

CAMERA-PROJECTOR 3D SCANNING OF A SEMI-SUBMERGED TERRAIN IN A FLUME

Ben Gorte¹, Andrés Vargas Luna², Beril Sirmacek¹

Delft University of Technology
¹Geoscience & Remote Sensing
²Hydraulic Engineering
Stevinweg 1, 2629 HS Delft, Netherlands
b.g.h.gorte@tudelft.nl

KEY WORDS: Camera-projector system, structured light photogrammetry, 3d reconstruction, phase estimation

ABSTRACT:

The paper presents a home-made 3d scanner, consisting of off-the-shelf components: a camera and a projector. It is intended for monitoring dynamics of riverbed morphology observed under laboratory conditions in a flume, which is currently under construction. Special attention is paid to satisfying high requirements concerning accuracy and precision despite a compact and versatile setup of the system. Preliminary results are shown.

1 INTRODUCTION

Research conducted so far in the morphological response of rivers has focused on bed level changes, sediment adaptation and width-adjustment dynamics. Laboratory experiments have shown their capability for analyzing the processes that interact in this morphological evolution (e.g. Friedkin, 1945; van Dijk et al., 2012). Traditional measurements in laboratory have been normally performed by means of bed profilers. This procedure consists in the collection of successive points across selected transects of the channel, using a rod or a laser as point gauges. This kind of measurements typically implies a slow rate of data acquisition if the number of cross sections is high, further reducing the survey frequency.

Moreover, when infrared lasers are used it is necessary to collect the data when the bed is dry because of refraction phenomena, whereas a few attempts have been made for measuring the bed level evolution in laboratory when the water is present. Within these attempts we can mention the close range digital photogrammetry technique (CRDP) which has been applied in the survey of submerged surfaces not only in flumes, but also in the field (Chandler et. al., 2001, Butler et.al.,2002).

Considering the effects of refraction at the air/water interface, high-resolution digital surface models (DSMs) have been obtained by means of through-water close range digital photogrammetry. However, the use of control points within the channel is required which in turn perturbs the flow. More recently, Smith et al. (2012) have applied terrestrial laser scanning (TLS) in the acquisition of DSMs of underwater objects. The inclusion of laser beams with green wavelengths (i.e. around 532 nm) eliminates the complication of water absorption and the laser refraction when infrared beams penetrate the air/water interface, but increases the costs due to the high price of the equipment.

As a new approach, we introduce in this paper the use of the structured light technique for obtaining point clouds from the channel bed in a flume with water flow.

2 CAMERA-PROJECTOR 3D SCANNERS

A projector-camera 3d scanner is a photogrammetric stereo system where one of the two cameras is replaced by a projector, a

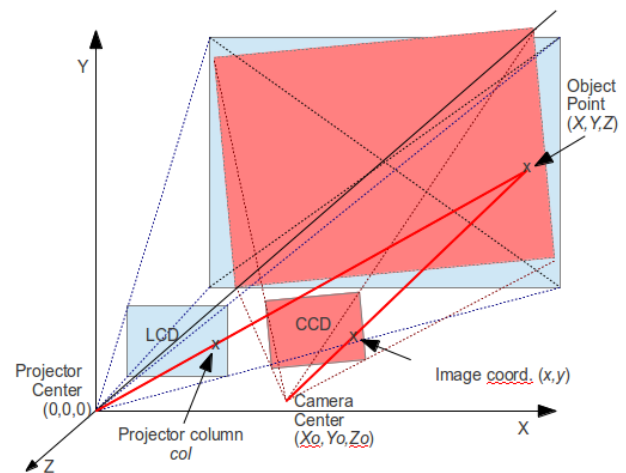


Figure 1: Projector-camera 3d scanner layout

device usually applied for Powerpoint presentations and in home cinemas. Instead of two there is only one sensor array in the remaining camera, and an LCD array generating an image in the projector. The challenge of finding correspondences is then reduced to creating patterns at known locations inside the projector array, and recognizing these in the camera image, whereby their positions in the camera image become known as well. By projecting and recording suitable, perhaps multiple, images, arbitrarily many correspondence can be obtained. The idea is by no means new, but is a variant of well-known structured light photogrammetry. Also the use of LCD projectors for this purpose has been demonstrated earlier (Lanman and Taubin, 2009, Sansoni and Realli, 2005). The current contribution is an effort to optimize the design of such a system in terms of convenience and practicality, as well as resolution and accuracy. We show the result in the application of monitoring a landscape in a flume, but we claim at the same time that we constructed a general-purpose, accurate, low cost 3d data acquisition system.

