Image-to-Geometry Registration on Mobile Devices – An Algorithmic Assessment

Christian Kehl[1,2] and Simon J. Buckley[1]

[1] UniResearch – Centre for Integrated Petroleum Research (CIPR),
P.O. Box 7810, N-5020 Bergen, Norway
{Christian.Kehl, Simon.Buckley}@uni.no
[2] University of Bergen – Department of Earth Sciences,
P.O. Box 7803, N-5020 Bergen, Norway
Christian.Kehl@geo.uib.no

Abstract: Although mobile image are ubiquitously available, their mapping to geometry without user intervention remains a challenge. The image mapping to existing geometry is used for texturing and annotation mapping across disciplines. The article presents a novel registration extension using mobile sensor data and details its full implementation on mobile platforms. The algorithm's accuracy and performance is compared to various existing Image-to-Geometry techniques in an urban- and geological setting. The achieved performance and accuracy demonstrate suitability of registration approach for field use, although improvements in GPS accuracy are necessary for some application purposes.

1 Introduction

Image-to-Geometry registration is an important step Texture Mapping and 3D Model annotations, which is used in applications for hydrocarbon reservoir modeling [1], archaeology [2], and city modelling [3]. Detailed models geometry within these domains can be captured with structured light, lidar or photogrammetry. Apart from the photogrammetric case, photographic colour information is captured separately from the geometric model. The photographs need to be accurately registered with the 3D geometry, which can be a challenging and time-consuming manual task. Real-world photographs are ubiquitously available due to advances in mobile imagery. This article addresses the challenge of semi- and fully automatic Image-to-Geometry registration for unreferenced photographs, focussing particularly on mobile device images. The result of this registration is a 3D pose (also known as extrinsic camera parameters), containing position and orientation in the coordinate system of the 3D geometry.

In Image-to-Geometry registration, we can distinguish between manual/semi-automatic (i.e. user-guided) and fully automatic methods. User-guided methods typically extract the pose automatically with manually-defined 2D-3D correspondences. The correspondences can be refined iteratively to improve the pose accuracy. The largest issue for these methods is the selection of 3D features by the user with given tools (e.g. Riegl RiSCAN, MeshLab), due to 3D interaction, object occlusion and scale.

Automatic pose estimation is typically a multi-stage process, consisting of a coarse registration (which can be obtained manually) and a fine registration that uses global optimization methods. The accuracy and success of automatic registration methods typically rely on good initial estimates, which can be difficult to obtain.

Our approach is based on the exploration of mobile device technologies, using the mobile camera and the range of integrated sensors for localization (e.g. GPS) and orientation. These sensors can provide the initial pose approximation for refining and automatically registering a captured photograph to a given 3D model.

The goal of this study is to propose an extension for automatic Image-to-Geometry registration using mobile device sensors. The obtained results are compared to given (manual) registration techniques with respect to accuracy and execution speed, using two case studies. Both registration techniques are implemented and tested on tablets and mobile phones.

2 Related Work

The overview of existing 2D-3D registration techniques is split in user-guided (manual) and (semi-) automatic approaches. We refer to Nöll et al. [4] for a survey on from *Augmented Reality* (AR) registration techniques. Although pose estimation for texture- and annotation mapping and AR seem to have much in common at first, the major difference is the inability to calibrate the environment with targets. The focus of our study is the mapping of natural scenes, with challenges coming from changing environmental conditions and complex real-world object textures.

a) User-Guided Image-to-Geometry Registration

One coarse, traditional approach for manual 2D-3D registration is the definition of correlated projection planes with a perspective 3-point algorithm [5]. The algorithm is very sensitive to noise and inaccuracies because it retrieves the 3 orientation- and 3 position parameters analytically. Generally successful registration procedures rely on numerical optimization methods using multiple points.

Multi-point based, user-guided 2D-3D registration procedures require the definition and correlation of controlpoints between the image and the geometry. Improved methods for the adapt user to define such points and correlations are rarely considered in the literature, although perceptual problems with 3D orientation and correlations are known.

