DIRECT IMAGE-TO-GEOMETRY REGISTRATION USING MOBILE SENSOR DATA

C. Kehl[a,b,*], S. J. Buckley[a], R. L. Gawthorpe[b], I. Viola[d], J. A. Howell[c]

[a] Uni Research AS - Centre for Integrated Petroleum Research (CIPR), Allégaten 41, 5007 Bergen, Norway -
(christian.kehl, simon.buckley)@uni.no
[b] Department of Earth Science, University of Bergen, Allégaten 41, 5007 Bergen, Norway -
(christian.kehl, rob.gawthorpe)@geo.uib.no
[c] Department of Geology & Petroleum Geology, University of Aberdeen, AB24 3UE Aberdeen, UK - john.howell@abdn.ac.uk
[d] Inst. of Computer Graphics and Algorithms, Technical University of Vienna, 1040 Vienna, Austria - viola@cv.tuwien.ac.at

Commission II, WG II/6

KEY WORDS: Image-to-Geometry, Automatic Pose Estimation, Mobile Devices, Registration Interfaces, Virtual Outcrop Geology

ABSTRACT:

Adding supplementary texture and 2D image-based annotations to 3D surface models is a useful next step for domain specialists to make use of photorealistic products of laser scanning and photogrammetry. This requires a registration between the new camera imagery and the model geometry to be solved, which can be a time-consuming task without appropriate automation. The increasing availability of photorealistic models, coupled with the proliferation of mobile devices, gives users the possibility to complement their models in real time. Modern mobile devices deliver digital photographs of increasing quality, as well as on-board sensor data, which can be used as input for practical and automatic camera registration procedures. Their familiar user interface also improves manual registration procedures. This paper introduces a fully automatic pose estimation method using the on-board sensor data for initial exterior orientation, and feature matching between an acquired photograph and a synthesised rendering of the orientated 3D scene as input for fine alignment. The paper also introduces a user-friendly manual camera registration- and pose estimation interface for mobile devices, based on existing surface geometry and numerical optimisation methods. The article further assesses the automatic algorithm's accuracy compared to traditional methods, and the impact of computational- and environmental parameters. Experiments using urban and geological case studies show a significant sensitivity of the automatic procedure to the quality of the initial mobile sensor values. Changing natural lighting conditions remain a challenge for automatic pose estimation techniques, although progress is presented here. Finally, the automatically-registered mobile images are used as the basis for adding user annotations to the input textured model.

1. INTRODUCTION

Textured surface models are used in a widening range of application domains, such a urban planning (Semmo and Döllner, 2015), cultural heritage (Potenziani et al., 2015), archaeology (Van Damme, 2015) and geological outcrop modelling (Howell et al., 2014, Rarity et al., 2014). After a model has been captured, often by means of terrestrial laser scanning (TLS) or photogrammetry, it may be desirable for domain experts to supplement the model with novel images to make initially-hidden features visible, or to add annotations. In the geosciences, researchers are often interested in annotating and interpreting available models for communicating insights, highlighting data anomalies and regions of interest, or for further processing in application-specific workflows. With recent advances in mobile devices (e.g. tablets, smartphones), the possibility for users to capture new images and create annotations using this new colour information is becoming increasingly important. Retexturing models, introducing supplementary textures, or annotating (i.e. interpreting, mapping) 3D structures in novel 2D images is based on retrieving accurate exterior orientations using Image-to-Geometry registration. Available tools to obtain the pose of a captured image commonly rely on complex, time-consuming, perceptually challenging and error-prone manual 2D-3D registration methods. This leads to a bottleneck in geoscientific workflows.

The goal of this study is to improve Image-to-Geometry registration for application workflows, by obtaining exterior orientations through reliable automatic- or less error-prone manual proced-

ures. Automatic pose estimation is a traditional photogrammetric research topic, where recent approaches provide applicable poses if initialised with coarse approximations for position and orientation. This paper presents an extension to automatic Image-to-Geometry procedures using the in-built sensor data from mobile devices, as well as a novel approach for manual registration, taking the perceptual challenges into account, as a robust automation backup. This manual approach makes use of standard mobile device user interfaces (smartphones and tablets), which are centered around touch-based, visual, and immediate feedback interaction using fingers or a stylus. The applicability of the proposed techniques is evaluated in detail on urban- and geological case studies, focussing on the achievable accuracy, sensitivity to user-defined parameters, and success in challenging imaging conditions. The impact of the proposed approaches to the tasks of model retexturing and 2D–3D annotation is discussed in the final section of the article.

2. RELATED WORK

A vast body of literature is available on Image-to-Geometry registration. This is commonly a multi-stage process, consisting of coarse estimation followed by fine alignment. A generic setup starts with an image or a collection of images, capturing a scene from varying, unknown angles and positions (possibly without overlap) relative to geometric data. The algorithmic goal is to recover each camera's exterior orientation parameters (i.e. pose). Camera parameters such as focal length and CCD chip dimensions are directly obtained during the photo capturing (e.g. using EXIF tags), or pre-calculated for static lens configurations (e.g.

---

*Corresponding author

during camera calibration).

Several registration approaches, emerging from recent advances in computer vision, are further discussed. We refer to the review of augmented reality (AR) pose estimation (Nöll et al., 2011) for further details. Our study objective differs from AR requirements, as marker-based pose estimation is often not viable in a geoscientific setting, due to general difficulties of placing targets in remote and inaccessible natural environments.