In the current paper we focus on the geometrical layout of the entire system, i.e. on the locations at which the camera and the projector are mounted w.r.t. the scene to be recorded. We will shortly address the resulting photogrammetric orientation and data processing procedures, but these can be considered standard pho-

togrammetric subjects. The main focus will be on obtaining optimal accuracy and precision within the chosen system geometry. The special requirements resulting from part of the scene being covered by water were our first motivation to start building a system “from scratch”. This gives us control over the entire system including the “core”, and allows us to build it up starting from the collinearity equations, where we will need to take the refraction at the air-water interface into account. These aspects are not yet covered in this paper, however.

3 CAMERA-PROJECTOR GEOMETRY

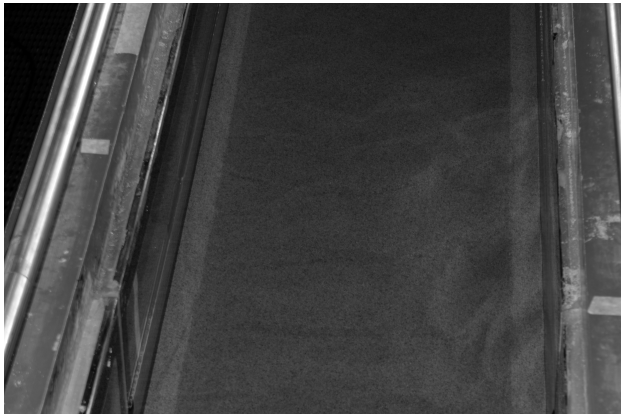


Figure 2: Flume used in the experiments

A fundamental choice in the design of a camera-projector system concerns the geometry, which can be *wide* or *narrow*. Within a wide geometry the distance between camera and projector is about the same as the expected depths to be measured, *i.e.* as the distances between the camera (or the projector) and the scene. A wide geometry roughly gives a *base to height* ratio between 0.5:1 and 1:1, which is generally considered optimal for accurate photogrammetric depth measurement. With a narrow geometry the camera and the projector are much nearer to each other, at a base of less than 5% of the maximum depth to be measured. The camera and projector are mounted side by side and their optical axes are approximately parallel. Not surprisingly this will lead to a lower depth accuracy. The paper investigates whether the narrow-geometry design is still a good choice, *i.e.* whether it allows to obtain certain pre-defined accuracy requirements. We will provide a theoretical argument for this; moreover, preliminary experiments were conducted in a tilting flume 0.40 m wide and 10 m long, with a bed composed by sand with a D50 of 0.5 mm at the Delft University of Technology (Fig. 2).

Monitoring a landscape partially covered by water in a flume imposes a particular limitation to the measurement geometry, in that we need to prevent the projector beam being reflected by the water straight into the camera. When using a wide geometry we, therefore, cannot have the projector and the camera at opposite sides of the flume, whereas with a narrow geometry the camera and projector cannot be mounted directly above the recorded flume area; instead, the system has to ‘look’ sideways, and therefore there will be varying distance between the scene and the camera/projector (beyond the height variations).

3.1 Trade-off

An advantage of a narrow geometry is that projector and camera can easily both cover the same area of the scene, provided that suitable lenses are chosen. Also when considering depth, more or less the same parts of the scene will be reached by both. It



Figure 3: Prototype camera-projector system

may happen that parts of the scene are not illuminated by the projector because of occlusion (shadow), but most of these are also not seen by the camera. In a wide geometry, in addition to shadowed areas (some of which may be seen by the camera) there may be illuminated, however unseen, areas.

In general the compactness of a narrow geometry system can be considered an advantage. As shown in Fig. 3 camera and projector are mounted rigidly with respect to each other and together form a mobile system, which after performing a relative orientation once, can be used at different locations without recalibration. With a wide geometry the camera and the projector will most likely be transported independently, requiring a new orientation procedure to be performed at every measurement location. Microsoft’s XBOX game controller Kinect, whose operation is based on recognizing its user’s body motion and is also widely being used for other 3d measurements, is probably the best-known example of a narrow geometry camera projector system.

