it permits to use conventional stereo correlation tools, by performing a very precise image rectification of each stereo pair. Although the acquisition system does not fit the pinhole camera model, which is necessary to make the rectification possible, the errors due to the pinhole assumption were shown to be negligible for small enough image sizes. Thus, the whole image can be treated by cutting it into small tiles that are processed independently.*

Keywords : *3D reconstruction, stereo, digital elevation models, remote sensing, pushbroom geometry*

1. INTRODUCTION

This paper presents a fully automatic 3D reconstruction pipeline for satellite images, meant to be modular and generic. This work was motivated by the recent availability of high resolution images from new satellites with stereo capabilities such as Pléiades. Even if all the the experiments described here were carried on Pléiades images, our work also applies to images from other satellites such as WorldView, Quickbird and Spot.

1.1. Pléiades stereo data

The Pléiades constellation is composed of two Earth observation satellites able to deliver images with a resolution of 70 cm and a swath width of 20 km. They operate in the same orbit and are phased 180° apart. Their unique agility allows to capture multiple views of the same target in a single pass. This permits nearly simultaneous stereo or tri-stereo acquisitions with a small base to height ratio, ranging from 0.15 to 0.8. The images have a pixel depth of 12 bits with a signal-to-noise ratio (SNR) greater than 90 dB. Pléiades, as many other Earth observation satellites, acquires images with a pushbroom sensor, which captures them line by line as the satellite moves. The calibration information describing the camera system is provided for all Pléiades images under the form of RPC functions.

1.2. The rational polynomial camera model

Each Pléiades image (EADS, 2012) is accompanied by a pair of functions, called RPC (Baltsavias and Stallmann, 1992; Dowman and Dolloff, 2000; Tao and Hu, 2001). These functions allow to convert from image coordinates to coordinates on the globe and back. The projection from object space to image plane is denoted by RPC : $\mathbf{R}^3 \to \mathbf{R}^2$, $(\varphi, \lambda, h) \mapsto (x, y)$, where 3-space points are represented by their spheroidal coordinates in the World Geodetic System (WGS 84). In that system a point of 3-space is identified by its latitude $\varphi \in [-90, 90]$, longitude $\lambda \in]-180, 180]$ and altitude h, in meters, above the reference ellipsoid. Its inverse, with respect to the first two components, is denoted by $\text{RPC}^{-1} : \mathbf{R}^3 \to \mathbf{R}^3$, $(x, y, h) \mapsto (\varphi, \lambda, h)$. It takes a point $\mathbf{x} = (x, y)$ in the image domain together with an altitude h, and returns the coordinates of the unique 3-space point $\mathbf{X} = (\varphi, \lambda, h)$ whose altitude is h and whose image is $\mathbf{x}$.

For example, the latitude component of the RPC^{-1} function for the image point (x, y) at altitude h is

$$\varphi_N = \frac{\sum_{i=1}^{20} C_i^{NUM,\varphi} \rho_i(x_N, y_N, h_N)}{\sum_{i=1}^{20} C_i^{DEN,\varphi} \rho_i(x_N, y_N, h_N)}.$$

where x_N, y_N, h_N, φ_N are normalized coordinates, C_i^* is the i-th coefficient of the polynomial and ρ_i produces the i-th factor of the polynomial.

For the sake of clarity, we shall denote by $\text{RPC}_u : \mathbf{R}^3 \to \mathbf{R}^2$ the projection function of the RPC model associated to image u, and by $\text{RPC}_u^{-1} : \mathbf{R}^3 \to \mathbf{R}^3$ the corresponding inverse function. Ideally, these functions should verify

$$\text{RPC}_u^{-1}(\text{RPC}_u(\varphi, \lambda, h), h) = (\varphi, \lambda, h)$$

and

$$\text{RPC}_u(\text{RPC}_u^{-1}(x, y, h)) = (x, y),$$

but as any model the rational polynomial camera model has a limited precision. In particular the two RPC functions are not exact inverses of each other. The errors due to concatenating the projection and inverse functions are negligible, being of the order of 10^{-7} degrees in WGS 84 coordinates, and $\frac{1}{100}$ of pixel in the image *i.e.* about 1 cm on the ground.