2.1 Manual, Correspondence-Based Methods

Manual approaches rely on user-defined image–geometry point correspondences. Spatial resection (Moffitt and Mikhail, 1980) and Direct Linear Transform (DLT) are classic, correspondence-based photogrammetric techniques for pose estimation. Tsai presented a camera calibration technique that incorporates extraction of position and orientation in its solution, giving a lower point bound for a robust pose estimation (Tsai, 1987). Non-linear global optimisations, such as Levenberg-Marquardt (LM), generate numerical approximations of the exterior orientation parameters. Recent analytical solutions (Lepetit et al., 2009) to the Point-n-Perspective (PnP) estimation (Quan and Lan, 1999) give robust, real-time estimates. Existing challenges for correspondence methods are the provision of accurate, well-distributed 3D markers for each camera view, and user-friendly methods for defining correspondences. Current registration software (Chen et al., 2008) and commercial tools demand correspondence definition in 3D, which is perceptually problematic to novice users (Brunnett et al., 2003), and therefore time-consuming, error-prone and inefficient. Correspondence-based methods are incorporated in complex algorithms (Pintus et al., 2011, Sottile et al., 2010), or serve as a comparison baseline.

2.2 Statistical Correlation-Based Methods

Statistical correlation-based methods register unorientated images with a synthesised rendering of a surface geometry. This requires a coarse pre-registration to obtain approximate pose parameters for viewport definition and rendering. The photographs and synthetic images are then statistically evaluated (e.g. using information value and entropy) and their Mutual Information (MI) is iteratively maximised to obtain a pose (Viola and Wells, 1997, Maes et al., 1997). MI has computational advantages over comparable methods due to rapid image generation using computer graphics (CG) (Pluim et al., 2003), and is also used for the fine alignment in complex image registration frameworks (Corsini et al., 2013).

Hybrid methods combine the simplicity and accuracy of correspondence based approaches with the speed and robustness of statistical approaches. Mutual Correspondences is such a hybrid method (Sottile et al., 2010). The initially proposed algorithm minimizes the reprojection error between keypoints of the synthetic scene and a photograph, and maximizes their mutual information. A comparable algorithm minimizes the keypoints' reprojection between the synthetic and real image with strict spatial correspondences, eliminating the need for MI maximisation (Bodensteiner et al., 2012). Both approaches depend on a priori coarse registration as initialisation, which is currently achieved via manual 2D–3D correspondence selection.

3. CONTRIBUTIONS

In this paper, we propose an extension to hybrid registration methods by using mobile sensor data as the coarse initialization for Image-to-Geometry fine alignment. The extension results in a computationally lightweight, ad-hoc, fully-automatic registration

procedure that can be used during fieldwork. In addition, we propose a simplified, user-friendly manual registration approach on mobile devices by mapping the 2D–3D correspondence selection to a 2D-2D task, and by exploiting increasingly-familiar mobile device interfaces. The paper further includes an in-depth study of feature matching parameters and illumination influence on the automatic registration accuracy. A novel combination of keypoint detection algorithm and pose optimization parameters is introduced and tested, which improves image registrations with moderate illumination differences. The method's applicability is demonstrated using case studies from urban- and geological field environments.

4. ASSISTED MANUAL REGISTRATION

Manual Image-to-Geometry methods are based on 2D–3D point correspondences, where the selection of occlusion-free and well-defined points in 3D can be problematic in natural scenes, particularly for novice users. In the presented workflow, 3D control points are defined as part of the surface model processing stage, prior to the model's distribution to practitioners. Points are defined at multiple, expert-selectable scales (i.e. varying distances to the object under study), depicting visually prominent, easily-recognisable object segments with respect to the chosen view. These 3D control points are stored along with the surface geometry, essentially as part of the object's metadata. Then, the textured geometry and control points are rendered from numerous viewpoints. While the viewpoints are currently selected manually, this process can benefit from automatic viewpoint selection algorithms (e.g. using MI (Feixas et al., 2009)) in the future. In this process, the control points are projected into the synthetic images as spherical glyphs, giving the basis for defining 2D–3D correspondences on the mobile device.

The rendered images and their point correspondences are loaded onto the mobile device for use by practitioners in their fieldwork. When capturing a new photograph from the in-built camera, the user selects one or more appropriate shared viewpoints. They are then prompted to define a 2D control point within the photograph using the mobile device interface shown in fig. 1. The interface then switches to the shared synthetic image, and the user selects the corresponding 3D control points as superimposed 2D marker on the rendered image. Because of selection inaccuracies on small screens, a distance-ranked marker list is created. The closest non-occluded marker to the user-selected point is chosen as correspondence. The user repeats the correspondence selection until enough links have been defined for exterior orientation estimates (minimum of seven (Tsai, 1987)). Finally, the pose estimation proceeds according to a LM or Efficient PnP (EPnP) scheme using the defined correspondences.

Registering the image via point markers demands easily identifiable, relatively accurate and well-distributed control points. Inaccuracies in the point marking propagate to the final pose estimation. Though it is possible to use fingers on the touch screen for marking, a stylus-enabled device is advantageous for defining the correspondence.

5. AUTOMATIC REGISTRATION AND MATCHING

The proposed automatic approach extends former pose estimation research, based on a Random Sampling Consensus (RANSAC) scheme of LM pose estimation using Scale-Invariant Feature Transform (SIFT) points. In our method, features are matched between the captured camera photograph and a rendering of the existing textured model, which delivers the 3D reference points. We use
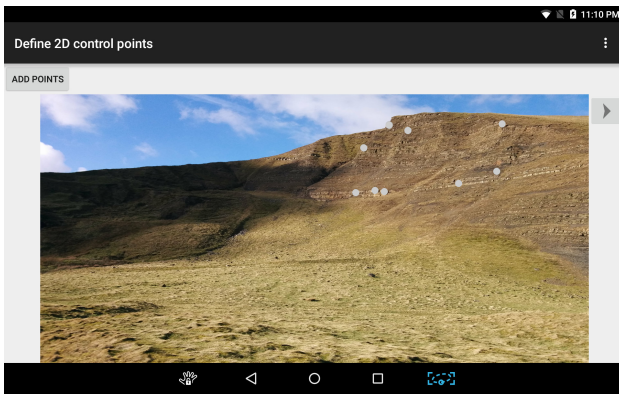
Figure 1: The manual registration interface, defining 2D correspondence points (grey circles) within the photograph, captured via smartphone.

the on-board sensor data from the mobile device to automatically determine the initial exterior orientation for the object rendering. This is in contrast to the user intervention required in comparable previous approaches and the manual interface presented in section 4 The full algorithm is outlined in fig. 2 and described below.

