**Image-to-Geometry Registration on Mobile Devices – Concepts, Challenges and Applications**

Melanie Kröhnert\*, Christian Kehl†‡, Herbert Litschke₮, Simon J. Buckley‡

\*Institute for Photogrammetry & Remote Sensing, TU Dresden, Helmholtzstr. 10, D-01069 Dresden, melanie.kroehnert@tu-dresden.de

†Aix Marseille Uniersité, CNRS, IRD, CEREGE UM34, Sedimentary Systems and Reservoir Development, Marseille, France

₮Hochschule Wismar, Philipp-Müller-Straße 14 , D-23966 Wismar

‡Uni Research AS – CIPR, Nygårdsgaten 112, NO-5008 Bergen

**Abstract.** text text text text text text text text text text text text text text text text text (Arial 8)

**1 Introduction**

Considering the worldwide distribution of 5.3 billion unique mobile subscriptions with a smartphone percentage of 56 % compared to global population of ~7.5 billon people [Ericsson,2017], a life without smartphones seems to be not imaginable nowadays. Smartphones with inbuilt cameras, powerful processing units and low-cost positioning systems seem to be very suitable wide-spread measurement devices that could be used for mobile mapping, measuring and visualisation purposes [Kröhnert2017]. Here, image-to-geometry intersection describes an essential topic for the translation of mobile captured image data into object space which allows for metric interpretation on call. Section 2 illustrates state-of-the-art concepts for solving the issue of image-to-geometry intersection whereas section 3 depicts challenges regarding the precise determination of camera´s intrinsic and extrinsic parameters and points out the issue of their geometric stability. Furthermore, difficulties related to image mapping under natural illumination in comparison to required accuracies for measuring purposes and performance are addressed. Section 4 and 5 treat existing applications for geo-monitoring and end with a short outlook of current research activities.

**2 Concepts**

The group of algorithmic concepts for registering images to available 3D object surfaces that has been demonstrated to work on outdoor cases in the past decade consists of mutual information (MI) [Guislain2016], horizon alignment [Baboud2011,<INSERT\_CITATION\_HERE], edge correlation [Boerner2016, Meierhold2009], point feature-based registration and hybrids thereof [Sottile2010].

MI [Maes1997] mutually registers two images (e.g. one synthesized from the available object shape) by calculating the joint histogram and derives thereof measures for information or entropy. Subsequently, candidate images are generated for adjacent vantage points, their mutual information is compared with the previous image, and the candidate maximising the shared information is taken as new starting point. The procedure is an iteratively maximises the MI until equilibrium. The concept is popular for outdoor- and indoor cultural heritage case studies [Corsini2009a, Corsini2013, Dellepiane2013]. The concept is not easily transferable to mobile devices as it relies on non-standard optimisation methods.

Wide-baseline outdoor images can benefit from horizon information. Baboud et al. [Baboud2011] presented an approach to localise register arbitrary outdoor photographs to an available terrain model by [...]. The technique was demonstrated on an alpine scenario. Despite the outdoor demonstration, the appearance of clear horizon separation is not guaranteed in actual geosciences field studies. Even outdoor cultural heritage applications demand close-up photography without horizon depiction.