Note that Pléiades images are provided by Airbus DS with both RPC and RPC^{-1} functions, while DigitalGlobe provides only the projection function RPC. The inverse RPC^{-1} has to be estimated from RPC. This is done iteratively.

1.3. Related works

Similarly to previous works (Wohlfeil et al., 2012; d'Angelo and Reinartz, 2012; d'Angelo and Kuschk, 2012; Kuschk, 2013), the pipeline presented here is fully automated. All tasks that used to be performed manually such as disparity range estimation, tie points selection for RPC refinement, and water masking, are performed automatically thanks to the proper use of SRTM data (Farr et al., 2007) and feature detectors such SIFT (Lowe, 2004). But unlike these works, ours does not include a particular stereo matching algorithm. Instead the main contribution of our work is that it gives a complete framework to evaluate any stereo matching algorithm (that works with stereo-rectified images) on pushbroom images. This framework is available online through a web interface and can be tested on any images, as soon as the associated RPC coefficients are provided.

In the next section we give an overview of the s2p pipeline, and in section 3 we present the results obtained from extensive experimentation carried out using images from Pléiades 1A and 1B.

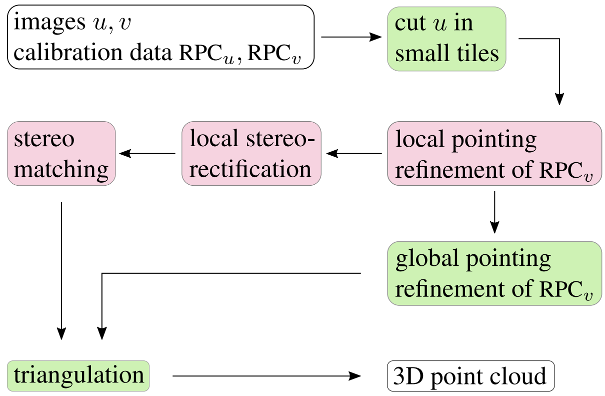


Figure 1: Stereo pipeline overview. The input is a pair of images with their respective rational polynomial camera models, and the output is a digital elevation model given as a georeferenced 3D point cloud. Green blocks are applied to the whole images, while pink blocks are applied on small independent tiles. They can be processed in parallel.

2. PIPELINE OVERVIEW

The philosophy of the proposed pipeline is to isolate the 3D reconstruction problem from the complexities associated to satellite imaging. To that aim the images are cut in small tiles. This permits to locally approximate the pushbroom geometry by an affine camera model, which in turn allows to stereo-rectify the tiles using standard computer vision tools (Hartley and Zisserman, 2004). The rectification error obtained on the tiles is below the tenth of pixel (de Franchis et al., 2014c). This allows to process the rectified tiles using off-the-shelf stereo matching algorithms.

The pipeline transparently deals with pointing inaccuracies, by estimating relative corrections for each tile without needing *ground control points* (de Franchis et al., 2014b) These local corrections are then combined in a global pointing correction for the entire image, which is used to perform a consistent 3D triangulation. The SRTM information is automatically incorporated to help identify corresponding regions in both images.

The proposed pipeline also handles tri-stereo datasets. In this case two stereo pairs are independently processed, the resulting elevation models are then merged to increase the coverage (see figure 2 for an example). This fully automatic pipeline is available online for testing (de Franchis et al., 2014a).

Figure 1 gives an overview of the processing pipeline for a stereo pair of images. The input images are cut in small tiles, to allow a very precise stereo-rectification. The optimal size of the tiles was shown to be 1000×1000 pixels (de Franchis et al., 2014c). Then for each tile the calibration data is refined (de Franchis et al., 2014b) and the images are stereo-rectified (section 2.1). Each stereo-rectified tile pair is matched using some standard stereo matching algorithm. The local refinements from all the processed tiles are combined to compute a global correction of the calibration. For each tile the triangulation uses the globally corrected calibration data, which is the same for all tiles. This ensures a perfect continuity between the 3D points computed from different tiles.