5.1 Mobile Sensor Orientation Estimate

Prior to image capture, the mobile device's magnetometer and accelerometer should be calibrated per field locality, using applications such as GPS Status [1]. Following this, new images can be acquired, and the global position and orientation are recorded at the time of capture. The global orientation using the *rotation vector* is referenced to magnetic north and earth's gravitational centre. In order to fuse the sensor data, the orientation, given as a quaternion, is corrected with the magnetic declination and the device orientation (e.g. portrait, landscape). A locally consistent pose is derived using the UTM coordinates of the recorded textured model. The varying coordinate systems are illustrated in fig. 3.
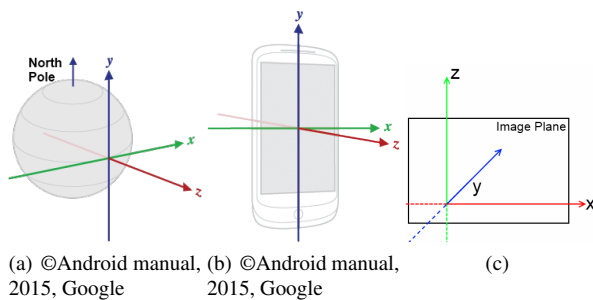

(a) ©Android manual, (b) ©Android manual, (c)
2015, Google   2015, Google

Figure 3: Schematic of coordinate systems defined in the Google Android operating system (a and b)[2] , and the system used later in this article (c).

Because of the low quality sensors commonly built into mobile devices, initial pose errors originate from both GPS and orientation measurements. A detailed overview of mobile device sensors and their calibration, error sources and error margins is available in the literature (Blum et al., 2013).

---

[1]GPS Status & Toolbox for Android - `http://play.google.com/store/apps/details?id=com.eclipsim.gpsstatus2`
[2]Sensor Event - Android Dev. `http://developer.android.com/reference/android/hardware/SensorEvent.html`

5.2 Image Synthesis

Using the given textured surface model and the initial pose, a synthetic image is generated on the Graphics Processing Unit (GPU) of the device using off-screen rendering. Camera interior orientation is used as input for the viewport definition. Furthermore, the utilization of the GPU allows adapting certain image parameters (e.g. contrast, brightness, gamma function) in real-time using post-processing shaders, which is potentially beneficial to the feature detection and mapping process reported later.

5.3 RANSAC-Based Pose Refinement

After obtaining the photo and the corresponding synthetic image, a feature detector (e.g. SIFT (Lowe, 2004), Speeded-Up Robust Features (SURF) (Bay et al., 2006), Maximally Stable Colour Regions (MSCR) (Forssén, 2007)) extracts 2D keypoints and descriptors. These descriptors are matched using symmetry- and homography constraints. Using raycasting, 3D points are obtained for matched keypoints in the photo. A PnP RANSAC scheme uses the 2D–3D correspondences to compute a pose. The pose with the biggest support is iteratively refined using LM. An example comparison between the captured photograph, the initial mobile sensor pose and the refined position and orientation is shown in fig. 4.
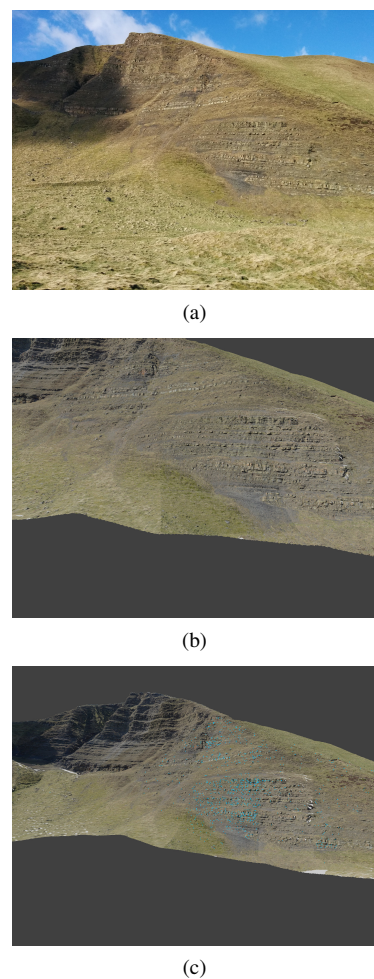

(a)

(b)

(c)