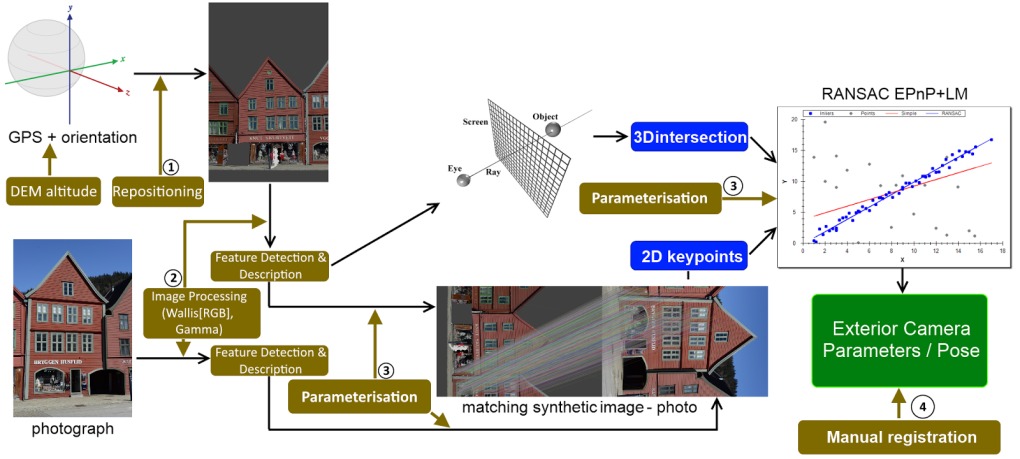
[ADD HERE EDGE ALIGNMENT OPTIONS]

The common approach for image registration onto coloured geometry on mobile devices is based on salient feature points of synthesised images and photos [Bodensteiner2012, Meek2013, Stelling2010]. The 2D features are supplemented with 3D information that can be used in a Point-n-Perspective (PnP optimization [Lepetit2009, Li2012], optionally refined via Levenberg-Marquardt [Madsen2004]. The whole registration process is illustrated in fig. 1. Research groups across domains, such as augmented reality [Gauglitz2014, Sweeney2015, Nuernberger2016], outcrop geology [Kehl2015a, Kehl2016b] and hydrology [Kröhnert2017, Eltner2017, Kröhnert2016], utilise this approach in mobile applications for localisation, tracking and interpretation purposes. Furthermore, the approach is integrated in common high-level concepts, such as simultaneous localisation and mapping (SLAM) and visual odometry [Caron2014], which are implemented on mobile systems, e. g. Google’s Project Tango [Google2016]. For applications with actual discrete geometry as surface representation, the 3D coordinates are supplied by raycast-intersections between the vantage point, the 2D point coordinate in the camera plane and the triangulated irregular network (TIN).

Modern applications increasingly use a point set representation of the object’s surface as this is directly provided by various 3D object scanning techniques. In these cases, the 3D coordinate computation needs to be adapted as rays cannot be intersected with zero-extent points. [VARIATION FOR F-B APPROACHES ON POINT SET/CLOUDS].

At last, hybrid (i.e. multi-modal) approaches exist that combine several of the above-outlined methods in one. One example is mutual correspondence, which combines the mutual information maximisation and point re-projection distance minimisation in a weighted manner [Sottile2010].

The remainder of this article focuses on feature-based registration as currently predominant concept on mobile devices due to its simple implementation, easy adaptability, generic applicability (e.g. in contrast to horizon alignment) and the achievable performance on even low-power devices.



**Fig. 1.** Main workflow for feature-based, mobile sensor-assisted image-to-geometry registration.

**3 Challenges**

Despite being tested on synthetic- as well as domain-specific case studies (e.g. cultural heritage, hydrology, geology) in past years, the available methods, in particular feature-based registration, are still far away from being fully-automatic. Applying the previous concepts still presents distinct challenges in real-world scenarios and for mobile device platforms, which are discussed in this section.

3.1 Camera intrinsics

Both, photogrammetric measurements and computer vision that use cameras for data acquisition for image processing tasks like visual odometry, 3D modelling or even monitoring should deal with the issue of camera calibration. Camera calibration comprises the determination of its intrinsic (principle point, focal length and skew) and several lens distortion parameters (e.g. radial and tangential distortion). To solve the issue of calibration, numerous techniques for camera modelling are well-known [Tsai1987, Zhang2000, Brown1971] and not further addressed in this paper.

