The rayCloud – A Vision Beyond the Point Cloud

Christoph STRECHA, Switzerland

Key words: Photogrammetry, Aerial triangulation, Multi-view stereo, 3D vectorisation, Bundle Block Adjustment

SUMMARY

Measuring structures in 3D based on Photogrammetry, i.e. purely based on images has become very important in the last years. Today, most of the available consumer maps (Google Maps, Bing Maps or Apple maps) are based on Photogrammetry and obtain their 3D structure by using stereo image correlation. Traditional photogrammetry software solutions provide a stereo display that allows users to extract 3D points. However, such displays use only two images. In this paper we introduce a novel 3D measurement interface that is based on multiple images, called rayCloud. We explain how this approach increases the accuracy of 3D measurements and demonstrate the variety of use for the typical photogrammetric workflow.

The rayCloud – A Vision Beyond the Point Cloud

Christoph STRECHA, Switzerland

1. INTRODUCTION

3D scene sensing can be achieved by active and more frequently also by passive, purely image based approaches. Active methods include time of flight laser scanning (LiDAR) and approaches based on structured light, as for instance used in Microsoft's Kinect sensor. While these approaches to reconstruct scenes in 3D work well, they have the disadvantage of requiring not only an explicit 3D sensor but also an imaging sensor to capture the visual appearance of the scene. Passive methods, on the contrary, rely purely on imagery and extract the 3D scene structure directly from images, based on the very same principle used by the human eyes to obtain 3D information.

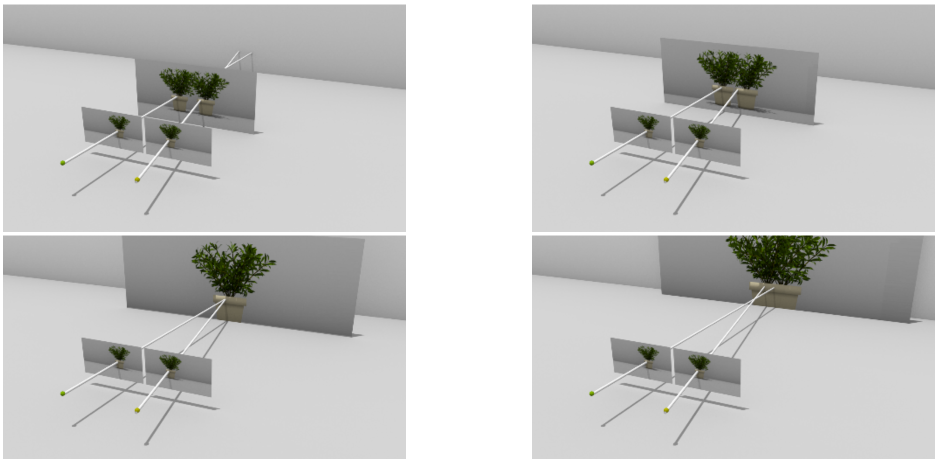

Figure 1 Stereo measurement by triangulation: *A scene is captured by two images. The knowledge of their respective camera centre allows to obtain a 3D point by finding the corresponding pixels and use them to intersect the two camera rays that go through these pixel locations (left-bottom). Similarly, stereo displays give the user the possibility to move a virtual plane such that the image projections onto that plane coincide.*

Traditional airborne photogrammetry requires a well-defined flight plan that is explicitly chosen to maximize the accuracy of stereo measurements. A manned aircraft will follow the flight plan to capture the images perfectly nadir, at the correct position and angle. Such a data acquisition results in high quality and high resolution images, including highly accurate data on their location and orientation that is acquired with GPS/GNNS and IMU.

If the images are captured by a light weight (<5kg) UAV, the philosophy of the whole acquisition process becomes very different. Due to its light weight character, the UAV will be much more sensitive to wind. Captured images might not be perfectly nadir and usually the position and orientation of the images is not known with high accuracy. In addition, images are sometimes blurred. To still achieve a high quality result with such imagery, the philosophy applied in this approach is to increase the redundancy of the data. More overlap, especially in flight direction, provides the needed base so that all images can be processed

automatically with high accuracy.
More redundancy requires also a different approach to measure 3D points. The traditional stereo measurements lack accuracy if an image pair with a relatively small baseline is used for measurements. The obvious extension is to move to a multi-view setting, i.e. use as many images as needed to insure accuracy. This approach is implemented in a user friendly way within the rayCloud.
Over the last decade, feature point descriptors such as SIFT [1] and similar methods [2] [3] have become indispensable tools in the Computer Vision and Photogrammetry communities to extract 3D structure from images only. They are extensively used to match images and compute relative camera poses. Given the relative camera pose, there have been a number of approaches [4] [5] [6] [7] [8] [9] developed to compute a dense 3D model. These approaches make it possible to extend traditional photogrammetry to terrestrial automated data processing. In such cases, measurements based on multi-view stereo give important additional advantages. Again, the images could potentially be rather small in baseline. In addition, the structure of the scene is more complex and one might need to measure 3D points based on different image pairs. The rayCloud concept provides important benefits also for this case.

