Digital stereo photogrammetry for grain-scale monitoring of fluvial surfaces: Error evaluation and workflow optimisation

Stéphane Bertin, Heide Friedrich, Patrice Delmas, Edwin Chan, Georgy Gimel'farb

Digital Stereo Photogrammetry for Grain-Scale Monitoring of Fluvial Surfaces: Error Evaluation and Workflow Optimisation

By Stephane BERTIN (sber081@aucklanduni.ac.nz), Heide FRIEDRICH (h.friedrich@auckland.ac.nz),
Department of Civil and Environmental Engineering
The University of Auckland

Patrice DELMAS (p.delmas@auckland.ac.nz), Edwin CHAN (ycha171@aucklanduni.ac.nz) and Georgy GIMEL'FARB (g.gimelfarb@auckland.ac.nz)
Department of Computer Science
The University of Auckland

Abstract

Grain-scale monitoring of fluvial morphology is important for the evaluation of river system dynamics. Significant progress in Remote Sensing and computer performance allows rapid high-resolution data acquisition, however, applications in fluvial environments remain challenging. Even in a controlled environment, such as a laboratory, the extensive acquisition workflow is prone to the propagation of errors in digital elevation models (DEMs). This is valid for both of the common surface recording techniques: digital stereo photogrammetry and terrestrial laser scanning (TLS). The optimisation of the acquisition process, an effective way to reduce the occurrence of errors, is generally limited by the use of commercial software. Therefore, the removal of evident blunders during post processing is regarded as standard practice, although this may introduce new errors. This paper presents a detailed evaluation of a digital stereo-photogrammetric workflow developed for fluvial hydraulic applications. The introduced workflow is user-friendly and can be adapted to various close-range measurements: imagery is acquired with two Nikon D5100 cameras and processed using non-proprietary "on-the-job" calibration and dense scanline-based stereo matching algorithms. Novel ground truth evaluation studies were designed to identify the DEM errors, which resulted from a combination of calibration errors, inaccurate image rectifications and stereo-matching errors. To ensure optimum DEM quality, we show that systematic DEM

errors must be minimised by ensuring a good distribution of control points throughout the image format during calibration. DEM quality is then largely dependent on the imagery utilised. We evaluated the open access multi-scale Retinex algorithm to facilitate the stereo matching, and quantified its influence on DEM quality. Occlusions, inherent to any roughness element, are still a major limiting factor to DEM accuracy. We show that a careful selection of the camera-to-object and baseline distance reduces errors in occluded areas and that realistic ground truths help to quantify those errors.

Keywords: Photogrammetry; digital; application; DEM/DTM; close range; high resolution; performance; accuracy

1. Introduction

1.1. Fluvial Morphology Remote Sensing

In situ characterisation of grain-scale fluvial morphology is challenging for hydraulic engineers and fluvial geomorphologists. In a riverine environment, the interactions between the sediment surface (characterised by the grain size distribution, the particle shapes and the arrangement of the particles) and the water flow significantly control the riverbed. Hence, an understanding of the river system dynamics, and associated habitats, relies on the ability to accurately describe the riverbed morphology.

Using 2.5D digital elevation models (DEMs) is increasingly becoming more common to represent the grain-scale surface morphology for gravel-bed rivers. This is enabled by the advent of new measurement techniques and improved PC performances. DEM analysis reveals the flow history (Mao et al., 2011, Ockelford and Haynes, 2013) and allows the parameterisation of the surface roughness for flow resistance equations (Smart et al., 2002, Aberle and Smart, 2003, Smith et al., 2011, Qin and Ng, 2012). DEMs are also essential for detailed computational fluid dynamics (CFD) simulations (Lane et al., 2002, Hardy, 2008, Hardy et al., 2009). In future, field collection of DEMs will help improve flood modelling by reducing the need to calibrate the surface roughness, a key parameter in flow simulations over rough surfaces.

Whilst both digital stereo photogrammetry and terrestrial laser scanning (TLS) support high-resolution data aquisition, measurement errors, which can affect data reliability and subsequent findings (Lane et al., 2005, Hodge et al., 2009), remain a major issue. For this reason, the majority of previously surveyed close-range fluvial environments are exposed

riverbeds in a controlled environment, such as the laboratory. At present, it is thus important to identify, quantify and reduce measurement error sources to improve the quality of grain-scale DEMs and allow using the techniques in more sophisticated experiments in future. For TLS applications, previous users scanned fluvial surfaces repeatedly to average out the errors (Hodge et al., 2009, Smith et al., 2012). However, there is still the need of significant post-processing in the form of filtering to obtain accurate metrics from exposed gravel beds in the field. Practical applications have shown that data post-processing should be avoided as it can introduce other errors (Hodge et al., 2009).

The most viable approach to minimise measurement errors is the optimisation of the data collection process. Workflow optimisation is of particular concern to stereo-photogrammetric users, as the DEM collection workflow is long, complex, and the source of various error types (Lane et al., 2000, Carbonneau et al., 2003, Bouratsis et al., 2013). However, improving data collection is generally limited by the use of commercial photogrammetric software.

To put the introduced workflow in context, we provide a review of previous hydraulic stereo-photogrammetric applications in the next Section. Attention is focused on the DEM collection workflow, the error sources and the solutions previously adopted.

1.2. Stereo Photogrammetry for Hydraulic Experiments

Stereo photogrammetry of fluvial environments covers a wide range of scales. DEM scales vary from several kilometres for the study of large braided rivers (Westaway et al., 2003), to several metres for mountainous streams (Bird et al., 2010) and to a mere metre for gravel-bed roughness characterisation (Butler et al., 2001, Bertin and Friedrich, 2014). The variety of stereo-photogrammetric applications is also reflected in the hardware and software/workflow selection for DEM reconstruction.

The advent of high-resolution digital cameras has led to the replacement of metric film cameras, allowing low-cost and versatile surveys (Chandler et al., 2001, Lane et al., 2001). Metric film cameras are provided with a calibration certificate that includes the parameters of the interior orientation (also called intrinsic parameters), although regular re-calibrations are recommended to ensure optimal accuracy (Cooper and Robson, 2001). Off-the-shelf digital cameras, as used in our experiments, require detailed calibration (see Section 3.4), if accurate metrics are to be extracted from imagery. In contrast, automatic stereo matching is now easier, which ensures an efficient DEM collection process and high data resolution. However, where a human operator previously ensured correct matching, automatic stereo matching now

Supprimé: When m

Supprimé: , o

relies on image quality, and a lack thereof can result in additional errors (Lane, 2000). Very recently, multi-view stereo (MVS) and structure-from-motion (SfM) photogrammetry started to be implemented in medium to large scale experiments (Westoby et al., 2012, Javernick et al., 2014). Even though these novel methods have the potential to better capture occlusions, they are not discussed here, as the workflow departs substantially from the more conventional binocular stereo (also called two-view) photogrammetry.

To date, most environmental and fluvial applications of stereo photogrammetry have relied on proprietary stereo-photogrammetric products. OrthoMAX module of Erdas Imagine® (later Leica photogrammetry suite, LPS, now IMAGINE Photogrammetry) is the most used commercial software, with AICON 3D Systems® a more recent product (Schmocker, 2011). Other commercial software, such as Trimble Inpho®, Intergraph ImageStation® and BAE Systems SOCET SET®, enable DEM reconstruction from stereo images, and have been tested for civil engineering and geomorphic purposes (González-Díez et al., 2014, Murillo-García et al., 2014, Stoter et al., 2015). As noted by Chandler et al. (2001), the use of proprietary software constrains the photogrammetric design. OrthoMAX requires conventional photogrammetric control targets to be placed in the region of interest. The control targets' 3D coordinates are recorded separately, using an independent device, and registered within the stereo model by bundle adjustment. This was shown to be a potential source of errors (Carbonneau et al., 2003). The tangential distortion is ignored in OrthoMAX for the calibration, which may be tolerable for high quality lenses and medium accuracy work (Fraser, 1997, Chandler et al., 2001). Furthermore, OrthoMAX's DEM reconstruction algorithm is limited, with substantial surface smoothing and poor results over rough surfaces and in occlusions (Chandler et al., 2001, Carbonneau et al., 2003), as it relies on area-based stereo matching. Substantial post-processing is needed when using those proprietary stereo-photogrammetric products for fluvial roughness studies (Carbonneau et al., 2003).

Attempts to optimise the DEM reconstruction process in OrthoMAX were first made by varying the DEM collection parameters, such as the minimum threshold of normalised cross-correlation and the template size for area-based stereo matching (Butler et al., 1998, Gooch et al., 1999). Butler et al. (1998, 2002) also changed the camera settings to obtain optimal exposures with maximum contrast; however, the image quality effect on stereo matching was not evaluated. As outlined by Aber et al. (2010), low image noise is expected to increase the DEM accuracy, which has been tested in specific application areas, such as stereomicroscopy. In Chandler et al. (2001), two different methods to obtain the calibration parameters were tried: (i) an in situ self-calibration with GAP software and (ii) an "on-the-

job" calibration with a 3D test field consisting of 70 retro-reflective targets, both resulting in similar DEM accuracy. Chandler et al. (2001) concluded that self-calibration is perhaps the preferable method, since it only requires the measurement of imagery used for the DEM extraction. However, the number and spatial arrangement of calibration control targets was found critical, with a need to have numerous (minimum of 15) and well-surveyed control targets evenly distributed throughout the x, y and z volume of the study site, which enables the recovery of reliable lens parameters (Chandler et al., 2001, Carbonneau et al., 2003).

Supprimé: repartition

More recently, stereo-photogrammetric solutions using non-proprietary algorithms are implemented in hydraulic experiments (e.g, Bouratsis et al. (2013) for the laboratory study of bridge pier scouring). The calibration parameters (including the tangential distortion) were obtained by using the freely accessible camera calibration toolbox for MATLAB® developed by Bouguet (2010). The stereo matching was performed on rectified images with a self-programmed correlation-based algorithm, using a window size of 35 x 35 pixels. Although a smooth surface was investigated, three geometrical filters were needed to process the data and remove blunders in a satisfactory manner. DEM errors were associated with the inadequate stereo setup design (baseline and the flying-height of cameras), which did result in substantial occlusions and thus stereo matching errors. Bouratsis et al. (2013) stressed the importance of image quality for stereo matching and suggested a structured light approach (whereby patterns are projected on the surface) to improve on the initially poor stereo matching results based on the riverbed texture only.

1.3. Paper Overview

This paper presents in detail a non-proprietary stereo-photogrammetric workflow developed for the grain-scale measurement of fluvial surfaces. The workflow is described and evaluated. Imagery acquired with two consumer-grade DSLRs and processed with non-proprietary algorithms form the basis of the DEM reconstruction. Details of the "on-the-job" MATLAB® calibration toolbox are presented and discussed. Dense scanline-based stereo matching is implemented and it is shown that it improves on the traditional area-based methods in terms of resolution and occlusion suitability. The presented stereo-photogrammetric technique was successfully applied to monitor the morphological changes of water-worked gravel beds, both in air and through-water (Bertin et al., 2013), with sub-millimetre sampling distances and vertical accuracies.