In case of typical consumer cameras like digital single lens reflection cameras (DSLR), the calibration is valid for one camera setting concerning focal aperture, length and focus adjustments and can be fixed easily by avoiding manual refocusing and aperture tuning. Today’s smartphones and tablets are largely equipped with inbuilt autofocussing cameras which are used by several apps (camera application, QR detection, AR games, etc.). Even if a camera application uses manual focussing, the camera will be refocused during each app start and affects the intrinsic (VERWEIS AUF DEHLING ARBEIT 🡪 AUFSTART EFFEKT). If a mobile smartphone camera should be used for measurements and thus should be calibrated, the mentioned problem can only be avoided if calibration and data acquisition are consecutive without closing the app. However, when running the camera app continuously the battery runs out quickly. Beside this, the devices temperature changes very fast when other apps, MEMS or GPS are started or closed from background processing tasks. BEISPIEL BRINGEN AIDA64

Obviously, this must influence the small inbuilt camera sensors and lenses due to hardware assembling and adhesive bond. VERWEIS AUF DEHLING ARBEIT 🡪 TEMP EFFEKT

3.2 Location- and Orientation Sensor Data Quality

The majority of today’s smartphones share inbuild microelectromechanical systems (MEMS) for orientation tasks from simple screen orientation determination up to navigation purposes. Commonly, MEMS comprise 3-axis accelerometers, 3-axis gyroscopes, magnetometers and gravity sensors. Barometers are increasingly integrated. It seems to be obvious that this kind of low-cost inertial measurement units (IMUs) cannot be comparable in resolution and stability due to lowest production cost. For Apple’s iPhone and Samsung’s Galaxy series (until 2014) it should be noted, that sensors share less than 5 % of the production costs and range between 1.60 and 7.00 USD (Lewin, 2014). In comparison to that, even light IMU’s for e.g. airborne applications like drone navigation amount to several hundred dollars. Nevertheless, due to various kinds of orientation sensors, sensor fusion and filtering help to solve the issue of noisy accelerometers and drifting gyroscopes that accumulate their errors respectively over time. Using their complementary characteristics, sensor fusion and implementation of filtering approaches could improve orientation accuracy and stability significantly (Kok et al., 2017, Pacha2015).

Figure 2 shows stability test results of the azimuth, pitch and roll angles remapped to landscape mode using the Android smartphone Samsung Galaxy S8 and a Kalman filtered fusion of the inbuild accelerometer, gyroscope and compass combined with the calibrated gyroscope (implementation according to Pacha2015). Furthermore, the azimuth includes magnetometer information pointing to the geographic north after correction by declination. Up to this, the smartphone was mounted on a tripod and installed apart from magnetic impurities like other smartphones or computers.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| rollpitchyaw | Time [sec] | σAzimuth [°] | σPitch [°] | σRoll [°] |
| 5 | 0,00572 | 0,00091 | 0,00205 |
| 10 | 0,00596 | 0,00301 | 0,00192 |
| 15 | 0,00605 | 0,00326 | 0,00417 |
| 20 | 0,00746 | 0,00681 | 0,00677 |
| 25 | 0,00860 | 0,01019 | 0,00891 |
| 30 | 0,01010 | 0,01108 | 0,01078 |
| 35 | 0,01186 | 0,01037 | 0,01210 |
| 40 | 0,01311 | 0,00985 | 0,01270 |
| 45 | 0,01291 | 0,01025 | 0,01222 |
|  | | | | |

**Fig. 2.** Stability test for orientation assessment using a Samsung Galaxy S8 smartphone and sensor fusion [Pacha2015] for increasing precision and stability regarding azimuth, pitch and roll angles.