Figure 4: Comparative image series between a captured photo (a) and its corresponding synthetic images using the initial pose (b) and the final pose (c) after optimisation.
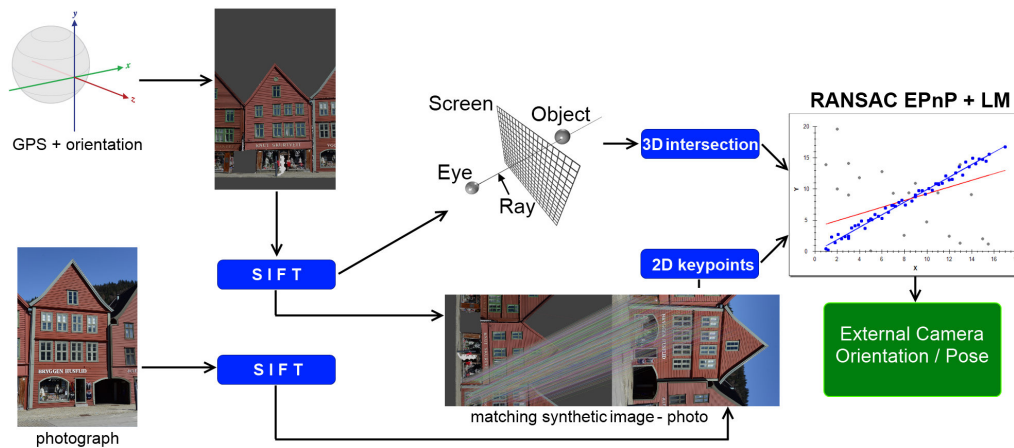
Figure 2: Outline of the proposed automatic registration algorithm, from capturing a new photograph using the device's on-board camera and sensor data (left) to the solved exterior orientation parameters (right).

6. EXPERIMENTS

The focus of the experiments is to assess the advantages and disadvantages of the automatic registration relative to manual Image-to-Geometry registration methods that are currently in use. Our interest is i) on the usability and sensitivity of mobile sensor pose input to the final registration, ii) the impact of tunable parameters during the image matching and pose estimation, and iii) the impact of changing lighting conditions on the final external orientation parameter accuracy.

The assessment uses two different datasets, both derived from time-of-flight TLS using the Riegl VZ-1000. The "Bryggen" dataset represents an urban use case, comprising the UNESCO World Heritage Site [3] late-medieval wharf of the former Hanseatic League in Bergen, Norway. The objects of interest within the Bryggen dataset were captured in high detail and consist of well-defined, easily-recognisable shapes. A challenge of the dataset is that image information contain reoccurring patterns (e.g. shop insignia, wood-panel facades), which can lead to mismatches of local image feature descriptors. Fig. 5 shows a captured dataset as coloured point set and textured surface model.

The second dataset, "Mam Tor", is a geological outcrop in the Peak District, Derbyshire, UK, with relevance for geological mapping. The area consists of a hillside with exposed sedimentary strata. The section of interest covers 280 metres (lateral) by 140 metres (vertical). Compared to the Bryggen dataset, less detail is available over comparable areas. In addition, the less well-defined object shape creates major challenges for the registration, as does vegetation within the scene and differing lighting conditions between textured model and mobile device image captures. Figure 6 shows the captured textured surface model and a selection of mobile device images to be registered.

The image texture is captured in both cases using a scanner-mounted, calibrated Nikon D800E camera. The laser-generated point set is processed according to established workflows (Buckley et al., 2008, Buckley et al., 2010), resulting in a textured triangulated irregular network (TIN). The image dataset for benchmarking the accuracy assessment is taken from the mounted Nikon camera with calibrated Nikkor 85mm prime lens (resolution: 7360x4920 pixels), while the mobile images are captured with a Google Nexus 5 smartphone (resolution 3262x2448 pixels).

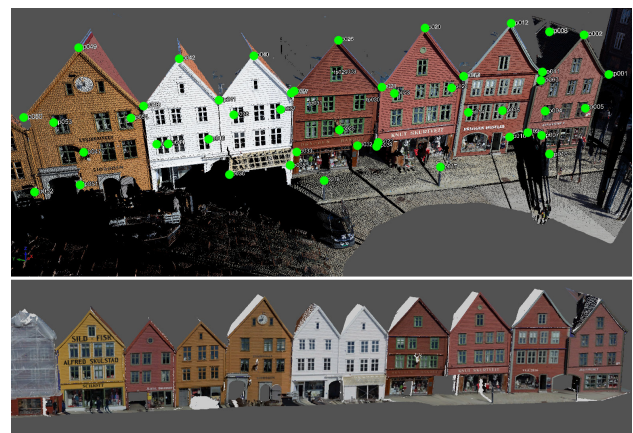[3]Bryggen World Heritage Center - http://whc.unesco.org/en/list/59

Figure 5: Rendering of the Bryggen dataset as coloured point set (top) and textured surface model (bottom), stretching along 40 m laterally. The green markers (top) highlight 3D control points for manual registration procedures.
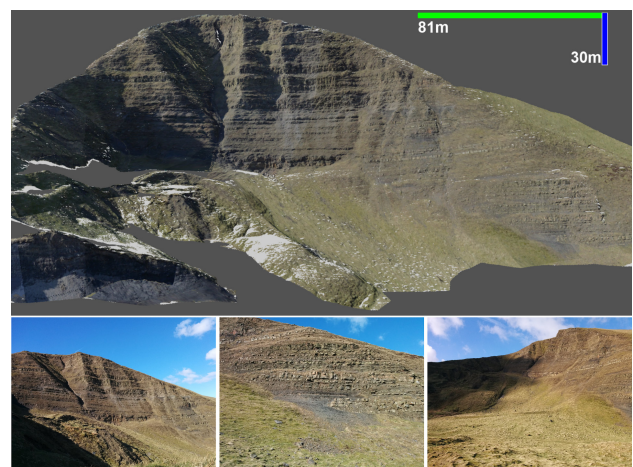

Figure 6: Overview rendering of the Mam Tor dataset as textured surface model (top, scale included) and a selection of supplementary photos to be incorporated (bottom).