2. 3D-POINT MEASUREMENTS FROM MULTIPLE VIEWS

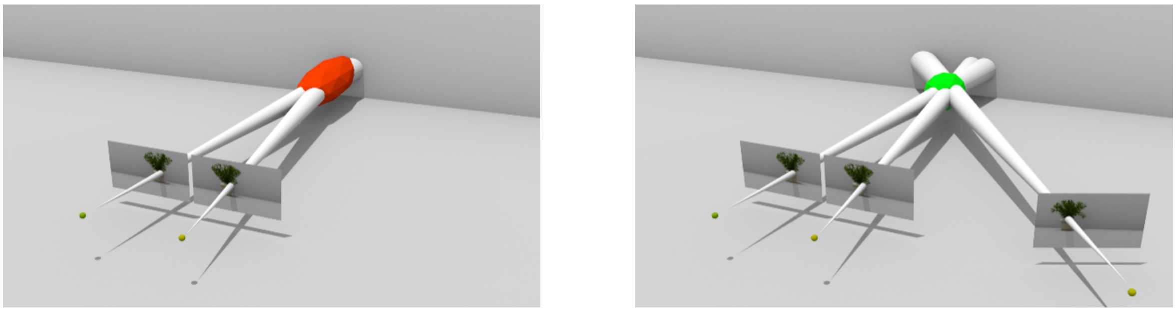

Figure 2 *Accuracy of stereo and multi-view stereo intersections.*

Figure 2: Schema of the accuracy improvement between traditional stereo measurements and the Pix4D's rayCloud concept. For a stereo setting, the accuracy in the viewing direction *dZ*is estimated as

$$dZ = dp\frac{Z^2}{fb}$$

where*Z* is the distance to the camera center, *f* is the focal length and *b* the baseline (distance between the two camera centers). The accuracy on the established correspondence (disparity accuracy) is *dp*. This accuracy can be achieved with the stereo system shown in **Figure 2**(left). The increase in accuracy for the multi-ray case is rooted in a larger baseline *b*, as shown in the drawing on the right.

Figure 3 and Figure 4 illustrate this situation in a screenshot of Pix4D's rayCloud editor. The middle image shows a 3D point cloud of a terrestrial reconstruction. The point cloud is the result of the automatic bundle adjustment that uses automatically matched, distinct keypoints through the image sequence. After clicking on a particular 3D point in the point cloud, the red rays show the 3D point connected to all camera centres where this point could potentially be visible. Zoomed patches around the projections in all images are show in the right part of the screenshot. The yellow circles indicate the measurement.