Registration setups with a global reference frame (i.e. not based on motion- and temporal sensor correlation) rely on common GPS data and absolute, geomagnetic orientation. Magnetic orientation is measurably influenced by magnetic impurities in close proximity, as often found in urban areas [Blum2013], as shown in fig. 3. This inhibits correct orientation in 3D inside cities, but for planar orientation it is less problematic. In most outdoor applications [Novakova2017, Cawood2017] and especially for 3D Image-to-Geometry registration, the sensor accuracy currently needs to be assessed on a per-case basis (see [Kehl et al 2016]). The GPS accuracy can be improved in urban areas via terrestrial network connection [Wang2012], which demands WiFi access. When dealing with raw GPS data, the location accuracy drops significantly. Lateral errors of up to 8 metres and vertical errors of tens of metres are realistic outdoor GPS limits (see Kehl et al. 2016 and Kehl et al. 2017c).

|  |  |
| --- | --- |
| headingShit | headingCorrect |

**Fig. 3.** Side-by-side comparison of geomagnetic heading readings from field compass and sensor-calibrated mobile device. The field compass is manufactured aluminium so to not disurb the measurements. The vast difference in mobile device accuracy stems from the iron bars below the wooden table (left), while remote places on bare rock provide satisfying results (right).

Resolving the geo-positioning accuracy limitation in the future may be a result of two major changes: (a) dropping differential GPS prices and (b) the computationally more manageable integration of real-time kinematics (RTK) and temporal sensor filtering (similar to sensor fusion for IMU approaches [Ligorio2013]) into the sensor software framework within Android. Currently, a comprehensible, user-driven re-positioning via DEMs resolves drastic sensor errors occurring outdoors [Kehl2017c]. These can be obtained via open-data media, e.g. Digital Earth Explorer[[1]](#footnote-1). Globally improving the sensor accuracy is the focus of intense investigations at the moment.

3.3 Image Mapping under Natural Illumination

Feature-based registration relies on an unambiguous, robust point-to-point correlation via distinctive feature vectors. As such, it is affected by distortions and content appearance changes, referred to as geometric- and radiometric variance. Radiometric variance, caused by environmental effects and natural illumination changes, still are still problematic for feature correlation. Recent studies have shown that some combinations of already-available techniques are robust against geometric- and radiometric distortion [Kehl2017]. Still, feature descriptors that account for the radiometric variance are in high demand. A major contribution to this research track would be the integration of local colour attributes to the feature space, as formerly used for image classification methods [Khan2013, VanWeijer2006].

****

**Fig. 4a.** Feature detection – MSER, descriptor – SIFT, matcher – BruteForce L1. Number of inlier matches: 11, distribution of matches: image centre, large radius.



**Fig. 4b.** Feature detection – SIFT, descriptor – SIFT, matcher – BruteForce L1. Number of inlier matches: 186, distribution of matches: image centre, small radius.

[IMAGE: DIFFERENCE BETWEEN SIFT AND MSCR-SIFT MATCHING

Melanie, perhaps you may want to add this example picture from your river-in-front-of-house scene here ?]

Water level monitoring using smartphones. Image-to-Geometry intersection for camera intrinsic and extrinsic parameter determination. MSER: less inliers, good distribution; SIFT more inliers, unstable distribution focussed on one single region.

Recent trends employ neural networks to train feature descriptors [XYZ], which appears to be a promising research direction. Furthermore, probabilistic pattern learning models (e.g. [..., XYZ]) may provide a selection of quickly-computable feature descriptions of which the highest-correlating model can be chosen for the registration procedure afterwards.

3.4 Accuracy Requirements in general

text text text text text text text text text text text text text text text text text text text text text text text text

text text text text text text text text text text text text text text text text text text text text text text text text

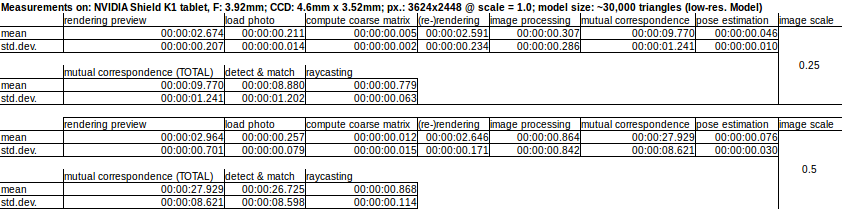
text text text text text text text text text text text text text text text text text (Arial 9, Blocksatz).