In the presented experiments, manual control points are evenly distributed in the image plane, ignoring extremely close- and far range areas in the image's depth plane (see fig. 5 top for Bryggen and fig. 7(a) for Mam Tor). Automatic control points are reasonably well-distributed, though mainly cover the exposed rock areas rather than vegetated slopes. Foreground objects receive the majority of control points due to their more distinct feature separation (as in fig. 7(b)). The orientation optimisation process does not demand strictly even distributions, although the control points need to represent variation within all three dimensions. Also, the point distribution should not be collinear.
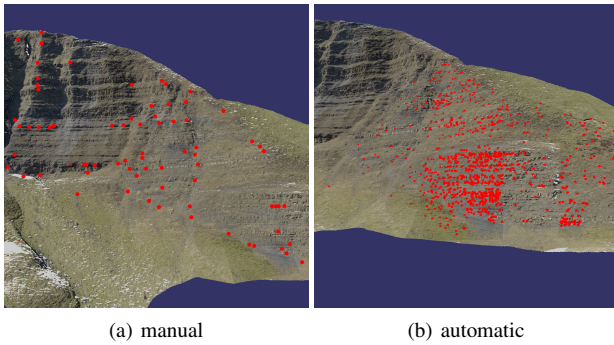

(a) manual (b) automatic

Figure 7: Comparison of manual (a) and automatic (b) control point distributions.

6.1 Accuracy Assessment

Images and pose parameters from the calibrated, scanner-mounted, globally-referenced camera serve as baseline for comparing other registration algorithms. The Bryggen dataset is used for the assessment due to its simplicity and accessibility. The mounting registration is compared to poses obtained via spatial resection, the proposed manual method using the projected markers in fig. 5, and the automatic registration with optimised parameters. The comparison metrics include the image reprojection error $\Delta_{pxl}[px]$ of pre-defined markers (see fig. 5), the rotational deviation of the orientation quaternion ΔQ, and the translational deviation of the position $\Delta \vec{t}[m]$. The deviation quaternion ΔQ is split into view direction vector $\Delta \vec{q}$ and view roll $\Delta \varphi(\vec{q})$, which is similar to the mobile device screen orientation. The clamped roll $\Delta \varphi(\vec{q}), \{-\pi, \pi\}$ is measured in rad, the direction vector $\Delta \vec{q}$ is unitless. The experiment results are given in tab. 1. The measured reprojection error for the baseline, camera-mounted configuration is due to the 2D–3D correspondences being human-selected within the photograph and the original lidar pointset, which means their selection accuracy is subject to human perception limitations. The automatic registration uses the default SIFT parameters by Lowe (2004) for feature matching in this experiment.

method	$\Delta_{pxl}[px]$	$\Delta \vec{q}$	$\Delta \varphi(\vec{q})$	$\Delta \vec{t}[m]$
mounting	35.21	0	0	0
spat. resection	32.42	0.00091	0.000738	0.053
manual reg.	43.35	0.01419	0.008197	0.080
automatic reg.	28.75	0.00021	0.000028	0.034

Table 1: Comparative accuracy of camera mounting, manual- and automatic registration with respect to Δ_{pxl}, rotational errors $\Delta \vec{q}$ and $\Delta \varphi(\vec{q})$, and $\Delta \vec{t}$.

6.2 Parameterization Sensitivity

The objective of the parametrization study, carried out on the Bryggen dataset, is to assess the impact of parameter variations in the keypoint matching and pose estimation procedure. The applied algorithm for feature detection and description is SIFT. The impact of SIFT internal parameters for keypoint matching has been assessed by Sima and Buckley, whose results are also applicable to the overall pose estimation (Sima and Buckley, 2013).

First, SIFT features are extracted in the photo and the rendered image using the method's default settings. In a second step, the two closest features are matched for each keypoint of one image within the respective other image. In step three, a ratio test is applied comparing the descriptor distances of both closest features. Iff the ratio between both distances is less than ratio parameter r_d, the matching is preserved. In step four, we assure a symmetric match between both image keypoints, reducing the closest two matches to one unique mapping. In the final step, the homography transformation is extracted from the matched keypoint pairs (introducing a confidence ratio parameter c_f), and only keypoints within a distance to the epipolar line $d(P, \vec{e}) \leq d_H$ are kept (with $\vec{e}$ being the epipolar line, P being a keypoint of the image in question, and d_H being the user parameter to study).

The RANSAC-based pose estimation procedure is controlled by the minimum point reprojection error E_p, which defines the optimisation target, the target ratio of inlier points after reprojection r_p, and the maximum number of iterations. The pose optimisation ends when the target ratio of 3D control points is within the given minimum reprojection error, or the maximum number of iterations is reached. In our experiments, the maximum number of iterations is fixed to 500. Table 2 shows the assessed value range for each parameter.

parameter	value range	std. value
c_f	$[0.85 : 0.01 : 0.99]$	0.92
d_H	$[0 : 0.5 : 15]$	7.5
r_d	$[0.55 : 0.01 : 0.84]$	0.7
E_p	$[0.5 : 0.5 : 3.0]$	2.0
r_p	$[0.6 : 0.02 : 0.8]$	0.7

Table 2: Overview of studied parameters, value ranges (given in the format of $[v_0 : \Delta v : v_N]$), and empirical standard values. Byggen

The influence of each parameter on the final pose quality is studied by observing the average pixel reprojection error, the average number of detected points, and the average percentage of correct correspondences. Based on the experiments (assessing each independent parameter combination for five Nexus 5 images of Bryggen), the following behaviour can be observed:

- the confidence ratio c_f is a stable parameter with moderate influence, while generally higher ratios result in better correspondences;

- the homography distance d_H is an unstable parameter with minor influence (weak correlation);

- the feature ratio r_d is a stable parameter with significant influence (strong correlation, see fig. 8(a)), where larger values yield better estimations;

- the allowed point reprojection error E_p (within the studied boundary) is an unstable parameter with significant influence;

- the target point inlier ratio r_p is a stable parameter with minor influence (weak correlation, see fig. 8(b)), where the studied value range always resulted in a applicable pose estimates. The trend of retrieved number of points and inlier percentage depicted in fig. 8 is representative for comparable radiometric setups of registering natural scene images.