2.1. Local stereo-rectification

The s2p stereo pipeline uses the standard computer vision approach for stereo-rectification (Hartley and Zisserman, 2004), which consists in estimating a fundamental matrix and then extracting from it a pair of homographies to rectify the images. It can be summarized as follows: given a list of at least 4 correspondences $(\mathbf{x}_i, \mathbf{x}'_i)_{i=1,\ldots,N}$ between the two views, the affine fundamental matrix $\mathtt{F}$ is estimated using the Gold Standard algorithm (Hartley and Zisserman, 2004). Then two rectifying affine transformations are extracted from $\mathtt{F}$ using the Loop and Zhang algorithm (Loop and Zhang, 1999). The fundamental matrix estimation requires only image matches, eliminating the need for ground control points and manual intervention.

Algorithm 1: Locally affine stereo-rectification of pushbroom images.

Data: $\text{RPC}_1, \text{RPC}_2$: RPC's of input images;
$x, y, w, h \in \mathbf{R}$: ROI coordinates in image 1
Result: $\mathtt{H}_1, \mathtt{H}_2$: rectifying homographies

```
1 estimate altitude range ;              // from RPCs or SRTM
2 compute N virtual matches (x_i, x'_i);  // N ≥ 4
3 estimate F from (x_i, x'_i) ;          // Gold Standard
4 estimate H_1 and H_2 ;                 // Loop Zhang
```

The problem of finding correspondences. A natural way to compute correspondences between the two views is to extract feature points, compute descriptors and match them, as done by SIFT (Lowe, 2004). But this may lead to a set of keypoints all lying on the same plane, *i.e.* on the ground. This configuration is degenerate and $\mathtt{F}$ cannot be computed from it. We found that a safer way to estimate $\mathtt{F}$ is to use the calibration data (Oh et al., 2010; Tao and Hu, 2001) to generate virtual correspondences between the two views.

Virtual correspondences generation. Given a region Ω in the reference image and an estimated altitude range $[h_m, h_M]$ for the associated 3-space points (*i.e.* points that were imaged into Ω) Ω is back-projected on the Earth surface thanks to RPC^{-1}. Let denote by $\Gamma = \text{RPC}^{-1}(\Omega \times [h_m, h_M]) \subset \mathbf{R}^3$ the back-projected domain, and denote by $(\mathbf{X}_i)_{i=1,\ldots,N}$ a regular sampling of Γ. Each 3-space point $\mathbf{X}_i$ is projected on the two images using the associated RPCs, leading to a virtual correspondence $(\mathbf{x}_i, \mathbf{x}'_i)$. The images contents at locations $\mathbf{x}_i$ and $\mathbf{x}'_i$ may not correspond, but $\mathbf{x}'_i$ is located on the epipolar curve of $\mathbf{x}_i$, and that is enough to estimate a fundamental matrix.

The locally affine stereo-rectification algorithm is summarized in Algorithm 1.

Dataset name	Nb. of images	Images dimensions (pix)
ST-PAUL_ST-LEU	2	$39\,164 \times 102\,038$
ST-DENIS_MAIDO	2	$39\,869 \times 116\,069$
SAINT-PIERRE	3	$38\,345 \times 38\,532$
ST-BENOIT	2	$39\,982 \times 39\,486$
ST-PIERRE_Volcan	2	$40\,000 \times 42\,192$
VOLCAN	2	$40\,000 \times 38\,582$
ST-BENOIT_STE-ROSE	2	$31\,595 \times 38\,802$
VOLCAN-ST-PHILIPPE	2	$31\,012 \times 38\,366$
ST-BENOIT-ST-JOSEPH	2	$39\,977 \times 77\,597$

Table 1: Nine Pléiades stereo datasets were used to cover the whole Réunion island.