In this study, DEM errors resulting from a combination of calibration errors, inaccurate image rectifications and stereo-mismatches are identified by purposely-designed tests. A framework to evaluate the calibration and the image rectification accuracy is presented. Relevant objects of known geometry, ground truths, were designed, produced and tested for the quantitative DEM quality assessment, enabling the identification of the error propagation. For this paper, we concentrate on presenting the evaluation of the technique for in-air measurement, with the presented concepts being valid also for through-water measurements.

This evaluation study will be useful to other researchers using non-proprietary digital stereo photogrammetry and optimisation for their projects, and for the development of camera calibration and image matching techniques.

2. Measurement Environment and Instrumentation

The workflow is designed for experiments in a hydraulic flume (Bertin and Friedrich, 2014). The flume in use is 19 m long, 0.45 m wide and 0.5 m deep, with a slope of 0.5%. For hydraulic experiments, a gravel bed is prepared over a one metre long full-width sediment recess (called the "test section"), located 14 m downstream from the inlet (Figure 1). During the experiments, the sediment is water-worked, and the evolving topography is recorded either through-water, or in air after the flume is drained. The data used for this paper are from in-air tests.

Two Nikon D5100 cameras with 16.2 Mpixel complementary metal oxide semiconductor (CMOS) sensors (4928 x 3264 pixels) and Nikkor 20 mm fixed-focus lenses are used for stereo-photogrammetric measurements. The cameras are rigidly attached above the test section, using a gantry-mounting system sliding on a rail (Figure 1). Baseline and height can be adjusted depending on the design needs. Even lighting conditions are ensured (Bertin et al., 2014). An Alienware laptop with Intel core CPU @ 2.20GHz, 8GB memory and MATLAB® 2013b allows on-the-spot processing.

3. DEM Reconstruction Workflow with Stereo Photogrammetry

3.1. Stereo-photogrammetric Design

The stereo-photogrammetric design is the first step in a stereo-photogrammetric project. It defines the theoretical accuracy achievable by the setup, and is thus critical to the measuring performance (Lane et al., 2001). In this step, the camera placement is optimised

for the application. Three parameters are adjusted: (i) the baseline between the two cameras, (ii) the orientation of the cameras, and (iii) the distance between the cameras and the gravel-bed. Additionally, a margin can be accounted for around the measurement window and removed from the DEM during analysis, as it is well known that DEM errors increase near the edges (Butler et al., 2002, Bertin et al., 2014). Compromises often have to be made when finalising the stereo-photogrammetric design. For example, increasing the baseline for a fixed flying height improves the theoretical depth resolution, but also increases the risk of occlusions. Similarly, increasing the flying height to account for a margin reduces the theoretical resolution.

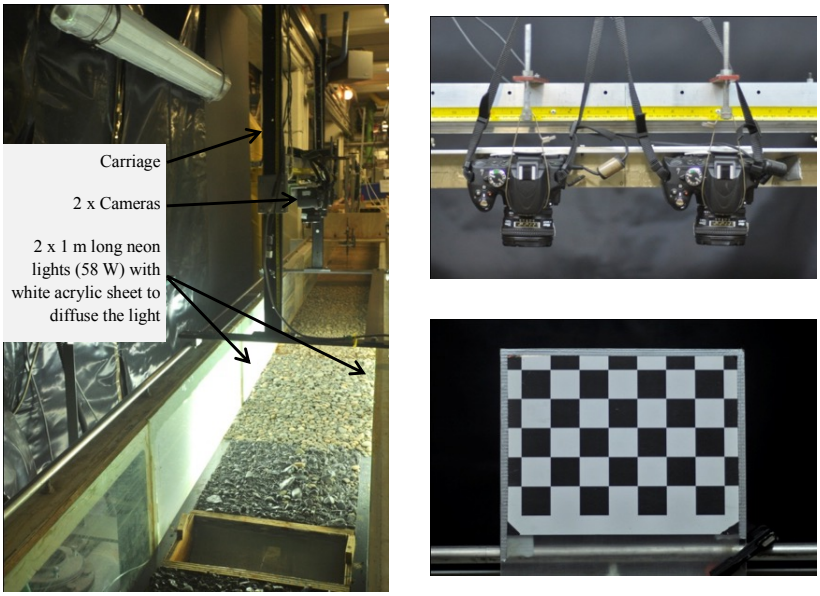


Figure 1. *Hydraulic flume with stereo photogrammetry installed for gravel-bed monitoring (main). The inserts on the right show how cameras are mounted onto the bar using cordons to secure the camera position (top), and the chequerboard with alternating black and white 30 mm squares used for the calibration and the evaluation of image rectification (bottom).*

To simplify image rectification and uniformly account for partial occlusions by gravel particles protruding and shadowing the surroundings, the camera pair in our experiments is mounted in canonical configuration, where the two optical axes are parallel with each other and perpendicular to the baseline (Figure 1). The standard central perspective projection

equations (pin-hole camera model) are used to determine the optimal flying height for the targeted measurement window size. Experimental heuristics are used to determine the optimal baseline. Not knowing the exact morphology of the studied area, such as encountered when recording a gravel bed, it is not possible to design the baseline in the same way as if one would know the occlusion structure.

The setup evaluated in this study was designed for a measurement window of size 450 x 450 mm, which is suited to study gravel-bed roughness (Mao et al., 2011, Ockelford and Haynes, 2013, Bertin and Friedrich, 2014). No margin was accounted for in this evaluation study, to enable the identification of the spatial distribution of stereo-photogrammetric errors. The baseline was set experimentally to 200 mm, requiring a distance between the cameras and the gravel-bed of 575 mm to cover the measurement window. The resulting common field of view (CFoV) between the two cameras, where 3D information can be extracted, is 478 x 450 mm in the object space, which results in a theoretical sampling distance of approximately 0.14 mm, which equals 1 pixel in the image domain. A theoretical depth resolution (distance between two disparity layers) of 0.39 mm can be achieved (Table I).

TABLE I. *Summary of the stereo-photogrammetric setup used (desired measurement window of size 450 x 450 mm). Computed values were determined using the pin-hole camera model and the rounded flying height.*

Rounded flying height (mm)	575
Baseline (mm)	200
CFoV (mm)	478 x 450
Range of disparity for 50 mm elevation range (pixels)	[1396-1523]
Overlap (%)	70
Pixel size (mm) / resolution (pixel/mm^2)	0.14 / 53
Sampling distance in DEM (mm)	0.25
Theoretical depth resolution or minimum measurable depth (mm)	0.39
Number of pixels in the 450 x 450 mm measurement window (-)	≃ 10,300,000

3.2. Setup Preparation

The cameras are connected to a computer, allowing remote control and live view using Nikon's Camera Control Pro 2 software, which is helpful to manually focus on the measured surface and to mechanically align the cameras. The latter is important to minimise the effect of image warping during image rectification to epipolar geometry. Easily identifiable feature points (e.g. chequerboard corners) are used to adjust the alignment until the corresponding

232 points are approximately on the same scanline. Once cameras are correctly focused and
233 mechanically aligned, it is necessary to ensure no modification is made to the physical setup
234 for the duration of the experiment.

235 3.3. Image Acquisition

236 Adequate selection of the camera settings for the application is important, as the crucial
237 element to a successful close-range stereo-photogrammetric process is attaining "good
238 images" (Matthews, 2008), where the term "good" refers to sharp images that have uniform
239 exposure with high contrast. Since the lighting environment constrains the cameras' settings,
240 it is important that both the lighting environment and the exposure settings are optimised
241 interdependently. To obtain the best quality images, the cameras are operated in manual
242 mode. We found that for our environment, using Nikkor 20 mm lenses, a combination of f/8
243 aperture (increased to f/10 or f/11 for selected tests) with a generic sensitivity value such as
244 ISO 200, ensures a good depth of field, a reduced vignette effect and uniform sharpness
245 across the image. Once the aperture and ISO are set, these settings should remain constant
246 throughout the acquisition of all images. For evaluation purposes we changed the shutter
247 speed, as presented later.

248 3.4. Calibration and Image Rectification

249 The calibration method of Zhang (1999), implemented in the calibration toolbox for
250 MATLAB® (Bouguet, 2010), is used to compute the camera calibration and pose parameters
251 (also called intrinsic and extrinsic parameters, respectively) from a series of stereo images
252 (called the "calibration images") of a planar chequerboard in different orientations (Figure 2).
253 Radial distortion (up to the fourth-order) and tangential distortion are also modelled during
254 calibration for each camera. For simplicity, "calibration parameters" hereafter refer to the
255 ensemble of the camera calibration parameters, the pose parameters and the distortion
256 coefficients. The chequerboard used in this study consists in alternating black and white 30
257 mm squares, prepared on a flat and rigid perspex plate (Figure 1), providing 40 (8 x 5)
258 control points (i.e. the chequerboard corners) per calibration image. A sufficiently accurate
259 checkerboard can be produced from off-the-shelf printers and other materials (Zhang, 1999).
260 Assuming a standard central perspective projection by the cameras, the calibration
261 toolbox extracts the chequerboard's corners, with distances between corners precisely known,
262 to compute a closed form solution for the calibration parameters and provide separately the

intrinsic and the extrinsic parameters of each camera, not including any lens distortion. The closed-form estimation of intrinsic parameters explicitly uses a constraint based on the orthogonality of vanishing points (Bouguet, 2010). Lens distortion is accounted for during the non-linear optimisation step, which minimises the total re-projection error (in the least squares sense) over all other calibration parameters, and solved with the Levenberg-Marquardt algorithm. The re-projection error (residuals between image measurements) is the difference in (x,y) pixel coordinates between the re-projected corners of the chequerboard, using the calibration results, and their measured locations with sub-pixel accurate corner detection. (x,y) represent the horizontal and vertical directions in an image, respectively.

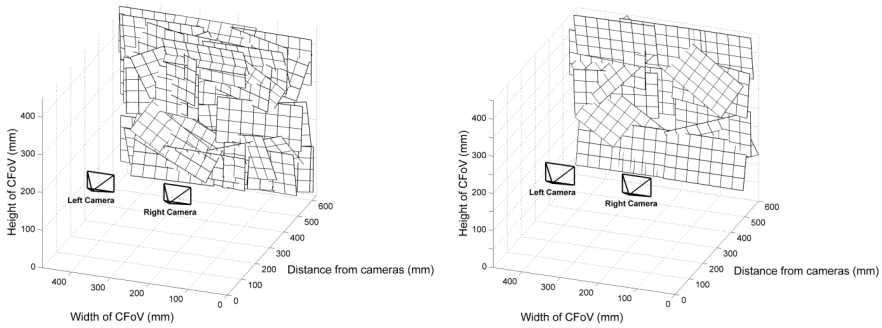


Figure 2. *Independent sets of (left) 40 "calibration images", the "parent" set in this research; and (right) 15 "rectification-check" images. The chequerboard corners are alternatively used as control points in calibration images, and as check points for image rectification in rectification-check images.*

A common procedure to refine the calibration data for each camera separately is to re-compute the chequerboard corners until a minimum re-projection error is attained (Bouguet, 2010, Bouratsis et al., 2013). This can be done manually for images where corner detection is initially poor, or automatically on all images. For the automatic re-computation, used thereafter, the toolbox re-computes the positions of every corner by using the re-projected grid as the initial guess locations for the corners.