As mentioned, at the downside of narrow geometry we find the sub-optimal base to height ratio, which inevitably leads to deterioration of depth accuracy and precision. This will be addressed in Section 5.

3.2 Relative Orientation

The general layout of the system is schematically shown in Fig 1, including the projection centers and the image areas of both projector and camera. The projector defines a local object coordinate system in meters, which has its origin at the projection center, X and Y axes parallel to the edges of the LCD array, and a Z -axis perpendicular to the array. After building up the system, a relative orientation procedure determines the location (X_0, Y_0, Z_0) and rotation $(\omega, \varphi, \kappa)$ of the camera w.r.t. the projector / the (X, Y, Z) -system (Fig. 3). After that, the system can

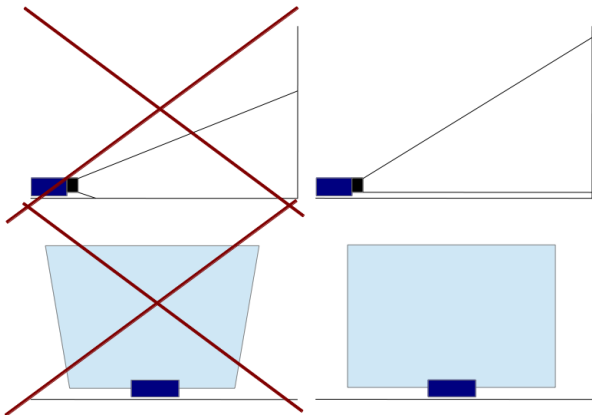


Figure 4: A projector beams “up” (top right) rather than horizontal (top left). Nevertheless the projected image is a rectangle (bottom right) rather than a trapezoid (bottom left). This is achieved by shifting the lens upwards w.r.t. the LCD array, giving a large value for vertical principal point offset.

be transported to different measurement locations without needing re-calibration. We implemented a self-calibrating procedure on the basis of a single checkerboard image projected on a flat surface by the projector and recorded by the camera, and a few measurements of sizes at the surface during the projection. This procedure yields focal lengths of camera f_c and projector f_p , as well as camera lens distortion parameters, in addition to the six relative orientation parameters mentioned before. Special attention is paid to the huge value of the projector’s vertical principal point offset, as illustrated in Fig. 4.

3.3 Data processing

The general principle of most camera-projector systems is that correspondences between pixel positions in the projector LCD and the camera sensor can be easily established, as projected patterns are recognized in the camera. We will describe this later in detail. We begin by assuming that at every camera image coordinate (x, y) the corresponding projector column col' is known - it differs from the usual column numbering by having the origin in the image center. We implicitly assume that correspondences are along epipolar lines, such that we do not need *row* positions as well, which would be less sensitive anyhow.

A camera at (X_0, Y_0, Z_0) transforms an object coordinate (X, Y, Z) into a camera coordinate $(x, y, -f_c)$:

$$\begin{pmatrix} x \\ y \\ -f_c \end{pmatrix} = sR^T \begin{pmatrix} Y - Y_0 \\ X - X_0 \\ Z - Z_0 \end{pmatrix}$$

Here s is an unknown factor and (X, Y, Z) can be anywhere along a line through the camera center and the image point. R implements the three orientation angles. We define an (X', Y', Z') system with the origin at the camera center, but having axes parallel to the projector’s. In that system the object point is at:

$$\begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix} = R \begin{pmatrix} x \\ y \\ -f_c \end{pmatrix}.$$

In the projector system the point would be along a line that connects the camera center and (X', Y', Z') :

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \lambda \begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix} + \begin{pmatrix} Y_0 \\ X_0 \\ Z_0 \end{pmatrix} \quad (1)$$

where λ specifies the (yet unknown) exact location on that line. Knowing the projector column col' we can state:

$$\frac{f_p}{col'} = \frac{X}{Z} = \frac{\lambda X' + X_0}{\lambda Z' + Z_0}$$

which can be re-written as:

$$\lambda(f_p X' - col' Z') = col' Z_0 - f_p X_0$$

and yields:

$$\lambda = \frac{col' Z_0 - f_p X_0}{f_p X' - col' Z'}$$

which can now be substituted into equation 1 to obtain a point in a 3d point cloud. substituted