3. RESULTS AND DISCUSSION

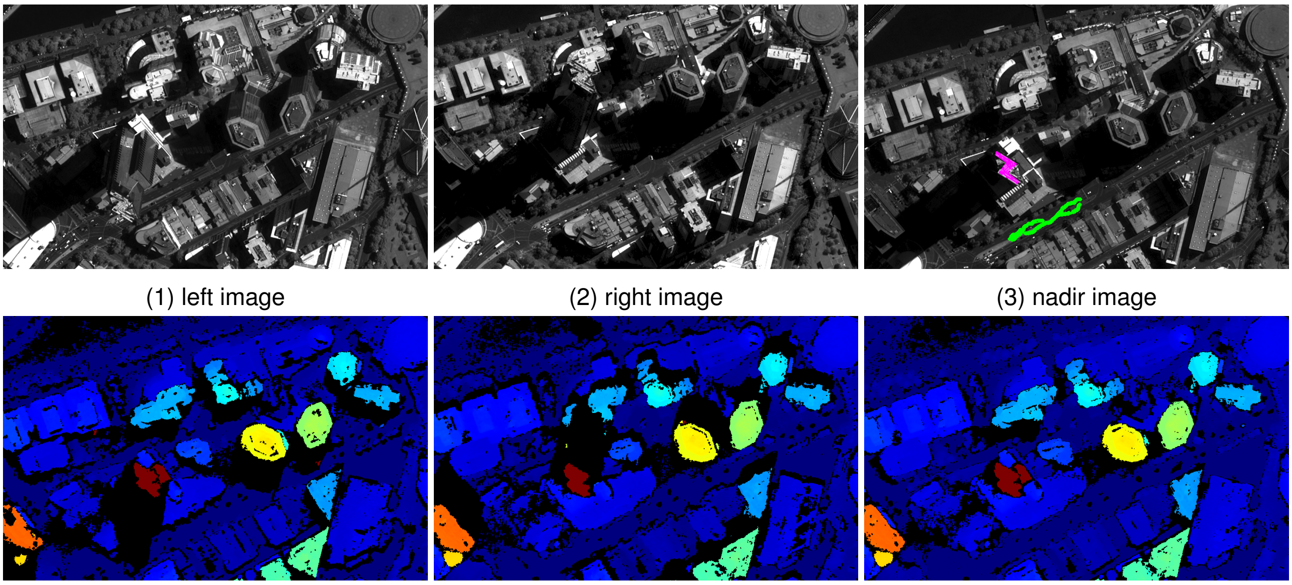
3.1. Melbourne tri-stereo

As a validation of the proposed pipeline we processed a region from a tri-stereo dataset of Melbourne. Our validation does not include ground control points. Thus we evaluated the relative precision by measuring the height of a known building. The Eureka Tower is a 297.3-metre skyscraper located in the Southbank precinct of Melbourne, which has been highlighted in figure 2 (1-3). The altitude estimates were computed by averaging the altitudes at the street level (yielding 16.15 ± 0.23 meters) and on the roof (yielding 312.97 ± 0.49 meters). Thus our estimated height of the Eureka Tower is 296.82 ± 0.72 meters.

Figure 2 (4-6) shows the elevation models obtained from the nadir-left and nadir-right pairs for the Melbourne dataset. Note that both images contain significant occluded regions in the vicinity of tall structures, however these regions are complementary. The fusion of both models exploits this complementarity to produce a denser elevation model.

3.2. Réunion island

As an illustration of the automatic power of the proposed method, the whole Réunion island was processed. As the island is larger than the swath width of Pléiades, several acquisitions were needed to cover the whole island. Table 1 lists the 9 stereo datasets that were used for the 3D reconstruction. The geographic positions of the images are shown on figure 3. The reference image of each dataset was cut into tiles of size 1000×1000, which led to a total number of approximately $20\,000$ tiles. All these tiles were processed automatically without any human intervention and without any parameter tuning. A subsampled version of the complete 3D point cloud is shown on figure 4, while a full resolution version of the "Piton de la fournaise" crater is shown on figure 5. Note that even if the coverage of the island is complete, some areas were hidden by clouds and consequently were not reconstructed by the algorithm. These areas are visible as blue holes in figure 4. The obtained digital elevation model was used by (Chen et al., 2014) to simulate landscape evolution and water run-off.