Ultimately, during the stereo calibration step, all calibration parameters are re-evaluated based on the calibration image presenting the minimum re-projection error on both cameras to provide a single set of extrinsic parameters for the stereo setup (Bradley and Heidrich, 2010).

Once the calibration parameters are known, all stereo images obtained with the setup in the calibrated configuration can be rectified to epipolar geometry to minimise y-parallax in

the rectified images. This way, the 2D search of corresponding pixels is reduced to 1D, which assists stereo matching efficiency. Image rectification (Fusiello et al., 2000) is included in the calibration toolbox. It should be noted that the toolbox transforms images to greyscale equivalents during rectification. The code is herewith modified to obtain rectified images in red/green/blue (RGB) format.

3.5. Stereo Matching and DEM Reconstruction

Stereo matching is performed on rectified images using the conventional and well-established computer stereovision algorithm called symmetric dynamic programming stereo (SDPS) (Gimel'farb, 2002), providing both dense disparity maps and ortho-images of the region covered by the CFoV. The disparity map is then transformed into the depth map using projective geometry and the calibration parameters (the so-called depth triangulation).

For each conjugate pair of epipolar lines across a stereo pair, the SDPS algorithm exhausts all continuous epipolar profiles of a single continuous surface to find the profile minimising the total mismatch between both the lines. Individual mismatches between the corresponding pixels, which represent binocularly visible surface points, are measured by the total absolute or squared difference between the pixel intensities. The pixel intensity is defined as the maximum tonal value between the red, green and blue channels, expressed on an 8-bit (0-255) scale. Individual partial occlusions, such that the surface points are observed only in a single image, are taken into account by a special additive penalty. In other words, if a pixel has no counterpart in the other image because of partial occlusion, the latter is penalised in order to regularise the overall matching score, the depth at this point being computed based on the assumption of a single continuous surface.

The total line-to-line mismatch is a linear combination of the total intensity mismatch between the corresponding points and a weighted total occlusion penalty for the binocularly observed and partially occluded surface points along the profile, respectively. Compared with previous stereo-photogrammetric applications in hydraulic experiments (see Section 1.2), the SDPS performs pixel-to-pixel signal matching, rather than a “feature-based” or “area-based” one. Hence, the obtained DEM is dense, contrary to the feature-based matching; the DEM’s smoothing will be lesser than in the area-based matching, and the sampling distance can be chosen, if necessary, as small as the pixel size at the object’s distance.

In application to a gravel surface, the more conventional feature-based stereo matchers cannot produce a dense depth map and also may be inaccurate due to the absence of easily

detectable and discriminable visual features. The gravel surfaces are highly textured with relatively small occluded areas at borders, ensuring a sufficiently high accuracy of the dense SDPS matcher. However, the employed line-to-line stereo matching needs the more accurate image rectification, as otherwise true correspondences between the pixels may not be found.

Post-processing by median filtering, a common practice in stereo-photogrammetric surveys (Carbonneau et al., 2003), is implemented in the SDPS algorithm and is automatically applied to eliminate blunders in the depth maps. Using the default filter width of 3 pixels and height of 11 pixels, each given pixel elevation is replaced by the median value over a 3 x 11 neighbourhood around the corresponding pixel position in the input image. The filter size is such that it is able to smooth the horizontal streaks of disparity error that may result from scanline-based stereo matching. It should be noted that one can opt for this post-processing filter to be skipped and thus true pixel size resolution depth information can be obtained.

In terms of matching accuracy, the SDPS algorithm is able to distinguish disparity layers up to 1 pixel. Since the disparity at a distance Z from the cameras is equal to bf/Zp the setup's theoretical depth resolution dz in metric units (also called minimum measurable depth; Table I) can be expressed using the following formula (only valid in case of canonical configuration):

$$dz = \frac{Z^2 p}{bf - Zp} \quad [1]$$

where (b, f, p and Z) are the baseline, focal length, pixel size in the sensor plane and camera-to-object distance (in mm), respectively.

To reconstruct the DEM, all pixels in a depth map are triangulated using projective geometry and the calibration results. A point cloud, containing the (x,y,z) coordinates of all points in metric units, is obtained. The points in the point cloud are not regularly spaced, as they are assigned to different disparity levels. Using the function 'gridfit' in MATLAB®, the point cloud is finally interpolated (using the default triangle interpolation method) onto regular grids with adjustable sampling distance, and represented as 2.5D DEMs. Previous tests showed that the interpolation scheme to transform a point cloud into a regular DEM has very little impact on the data accuracy and that the selection of the sampling distance is critical (Bertin et al., 2014). DEM quality is best preserved when the raw point cloud is interpolated onto a grid with spacing close to the sampling distance of the point cloud (Hodge et al., 2009, Bertin et al., 2014). In this study, raw point clouds were interpolated onto

Supprimé: This can be related to

orthogonal grids with 0.25 mm sampling distance (Table I). Our available computer performance for subsequent analysis was the limiting factor for choosing this resolution.

4. Errors in Digital Stereo Photogrammetry

4.1. Identification and Quantification

An overview of error causes is provided in Table II. The paper focuses on errors that arise from the stereo-photogrammetric processing steps, i.e. the calibration, the image rectification and the stereo matching. Table II also summarises the identification methods and the implemented solutions to reduce or even eliminate the errors. The error sources and the identification methods are discussed in detail below.

4.1.1. Calibration Error

The calibration toolbox allows evaluating the quality of the calibration based on the numerical uncertainties of the estimated parameters and the re-projection error for each camera. The numerical uncertainties are approximately three times the standard deviation of the parameters after optimisation and are thus equivalent to the precisions of the parameters from the calibration. The toolbox provides the standard deviation of estimated re-projection error over all calibration images along the x and y directions, for both the left and right cameras (four parameters).

4.1.2. Rectification Error

For image rectification, the essential parameters to consider are the relative translation and rotation between the two cameras, as well as the calibrated principal distances and the positions of the principal points for each camera. Any calibration error on the above parameters will prevent a correct image rectification and will result in residual y-parallax in the rectified images, called rectification error. Although the two steps should be regarded as separate, the removal of image distortion is performed simultaneously with the image rectification.

To compute the rectification error, an independent set of calibration images, called the "rectification check" images (Figure 2), is recorded, providing check points (i.e. the chequerboard corners, 40 per images) covering the whole CFoV. Per definition, the rectification error is the absolute difference in vertical (y) pixel coordinates between the measured check points of the left-hand and right-hand rectification-check images after image rectification. For the quantitative evaluation of the calibration quality in this study, the mean

(unsigned) rectification error and the maximum (unsigned) rectification error were computed as the average and the maximum of the rectification errors over all check points in all rectification check images (i.e. 600 check points covering the CFoV), respectively. The maximum rectification error at the 99.7 % level was computed as the sum of the mean error and three times the standard deviation of error.

4.1.3. DEM Quality

Ground truths were essential to DEM evaluation to understand the links between calibration, image rectification, stereo matching and thus the resulting DEM error. The latter is evaluated by measuring a ground truth and comparing measured and "truth" elevations on a point-by-point basis, after mutual alignment. The mean unsigned error (MUE), the standard deviation of error (SDE), and the maximum unsigned error were used for the evaluation in this study. At several occasions, the maximum error at the 99.7 % level was computed as the sum of the MUE and three times the SDE.

Two different ground truths were used (Figures 3 and 4). Both of them lacked initially the texture and contrast of a real gravel-bed (Bertin et al., 2014). To compensate for this, the ground truth surfaces were finished with a light layer of paint, and the addition of differently coloured fine sand (diameter less than 0.2 mm, which is smaller than the 0.39 mm theoretical depth resolution achievable by the setup). For the assessment of the measured DEMs, manufacturing imprecisions and the effect of the surface finishing were neglected.

The first ground truth, hereafter referred to as the "gravel-bed model", was prepared with a 3D printer based on the DEM of a water-worked gravel bed as measured during a previous experiment. The ground truth elevations are produced with the manufacturer's specified accuracy range of 0.033 to 0.066 mm every 0.25 mm in both directions over a surface of size 296 x 184 mm (Figure 3). The preparation of the realistic ground truth and the method to automatically align measured and "truth" data before elevation differences can be computed on a point-by-point basis at more than 800,000 known locations is described in Bertin et al. (2014).

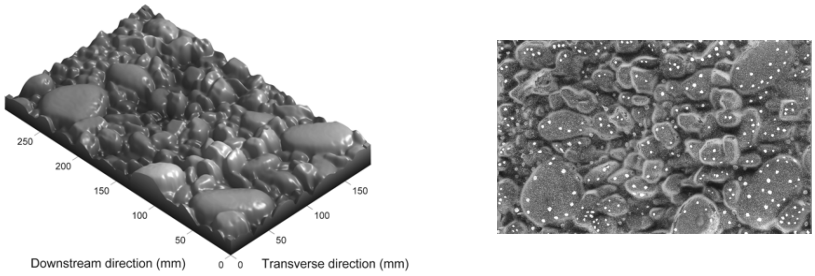


Figure 3. *"Truth" DEM of the gravel-bed model (left); greyscale ortho-image of the model after the surface was finished to maximise the stereo matching performance (right). Elevations are known at every 0.25 mm over 296 x 184 mm. DEM quality is evaluated by measuring the ground truth and comparing measured and "truth" elevations, after mutual alignment, at over 800,000 locations.*

A second ground truth, called the "hemisphere model" (Figure 4), the same size as our measurement window (450 x 450 mm), was prepared to evaluate the full spatial distribution of DEM errors. A digital model was prepared with MATLAB®, taking the form of a "truth" DEM, with elevations known at every 0.25 mm (Figure 4). To produce the actual model, an aluminium sheet (450 x 450 mm) was machined with a DOOSAN DNM 650 milling machine, 30 mm thick and perfectly flat. 121 adjoining cylinders (40 mm in diameter, 20 mm deep) were drilled onto the plate to allow ping pong balls (40 mm diameter) to be fitted half-way through. When reversed and laid down on a surface plate (flatness guaranteed with accuracy of 250 nm), no rocking due to uneven ping pong balls' protrusions was observed, satisfying the criteria of a precise ground truth.