After establishing 2D-3D relations the external camera parameters can be inferred using analytical methods for a *Perspective-n-Point* (PnP) solution (i.e. space resection and *Direct Linear Transform* (DLT)) from the correlations. Tsai's method for camera calibration includes the estimation of external camera parameters and defines a lower bound of necessary correlations [6]. Using a non-linear solver, such as the *Levenberg-Marquardt* (LM) method, has shown to be a good starting point for pose estimation [7]. Recent advances towards closed-form solutions for the PnP problem, as shown by Lepetit et al. [8], robustly estimate pose parameters in linear time constrains. Common to all user-guided methods is their sensitivity to noise, controlpoint- and camera calibration inaccuracies.

b) (Semi-) Automatic Image-to-Geometry Registration

Automatic Image-to-Geometry registration is a multi-stage process of coarse- and fine registration. *Structure-from-Motion* (SfM) techniques reconstruction [9] - and *Iterative Closest Points* (ICP) registration [5] techniques work on small and mid-scale models. This scale limitation and the large computational effort are unsuitable for registering natural landscapes, which is the goal of our study.

Mutual Information (MI) [10] [11] can be used to render a 3D scene and visually match (and register) the scene with a given real-world photograph, as proposed by Corsini et al. [12]. GPU rendering capabilities allow generating scale-independent synthetic scenes very rapidly. A challenge of obtaining good initial estimates is currently solved via user interaction. Mobile sensors may replace the manual intervention in the registration approach.

Another method, analogous to manual registration techniques, relies on the point matching of 2D features and 3D extremal points [13]. Drawbacks of this approach are the complex feature descriptor mapping and the noise sensitivity (both for 3D extremal points and 2D image keypoints).

Mutual Correspondences (MC) procedures take inspiration from the synthetic scene generation of MI 3D registration, the matching of 2D-3D feature points, and the numerical computation of external camera parameters. Sottile et al. initially proposed minimizing information discrepancies (MI) and the reprojection error of feature points [14]. Bodensteiner et al. developed a MC registration technique without the information minimization term, focussing on improving the 2D-projected 3D feature point matching using self-similarity measures [15]. Our research approach is comparable to this approach, but focuses on replacing the manual pose initialization with mobile sensor data.

3 Contribution

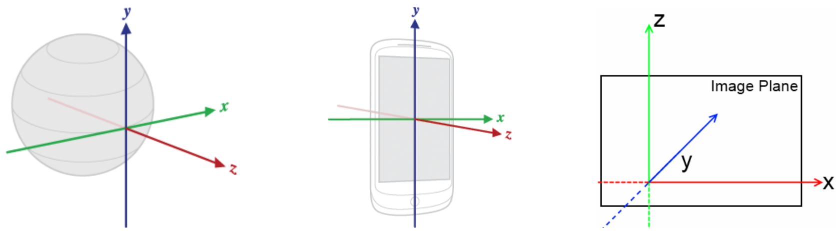
This article presents four contributions to the challenge of Image-to-Geometry mapping for texturing purposes. First, we present an extension to a Mutual Correspondence approach inspired by Bodensteiner et al. [15] for solving the initial pose estimation using mobile sensor data, eliminating the need for user intervention. Second, we present a complete implementation of this fully automatic Image-to-Geometry workflow on mobile devices, such as smartphones and tablets. Third, we compare common manual registration approaches with the novel automatic registration workflow in terms of final pose accuracy and processing time. Our focus is on geosciences use cases (i.e. urban environments and virtual outcrop geology).

4 Mobile Devices and Location-Orientation Sensors

Estimating the initial pose for the image- and geometry alignment can be done algorithmically (Corsini et al. use a 3D planar domain decomposition with the 4PCS algorithm [12]) or via user interventions (Pintus et al. demand the user selection of 7 to 12 correspondence pairs [16]). The output is an orientation and a translation, so that:

$$K = [R_{3,3}, T_3]$$

Mobile devices are equipped with sensors that measure location, using the Global Navigation Satellite System (GNSS), and orientation, using accelerometers and magnetometers, in a global reference frame. The location coming from a GNSS can be used as translation vector T_3 after converting it to the local coordinate system of 3D model. The rotation matrix is obtained in the coordinate frame of the mobile device (see Fig. 1). Because the x-axis is just derived from magnetic- and gravitational measurements, the individual rotational components of the matrix have different error margins and error sources. A thorough analysis is given by Blum et al. [17]. We continue rotational calculations with quaternions due to their compact format, easier handling and due to numerical accuracy. A quaternion within the local coordinate system is extract using the original quaternion, the mobile device orientation and the magnetic declination. The resulting transformation matrix is used as a coarse registration matrix.