3.5 Performance Requirements

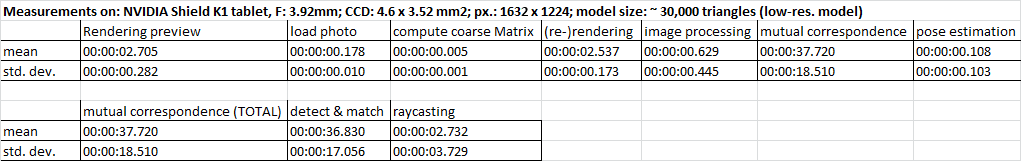
The methods being employed for image-to-geometry registration on mobile devices are depended on the required performance (i.e. computation) time of the algorithm as well as the methodological constraints of the application domain. While mobile graphics and computations made significant advances in recent years (see [Garcia2015, Kehl2015a, Agus2017]), a major problem is to scale up lab-sized results (with respect to 3D model- and image size) to actual application demands. This is because mobile devices are memory-limited – in strict contrast to most desktop- and laptop computing platforms.

In cases where the application constraints allow to use the mobile device as plain input sensor and output presentation platform, it is common to use the WiFi connection for image- and sensor data transmission while the actual processing is being done on a remote server. Examples for this approach can be found in SfM reconstruction [Fritsch2015], mobile rendering [Ponchio2016]. [MELANIE: HOW DO YOU EMPLOY THE WEB ARCHITECTURE? – or is it currently only ‘in-plan’]. The possibility of using network connectivity also reduces the energy consumption of the registration process on the mobile devices itself, which makes sensor tracking more viable for increasing the location- and orientation accuracy. The specific challenge is then to define a trade-off between network transmission load and tasks that are done locally on the device.

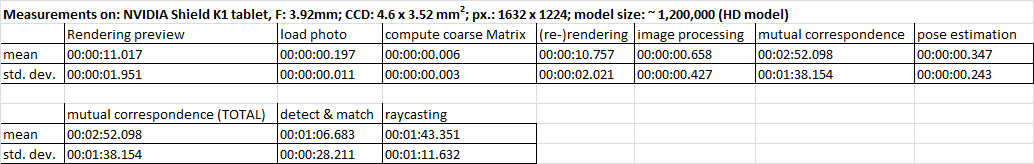
In cases where network connectivity is not available or not being used, performance is a much more limiting factor of what can algorithmically be achieved. The computational complexity of feature-based geometry is linked to the image resolution (see fig. 3-5) and size of the 3D surface model. The effect is commonly mitigated on desktop hardware due to CPU vectorisation, SSE instructions and GPU-based image filtering, which are not available on mobile hardware architectures. Lower-resolution images speed up the calculation but also result in less-accurate feature matching in general. Furthermore, the rendering of the 3D model – marginally contributing to the computation runtime on desktops – is slow and restrictive (in terms of model size) on mobile devices, meaning that it contributes majorly to the algorithmic runtime (see table XYZ). The computational complexity relates to a given energy budget used to register an image (see tablet XYZ), which is a drawback for extended utilisation of the technology for some application domains.



**Tab. 1.** Computation performance, comparing different image resolutions scales and their computational impact while keeping the model size fixed.