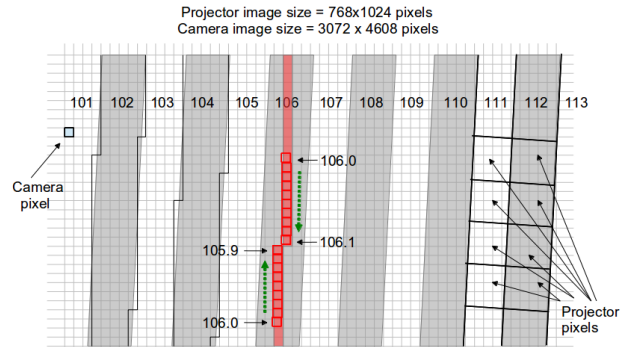


Figure 5: Spatial resolutions of camera and projector. The camera pixels indicated in red, at the centers of stripes in the projected patterns, are selected for further processing into the final point cloud. The corresponding projector column number is computed with sub-pixel precision.

4 DEPTH RESOLUTION IN A NARROW GEOMETRY SYSTEM

In a *wide* geometry system the precision of locations in object space is in the order of the precision at which location and parallax can be measured in the image, times the spatial resolution of the imagery. With a base-to-height ratio between 0.5:1 and 1:1 accuracies in depth and planimetry are about equal and proportional to spatial resolution, and therefore proportional to the distance between the camera(s) and the scene.

With a *narrow* geometry, assuming parallel optical axes, it is usual to speak about disparity rather than parallax. It is the difference between the locations where a point in the scene is represented in both images, expressed in pixels. The disparity x equals the distance B (for base) between the camera and the projector, divided by the pixel size $p(Z)$ at that distance Z (between the camera and the scene point), which in turn depends on the focal length f measured in sensor pixels:

$$x = \frac{B}{p(Z)} = \frac{Bf}{Z}.$$

A sensible measure for the depth resolution dZ at depth Z , where the disparity equals x pixels, is obtained by looking how far depth

has to increase before a disparity of $x - 1$ would occur. This links the measurement x and its sensitivity to the wanted parameter Z . Since

$$\frac{dZ}{dx} = -\frac{Bf}{x^2} = -\frac{Z^2}{Bf}$$

the depth resolution deteriorates with the squared of the distance (Khoshelham and Oude Elberink, 2012), which obviously restricts the maximum distance that a narrow geometry system can measure within given accuracy requirements. We will argue that this is tolerable, as the maximum distance is anyhow restricted by other factors, such as radiometric signal-to-noise ratio, and depth of field. However, certain refinements in disparity measurement are necessary, as follows.

Binary and Gray Encoding Computing disparity, i.e. the difference in column position between the camera image and the projector "image" of any object point, requires knowing the projector column number at a camera pixel. This is achieved by coding the projector column number in a series of projected images, having these recorded one by one by the camera, and analysing the recorded sequence at each camera pixel.