(a) ©Google (b) ©Google (c)

Fig. 1 Used coordinate systems in the system. The global rotation matrix is given in the mobile coordinate system (a), where the y-axis points to the magnetic north (top-edge of the device in native orientation), the z-axis points away from the center of gravity (out of the device screen), and the x-axis is obtained as yz-axis cross product (b). The used coordinate system aligns with common geographic system, with the xy-plane being ground-parallel while the z-axis delimits altitudes (c).

5 Automatic Image-to-Geometry Mobile Registration using Mutual Correspondences

The MC input is given by the mobile photograph, a textured 3D model, and the calibrated camera model including the focal length of the view. As discussed by Bodensteiner et al. [15], a coloured point set can also be used for the registration, though visual saliency within the coloured point set demands a more complex processing of keypoint matches. Subsequently, the 3D scene is rendered using the coarse alignment transformation matrix as view matrix. Then, keypoints are extracted from the synthetic- and real-world image using SIFT- or SURF feature descriptors, and matches using a RANSAC approach. The matches keypoints of the synthetic image are used as a starting point for Raycasting in order to obtain their 3D correspondence points. We give preference to Raycasting over the depth buffer values due to the limited depth mapping accuracy. The 3D raycasted intersections of the model are subsequently related to the 2D keypoints of the photo. Taken the undistorted photo keypoint coordinates and the 3D pointset, we can use established numerical optimization methods such as Levenberg-Marquardt (LM) and EPnP (Efficient PnP) to determine the final pose. The workflow is shown in Fig. 2.

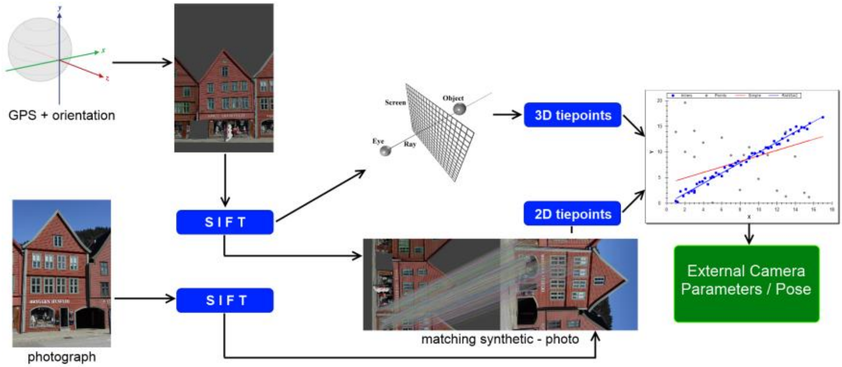


Fig. 2 Image-to-Geometry algorithm, based on mobile data (left image side), Mutual Correspondence (mid-part) and PnP pose estimation (right side).

6 Mobile Platform Implementation

The presented Image-to-Geometry approach is implemented on Google's Android platform. Android applications are based on Java, running in an adapted *Java Virtual Machine* (JVM). Additional computational packages for graphics and computer vision are compiled as C++ libraries and communicate with the Java application through *Java Native Interface* (JNI) wrappers.

We use the wrapped OpenSceneGraph package "osgAndroid"[1] for graphics, giving access to various geometry- and image formats. Additionally, it provides Level-of-Detail Out-of-Core functionality for rendering very large models. The application also interfaces OpenCV for the tasks of pose estimation and point reprojection. A local package provides access to SIFT and SURF implementations.

Mobile devices have relatively powerful processors and network interfaces, but the limited local memory size prohibits using large models. Moreover, because of the JVM-managed memory, we encountered situations where large data failed to load, as the JVM disposed loaded data during the computation. Therefore, some parts of the algorithm are implemented entirely in C++ to circumvent the memory restrictions.