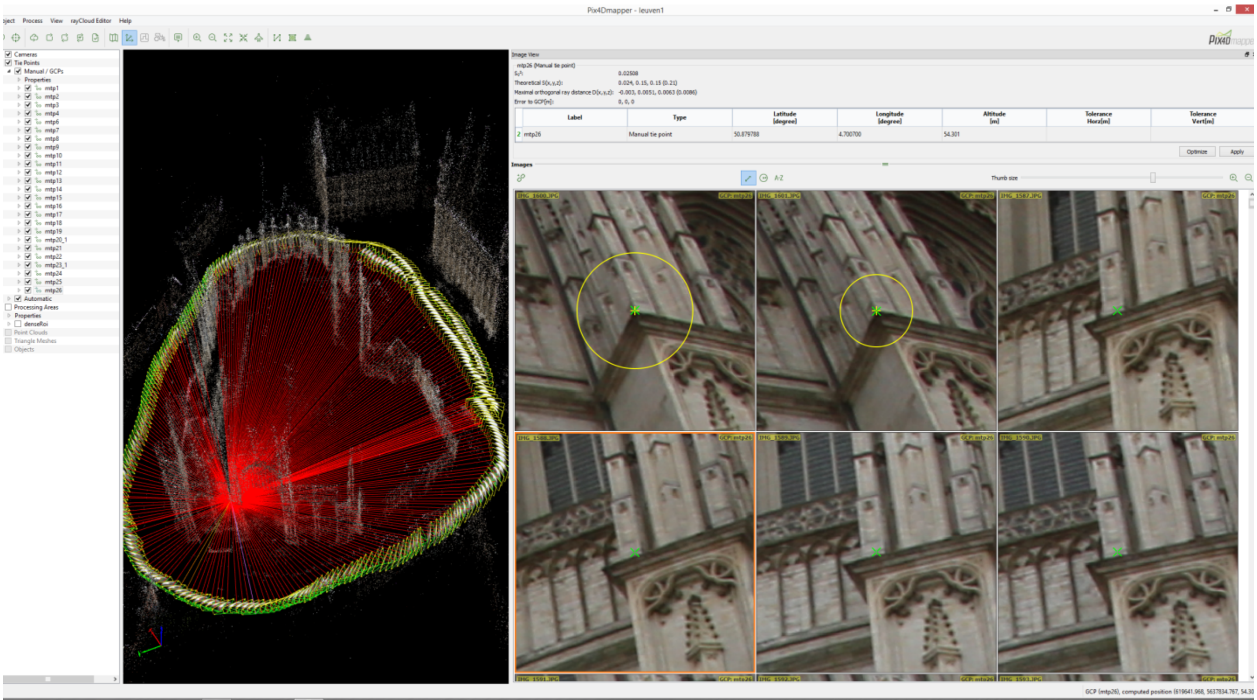

Figure 3 *Pix4D rayCloud: measuring point in stereo.*

Figure 3 shows the case of two view stereo: in two images a point was measured (yellow circle), these measurements are used to find the 3D point by intersecting the rays from the camera centres that go through the pixel measurements (yellow). The resulting optimal 3D point is then projected inside all images (green cross). In Figure 3 one can see that this re-projection (green cross) is actually off the corresponding point in the other images. This indicates a larger error for this 3D point. As previously explained, this is due to the relatively short baseline between the image centres.

The situation improves drastically if, in addition to the two measurements in Figure 3, an additional measurement is provided by the user (see Figure 4). Due to the larger baseline, the three measurements lead to a more accurate 3D intersection. It is clearly visible that now the resulting re-projected point (green cross) perfectly matches the other images.

3.