| binary code | decimal value | gray code | dec val | gray val |
|-------------|---------------|-----------|---------|----------|
| 0 0 0 0 0 | 0 | 0 0 0 0 0 | 0 | 0 |
| 0 0 0 0 1 | 1 | 0 0 0 0 1 | 1 | 1 |
| 0 0 0 1 0 | 2 | 0 0 0 1 1 | 3 | 2 |
| 0 0 0 1 1 | 3 | 0 0 0 1 0 | 2 | 3 |
| 0 0 1 0 0 | 4 | 0 0 1 1 0 | 6 | 4 |
| 0 0 1 0 1 | 5 | 0 0 1 1 1 | 7 | 5 |
| 0 0 1 1 0 | 6 | 0 0 1 0 1 | 5 | 6 |
| 0 0 1 1 1 | 7 | 0 0 1 0 0 | 4 | 7 |
| 0 1 0 0 0 | 8 | 0 1 1 0 0 | 12 | 8 |
| 0 1 0 0 1 | 9 | 0 1 1 0 1 | 13 | 9 |
| 0 1 0 1 0 | 10 | 0 1 1 1 1 | 15 | 10 |
| 0 1 0 1 1 | 11 | 0 1 1 1 0 | 14 | 11 |
| 0 1 1 0 0 | 12 | 0 1 0 1 0 | 10 | 12 |
| 0 1 1 0 1 | 13 | 0 1 0 1 1 | 11 | 13 |
| 0 1 1 1 0 | 14 | 0 1 0 0 1 | 9 | 14 |
| 0 1 1 1 1 | 15 | 0 1 0 0 0 | 8 | 15 |
| 1 0 0 0 0 | 16 | 1 1 0 0 0 | 24 | 16 |
| 1 0 0 0 1 | 17 | 1 1 0 0 1 | 25 | 17 |
| 1 0 0 1 0 | 18 | 1 1 0 1 1 | 27 | 18 |
| 1 0 0 1 1 | 19 | 1 1 0 1 0 | 26 | 19 |
| 1 0 1 0 0 | 20 | 1 1 1 1 0 | 30 | 20 |
| 1 0 1 0 1 | 21 | 1 1 1 1 1 | 31 | 21 |
| 1 0 1 1 0 | 22 | 1 1 1 0 1 | 29 | 22 |
| 1 0 1 1 1 | 23 | 1 1 1 0 0 | 28 | 23 |
| 1 1 0 0 0 | 24 | 1 0 1 0 0 | 20 | 24 |
| 1 1 0 0 1 | 25 | 1 0 1 0 1 | 21 | 25 |
| 1 1 0 1 0 | 26 | 1 0 1 1 1 | 23 | 26 |
| 1 1 0 1 1 | 27 | 1 0 1 1 0 | 22 | 27 |
| 1 1 1 0 0 | 28 | 1 0 0 1 0 | 18 | 28 |
| 1 1 1 0 1 | 29 | 1 0 0 1 1 | 19 | 29 |
| 1 1 1 1 0 | 30 | 1 0 0 0 1 | 17 | 30 |
| 1 1 1 1 1 | 31 | 1 0 0 0 0 | 16 | 31 |

Figure 6: Binary and Gray encoding schemes

The straightforward approach is that each image in the sequence is one bit plane in the binary encoding of column numbers. In a projector of 1024 columns, as in the presented example, the encoding of column numbers between 0 and 1023 requires 10 bits: a sequence of ten black-and-white images. The left side of Fig. 6 illustrates this for 5 (instead of 10) bitplanes encoding numbers between 0 and 31. The procedure is to generate the bitplanes in the projector one by one as black-and-white images: 0 = black, 1 = white. The ten projections onto the scene are recorded using the camera. The images are subsequently analysed. At each pixel it is first decided whether it is dark or light, by comparing it with an "average" image A providing the (location dependent) threshold, to be discussed below. This will tell whether a black (0) or a white (1) projection occurred at that spot, independently of local reflection and illumination conditions. The ten bits thus obtained at every camera image position will yield the number

of the projector column, whereby correspondence is established. Our camera has a much higher resolution (3072 x 4608 pixels) than our projector (768 x 1024). As we choose camera and projector positions and orientations, and focal lengths, in such a way that the scene coverages have maximum overlap, each projector column spans a narrow region, with a width of 4-5 pixels, in the camera image. When ignoring any effect of blurring, pixels that are completely 'inside' a stripe would be assigned projector column values reliably. Other camera pixels would be 'intersected' by the boundary between adjacent stripes. For those pixels, one should hope that the resulting bit pattern corresponds to either of the two values.

A refinement to support this is the Gray encoding scheme, which is illustrated at the right side of Fig. 6. It was termed 'reverted binary code' by its inventor Frank Gray in a patent application of 1947, and it found many applications in communication technology. The mapping from binary to Gray and back can be implemented via lookup tables. Gray encoding is considered advantageous because any two neighboring values have only one bit different, i.e. the Hamming distance between any two neighboring values equals 1 (Lanman and Taubin, 2009). Unfortunately, this does not imply that a wrongly-transmitted bit gives a unit difference in the transmitted value, as for any given value ten different bit errors may occur, and only two of these will yield a neighboring value; in the eight remaining cases (much) larger error will result. A perhaps more relevant advantage of Gray encoding for our application is that none of the bitplanes alternates like the least significant bit of binary encoding does (0101010101). Instead, the plane with the maximum spatial frequency is 0110011001. As depth of field of the projector is an issue, the maximum spatial frequency is limited by the modulation transfer function at the near and the far ends of the distance range. Note that the second bitplane of Gray encoding has the same frequency as the first, but is shifted one position (0011001100 resp. 0110011001). Fig. 7 shows the image of decoded projector column numbers in the camera imagery. Values increase from low (blue) at the left side of the image, to high (red) at the right side as expected. The transitions are supposed to be smooth, except at depth jumps in the scene; however, a few more irregularities seem to occur as well.