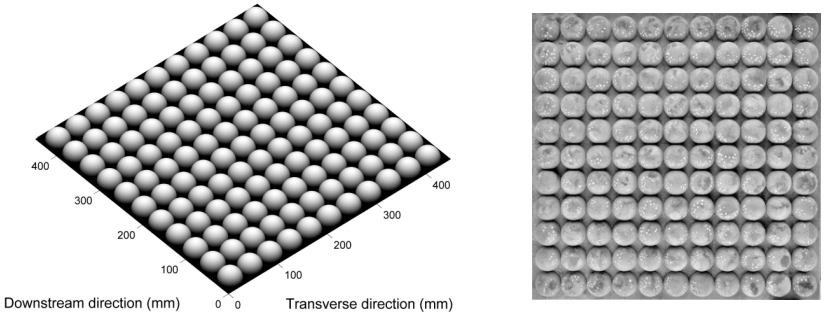


Figure 4. ***"****Truth" DEM of the hemisphere model (left); greyscale ortho-image of the model after the surface was finished to maximise the stereo matching performance (right). Elevations are known at*

426 *every 0.25 mm over 450 x 450 mm. DEM quality is evaluated by measuring the ground truth and*
427 *comparing measured and "truth" elevations, after mutual alignment, at over 3,200,000 locations.*

428 **4.1.4. Dome Effect**

429 The presence of systematic errors in DEMs obtained with stereo photogrammetry, taking
430 the form of a cubic surface centred on the photo base, called a dome, was noticed in previous
431 studies and explained by the incorrect removal of the radial distortion (Chandler et al., 2005,
432 Wackrow and Chandler, 2011). To detect the presence of domes in our data, a flat surface
433 (450 x 450 mm, 250 nm specified flatness) is measured with the setup. After alignment of the
434 obtained DEM with the photo base, the dome effect is estimated by fitting a bi-quadratic
435 surface to the detrended DEM by least-squares.

436 ***TABLE II.*** *A summary of the causes of stereo-photogrammetric errors evaluated in the paper, the identification methods and the solutions adopted.*

Causes	Description	Identification and quantification	Mitigation / *removal
Inaccurate calibration (IP/EP/distortion)	Conjugate points are not a same scan line after image rectification (rectification error or y-parallax), which may result in systematic matching errors	Compute rectification error Ground truth measurement	Optimise calibration strategy (number spatial arrangement of control points) Apply a margin during design *Median filtering
	Dome effect in the measured elevations due to the introduction of x-parallax in the stereo model	Measurement of a flat surface Ground truth measurement	Optimise calibration strategy (spatial arrangement of control points) Apply a margin during design *Bi-quadratic surface detrending if dor shape is known
Image quality and stereo matching	Gross errors due to incorrect matches	Internal reliability test using different imagery Ground truth measurement	Optimise image quality (change camer settings and/or lighting environment, u image processing techniques) *Median filtering

437 IP for intrinsic parameters. EP for extrinsic parameters.

4.2. Workflow Optimisation Methodology

Using the identification/quantification methods presented in Table II, we evaluated each DEM reconstruction step and quantified errors for each step and configuration. In test 1, the effect of the number of calibration images and the effect of automatically re-computing the corners to refine the calibration data were evaluated. The influence of the 3D spatial distribution (both horizontal and vertical) of the control points was evaluated in test 2. Test 3 was designed to provide a better understanding of what is image quality and how it influences stereo matching.

Unless otherwise mentioned, chequerboard calibration images were obtained inside the illuminated hydraulic flume, using manual focus, f/8 aperture setting, 1/50 s shutter speed, ISO 200 and JPEG (1:4) fine image format for both cameras. The chequerboard was held in position by supports underneath, to avoid any movement. Stereo images of the gravel-bed and hemisphere models were recorded in JPEG (1:4), using 1/20 s shutter speed, and transformed with Retinex. Detailed information on Retinex is given in Section 4.2.3.

4.2.1. Test 1: Calibration Quantity and Automatic Corner Re-computation

40 stereo images of the chequerboard were recorded, with the chequerboard in different positions and orientations, with care being taken to have control points (chequerboard corners) distributed throughout the CFoV (Figure 2). A maximum of 40 was chosen to verify the statement that 20 to 30 views are required (Bradley and Heidrich, 2010). Furthermore, performing a calibration using 40 calibration images still allows for a time-efficient process, which is tolerable for applications in both the laboratory and the field. Smaller subsets containing 10, 20 and 30 calibration images were created by randomly selecting images from the "parent" set of 40 calibration images. The chosen subsets ensured a decent CFoV chequerboard coverage. The calibration toolbox algorithms were successively applied for all calibration image sets individually, with and without automatic corner re-computation, resulting in eight different sets of calibration parameters, for which the stereo-photogrammetric errors were computed.

4.2.2. Test 2: Calibration Spatial Distribution

Two subsets, both containing 15 calibration images but covering different CFoV areas, were created from the "parent" set of 40 calibration images to enable the evaluation of the effect of the control points' planar distribution during calibration. The first subset comprised images presenting control points on one half of the CFoV only (Figure 9). The obtained

calibration data is hereafter referred to as the "partial" calibration. The second subset, called the "complete" calibration, presented control points distributed throughout the CFoV (Figure 9). Stereo-photogrammetric errors were computed for both calibration subsets. Based on the results of test 1, corner re-computation during calibration was not used in this test.

To investigate on the importance of the calibration distance (i.e. the distance between the cameras and the control points), the rectification error was measured at distances that encompass the calibration distance (Figure 12). Calibration was completed using control points distributed throughout the CFoV at distances from the cameras between 510 mm and 630 mm. The rectification error was measured at distances from the cameras between 430 mm and 700 mm. To avoid any bias due to different horizontal arrangements of the check points at the various distances, only check points in the central region of the CFoV were utilised. To enable this test, the image acquisition settings were changed to f/20 aperture and 1/20 s shutter speed.

Supprimé: repartitions

4.2.3. Test 3: Image Quality

Stereo images of the gravel-bed model were recorded with five different shutter speeds (1/5 s, 1/10 s, 1/20 s, 1/40 s and 1/60 s) in both RAW and JPEG (1:4) image formats. The other camera settings (f/8, ISO 200) remained identical to those used for the acquisition of the calibration images. We purposely obtained images with varied histograms (Figure 13). Before the stereo matching, RAW images were exported to JPEG using the default encoding parameters and the two lowest compression rates applicable (JPEG (1:2.6) and JPEG (1:15) for "highest quality" and "high quality" JPEG compressions, respectively). The stereo images directly recorded in JPEG (1:4) were subject to transformations, such as greyscale and Retinex. Retinex is a popular human-perception-based image processing technique, increasing the sharpness and compressing the dynamic range of images by reducing flares and enlightening shadows. The Retinex theory (Land, 1986) postulates that a given image can be decomposed into two different images, a reflectance image and an illumination image, on which transformations (e.g. normalising illumination and increasing the local contrast in the dark zones) can be operated to mimic direct scene viewing. Following the emergence of the Retinex theory, different Retinex algorithms were developed; the most widely used being the multi-scale Retinex with colour restoration (MSRCR) algorithm that combines dynamic range compression and tonal rendition (Jobson et al., 1997). Retinex methods have common advantages that they do not require training images and have low computational complexity. A possible defect is the colour greying of images, either globally or in regions where the

grey-world assumption is violated (Rahman and Woodell, 2004). Retinex was used for medical image enhancement (Meng et al., 2012), shadow detection and removal (Sun et al., 2008), illumination-invariant face recognition (Park and Kim, 2008) and point selection in 3D shape registration (Liu et al., 2014). To the best of our knowledge, Retinex was never employed to improve image quality before stereo matching. Using GIMP® (http://www.gimp.org/), the MSRCR algorithm was uniformly applied with the following parameters: 240 (scale), 3 (scale division) and 4 (dynamic). These parameters were chosen to provide the best DEM accuracy based on preliminary tests.

To enable a realistic comparison between obtained DEMs of the gravel-bed model, where image quality is the only difference, the same calibration data was used to rectify all images and to triangulate all depth information, leading to DEM reconstruction.

5. Results and Discussion

5.1. Test 1: Calibration Quantity and Corner Re-computation

Figure 5a shows the numerical uncertainties on the baseline after calibration with different numbers of calibration images, with and without automatic corner re-computation. The uncertainty/precision on the camera baseline was calculated as the square root of the sum of the squared (x,y,z) uncertainties given by the toolbox. Other parameters, such as the focal length (or principal distance), the principal point and the distortion coefficients are not shown, but presented a similar shape of uncertainties. The precisions of the calibrated parameters are reduced with increasing number of calibration images, with a substantial jump between 10 and 20 images (Figure 5a). Corner re-computation also reduced the uncertainties, resulting in around 20 % improvement for all data sets.

The re-projection error presented in Figure 5b is the average of the re-projection errors given by the toolbox in the horizontal (x) and vertical (y) directions in an image, averaged between the left and the right camera. It represents the overall precision with which the calibration is able to re-project the chequerboard corners on the calibration images. Figure 5b shows that the re-projection error is significantly reduced (about 3-fold) by automatically re-computing the corners during calibration.

The rectification error (Figure 5c) is reduced with increasing calibration images. A minimum rectification error was obtained with 30 images, after which the rectification error slightly increased again. Although re-computing the corners during calibration improved the precisions of the calibration parameters (Figure 5a), it did not improve image rectification.

This suggests that the automatic corner re-computation had minimal effect on the calibration parameters, and thus on image rectification. For all calibrations, the mean and the standard deviation of rectification error are small (about 0.2 pixel and smaller), indicating that most conjugate points are on the same scanline after image rectification, ensuring minimal systematic matching error throughout the CFoV.