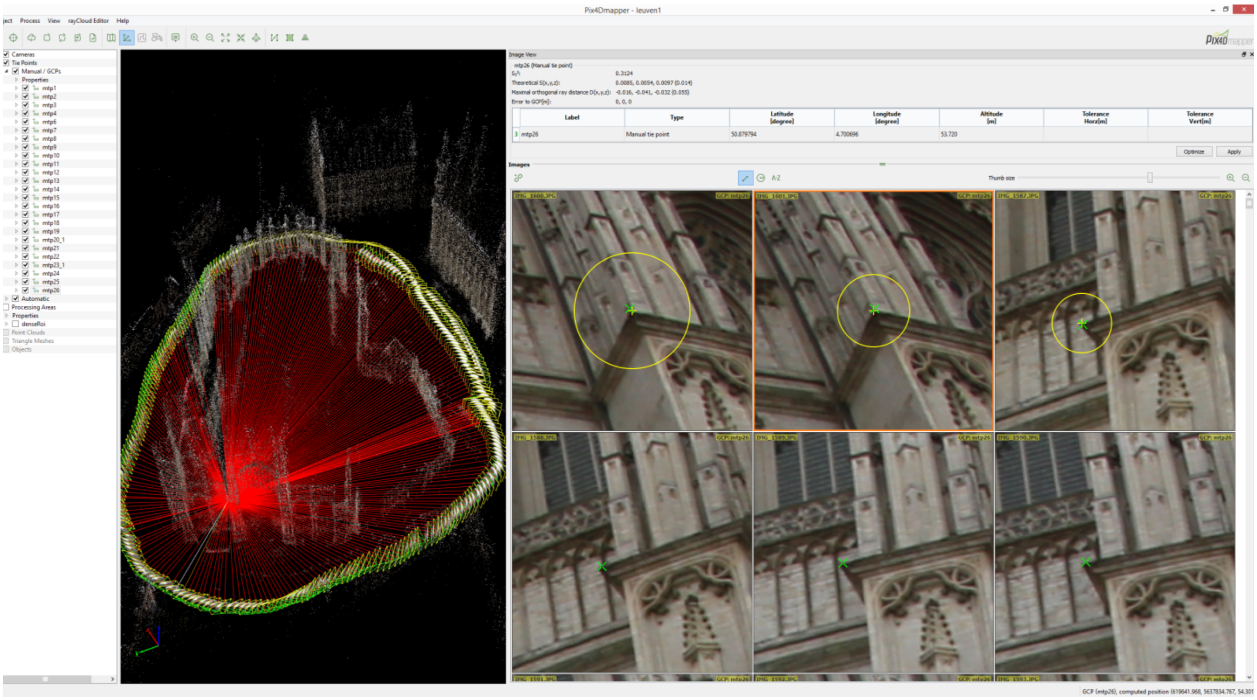

Figure 4 *Pix4D rayCloud: measuring points with multi-view stereo*

3. COMPARISON WITH LIDAR

In this section we compare a photogrammetry based survey from an UAV with the results of a terrestrial laser scan.
The histogram below shows a similar statistical analysis of the elevation differences between the DSM produced by Pix4D and the surface scanned with LIDAR. Some 2.7 Mio points were compared.

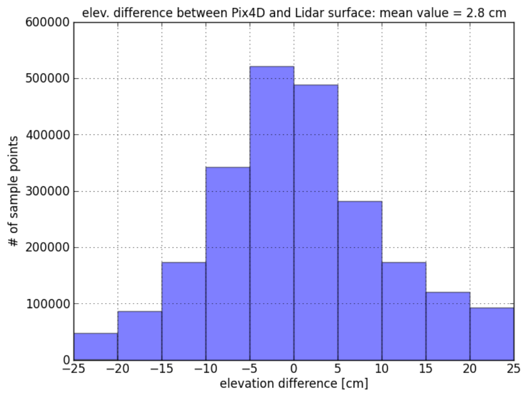

Figure 5 *Elevation difference histogram between LIDAR based surface points and Pix4D DSM surface.*

The mean deviation between the Pix4D surface and the LIDAR surface is about 3 cm. 2/3 of the test points lie within 2 GSD and ¾ of the test points are within 3 GSD, reaching the same level as best possible results theoretically achievable with any photogrammetry method, even when using lower quality UAV imagery. While for the considered application such an error is marginal, more precise results can be obtained using either a higher quality camera or reducing the flight altitude.

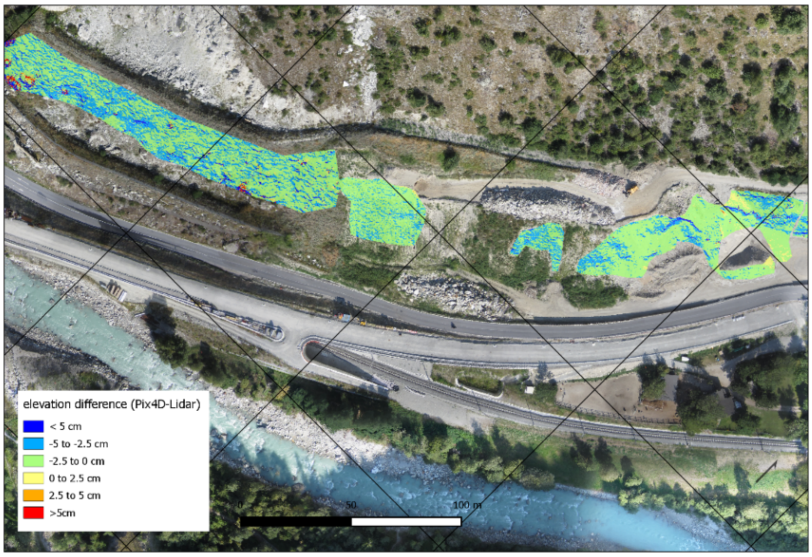

Figure 6*Top view with colour coded deviation betweenLIDAR survey zones and Pix4D DSM.*

From surfaces to volumes

For the ***relative volume comparison***, we designated for site A and B three distinct, small-scale stockpiles to calculate their volumes. Using a standard GIS software, the calculation was performed by subtracting the actual stockpile surfaces from a theoretical reference plane at the base of the stockpiles. The GNSS points were used to build a triangular irregular network (TIN) surface while the initially created 5cm DSM grids were used for the UAV photogrammetry and LIDAR data. A summary of the results are shown in the table below.