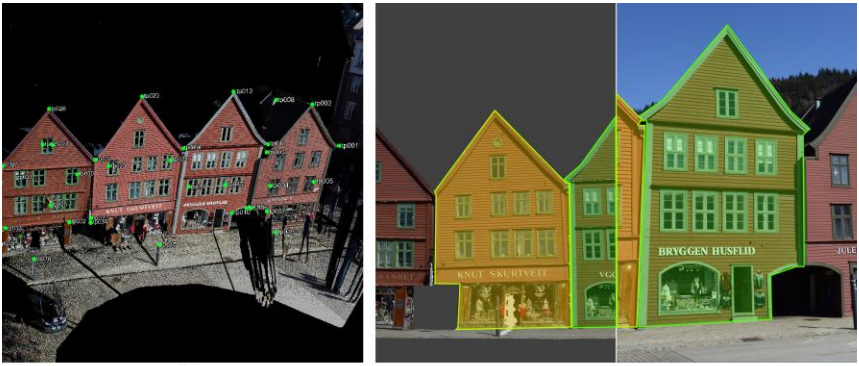
7 Method Assessment based on Geoscience Use Cases

a) Urban Environment Use Case – Tyske Bryggen / Bergen / Norway

Tyske Bryggen (referred to as "Bryggen") is a late-medieval building complex of the

[1] osgAndroid - http://github.com/miragetech/osgAndroid

Hanse at the port of Bergen. Originally conceived as a trading post for fish and goods, it is nowadays used for restaurants and (traditional) shops (see Fig. 3). The dataset was collected using a Riegl LMS-Z420i Terrestrial Laser Scanner (TLS) and a mounted Nikon camera for the image texture. The mounting serves as a baseline measurement for comparing different registration techniques.



(a) (b)

Fig. 3 Overview of Tyske Bryggen as rendering. 3D markers are highlighted in green (a). The model is rendered from an initial position with very few overlap (b, synthetic image left, photograph right).

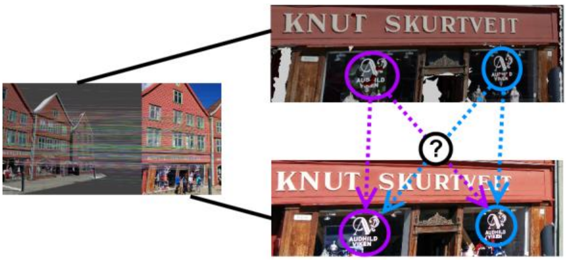


Fig. 4 Re-occuring patterns are an exemplary challenge when establishing correspondences in real-world urban datasets.

An automatic registration of additional photographs with this dataset is challenging due to re-occurring patterns within the image (see Fig. 4). Fig. 5 shows an accuracy comparison between several automatic and manual methods.

For automatic registration assessment on mobile devices, we used the Google Nexus 5 smartphone and the NVIDIA Shield Tablet for comparison. The current processing pipeline of our models stores high-density geometry and JPEG-compressed images, where the image decompression is done in the rendering phase. While the graphics-targeted NVIDIA tablet supports OpenGL 3.2 profile functions (including DCT decompression), common smartphones and tablets only support Embedded OpenGL 2 profile (without DCT decompression). Therefore, we used different configurations of geometry and imagery during the runtime assessment (see Table 1 for average runtime measurements).