- a parameter optimization, as given in this study, can yield registration accuracy improvements on a per-dataset basis
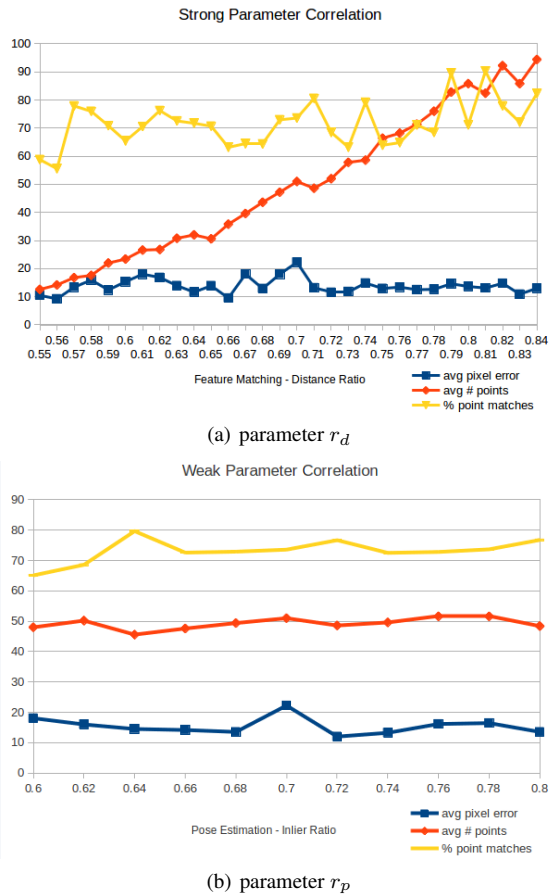
method	$\Delta_{pxl}[px]$	$\Delta\vec{q}$	$\Delta\varphi(\vec{q})$	$\Delta\vec{t}[m]$
manual reg.	31.78	0	0	0
mobile sensors	<100	0.203	0.0916	45.33
automatic reg.	9.92	0.018	0.0086	8.76

Table 3: Comparative accuracy of initial sensor data, manual- and automatic registration with respect to Δ_{pxl}, rotational errors $\Delta\vec{q}$ and $\Delta\varphi(\vec{q})$, and $\Delta\vec{t}$. Mam Tor, March 2015

when lighting conditions were significantly different to the textured model input images. Moreover, due to rainfall in preceding days, the mudstone beds in the outcrop are significantly darker in colour. Keypoint matching and pose retrieval failed with the optimised parameter set of sec. 6.2, as shown in fig. 9(a). We apply a gamma adaptation as a post-processing shader to the synthetic image because of the stronger reflectance of the rock. This improves the keypoint matching, though pose extraction still fails (see fig. 9(b)). Only by applying the post-processing and changing the keypoint extraction to MSCR - a feature extraction algorithm considering colour information instead of just greyscale values - the automatic procedure is able to retrieve an acceptable pose for 4 out of 20 images, as shown in fig. 9(c).

7. DISCUSSION

As can be seen from the measurements in sec. 6.1 and 6.3, all methods are able to retrieve accurate external camera parameters with a reprojection error of less than 1% of the diagonal image resolution, a positional error between 8 cm (for calibrated, mounted, high-resolution camera images) and 9 m (for uncalibrated, noisy, handheld mobile device images), and accurate rotational parameters. Using the raw mobile device sensor data for image registration is insufficient, as this leads to positional errors in the retrieved pose of up to 45 m within the presented experiments. In urban environments, reoccurring texture patterns within the scene lead to some incorrect point correspondences being found. However, the automatic Image-to-Geometry registration is able to retrieve a pose closest to the fixed camera mounting, yielding the smallest reprojection error of all methods.

In the geological setting, mobile device images are easily registered using the novel manual registration approach. Particularly, with a growing number of images to register, the novel workflow promises significantly less processing time due to the simple 2D-2D matching. We observed that a trained operator takes close to 20 minutes for selecting the minimum number of correspondences per image (for PnP pose estimation), where the mobile device approach demands between 5 and 7 minutes per image. The automatic registration is able to consistently estimate acceptable pose parameters for mobile devices images when the lighting conditions of the textured surface model and the acquired additional images are similar. The major issues for the automatic registration are twofold: First, the acquired GPS position data are often inaccurate using the built-in sensor. Using the GPS data demanded visual checks and minor manual corrections before running the automatic registration. Apart from insufficient GPS reception and noise, a major error source (and thus source of correction) is the lack of a built-in geoid model for referencing altitude measurements relative to the orthometric textured model. In the future, this can be bypassed by including a geoid model or snapping the height value to a local elevation model. Secondly, changing lighting conditions are largely problematic to feature-based pose estimation, where the use of colour feature descriptors and appropriate post-processing shaders improve the process to a certain degree. Still, more significant improvements

(a) parameter r_d

(b) parameter r_p

Figure 8: Graphs showing the difference between strongly-correlated parameters (such as r_d in (a)) and weakly-correlated parameters (such as r_p in (b)) by observing average pixel errors, number of detected features and ratio of correct correspondences.

6.3 Lighting Sensitivity

Using the Mam Tor dataset, we can assess the impact of lighting conditions on the automatic registration, as two image sets, captured in different field campaigns (March and September 2015), are available. Lighting conditions have a major impact on the registration due to the feature descriptor extraction and matching. In the first assessment, we use the March dataset, which was captured simultaneously with the textured model in identical lighting conditions. The measured metrics are equal to Section 6.1, but the comparison baseline is the manual mobile registration. The results are presented in tab. 3.