(1) left image (2) right image (3) nadir image

(4) DEM from nadir-left pair (5) DEM from nadir-right pair (6) fused DEM

Figure 2: Tri-stereo Pléiades images of Melbourne (1-3), the roof and street areas used for the evaluation are shown in (3). The elevation maps obtained by taking only two images are shown in (4) and (5), while (6) corresponds to the fusion with outlier filtering. Black areas represent rejected pixels.

Figure 3: Coverage of the Réunion island by Pléiades stereo images. Eight stereo pairs and one tri-stereo dataset were used. Each rectangle corresponds to a dataset. The larger image has dimensions of approximately $40\,000 \times 120\,000$ pixels.

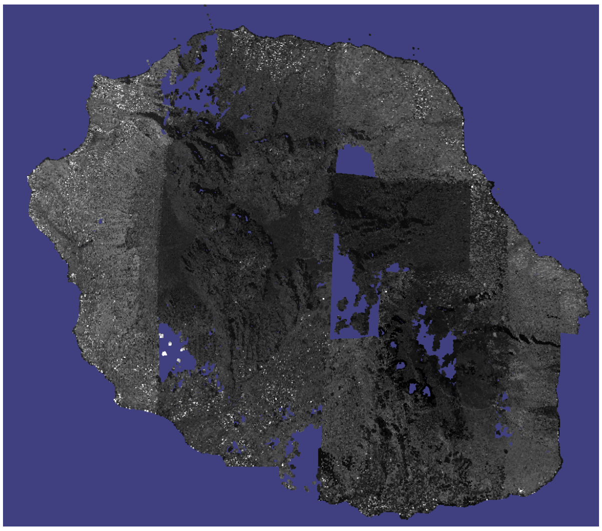


Figure 4: 3D reconstruction of the whole Réunion island obtained with s2p from 9 Pléaides stereo datasets. The points are colored with the original panchromatic channel of the respective reference images. The input images were downsampled 16 times.

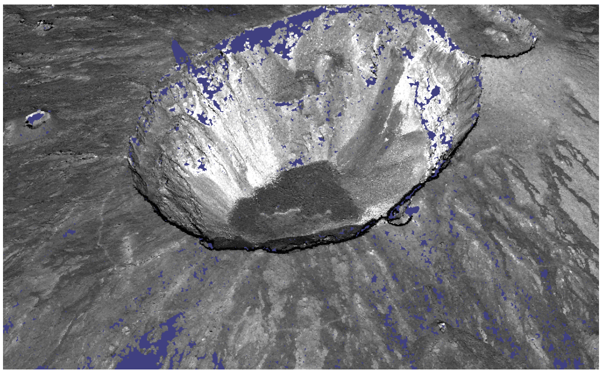


Figure 5: Crater of the "Piton the la fournaise", located on the eastern side of Réunion island. It was obtained with s2p from one Pléiades stereo pair at full resolution.