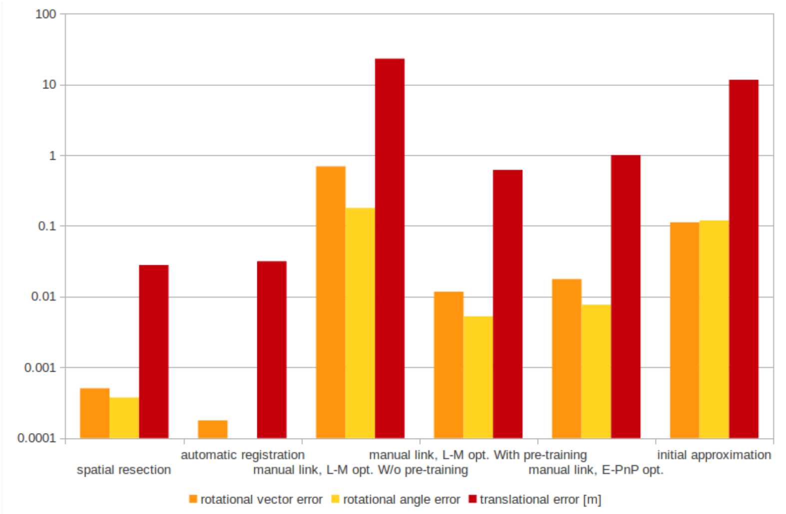


Fig. 5 Chart giving the average error over 2 Bryggen, compared to the scanner mounting, for vectorial and angular part of rotation quaternion and positioning error (translational).

Focal length: 3.92mm, Image sensor 4.6mm x 3.52 mm, original image size: 3262x2448 pixels; registration done on half image size								Timings [min:sec.msec]			
Device	**3D Geometry**	**Texture form.**	**Keypoints Photo**	**Keypoints Rendering**	**2D-3D corresp.**	**Sensor Mat.**	**Rendering**	**Detect & Match**	**Raycasting**	**Pose Est.**	**Inliers [%]**
NVIDIA Shield tablet	full model	DCT image	30375.8	22163.6	400	00:00.010	00:04.558	01:54.538	00:14.960	00:00.273	12.61831
NVIDIA Shield tablet	Subsample 20%	DCT Image	30375.8	18401.6	309.2	00:00.009	00:00.904	01:41.362	00:08.675	00:00.198	11.00416
NVIDIA Shield tablet	Subsample 20%	RGB Image	30375.8	17146	330.2	00:00.006	00:02.448	01:40.594	00:09.844	00:00.252	14.13055
Google Nexus 5 phone	Subsample 20%	RGB Image	30375.8	17156	329.6	00:00.010	00:08.446	02:13.390	00:19.613	00:00.308	0.13617

Table 1 Runtime experiments for the automatic registration pipline on smart devices with varying textures and geometries (full model ~3.9 mil. triangles).

b) Geological Use Case – Mam Tor / Peak District / United Kingdom

Mam Tor is a hillside outcrop representing our VOG target application. Outcrop models are used as geological study analogues for subsurface oil- and gas reservoirs. Geologists can map sediment depositions in the field from surface-accessible outcrops, which serve as input for geostatistics (i.e. *Multi-Point Statstics* (MPS) to model subsurface reservoirs. This intended use case also explains the focus on mobile implementations: in urban environments, the workflow can be realized as a web-application that is easily accessible to mobile platforms through WiFi. In contrast to that, geologists in the field rarely have WiFi access, which means that data and processing need to be realized on the smartphone or tablet itself.

The given pictures (see example Fig. 6) are taken in the field, independently from the scanner, which is why we take the accuracy of the manual registration as baseline for the study. Fig. 7 shows the rotational and translational deviation of the automatic matching and the initial mobile parameters. Table 2 shows the average runtime for the image registration of 12 images.

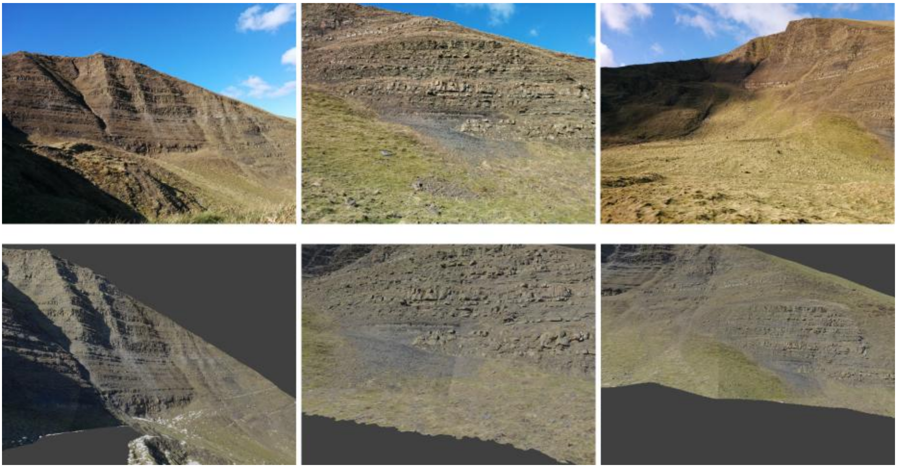


Fig. 6 Overview of the Mam Tor use case, showing the real-world geologic mobile photographs (top row) and the corresponding Virtual Outcrop rendering from the (error-corrected) GPS-orientation logged initial pose estimate (bottom row).