The automatic procedure is able to retrieve acceptable poses for all ten images, acquired via mobile devices, at equal lighting conditions. The September dataset consists of 20 images captured
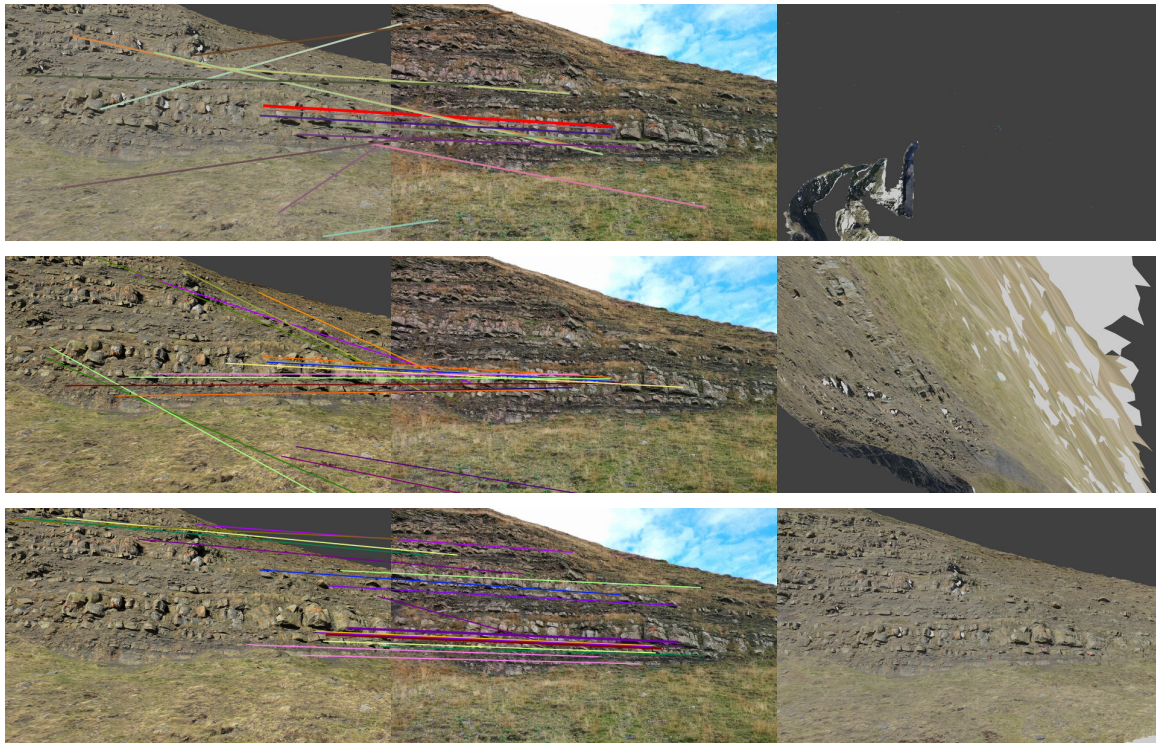
Figure 9: Results of pose estimation for changing lighting conditions for SIFT with default parameters (top), SIFT-optimised (middle) and MSCR-optimised (bottom) keypoint matching. Each example shows the mobile sensor-based rendering (left), the captured mobile image (middle), and their 2D correspondences, and a rendering from the calculated pose (right). With successful pose estimation, the viewport of the right-hand image should match the input photograph (middle).

are potentially to be gained by using the available geometry information for combined statistical pose estimation (Sottile et al., 2010).

The resulting exterior orientation parameters of the presented methods can be used in the projective mapping of areas of interest, such as sediment bedding and faults in geological applications (see fig. 10), as well as for adding new images to the existing textured models (fig. 11). This can be useful where initial texture quality is poor (lighting conditions, obliquity or low resolution) or 2D features have changed since the initial 3D model acquisition (new coat of paint for the urban environment). For the latter, the registration procedure is flexible enough to also register handheld, consumer-grade or SLR camera images to a given textured model (on desktop computers), thus replacing the need for a lengthy manual registration of additional photographs: a user can manually define a small number of coarse overview poses and link them to the freely-captured images. With these coarse initial pose estimates, the proposed automatic Image-to-Geometry procedure is able to automatically register the images, leading to a significant reduction in manual input for texturing workflows (Sima, 2013).

(a) 2D annotation, Nexus 5 image   (b) 3D annotation, projected on model
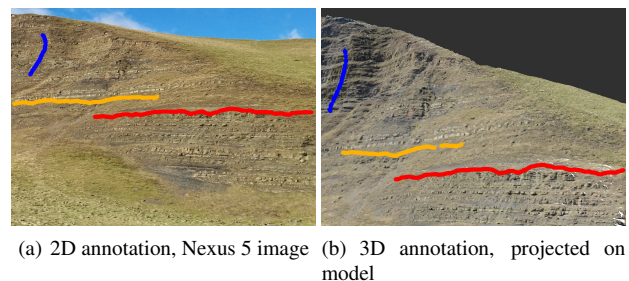
Figure 10: Geological annotation mapping of a virtual outcrop model from mobile device imagery.

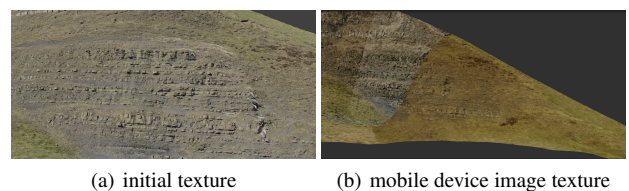(a) initial texture   (b) mobile device image texture

Figure 11: Re-texturing a textured model with mobile device (Nexus 5) image collection.

ACKNOWLEDGEMENTS

and Riegl LMS GmbH is thanked for continued software support.

REFERENCES

Bay, H., Tuytelaars, T. and Van Gool, L., 2006. SURF: Speeded Up Robust Features. In: *Computer vision–ECCV 2006*, Springer, pp. 404–417.