Site	**Pix4D** volume	**GNSS** volume	**LIDAR** volume	Difference **Pix4D – GNSS/LIDAR**
A - Central	16’591 m^3	16’238 m^3	-	+353 m^3 (+2%)
A - West	16’657 m^3	16’173 m^3	-	+484 m^3 (+3%)
B	138’635 m^3	-	138’831 m^3	-196 m3 (-0.1)
Total 3D surface area (A-Central + B-West + B):				

In addition to this relative volume comparison between different methods, we can also estimate an ***absolute error potential*** of the Pix4D volume. Comparing the Pix4D surface against the GNSS single points previously or the LIDAR surface, we know that the mean overall deviation lies between -2 to 4 cm. Taking into consideration the randomly selected sample locations and the sample number, we can assume that this range is not only valid for some points but the complete surface. This leads to the conclusion that we expect a maximal volume error between 440 and 1’330 m^3 for the total investigated stockpile areas, equalling to 0.3 to 0.8 % of the total volume. While this very simplistic conclusion is admittedly only valid for the specific stockpile scenario, it nevertheless indicates a rough estimation on the maximal error cap for our test case.

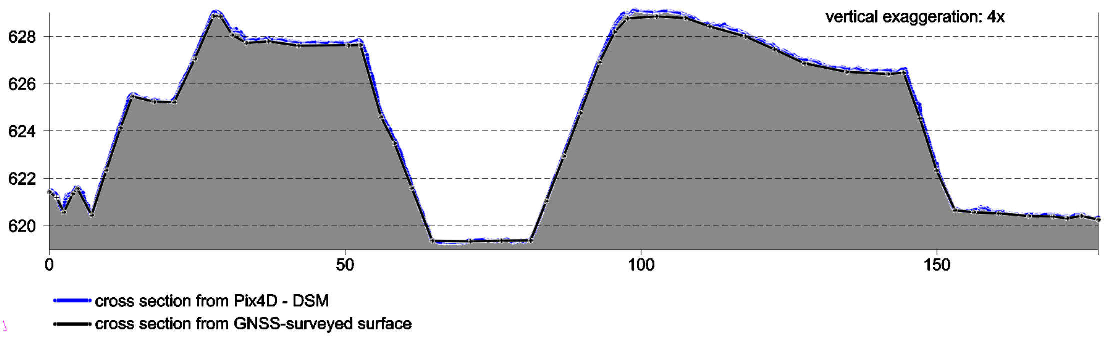

Figure *7 Comparison of stockpile cross sections between Pix4D DSM and GNSS surveying points.*

REFERENCES

[1] D.G.Lowe, „Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision,* Bd. 2, pp. 91-110, 2004.

[2] K. Mikolajczyk and C. Schmid, An Affine Invariant Interest Point Detector, ECCV, 2002.

[3] C. Strecha, A. M. Bronstein, M. M. Bronstein and P. Fua, „LDAHash: Improved Matching with Smaller Descriptors," *TPAMI,* Bd. 34, 2012.

[4] C. Strecha, T. Tuytelaars, L. Van Gool, „Dense Matching of Multiple Wide-baseline Views," 2003.

[5] H. Hirschmüller, „Stereo Processing by Semiglobal Matching and Mutual Information," *TPAMI,* pp. 328-341, 2008.

[6] S. Agarwal, Y. Furukawa, N. Snavely, B. Curless, S.M. Seitz, R. Szeliski, „Reconstructing Rome," *Computer vol 43, no 6,* p. 40.47, 2012.

[7] C. Strecha and W. von Hansen and L. Van Gool and P. Fua and U. Thoennessen, „On Benchmarking Camera Calibration and Multi-View Stereo for High Resolution Imagery," 2008.

[8] D. Scharstein and R. Szeliski, „A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms," Bd. 47, 2002.

BIOGRAPHICAL NOTES

Dr. Christoph Strecha received a PhD degree from the Catholic University of Leuven (Belgium) in 2008 under the supervision of Prof. Luc Van Gool for his thesis on multi-view stereo. He then worked as a post-doc and was co-chair of ISPRS Commission III/1. In 2011 he founded Pix4D, a Swiss company which develops and markets software for fully automatic production of survey grade 3D models and orthomosaics from UAV, aerial and terrestrial images.

CONTACTS

Dr. Christoph Strecha
Pix4D
EPFL Innovation Park – Building D
1015 Lausanne
SWITZERLAND
Tel. +41 21 552 05 90
Email: christoph.strecha@pix4d.com
Web site: www.pix4d.com