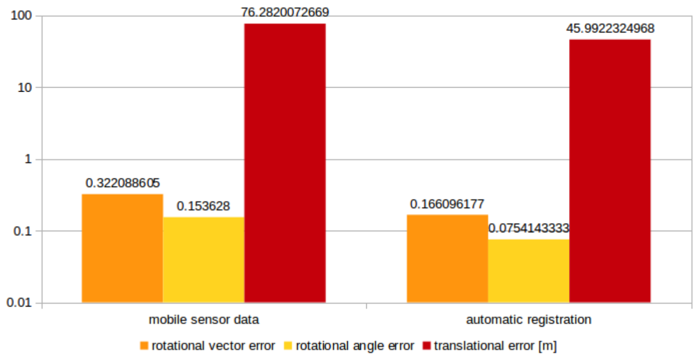


Fig. 7 Chart giving the average error over 9 image of Mam Tor, compared to a manual registration technique, for vectorial and angular part of rotation quaternion and positioning error (translational).

Focal length: 3.92mm, Image sensor 4.6mm x 3.52 mm, original image size: 3262x2448 pixels; registration on half image size; avg of 12 photos						Timings [min:sec.msec]				
Device	**3D Geometry**	**Texture form.**	**2D-3D corresp.**	**Sensor Mat.**	**Rendering**	**Detect & Match**	**Raycasting**	**Pose Est.**	**TOTAL**	**Inliers [%]**
NVIDIA Shield tablet	Subsample 25%	DCT Image	27.82	00:00.009	00:08.380	00:55.382	00:09.067	00:00.296	01:13.960	10.87083
NVIDIA Shield tablet	Subsample 25%	RGB Image	33.36	00:00.007	00:08.807	00:57.046	00:08.116	00:00.369	01:12.367	13.16742
Google Nexus 5 phone	Subsample 25%	RGB Image	30.55	00:00.010	01:24.948	01:11.091	01:34.854	00:00.526	04:12.630	13.76332

Table 2 Timings as in table 1, for Mam Tor dataset (full size: ~60 mil. triangles)

8 Discussion

As observed in the “Bryggen” case, we are often able to retrieve a reasonable pose with most algorithms from a rather generic starting point, given that the object of interest is in view. The data suggest that a generic non-linear LM optimization without parameter initialization fails when getting stuck in local optima of the 6D parameter space. This can be avoided by choosing a reasonable starting point for each parameter. The automatic registration procedure delivers an accurate final pose with the smallest errors in parameter space. This is an advantage over manual registration methods, where the 2D/3D perception of the user reaches natural accuracy limits. Despite this, image-based registration techniques can fail in urban environments due to re-occurring patterns and occluding pedestrians.

In the geologic use case, the presented images were captured in the same acquisition campaign as the textured 3D model, which means that illumination- and weather conditions were comparably constant. Height-corrected mobile sensor data are sufficiently accurate to automatically obtain a correct position and orientation in this setting. On the other hand, the results also make the large deviation of mobile sensor data and actual position/orientation apparent: Particularly comparing the GPS data with the real position yield an average deviation of 52.42 metre in longitude, 17.56 metre in latitude and 8.86 metre in altitude. Moreover, our mobile positions were obtained with an external GPS because initial experiments in a controlled environment have show altitude deviations of the mobile device build-in GPS of $\pm 15m$ up to $\pm 80m$ (in line with available literature [17]).

Registering images with a textured 3D model using the presented algorithm takes between 1 minute (tablet) and 5 minutes (phone). This makes the presented Image-to-Geometry procedure very well applicable to field use. The energy efficiency of the implementation can be improved, as the 3D rendering and the large image dimensions demand a certain processing power. During the experiments, the device batteries emptied in 1 hour. Still, the results show that the automatic, visual registration technique can even be realized on smartphones with limited rendering capabilities. In the future, we focus on assessing the impact of different ambient illumination conditions on the registration procedure.