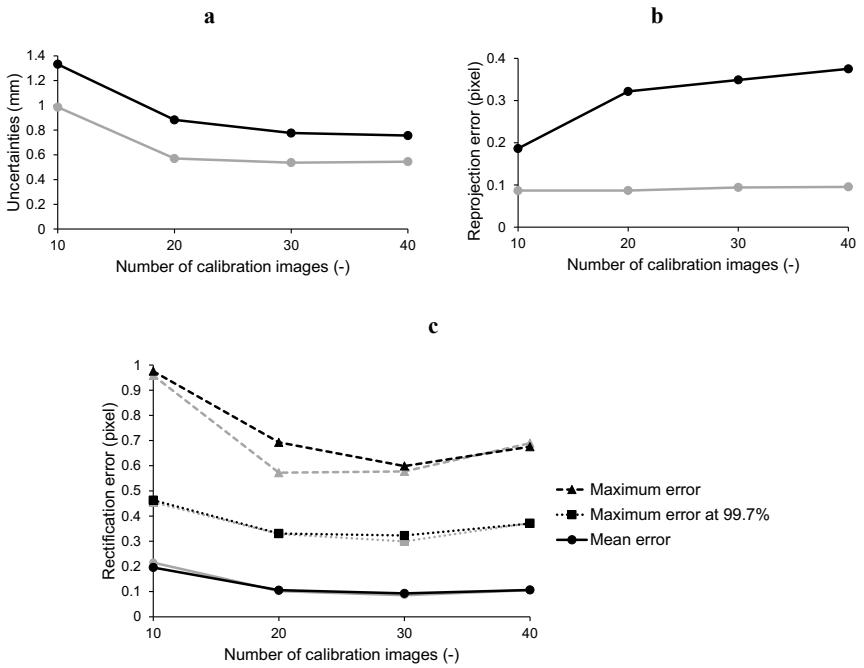


Figure 5. *Influence of the automatic corner re-computation and the number of calibration images on (a) the numerical uncertainties on the calibrated baseline; (b) the re-projection error; and (c) the rectification error. Calibrations with corner re-computation (grey), calibrations without corner re-computation (black).*

Figure 6 presents the DEMs of the gravel-bed and hemisphere models obtained with the "parent" calibration (40 images) and their associated DEMs of difference (DoDs) after comparison of the DEMs with the "truth" elevations. Figure 6 shows clearly that most DEM errors occur at the edges and troughs, due to occlusions. Edges and troughs also cause poor image correspondences because of shadows and uniform colouring of the surface (i.e. low-textured areas). For occlusions, the at-a-point elevation is based on the assumption of a

continuous surface, generally resulting in over-smoothed deep holes. The DoD of the hemisphere model shows that DEM errors increase both in frequency and magnitude with the radial distance from the centre of the CFoV because of occlusions, since oblique rays are less able to reach areas shadowed by higher grounds. This is not observed in the DEM of the gravel-bed model, which can partly be attributed to its smaller size. Additionally, the decision to set the baseline parallel to the downstream direction resulted in errors preferentially aligned with the transverse direction. When recording surfaces similar to the hemisphere model, increasing the setup flying-height would likely reduce the occlusions and improve the DEM quality. A poorer resolution would result as a consequence, as mentioned previously.

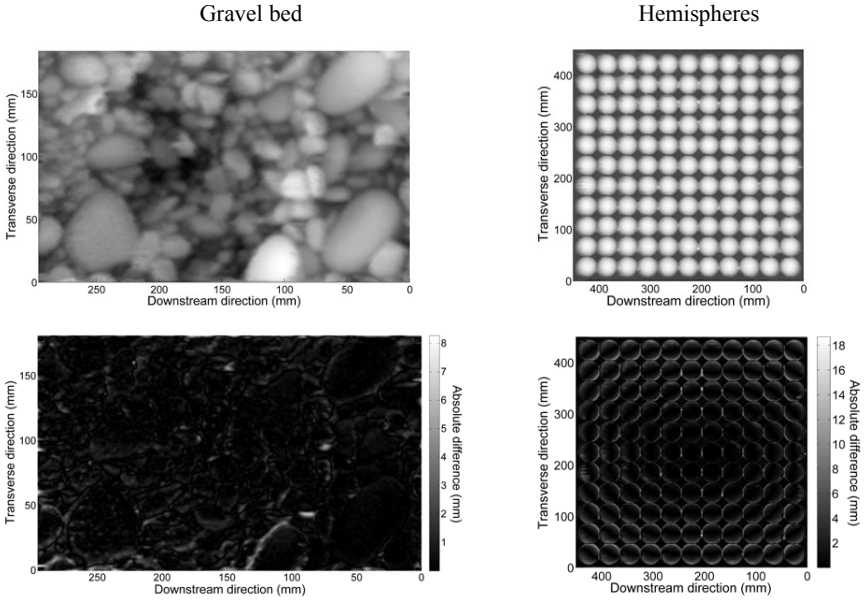


Figure 6. *(Top) measured DEMs, aligned with the truth DEMs; and (bottom) associated DoDs (absolute differences are shown for clarity). Calibration was performed with the "parent" set of images without corner re-computation; stereo matching was done on JPEG (1:4) images obtained with 1/20 s and transformed with Retinex.*

Figure 7 shows the quantitative evaluation of the DEM quality, computed over the gravel-bed and hemisphere models, for the eight calibrations tested. All gravel-bed model DEMs have optimal accuracy with MUE equal to 0.51 ± 0.03 mm and a maximum absolute error of 8.90 ± 0.7 mm. In Bertin et al. (2014), a MUE of 0.43 mm and a maximum absolute

error of 8.16 mm was obtained with a baseline distance of 250 mm and a flying-height of 636 mm, suggesting that the DEM errors can be reduced by changing the stereo-photogrammetric design. As Figure 6 suggested, the DEMs of the hemisphere model are of lesser quality than the DEMs of the gravel-bed model, with an overall three-fold increase in the MUE and two-fold increase in the maximum absolute error. Figure 7 shows that the MUE and the maximum error at the 99.7 % level tend to decrease with the number of calibration images for all DEMs. This may be explained by the joint reduction of the uncertainties on the calibration parameters and the rectification error with the number of calibration images, which was observed in Figure 5a and Figure 5c. Similarly to previous observations (Figure 5c), the effect of the corner re-computation on the DEM quality is small, and no clear trend is observable. We nevertheless believe that corner re-computation is useful when chequerboard image quality prevents accurate corner detection, e.g. in the case of through-water photogrammetry.

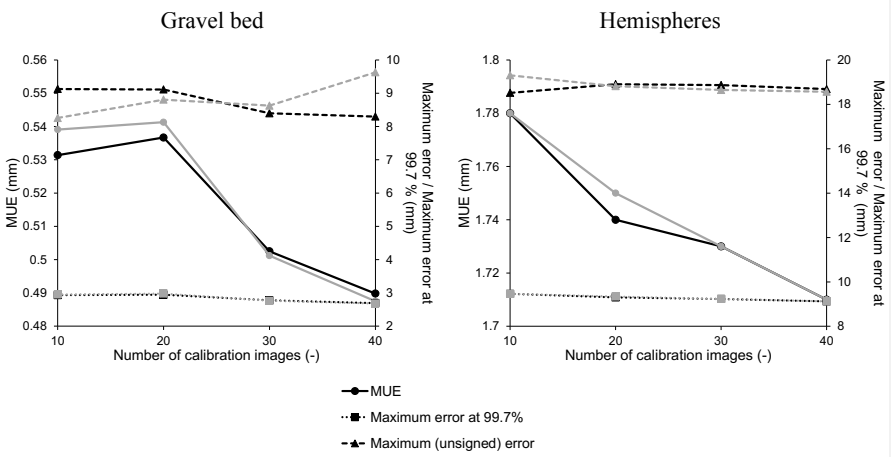
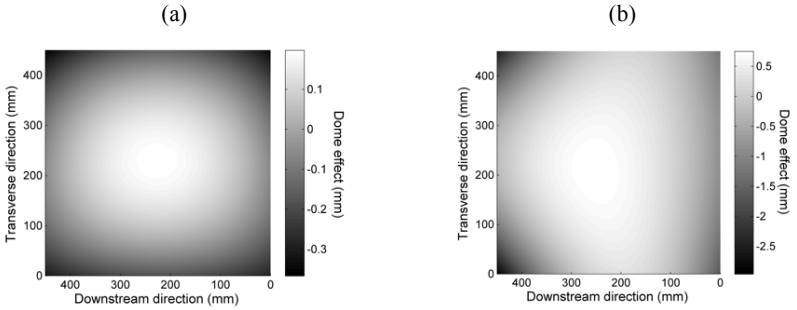


Figure 7. *Influence of the automatic corner re-computation and the number of calibration images on the DEM quality. Calibrations with corner re-computation (grey), calibrations without corner re-computation (black).*

A small dome (elevation span of 0.64 mm ± 0.08; mean absolute elevation when centred on zero of 0.12 mm ± 0.02) was noticed in the DEM of a flat surface (450 x 450 mm) for all calibrations (Figure 8a). Increasing the number of calibration images and re-computing the corners during the calibration did not reduce the dome effect.

591 Wackrow and Chandler (2011) showed that the mildly convergent imagery provides
592 means to reduce the dome effect. Similarly to the optimisation of the baseline distance,
593 deciding between a convergent and a vertical image configuration is difficult and is surface-
594 dependent.

595 A counter measure to minimise the dome effect in DEMs is the application of a margin
596 during the design to limit the DEM analysis to a more central region where the dome is
597 smaller. But again, to be effective, the minimisation of the dome effect by accounting a
598 margin has to be balanced against a deteriorated depth resolution when the camera-to-object
599 distance is increased.



600 ***Figure 8.*** *Dome observed in the DEM of a flat surface (450 x 450 mm). (a) The dome effect was*
601 *similar for all calibrations tried in test 1. (b) Dome in the case of the "partial" calibration (test 2).*
602 *The dome was measured by fitting a bi-quadratic surface to the DEM by least-squares, after the DEM*
603 *was aligned with the photo base.*

604 5.2. Test 2: Calibration Spatial Distribution

605 Figure 9 and Table III show that both the rectification error and the DEM error are
606 increased in the regions where control points were missing during calibration. This is valid
607 for the "partial" calibration, but is also visible for the "complete" calibration. For the "partial"
608 calibration, rectification error statistics are severely increased (about 10 times) between
609 regions with control points (called the "calibration region") and regions without control
610 points (Table III). This in turn results in a two-fold increase in the DEM error between the
611 two regions (Table III). The degradation of the DEM quality is less than the difference in
612 rectification error would suggest. Firstly, only large rectification errors (> 0.5 pixel) have an
613 impact on the stereo matching, and thus on the DEM quality, since SDPS stereo matching is
614 performed along lines of 1 pixel width. Secondly, point cloud to DEM transformation, during

which the DEM errors due to solitary large rectification errors are smoothed by interpolation, reduces the final DEM error. And thirdly, DEM errors are bounded by the relief of the hemisphere model.