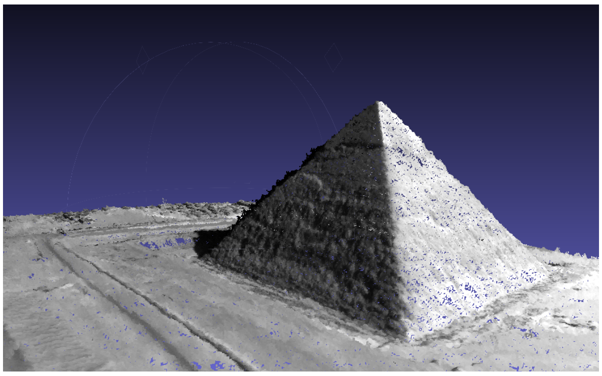


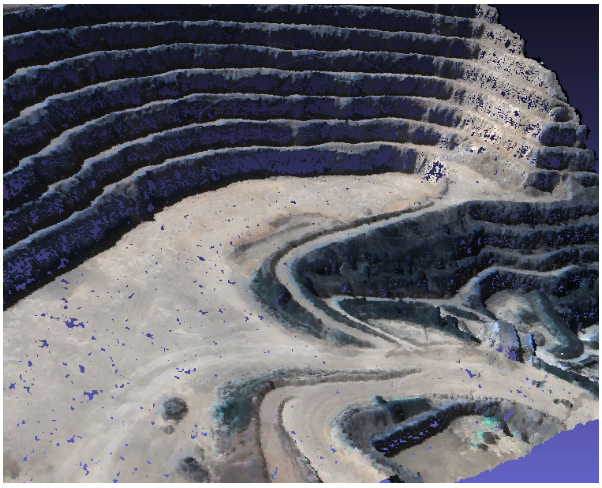
Figure 6: The Pyramid of Khafre, as computed by the proposed pipeline. To obtain this point cloud, the user clicked a single time on the appropriate place of the map.

4. CODE AND ONLINE DEMO

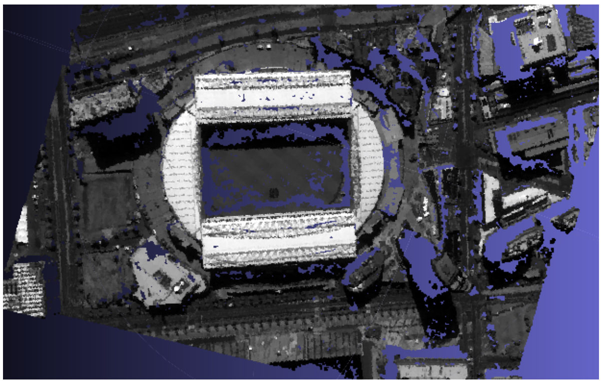
The stereo pipeline described here is completely implemented and will be released as open-source software. It can be tested online (de Franchis et al., 2014a) thanks to the demo framework of the IPOL journal (IPOL, 2010). Several stereo datasets from Pléiades and a stereo pair from WorldView-1 are available for testing. Users can also upload their own stereo datasets and run s2p on it. The implementation is compatible with all the stereo datasets provided by Airbus DS and DigitalGlobe, and thus could be tested on images from WorldView-2, QuickBird-1 and Spot-6. Figure 6 shows the summit of Mont Blanc. This point cloud was obtained after a single click on a Pléiades image. These and other reconstructions, as those shown on figure 7 can be performed online (de Franchis et al., 2014a).

ACKNOWLEDGEMENTS

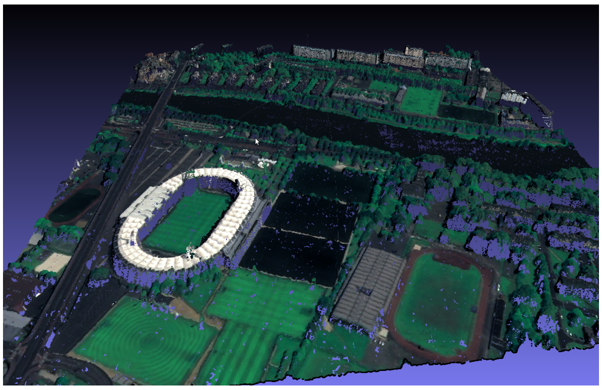
Work partially supported by Centre National d'Etudes Spatiales (MISS Project), European Research Council (Advanced Grant Twelve Labours), Office of Naval Research (under Grant N00014-97-1-0839), Direction Générale de l'Armement, Fondation Mathématique Jacques Hadamard and Agence Nationale de la Recherche (Stereo project).