Blum, J. R., Greencorn, D. G. and Cooperstock, J. R., 2013. Smartphone sensor reliability for augmented reality applications. In: *Mobile and Ubiquitous Systems: Computing, Networking, and Services*, Springer, pp. 127–138.

Bodensteiner, C., Hebel, M. and Arens, M., 2012. Accurate single image multi-modal camera pose estimation. In: K. Kutulakos (ed.), *Trends and Topics in Computer Vision*, Lecture Notes in Computer Science, Vol. 6554, Springer Berlin Heidelberg, pp. 296–309.

Brunnett, G., Hamann, B., Müller, H. and Linsen, L., 2003. *Geometric Modeling for Scientific Visualization*. Mathematics and Visualization, Springer Berlin Heidelberg.

Buckley, S. J., Howell, J. A., Enge, H. D. and Kurz, T. H., 2008. Terrestrial laser scanning in geology: data acquisition, processing and accuracy considerations. *Journal of the Geological Society* 165(3), pp. 625–638.

Buckley, S. J., Schwarz, E., Terlaky, V., Howell, J. A. and Arnott, R., 2010. Combining Aerial Photogrammetry and Terrestrial Lidar for Reservoir Analog Modeling. *Photogrammetric Engineering & Remote Sensing* 76(8), pp. 953–963.

Chen, B., Ramos, G., Ofek, E., Cohen, M., Drucker, S. and Nistér, D., 2008. Interactive techniques for registering images to digital terrain and building models. Technical report, Technical report, Microsoft Research.

Corsini, M., Dellepiane, M., Ganovelli, F., Gherardi, R., Fusiello, A. and Scopigno, R., 2013. Fully Automatic Registration of Image Sets on Approximate Geometry. *International journal of computer vision* 102(1-3), pp. 91–111.

Feixas, M., Sbert, M. and González, F., 2009. A unified information-theoretic framework for viewpoint selection and mesh saliency. *ACM Transactions on Applied Perception (TAP)* 6(1), pp. 1.

Forssén, P.-E., 2007. Maximally Stable Colour Regions for Recognition and Matching. In: *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, IEEE, pp. 1–8.

Howell, J. A., Martinius, A. W. and Good, T. R., 2014. The application of outcrop analogues in geological modelling: a review, present status and future outlook. *Geological Society, London, Special Publications* 387, pp. SP387–12.

Lepetit, V., Moreno-Noguer, F. and Fua, P., 2009. EPnP: An Accurate O(n) Solution to the PnP Problem. *International Journal of Computer Vision* 81(2), pp. 155–166.

Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60(2), pp. 91–110.

Maes, F., Collignon, A., Vandermeulen, D., Marchal, G. and Suetens, P., 1997. Multimodality image registration by maximization of mutual information. *Medical Imaging, IEEE Transactions on* 16(2), pp. 187–198.

Moffitt, F. and Mikhail, E., 1980. *Solutions Manual to Accompany Photogrammetry*. Harper & Row.

Nöll, T., Pagani, A. and Stricker, D., 2011. Markerless camera pose estimation-an overview. In: *OASIcs-OpenAccess Series in Informatics*, Vol. 19, Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

Pintus, R., Gobbetti, E. and Combet, R., 2011. Fast and robust semi-automatic registration of photographs to 3d geometry. In: *Proceedings of the 12th International conference on Virtual Reality, Archaeology and Cultural Heritage*, Eurographics Association, pp. 9–16.

Pluim, J. P., Maintz, J. A. and Viergever, M. A., 2003. Mutual-information-based registration of medical images: a survey. *Medical Imaging, IEEE Transactions on* 22(8), pp. 986–1004.

Potenziani, M., Callieri, M., Dellepiane, M., Corsini, M., Ponchio, F. and Scopigno, R., 2015. 3dhop: 3d heritage online presenter. *Computer & Graphics* 52, pp. 129–141.

Quan, L. and Lan, Z., 1999. Linear n-point camera pose determination. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 21(8), pp. 774–780.

Rarity, F., Van Lanen, X., Hodgetts, D., Gawthorpe, R., Wilson, P., Fabuel-Perez, I. and Redfern, J., 2014. Lidar-based digital outcrops for sedimentological analysis: workflows and techniques. *Geological Society, London, Special Publications* 387(1), pp. 153–183.

Semmo, A. and Döllner, J., 2015. Interactive Image Filtering for Level-of-Abstraction Texturing of Virtual 3D Scenes. *Computers & Graphics* 52, pp. 181–198.

Sima, A. A., 2013. An Improved Workflow for Image- and Laser-based Virtual Geological Outcrop Modelling. Phd thesis, University of Bergen. 180p.

Sima, A. A. and Buckley, S. J., 2013. Optimizing sift for matching of short wave infrared and visible wavelength images. *Remote Sensing* 5(5), pp. 2037–2056.

Sottile, M., Dellepiane, M., Cignoni, P. and Scopigno, R., 2010. Mutual correspondences: An hybrid method for image-to-geometry registration. In: *Eurographics Italian Chapter Conference*, pp. 81–88.

Tsai, R. Y., 1987. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *Robotics and Automation, IEEE Journal of* 3(4), pp. 323–344.

Van Damme, T., 2015. Computer vision photogrammetry for underwater archaeological site recording in a low-visibility environment. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 1, pp. 231–238.

Viola, P. and Wells, W. M., 1997. Alignment by maximization of mutual information. *International journal of computer vision* 24(2), pp. 137–154.

---

[4]Sedimentary Architecture of Field Analogues for Reservoir Information (SAFARI) - `www.safaridb.com`

[5]Virtual Geoscience Conference - `www.virtualoutcrop.com/vgc2016`