Acknowledgements

This research is part of the “VOM2MPS” project (grant number 234111/E30), funded by the Norwegian Research Council and the FORCE consortium through Petromax 2 and SAFARI. Data are provided and used by the SAFARI project[2] for mapping of sedimentary depositional system on Virtual Outcrop Models. The authors also thank John A. Howell, Robert L. Gawthrope, Ivan Viola and Sophie Viseur for the insightful discussions during the research period.

References

[1] J. A. Howell, A. W. Martinius and T. R. Good, "The Application of Outcrop

[2] SAFARI – www.safaridb.com

Analogues in Geological Modeling: A Review, Present Status and Future Outlook," *The Geological Society, London, Special Publications,* vol. 387, no. 12, 2014.

[2] T. Van Damme, "Computer Vision Photogrammetry for Underwater Archaeological Site Recording in a Low-Visibility Environment," *ISPRS Archives,* vol. 1, pp. 231-238, 2015.

[3] A. Semmo and J. Döllner, “Interactive Image Filtering for Level-of-Abstraction Texturing of Virtual 3D Scenes,” *Computer & Graphics,* vol. 52, pp. 181-198, 2015.

[4] T. Nöll, A. Pagani and D. Stricker, "Markerless Camera Pose Estimation - An Overview," *OASIcs - OpenAccess Series in Informatics,* vol. 19, 2011.

[5] P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 14, no. 2, pp. 239-256, 1992.

[6] R. Y. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," *IEEE Journal of Robotics and Automation,* vol. 3, no. 4, pp. 323-344, 1987.

[7] P. J. Neugebauer and K. Klein, "Texturing 3D Models of Real World Objects from Multiple Unregistered Photographic Views," *Computer Graphics Forum,* vol. 18, no. 3, pp. 245-256, 1999.

[8] V. Lepetit, F. Moreno-Noguer and P. Fua, "EPnP: An Accurate O(n) Solution to the PnP Problem," *International Journal of Computer Vision,* vol. 81, no. 2, pp. 155-166, 2009.

[9] N. Snavely, S. M. Seitz and R. Szeliski, "Photo Tourism: Exploring Photo Collections in 3D," *ACM Transactions on Graphics (TOG),* vol. 25, no. 3, pp. 835-846, 2006.

[10] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal and P. Suetens, "Multimodality Image Registration by Maximization of Mutual Information," *IEEE Transactions on Medical Imaging,* vol. 16, no. 2, pp. 187-198, 1997.

[11] P. Viola and W. M. Wells, “Alignment by Maximization of Mutual Information,” *International Journal of Computer Vision,* vol. 34, no. 2, pp. 137-154, 1997.

[12] M. Corsini, M. Dellepiane, F. Ganovelli, R. Gherardi, A. Fusiello and R. Scopigno, "Fully Automatic Registration of Image Sets on Approximate Geometry," *International Journal of Computer Vision,* vol. 102, no. 1-3, pp. 91-111, 2013.

[13] J.-P. Thirion, “Extremal Points: Definition and Application to 3D Image Registration,” *IEEE Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 587-592, 1994.

[14] M. Sottile, M. Dellepiane, P. Cignoni and R. Scopigno, "Mutual

Correspondences: An Hybrid Method for Image-to-Geometry Registration," *Eurographics Italian Chapter Conference,* pp. 81-88, 2010.

[15] C. Bodensteiner, M. Hebel and M. Arens, "Accurate Single Image Multi-Modal Camera Pose Estimation," *Trends and Topics in Computer Vision,* vol. 6554, pp. 296-309, 2012.

[16] R. Pintus, E. Gobbetti and R. Combet, "Fast and Robust Semi-Automatic Registration of Photographs to 3D Geometry," *Proceedings of the 12th International Conference on Virtual Reality, Archaelogy and Cultural Heritage,* pp. 9-16, 2011.

[17] J. R. Blum, D. G. Greencorn and J. R. Cooperstock, "Smartphone sensor reliability for augmented reality applications," in *Mobile and Ubiquitous Systems: Computing, Networking and Services*, 2013.