**Tab. 2.** Computation performance for registering half-resolution images to low-resolution models (running on the majority of tablets interactively)



**Tab. 3.** Computation performance for registering half-resolution images to high-resolution models, only facilitated by tablets with dedicated GPUs (e.g. NVIDIA Tegra chipsets)

ZUKÜNFTIG TESTS AUF NEXUS5

3.6 Device Variability (should be at beginning?! – yes, we can do that.)

Melanie, please add here things you may want to address in respect to device variability

Text text text text text text text text text text text text text text text text text text text text text text text text

text text text text text text text text text text text text text text text text text text text text text text text text

A last point variable between the given Android device range concerns its graphics computing capabilities. The rendering of a given 3D model is done on the GPU via OpenGL. Apart from the natural hardware performance differences, the employed graphics chips (e.g. Qualcomm Adreno, ARM Mali, NVIDIA Tegra) on support different rendering instructions, in particular with texture support. For textured surface models (often used in feature-based matching), on-chip texture decompression makes a significant difference in rendering speed (supported by Mali and Tegra). Also, whereas the majority of tablet brands use Qualcomm’s system-on-a-chip (SoC) architecture, where CPU and GPU share the same memory, other device provide dedicated graphics memory that speeds up rendering significantly. On top of the pure rendering-related differences, some graphics processors (e.g. Mali and Tegra) provide non-graphics related GPU Computing capabilities via OpenCL and CUDA, which allows for drastic runtime reductions for future, optimised image-to-geometry systems [Heymann2007, Hudelist2014]).

**4 Applications**

Volunteered Geographic Information (VGI) becomes growing attention concerning multiple areas of geosciences. Section 4 highlights two important topics which could be highly improved by complementary VGI. Crowdsourcing seems to be very helpful for mobile data acquisition in case of “geo-events” (like floods or landslides) with high spatial and temporal resolution where they are needed most. Thus, it is obvious that captured data by smartphones -comprising image, position and orientation data- must be fused and translated from the device to the object space. This is where the mobile image-to-geometry intersection comes in. In the following, two major fields are presented that profit from mobile data acquisition depending on the mentioned topic.

4.1 Hydrology

Regarding worldwide increasing flood hazards, there is an enormous increase in the importance of river monitoring including flow velocities, water levels and river cross sections. Conventional gauging stations are most solely installed which would lead to an insufficient coverage of hydrographic data when they are most needed – like sudden flood events. Even small running water catchments can turn in devastating streams that pose threat to the environment and to human health. With aid of the public, mobile data could be acquired quickly and could be used for spatio-temporal densification of hydrologic data [Kroehnert2017, Eltner2016, Kroehnert2016]. Short time lapse smartphone image sequences with known initial pose and orientation, could be used to determine the river line that could be further intersected with prevailing object data. In that way, the detected shore could be transferred into several water levels . For this, the shore area should be visible inside of the image and must provide sufficient information for image-to-geometry intersection in terms of buildings, stones or other artificial objects (see chapter 2).

4.2 Petroleum Geology

The feature-based registration on textured triangle meshes is integrated in the Geological Registration and Interpretation Toolkit (GRIT), a 2D-3D mobile application for smartphones and tablets to study geological rock exposures (i.e. outcrops). Geological studies include several purposes ranging from sedimentary architecture reconstruction (e.g. SAFARI project [Dreyer1993, Howell2014]) over structural studies for flow analysis [Lamarche2012] to structural studies for geothermal prospect evaluation. In the case of GRIT, the technology has been applied on sedimentology case studies at Mam Tor, Derbyshire, UK [Kehl2016a, Kehl2017a] and an oil reservoir reservoir analogue study at the Saltwick Formation, Whitby, North Yorkshire, UK [Kehl2017c] (fig. 6).

[IMAGE 3D INTERPRETATION OF MAM TOR – already shown sooo often, perhaps rather show something new]



**Fig. 6.** 3D-registered interpretations of the Saltwick Formation (North Yorkshire, UK) geological case study highlighting the sandstone channel architecture of the North Sea cliff section outcrop.

A case study at Calvisson, department Gard, France was used to gain insight into fracture networks [Bisdom2014] and their interconnectivity to more reliably predict multi-phase flow (i.e. combined flow of multiple, heterogeneous fluids) in subsurface reservoirs with strong deformation patterns [Lamarche2017]. Further investigations are planned in the soon future.