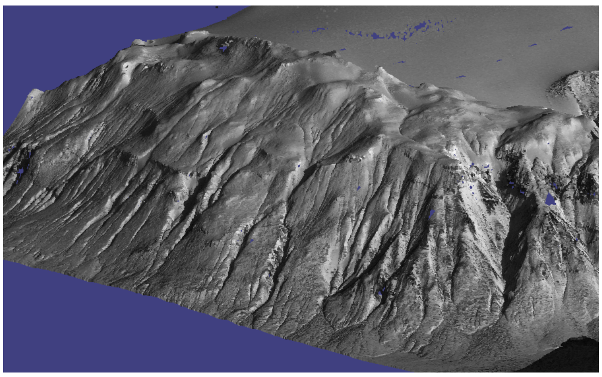
(1) Terrace in the Massif des Calanques (Marseille)



(2) Docklands Stadium (Melbourne)



(3) Stadium Municipal (Toulouse)



(4) Glacier in Iceland

Figure 7: 3D point clouds automatically generated from Pléiades tri-stereo datasets, without any manual intervention, with the s2p stereo pipeline. Its implementation can be tested online through a web browser.

References

Baltsavias, E., Stallmann, D., 1992. Metric information extraction from spot images and the role of polynomial mapping functions. Int. Arch. Photogramm. Remote Sens. 29, 358–364.

Chen, A., Darbon, J., de Franchis, C., Facciolo, G., Meinhardt, E., Michel, J., Morel, J.-M., 2014. Numerical simulation of landscape evolution and water run-off on digital elevation models obtained from pléiades. In: Revue Francaise de Photogrammétrie et de Télédétection.

d'Angelo, P., Kuschk, G., Jul. 2012. Dense multi-view stereo from satellite imagery. In: Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International. IEEE, pp. 6944–6947.

de Franchis, C., Facciolo, G., Meinhardt-Llopis, E., 2014a. s2p ipol demo. `http://dev.ipol.im/~carlo/s2p`.

de Franchis, C., Meinhardt-Llopis, E., Michel, J., Morel, J.-M., Facciolo, G., July 2014b. Automatic sensor orientation refinement of pléiades stereo images. In: Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International.

de Franchis, C., Meinhardt-Llopis, E., Michel, J., Morel, J.-M., Facciolo, G., Oct 2014c. On stereo-rectification of pushbroom images. In: Image Processing (ICIP), 2014 21th IEEE International Conference on.

Dowman, I., Dolloff, J., 2000. An evaluation of rational functions for photogrammetric restitution. Int. Arch. Photogramm. Remote Sens. 33, 254–266.

d'Angelo, P., Reinartz, P., 2012. Dsm based orientation of large stereo satellite image blocks. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci 39 (B1), 209–214.

EADS, A., October 2012. Pléiades Imagery User Guide V 2.0.

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., et al., 2007. The shuttle radar topography mission. Rev. Geophys. 45 (2).

Hartley, R., Zisserman, A., Apr. 2004. Multiple View Geometry in Computer Vision, 2nd Edition. Cambridge University Press.

IPOL, 2010. Image processing on line. `http://www.ipol.im`.

Kuschk, G., Feb. 2013. Large scale urban reconstruction from remote sensing imagery. In: 3D-ARCH 2013 - 3D Virtual Reconstruction and Visualization of Complex Architectures. Vol. XL-5/W1. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.

Loop, C., Zhang, Z., 1999. Computing rectifying homographies for stereo vision. In: Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Vol. 1. IEEE Comput. Soc, pp. 125–131.

Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. 60 (2), 91–110.

Oh, J., Hee Lee, W., Toth, C. K., Grejner-Brzezinska, D. A., Lee, C., Dec. 2010. A Piecewise Approach to Epipolar Resampling of Pushbroom Satellite Images Based on RPC. Photogramm. Eng. Remote Sens. 76 (12), 1353–1363.

Tao, C., Hu, Y., 2001. A comprehensive study of the rational function model for photogrammetric processing. Photogramm. Eng. Remote Sensing 67 (12), 1347–1357.

Wohlfeil, J., Hirschmüller, H., Piltz, B., Börner, A., Suppa, M., Jul. 2012. Fully automated generation of accurate digital surface models with sub-meter resolution from satellite imagery. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci XXXIX-B3, 75–80.