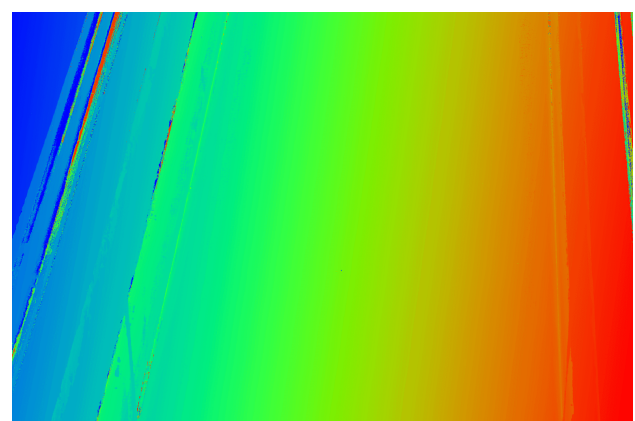


Figure 7: Projector column numbers after decoding the camera image sequence.

Subpixel estimation

In order to improve resolution at long distances we want to estimate disparity with sub-pixel precision. In the above description we already noticed that the camera resolution is higher than the projector's, and therefore the integer projector column number, as it appears from decoding the Gray patterns, yields a coarse

line pair) in the camera image is 9 pixels. The obtained phase difference α will be between $-\pi/2$ and $\pi/2$ and after scaling this to $[-0.5 .. 0.5]$ it can be used as a correction to the integer column numbers obtained in the Gray decoding step. Care has to be taken at the 'transitions' between neighboring columns: since the two processes (Gray decoding and phase estimation, respectively yielding the integer and fractional parts of the same value) are rather independent, there the results can be "out of sync": we might add 0.48, for example, whereas in fact 0.52 should be subtracted. Since the reliability of depth measurements is more important than their density (we have more camera pixels than the number of points we want), we only select pixels where the phase correction is small; these are pixels at the centers of the projected stripes (the red pixels in 5), where Gray decoding is most reliable.

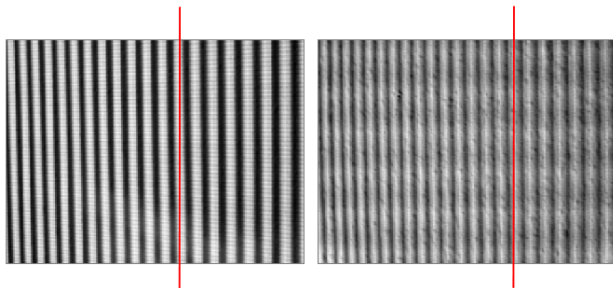


Figure 10: Two highest-frequency patterns of different quality due to depth-of-field influence.

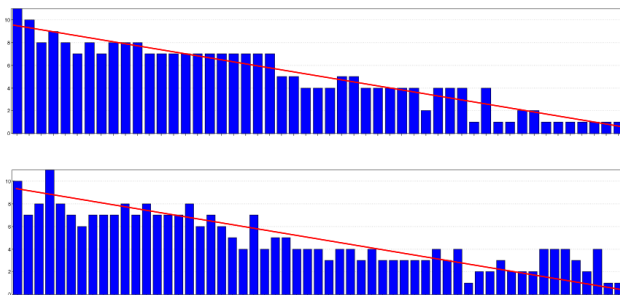


Figure 11: Phase estimate accuracy (see text).

5 RESULTS AND DISCUSSION

The result of phase estimation is shown in Figures 10 and 11. The first shows two sample high frequency patterns of different quality due to depth-of-field influences. When going from top to bottom along the red vertical lines phase is shifting very gradually (linearly) in both cases over a range of about one camera pixel. The resulting phase estimates are shown as blue bars in Fig. 11 (the upper and lower graphs correspond to the left and right image of Fig. 10 respectively). Also shown, by the red diagonal lines, are the correct values at all positions. The average difference (standard deviation) between the red lines and the blue bars is less than 0.15 camera pixel. Part of the reconstructed point cloud of the bed at the bottom of the flume (Fig. 2) is shown in Fig. 12. It shows a rather smooth result with little noise, in which the micro structure of the relief can be seen.