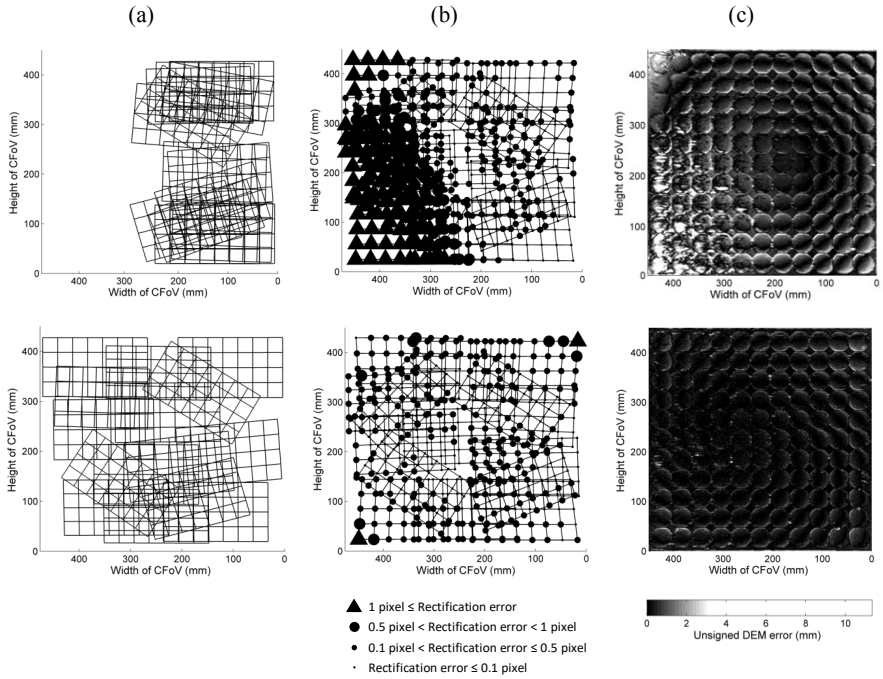


Figure 9. *Plan-view of control points location in calibration images (a); associated rectification error (b); and associated DEM error over the hemisphere model (c). The top row corresponds to the "partial" calibration. The bottom row corresponds to the "complete" calibration. In order to emphasise the effect of a "partial" calibration, the internal reliability of the DEMs was evaluated by doing a comparison with the DEM obtained with the "parent" calibration, not the "truth" DEM.*

A closer look at Figure 9 and Table III shows that the "partial" calibration not only affected the point measurements outside the "calibration region", but to a lesser extent has a global effect over the whole measurement window. Indeed, the DEM error inside the "calibration region" in the case of the "partial" calibration (MUE of 0.66 mm and SDE of 0.90 mm) is larger than the overall (over the whole CFoV) DEM error in the case of the "complete" calibration (MUE of 0.30 mm and SDE of 0.42 mm). Since the rectification error was only increased outside the "calibration region" for the "partial" calibration, we conclude

630 that there must be other DEM error sources in addition to those from inaccurate image
631 rectification.

632 ***TABLE III.*** *Rectification error and DEM error (using the hemisphere model) in the case of the*
633 *"partial" and the "complete" calibrations. In order to emphasise the effect of a "partial"*
634 *calibration, the internal reliability of the DEMs was evaluated by doing a comparison with the DEM*
635 *obtained with the "parent" calibration, not the "truth" DEM.*

	Spatial distribution of control points	Partial coverage of CFOV			Complete coverage of CFoV
	Region where errors are calculated	Inside calibration region	Outside calibration region	Whole CFoV	Whole CFoV
Rectification error (pixel)	Mean	0.13	1.18	0.60	0.14
	Standard deviation	0.12	1.25	0.99	0.08
	Maximum	0.76	10.11	10.11	0.78
DEM error (mm)	MUE	0.66	1.09	0.88	0.30
	SDE	0.90	1.51	1.25	0.42
	Maximum (unsigned)	5.60	11.34	11.34	4.21

636

637 Figure 10 presents the complete distortion model, the intrinsic parameters and the re-
638 projection errors for the "partial" and the "complete" calibrations. Observations for both
639 cameras were identical; therefore only results for the left camera are presented. Firstly, the re-
640 projection errors are smaller for the "partial" calibration, which disagrees with the
641 rectification error and the DEM error observations (Table III). This suggests that the re-
642 projection error is not adequate to characterise the calibration effect onto the stereo matching.
643 Other studies showed that the rectification error is more suitable than the re-projection error
644 to evaluate the quality of a calibration for binocular systems (Bradley and Heidrich, 2010,
645 Bertin et al., 2012). Secondly, the shape of the distortion model is modified for the "partial"
646 calibration, with a lateral displacement and a change in curvature. The computed distortion

coefficients and the profiles of radial and tangential distortion confirm this change, showing an increase in distortion for the "partial" calibration (Figure 10).

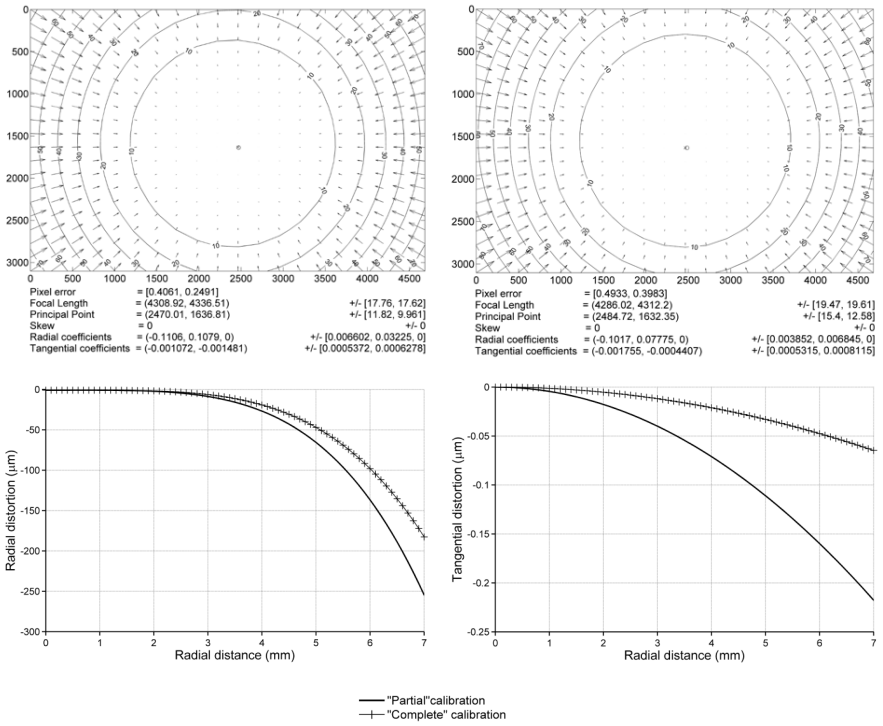


Figure 10. *(Top row) Complete distortion model, intrinsic parameters and re-projection errors obtained during calibration of the left camera for (left) the "partial" calibration; and (right) the "complete" calibration. (Bottom row) Profiles of radial (left) and tangential (right) distortion of the left camera. Similar observations were made with the right camera.*

Resulting from the inaccurate distortion model obtained with the "partial" calibration, a large dome (elevation span of 3.71 mm and mean absolute elevation when centred on zero of 0.47 mm; compared with 0.64 mm and 0.12 mm, respectively, for a "complete" calibration) was observed in the DEM of a flat surface (Figure 8b), explaining why DEM errors not only increased outside the "calibration region", but globally over the CFoV.

Figure 11 presents the results obtained with the "parent" calibration. We recall the "parent" calibration was completed using 40 calibration images, which was the largest image set in this study. As shown, it is difficult to place the chequerboard (and thus obtaining

control points) to cover the whole CFoV (Figure 11). Rectification errors larger than 0.5 pixel are found in the two corners of the measurement window where only one control point was available for calibration. Hence, to enable correct stereo matching throughout the measurement window, it is essential to obtain sufficient control points (our results suggest at least two).

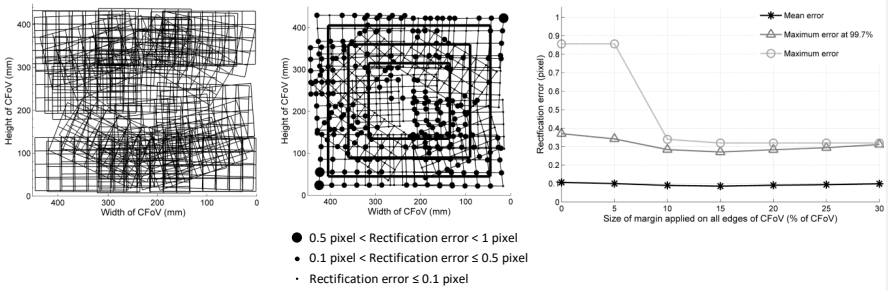


Figure 11. *Plan-view of control points location in the case of the "parent" calibration: 40 calibration images (left); and associated rectification error (middle). The lines delimit the regions after a margin of size 10, 20 and 30% the size of the CFoV, respectively, has been applied. The graph on the right shows the rectification error as a function of the margin size.*

Accounting for a margin around the CFoV, where the rectification error is not counted, can be helpful to remove marginal pixels with large rectification errors (Figure 11), thus reducing the risk to have large DEM errors at these same locations. Median filtering during the stereo matching may also smooth the effects of solitary large rectification errors.

With no loss in accuracy, image rectification to epipolar geometry can be performed for points with a camera distance outside the range of control points (Figure 12). As previously found, control points should be evenly distributed throughout the 3D volume covered by the object of interest (Chandler et al., 2001, Carbonneau et al., 2003). Figures 9 and 12 show that in our case control points only need to cover the 2D area of interest. For hydraulic experiments this implies that calibration images can be obtained above the gravel-bed surface, without increasing the rectification error at the distance of the gravel bed. Subsequent stereo matching performance will not be decreased. This is a major advantage, as the setup does not need to be moved for the calibration, which minimises the risk to alter the validity of the calibration over time. Our results support and extend the early findings of Weng et al. (1992), who showed that calibrated cameras are able to measure points at distances beyond the range of control points with minimal (1.3 %) loss in accuracy. However,

caution needs to be taken when generalising these results to a much larger range of camera distances, as only a narrow distance strip was tested.

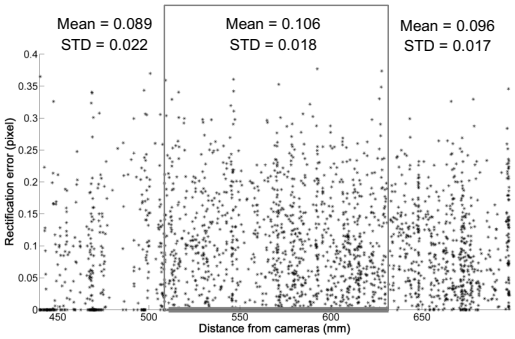


Figure 12. *Rectification error, determined at each corner of the chequerboard in 65 rectification-check images (i.e. at 2600 check points), versus the distance from the cameras. The grey markers on the horizontal axis and the box delimit the range of control points' distances during calibration.*

5.3. Test 3: Image Quality

Figure 13 presents the original images of the gravel-bed model, directly recorded in JPEG (1:4) using various shutter speeds (other camera settings remained constant), and their Retinex equivalents, all cropped to the size of the model (296 x 184 mm), with their respective luminosity histograms. Pixel luminosity corresponds to the sum of 0.299, 0.587, and 0.114 times the intensity in the red, green and blue tones, respectively. As evidenced in Figure 13, the MSRCR algorithm implemented in GIMP® is able to re-centre the distribution of pixel luminosity for all images tested. Whilst original images look very different, the images transformed with Retinex are very similar visually. As observed in previous works, a colour greying happened during the Retinex transformations (Jobson et al., 1997, Rahman and Woodell, 2004).

Table IV presents the MUE computed after differentiation of the DEMs reconstructed with the images obtained for the test and the "truth" DEM of the gravel-bed model. We recall that the same calibration data was used throughout the test to ensure a consistent comparison between the DEMs, where image quality is the only difference. HHQ and HQ correspond to images recorded in RAW exported to "highest quality" (1:2.6 compression ratio) and "high quality" (1:15 compression ratio) JPEG, respectively.