Additionally, the mobile device application is envisaged to be used on further geothermal research studies in Mexico, as well as sedimentary studies on Spitzbergen, Norway for carbon capture- and storage (CCS) projects of local coal power plants (see Senger et al. [Senger2015] for initial studies on the subject).

**Acknowledgements**

The VOM2MPS (i.e. the research of Uni Research-affiliated authors) is funded as a Petromaks 2 project (no. 234111/E30) by the Research Council of Norway and the Force consortium, while data (e.g. Yorkshire datasets) are provided by SAFARI ([www.safaridb.com](http://www.safaridb.com)). Data from Calvisson where collected and provided through an industry-funded research project by TOTAL.

Gratefully, acknowledge is given to the European Social Fund (ESF) and the Free State of Saxony for their financial support on a grant.

**References**

1. Ericsson (2017). Ericsson Mobility Report. August 2017. Stockholm: Ericsson. Retrieved September 02, 2017 from

https://www.ericsson.com/assets/local/mobility-report/documents/2017/ericsson-mobility-report-interim-update-august-2017.pdf

2. Kröhnert, M. & Meichsner, R. (2017). Segmentation of environmental time lapse image sequences for the determination of shore lines captured by hand-held smartphone cameras. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., IV-2/W4, 1-8, https://doi.org/10.5194/isprs-annals-IV-2-W4-1-2017, 2017.

3. Heymann, S., Müller, K., Smolic, A., Fröhnlich, B. & Wiegand, T. (2007). SIFT Implementation and Optimization for General-Purpose GPU. Winter School of Computer Graphics (WSCG). Eurographics.

4. Lewin, S. (2014). The smartphone recipe. IEEE Spectrum, 51(12): p 80.

5. Kok, M., Hol, J. D., & Schön, T. B. (2017). Using inertial sensors for position and orientation estimation. arXiv preprint arXiv:1704.06053.

6. Pacha, A. (2015). Sensor Fusion for Robust Outdoor Augmented Reality Tracking on Mobile Devices. GRIN Verlag, USA.

7. Castro Degrossi, L., Porto de Albuquerque, J., Restrepo-Estrada, C. E., Mobasheri, A., & Zipf, A. (2017/08). Exploring the geographical context for quality assessment of VGI in flood management domain. Conference Paper. 23rd Americas Conference on Information Systems. Boston, USA.

8. Eltner, A., Sardemann, H., Kröhnert, M. & Schwalbe, E. (2017). Camera based low-cost system to monitor hydrologic parameters in small catchments. EGU General Assembly Conference Abstracts 19.

9. Kröhnert, M. (2016). Automatic waterline extraction from smartphone images. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLI-B5, 857-863.

10. Boerner, R. & Kröhnert, M. (2016). Brute force matching between camera shots and synthetic images from point clouds. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLI-B5, 771-777.

11. Stelling, N.; Spehr, B.; Schilling, A.; Maas, H.-G.; Gumhold, S. (2010). Automatic feature matching between digital images and 2D representations of a laser scanner point cloud. In: International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences Vol. 38, Part 5.

12. Meierhold, N. & Schmich, A. (2009). Referencing of images to laser scanner data using linear features extracted from digital images and range images. In: International Archives of Photogrammetry, Remote Sensing and Spatial Information Science Vol. XXXVIII-3/W8.

13. Brown, D. C. (1971). Close-range camera calibration. In Photogrammetric Engineering, 37(8), 855-866.

14. Zhang, Z. (2000). A flexible new technique for camera calibration. IEEE Transactions on pattern analysis and machine intelligence, 22(11), 1330-1334.

15. Tsai, R. (1987). A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. IEEE Journal on Robotics and Automation, 3(4), 323-344.

1. Digital Earth Explorer – https://earthexplorer.usgs.gov [↑](#footnote-ref-1)