6 CONCLUSION

We started developing a camera-projector system that uses structured light photogrammetry to monitor the dynamics of a riverbed in a flume. The advantage of an in-house development is to have full control over various aspects, amongst which is the fact that

the riverbed will be partly submerged, whereby refraction at the water surface has to be considered. We constructed a versatile and compact system, and preliminary results demonstrate that the loss of accuracy caused by having a 'narrow' geometry can be counteracted by carefully aiming at sub-pixel disparity measurements.

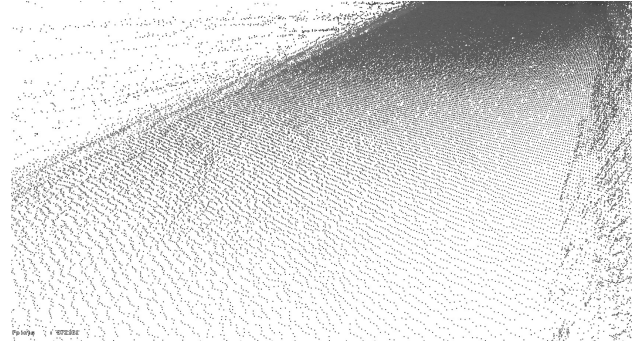


Figure 12: Reconstructed point cloud

REFERENCES

- Bräuer-Burchardt, Christian, Max Möller, Christoph Munkelt, Matthias Heinze, Peter Kühmstedt and Gunther Notni (2012). Determining Exact Point Correspondences in 3D Measurement Systems Using Fringe Projection – Concepts, Algorithms and Accuracy Determination, Applied Measurement Systems, Prof. Zahurul Haq (Ed.), ISBN: 978-953-51-0103-1, InTech, DOI: 10.5772/37591.
- Butler, J.; Lane, S.; Chandler, J. & Porfiri, E. (2002), 'Through-Water Close Range Digital Photogrammetry in Flume and Field Environments', *The Photogrammetric Record* 17(99), 419-439
- Büttgen, Bernhard, Thierry Oggier, Michael Lehmann, Rolf Kaufmann, Felix Lustenberger, CCD/CMOS Lock-In Pixel for Range Imaging: Challenges, Limitations and State-of-the-Art, In Proceedings of 1st Range Imaging Research Day, Zürich 2005.
- Chandler, J. H.; Shiono, K.; Rameshwaren, P. & Lane, S. N. (2001), 'Measuring Flume Surfaces for Hydraulics Research Using a Kodak DCS460', *The Photogrammetric Record* 17(97), 39-61.
- Dijk, W. M. v.; Lageweg, W. I. v. d. & Kleinhans, M. G. (2012), 'Experimental meandering river with chute cut-offs', *Journal of Geophysical Research* 117(F3), F03023, doi:10.1029/2011JF002314.
- Friedkin, J. (1945), 'A laboratory study of the meandering of alluvial rivers', Technical report, Vicksburg, MS: US Army Corps of Engineers, US Waterways Experiment Station.
- Khoshelham, K. and Oude Elberink, S.J. (2012) Accuracy and resolution of Kinect depth data for indoor mapping applications. In: *Sensors : journal on the science and technology of sensors and biosensors : open access*, 12 (2012)2 pp. 1437-1454.
- Lanman, Douglas, Gabriel Taubin, SIGGRAPH 2009 and SIGGRAPH Asia 2009 Courses, Build Your Own 3D Scanner: Optical Triangulation for Beginners.
- Oggier, Thierry, Image Sensor Technologies for 3D Time-of-flight Range Imaging. in Proc. of 2009 International Image Sensor Workshop, Bergen, NORWAY, June 22-28, 2009.
- Sansoni, Giovanna and Elisa Redaelli, A 3D vision system based on one-shot projection and phase demodulation for fast profilometry, *Meas. Sci. Technol.* 16 (2005) 1–10 doi:10.1088/0957-0233/16/0/000.