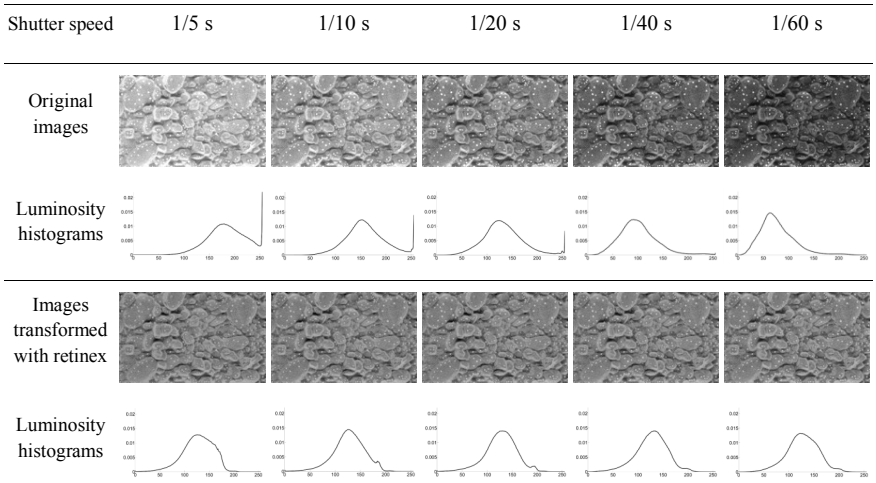


Figure 13. *Left-hand images of the gravel-bed model, cropped to the size of the ground truth (296 x 184 mm), and their associated luminosity histograms. Only the shutter speed was changed, other settings (ISO, aperture) remained constant. The original images, recorded in JPEG (1:4), and their Retinex transformations are presented.*

Image transformations (greyscale and Retinex) have a considerable impact on the stereo matching and thus on the DEM accuracy (Table IV). The most accurate DEMs were those obtained with images transformed with Retinex, with an average 40 % reduction of the MUE compared with the DEMs reconstructed using the original images. This finding is significant, as to the authors' knowledge it is the first time the MSRCR algorithm is implemented to facilitate the stereo matching. In contrast, the stereo matching performed on the greyscale equivalent images provided the worst results. We therefore recommend other users of the calibration toolbox to modify the code to obtain rectified images in RGB format to enable superior stereo matching results.

Table IV shows that RAW to JPEG transformation reduced the DEM quality in terms of the MUE (about 5 % for HHQ, 10 % for HQ), compared to DEMs collected using images directly recorded in JPEG (1:4). It suggests that the in-camera and the camera software JPEG encodings are different. Here, the in-camera JPEG compression provided better quality images for the stereo matching. We still believe RAW imagery will be useful for other studies, where there is no a-priori knowledge of the optimum stereo matching camera settings, since exposure changes can be made on RAW images.

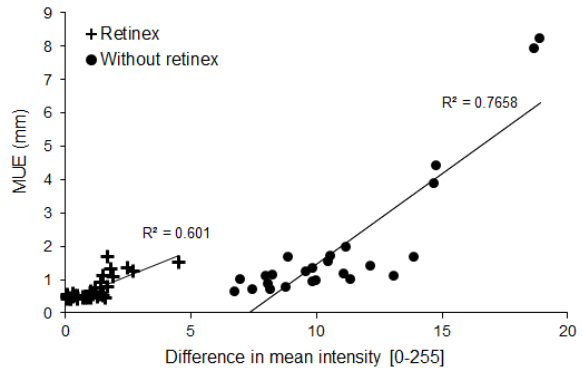
TABLE IV. *Mean unsigned error (MUE) in millimetre measured on the DEMs of the gravel-bed model, reconstructed using images recorded with various shutter speeds (other camera settings constant) in both JPEG(1:4) and RAW format. Images recorded in RAW were converted to HHQ and HQ JPEG (JPEG (1:2.6) and JPEG (1:15), respectively) before the stereo matching. Original images directly recorded in JPEG (1:4) were transformed with Retinex and greyscale. To allow a realistic comparison between DEMs, where the differences are only due to the image quality, the image rectification prior to the stereo matching and the depth triangulation were performed using the same calibration data. Image file size is indicated, and was determined on the left-hand images recorded with 1/20 s shutter speed.*

Transformation \ Shutter speed (s)	1/5	1/10	1/20	1/40	1/60
Retinex (7.8Mb)	0.54	0.51	0.49	0.49	0.51
Original (6.6 Mb)	0.96	0.69	0.62	1.08	1.64
HHQ (9.1 Mb)	1.02	0.72	0.64	1.12	1.59
HQ (1.8 Mb)	1.09	0.73	0.66	1.13	1.57
Greyscale (4.7 Mb)	1.43	1.16	0.98	2.01	2.57

The tested shutter speed camera setting impacted the stereo matching performance and the DEM quality substantially (Table IV). All other camera settings were unchanged in this test. The stereo pair recorded with 1/20 s shutter speed always provided the best DEMs in terms of the MUE. It is important to note that this finding is specific to this study, as the shutter speed selection depends on various parameters, such as the lighting environment and the shininess of the surface. For all images except the Retinex equivalents, the correct selection of the shutter speed was essential, since a change of the shutter speed from 1/20 s to 1/60 s resulted in DEMs with a large difference in MUE (about 60 % change). For images transformed with Retinex, the representation of the MUE for various shutter speeds is less widely spread (Table IV), suggesting that a larger range of shutter speeds has a lesser effect on stereo matching. Still, a 9 % difference in MUE is observed between two DEMs collected using Retinex imagery (DEMs reconstructed with 1/5 s and 1/20 s, respectively). Therefore, whilst image transformation techniques like Retinex have the potential to facilitate the stereo matching, the careful selection of the camera settings must still be regarded as a key element

752 in a stereo-photogrammetric project. According to the results obtained in this test, a good
753 image for stereo matching is represented by a luminosity histogram with a pronounced
754 central peak before any transformation.

755 Figure 14 presents a summary of the MUE for 62 gravel-bed model DEMs. The
756 horizontal axis shows the difference in mean pixel intensity between the two images,
757 calculated over the size of the gravel-bed model (296 x 184 mm). Even in a controlled
758 environment such as the laboratory, exposure difference between the left and right images
759 forming a stereo pair may be observed, often due to the lighting source. Neon lights (Figure
760 1), which were used for the study, have the advantage of producing an even illumination of
761 the test section. The known disadvantage is that they cause a "flicker effect" with light pulsed
762 at twice the supply frequency. Figure 14 shows clearly that the difference in mean pixel
763 intensity between the two images forming a stereo pair is reduced when transformed with
764 Retinex.



765

766 ***Figure 14.*** *Effect of the image quality (here described in terms of the difference in mean pixel*
767 *intensity between the two images forming a stereo pair) on the DEM accuracy, represented by the*
768 *MUE. The results obtained on 62 DEMs of the gravel-bed model (37 Retinex and 25 original images)*
769 *are summarised in this graph.*

770 Furthermore, the results show that the MUE and the difference in mean pixel intensity are
771 correlated, with a distinct relationship for each image type (i.e. with or without Retinex). For
772 both original images and their Retinex equivalents, the MUE increases with an increased
773 difference in mean pixel intensity between the two images used for the stereo matching. To
774 evaluate image quality, the difference in mean pixel intensity between the two images

forming a stereo pair, needs to be an important parameter to judge the pair's suitability for stereo matching.

6. Conclusions

This paper presented a digital stereo-photogrammetric workflow for close-range measurement, in application to grain-scale fluvial surfaces. Processing steps (calibration, image rectification and stereo matching) were introduced and discussed. A controlled laboratory evaluation was carried out to identify the cause of processing errors and to find means to reduce their impact on DEM quality.

The introduced workflow can be described as user-friendly, low-cost and versatile. Imagery was acquired with two consumer-grade DSLRs and processed using non-proprietary algorithms. Compared with previous fluvial applications, "on-the-job" calibration using stereo images of a 2D chequerboard does not require the placement of fixed control targets on the riverbed surveyed with an independent device, and enables both radial and tangential distortion coefficients to be recovered. Scanline-based stereo matching resulted in dense DEMs with the possibility to have sampling distances as small as the pixel size at the object's distance, which improves substantially on traditional area-based methods. Limited post-processing was applied and prevented the introduction of new errors.

Consequently, our DEMs have high point densities (0.25 mm sampling distance) and high vertical accuracies (MUE close to 0.5 mm), which are required to realistically represent water-worked riverbed roughness. To the authors' knowledge this represents one of the most precise and accurate topography measurement solution for hydraulic research, which rivals results obtained with laser-scanning (Hodge et al., 2009).

Using the presented evaluation strategy, it is shown that errors are systematically introduced to the workflow, starting with the calibration. Evenly distributed control points throughout the measurement window improve DEM quality, by ensuring a smaller than 0.5 pixel homogeneous rectification error and removing most of the optical imagery distortion. In general terms, an increasing number of calibration images improved DEM quality, whilst re-computing corners had no observable effect outside the re-projection error reduction. Following on previous works, our studies confirm that the calibration quality for binocular systems is best expressed in terms of the rectification error, rather than the more commonly used re-projection error (Bradley and Heidrich, 2010, Bertin et al., 2012). Image quality, a function of the dynamic range in the imagery, as well as the similarity between the two

images forming a stereo pair, was shown critical to stereo-matching success. We presented how image transformation techniques, such as Retinex, provide means to heighten the similarity between the images forming a stereo pair and thus to improve stereo matching. Given suitable calibration parameters, image enhancement and stereo matching were the steps where optimisation resulted in the most significant DEM accuracy improvement in our study. We therefore recommend considering these steps carefully to achieve optimal accuracy.

Ongoing work is required on how best to address the existence of occlusions, which are still a major source of photogrammetric errors. Recent studies use multi-view stereo (MVS) and structure-from-motion (SfM) photogrammetry techniques for medium to large scale experiments (Westoby et al., 2012, Javernick et al., 2014), but not for the very small scales. Those techniques are potentially more suited to capture occlusions. In the present study, occlusion errors could not be removed completely, although calibration and image quality optimisation improved stereo-matching results. Occlusions are the direct consequence of the photogrammetric design. The use of realistic ground truths, such as the ones presented in this study, is shown to be useful for the experimental determination of the optimum stereo-photogrammetric design. The next step is acquiring multiple depth maps of the same scene from different viewpoints, which can then be fused to improve data quality.

Acknowledgement

The authors would like to thank the Editor-in-Chief, Derek Lichti, and three anonymous reviewers for their valuable comments, which helped to improve the paper.

References

Aber J. S., Marzolff I. and Ries J. B., 2010. Chapter 6 - Cameras for small-format aerial photogrammetry. In Small-Format Aerial Photography. J. S. A. M. B. Ries. Amsterdam, Elsevier, 69-80.

Aberle J. and Smart G. M., 2003. The influence of roughness structure on flow resistance on steep slopes. Journal of Hydraulic Research, 41(3), 259-269.

Bertin S. and Friedrich H., 2014. Measurement of gravel-bed topography: evaluation study applying statistical roughness analysis. Journal of Hydraulic Engineering, 140(3), 269-279.

Bertin S., Friedrich H., Chan E. and Delmas P., 2012. The development and internal assessment of a high-resolution, non-proprietary, stereo-photogrammetric setup for hydraulic experiments. Proceedings of the 27th Conference on Image and Vision Computing New Zealand, Dunedin, New Zealand, ACM, 515-520.

Bertin S., Friedrich H., Delmas P. and Chan E., 2013. The use of close-range digital stereo-photogrammetry to measure gravel-bed topography in a laboratory environment. Proceedings of the 35th IAHR Congress, Chengdu, China, 12 pages.

Bertin S., Friedrich H., Delmas P., Chan E. and Gimel'farb G., 2014. DEM quality assessment with a 3D printed gravel bed applied to stereo photogrammetry. Photogrammetric Record, 29(146), 241-264.

Bird S., Hogan D. and Schwab J., 2010. Photogrammetric monitoring of small streams under a riparian forest canopy. Earth Surface Processes and Landforms, 35(8), 952-970.

Bouguet J.-Y. 2010. http://www.vision.caltech.edu/bouguetj/calib_doc/. [Accessed: 1st July 2013].

Bouratsis P., Diplas P., Dancey C. L. and Apsilidis N., 2013. High-resolution 3D monitoring of evolving sediment beds. Water Resources Research, 49(2), 977-992.
Bradley D. and Heidrich W., 2010. Binocular camera calibration using rectification error. Proceedings of the 2010 Canadian Conference on Computer and Robot Vision (CRV), Ottawa, Ontario, Canada**,** 183-190.
Butler J. B., Lane S. N. and Chandler J. H., 1998. Assessment of DEM quality for characterizing surface roughness using close range digital photogrammetry. Photogrammetric Record, 16(92), 271-291.
Butler J. B., Lane S. N. and Chandler J. H., 2001. Characterization of the structure of river-bed gravels using two-dimensional fractal analysis. Mathematical Geology, 33(3), 301-330.
Butler J. B., Lane S. N., Chandler J. H. and Porfiri E., 2002. Through-water close range digital photogrammetry in flume and field environments. Photogrammetric Record, 17(99), 419-439.
Carbonneau P. E., Lane S. N. and Bergeron N. E., 2003. Cost-effective non-metric close-range digital photogrammetry and its application to a study of coarse gravel river beds. International Journal of Remote Sensing, 24(14), 2837–2854.
Chandler J., Fryer J. and Jack A., 2005. Metric capabilities of low-cost digital cameras for close range surface measurement. Photogrammetric Record, 20(109), 12-26.
Chandler J., Shiono K., Rameshwaran P. and Lane S., 2001. Measuring flume surfaces for hydraulics research using a Kodak DCS460. Photogrammetric Record, 17(97), 39-61.
Cooper M. A. R. and Robson S., 2001. Theory of close range photogrammetry. In Close range photogrammetry and machine vision. K. B. Atkinson. Caithness, Whittles. **2,** 371.
Fraser C. S., 1997. Digital camera self-calibration. ISPRS Journal of Photogrammetry and Remote Sensing, 52(4), 149-159.
Fusiello A., Trucco E. and Verri A., 2000. A compact algorithm for rectification of stereo pairs. Machine Vision and Applications, 12(1), 16-22.
Gimel'farb G., 2002. Probabilistic regularisation and symmetry in binocular dynamic programming stereo. Pattern Recognition Letters, 23(4), 431-442.
González-Díez A., Fernández-Maroto G., Doughty M. W., Díaz de Terán J. R., Bruschi V., Cardenal J., Pérez J. L., Mata E. and Delgado J., 2014. Development of a methodological approach for the accurate measurement of slope changes due to landslides, using digital photogrammetry. Landslides, 11(4), 615-628.
Gooch M. J., Chandler J. H. and Stojic M., 1999. Accuracy assessment of digital elevation models generated using the Erdas Imagine Orthomax digital photogrammetric system. Photogrammetric Record, 16(93), 519–531.
Hardy R. J., 2008. Geomorphology fluid flow modelling: can fluvial flow only be modelled using a three-dimensional approach? Geography Compass, 2(1), 215-234.
Hardy R. J., Best J. L., Lane S. N. and Carbonneau P. E., 2009. Coherent flow structures in a depth-limited flow over a gravel surface: The role of near-bed turbulence and influence of Reynolds number. Journal of Geophysical Research, 114(F01003), 18 pages.
Hodge R., Brasington J. and Richards K., 2009. In situ characterization of grain-scale fluvial morphology using Terrestrial Laser Scanning. Earth Surface Processes and Landforms, 34(7), 954-968.
Javernick L., Brasington J. and Caruso B., 2014. Modeling the topography of shallow braided rivers using Structure-from-Motion photogrammetry. Geomorphology, 213, 166-182.
Jobson D. J., Rahman Z. u. and Woodell G. A., 1997. A multiscale retinex for bridging the gap between color images and the human observation of scenes. IEEE Transactions on Image Processing, 6(7), 965-976.
Land E. H., 1986. An alternative technique for the computation of the designator in the retinex theory of color vision. Proceedings of the National Academy of Sciences, 83(10), 3078-3080.
Lane S., Chandler J. and Porfiri K., 2001. Monitoring river channel and flume surfaces with digital photogrammetry. Journal of Hydraulic Engineering, 127(10), 871-877.
Lane S. N., 2000. The measurement of river channel morphology using digital photogrammetry. Photogrammetric Record, 16(96), 937-961.
Lane S. N., Hardy R. J., Elliott L. and Ingham D. B., 2002. High-resolution numerical modelling of three-dimensional flows over complex river bed topography. Hydrological Processes, 16(11), 2261-2272.
Lane S. N., James T. D. and Crowell M. D., 2000. Application of digital photogrammetry to complex topography for geomorphological research. Photogrammetric Record, 16(95), 793-821.
Lane S. N., Reid S. C., Westaway R. M. and Hicks D. M., 2005. Remotely sensed topographic data for river channel research: the identification, explanation and management of error. In Spatial Modelling of the Terrestrial Environment, John Wiley & Sons, Ltd**,** 113-136.
Liu Y., Martin R. R., de Dominicis L. and Li B., 2014. Using retinex for point selection in 3D shape registration. Pattern Recognition, 47(6), 2126-2142.
Mao L., Cooper J. R. and Frostick L. E., 2011. Grain size and topographical differences between static and mobile armour layers. Earth Surface Processes and Landforms, 36(10), 1321-1334.
Matthews N. A., 2008. Aerial and close-range photogrammetric technology: providing resource documentation, interpretation, and preservation. U.S. Department of the Interior, Bureau of Land Management, National Operations Center, Denver, Colorado.
Meng Q., Bian D., Guo M., Lu F. and Liu D., 2012. Improved multi-scale retinex algorithm for medical image enhancement. In Information Engineering and Applications. R. Zhu and Y. Ma. London, Springer. 154**,** 930-937.
Murillo-García F., Alcántara-Ayala I., Ardizzone F., Cardinali M., Fiourucci F. and Guzzetti F., 2014. Satellite stereoscopic pair images of very high resolution: a step forward for the development of landslide inventories. Landslides, 1-15.
Ockelford A.-M. and Haynes H., 2013. The impact of stress history on bed structure. Earth Surface Processes & Landforms, 38(7), 717-727.
Park Y. K. P. S. L. and Kim J. K., 2008. Retinex method based on adaptive smoothing for illumination invariant face recognition. Signal Processing, 88(8), 1929-1945.

913 Qin J. and Ng S., 2012. Estimation of effective roughness for water-worked gravel surfaces. Journal of Hydraulic
914 Engineering, 138(11), 923-934.
915 Rahman Z. u. and Woodell G. A., 2004. Retinex processing for automatic image enhancement. Journal of Electronic
916 Imaging, 13, 100-110.
917 Schmocker L., 2011. Application of a videometric measurement system to investigate spatial dike breach. In Experimental
918 Methods in Hydraulic Research. P. Rowinski. Berlin, Springer. 1**,** 247-254.
919 Smart G., Duncan M. and Walsh J., 2002. Relatively rough flow resistance equations. Journal of Hydraulic Engineering,
920 128(6), 568-578.
921 Smith M., Vericat D. and Gibbins C., 2012. Through-water terrestrial laser scanning of gravel beds at the patch scale. Earth
922 Surface Processes and Landforms, 37(4), 411-421.
923 Smith M. W., Cox N. J. and Bracken L. J., 2011. Terrestrial laser scanning soil surfaces: a field methodology to examine soil
924 surface roughness and overland flow hydraulics. Hydrological Processes, 25(6), 842-860.
925 Stoter J., Roensdorf C., Home R., Capstick D., Streilein A., Kellenberger T., Bayers E., Kane P., Dorsch J., Woźniak P.,
926 Lysell G., Lithen T., Bucher B., Paparoditis N. and Ilves R., 2015. 3D Modelling with National Coverage:
927 Bridging the Gap Between Research and Practice. In 3D Geoinformation Science. M. Breunig, M. Al-Doori, E.
928 Butwilowski, P. V. Kuper, J. Benner and K. H. Haefele, Springer International Publishing**,** 207-225.
929 Sun J., Du Y. and Tang Y., 2008. Shadow detection and removal from solo natural image based on retinex theory. In
930 Intelligent Robotics and Applications. C. Xiong, Y. Huang, Y. Xiong and H. Liu. Berlin Heidelberg, Springer.
931 5314**,** 660-668.
932 Wackrow R. and Chandler J. H., 2011. Minimising systematic error surfaces in digital elevation models using oblique
933 convergent imagery. Photogrammetric Record, 26(133), 16-31.
934 Weng J., Cohen P. and Herniou M., 1992. Camera calibration with distortion models and accuracy evaluation. IEEE
935 Transactions on Pattern Analysis and Machine Intelligence, 14(10), 965-980.
936 Westaway R. M., Lane S. N. and Hicks D. M., 2003. Remote survey of large-scale braided, gravel-bed rivers using digital
937 photogrammetry and image analysis. International Journal of Remote Sensing, 24(4), 795-815.
938 Westoby M. J., Brasington J., Glasser N. F., Hambrey M. J. and Reynolds J. M., 2012. 'Structure-from-Motion'
939 photogrammetry: a low-cost, effective tool for geoscience applications. Geomorphology, 179(0), 300-314.
940 Zhang Z., 1999. Flexible camera calibration by viewing a plane from unknown orientations. Proceedings of the Computer
941 Vision, 1999. 7th IEEE International Conference on Computer Vision.**,** 666-673.

942

943