REGISTRATION OF TERRESTRIAL LASER SCANS VIA IMAGE BASED FEATURES

Shahar Barnea, Sagi Filin

Dept. of Transportation and Geo-Information Eng., Technion – Israel Institute of Technology, Haifa 32000, Israel
{barneas, filin}@tx.technion.ac.il

Commission V, WG 3

KEY WORDS: Registration, Feature Extraction, Terrestrial Laser Scanner, Point Cloud, Algorithms

ABSTRACT:

The complexity of natural scenes and the amount of information acquired by terrestrial laser scanners turns the registration among scans into a complex problem. This problem becomes even more complex when considering the relatively low angular resolution of terrestrial scanner compared to images, the monotonicity of manmade surfaces that makes the detection of corresponding objects difficult, and the lack of structure of vegetated objects that makes the detection of meaningful features difficult. Since most modern scanners are accompanied with consumer cameras of relatively high quality, it stands to reason making use of the image content for the registration process. Such alternative will benefit from the large body of image based registration work that has been carried out for several decades and therefore has the potential of providing an alternative and simple approach for the registration of pairs and multiple scans simultaneously. In this paper, we study the registration of terrestrial scans via image-based information. For this purpose, we propose an efficient autonomous model that supports the registration of multiple scans. Following the presentation of the model, we analyze its application to outdoor, complex scenes, ones that are common to find in actual laser scanning projects.

1. INTRODUCTION

Terrestrial laser scanners are rapidly becoming a standard technology for 3D modeling in surveying and engineering projects. In most cases, the acquisition of several scans is needed to obtain full scene coverage, and therefore requires the registration of the individual scans into one global reference frame. For the registration, the common practice involves the deployment of artificial targets in the scene as tie objects, with typical targets having the form of spheres, which are easily distinguishable, or reflectors whose high-energy return eases their detection. Following the detection of the tie objects, the rigid body transformation between the coordinate systems can be solved. To avoid manual intervention in the registration process, a growing body of work addresses the problem of autonomous registration in relation to both range images and terrestrial laser scans. The commonly studied model usually involves variants of the Iterative Closest Point (ICP) algorithm family (Besl and McKay, 1992; Chen and Medioni, 1992) that differ in the features toward which distances are minimized (see e.g., Rusinkiewicz and Levoy, 2001), and the numerical framework that is being used (e.g., Mitra et al., 2004; Pottmann et al., 2006). Dalley and Flynn (2002) sort the iterative algorithms by their robustness to initial pose parameters, rate of convergence, and by their sensitivity to outliers. For reasons such as existence of local extrema in the solution space, existence of outliers, occlusions, and lack of information regarding the point distribution in the object space, no guaranty can be given that convergence to the actual solution is reached unless the iterations begin close enough.

As the iterative methods require good initial pose parameters, autonomous techniques for their approximation have been proposed for range images of relatively simple objects, with well-defined shape and structure, and high-level of connectivity (see e.g., Gelfand et al., 2005; Huber, 2002; Huang et al., 2006). A small number of works address the actual complexity of terrestrial laser scans. Bae and Lichti (2004) are using a variation in curvature as the matching criterion on local points. This requires the computation of the normal vector and the curvature itself. Dold and Brenner (2006) propose an autonomous matching procedure that is based on planar patches. Following their extraction, patches from different scans are matched subject to geometric constraints. Gruen and Akca (2005) present a least squares matching based registration scheme. The reported algorithm is more stable than the classic ICP, but still requires an initial transformation.

The registration of terrestrial laser scans can be aided by the images that are usually acquired simultaneously with the range data. Images enjoy high spatial resolution, and record color content of the scene, which is usually very rich and diverse. The role of image content for realistic texture rendering suggests that the tight link between the two sensors is only due to increase. As such, it provides an alternative candidate to form the registration process of laser scans. Image based registration also benefits from the vast amount of research that has been devoted to the registration problem. Registration of laser scans supported by images received indeed some attention in recent years. Ulirch et al. (2003) define a framework to integrate image information with scanning data. Kang et al. (2007) propose using the Moravec operator and cross correlation as a means to find point correspondence between images and use those for the registration phase. Al-Manasir and Fraser (2006) suggest using relative orientation between images for scans registration supported by the placement of artificial, signalized, targets. Seo et al. (2005) present an approach that uses image-matching schemes on relatively small scenes acquired by a table scanner. Finally, Liu et al. (2006) consider a more general framework with no rigid attachment between the camera and the scanner but with the imposition of some specific geometric constraints.

Since image-based content is only an integral part of most laser scanning systems, it stands to reason investigating the potential in the registration of laser scans using intensity information. Normally, such registration will be purely image based (e.g., via bundle adjustment), where images will be mutually matched and simultaneously solved. However, laser-scanning projects usually acquire data from a relatively wide base, and therefore, especially in open scenes, only a limited number of images overlap between scans, particularly for establishing a strong

Figure 1. Top: panoramic view of the scanned scene as acquired by a camera mounted on the scanner (for the original images see Figure 5), Bottom: Polar representation of terrestrial laser scans; the horizontal and vertical axes of the image represent the values of θ, φ respectively and intensity values as distances ρ (bright=far). "No-return" and "no-reflectance" pixels are marked in red.

photogrammetric image block. Additionally, image based registration will relate to object space by up to a scale factor. Therefore, establishing this link requires a subsequent registration, and if autonomous registration is of concern, such registration should relate to the laser point cloud.

The approach proposed here is based on using the direct relation between the acquired images and the laser data (see Fig. 1), but instead of solving a block of images it solves a set of rigid body transformations, which are more robust, efficient, and require a small subset of points. The model applies to the registration of pair of scans as well as multiple scans and assumes no support in the form of artificial targets or a priori scanning pose parameters. Essentially the assumption is that a digital camera is attached to the laser scanner equipment and is calibrated with respect to it. Our objective is to utilize both the relatively robust geometric models for the registration of 3D scans with the powerful techniques of keypoint image matching as a means to generate the initial set of correspondences. Our aim is to develop an algorithm that can handle the data volume and the expected complexity of the scanned scenes. To make the registration more reliable and robust we make use of the known calibration between the laser scanner and the imaging system to treat the problem in a dual manner – extracting features and matching them in 2D image space but computing the actual transformation between the scanners, in 3D space. With the proposed model, we test the applicability of the model to the registration of terrestrial laser scans. We analyze the advantages and disadvantages of image supported terrestrial laser scans registration. The results provide an insight into how these sources of information can be used jointly for the registration of terrestrial laser scans.

2. METHODOLOGY

Generally, there are two reference frames involved in the model – the image reference frame (and there are n images acquired per scan), and the scanner reference frame. Essentially, our objective is to recover the scanner pose parameters, using the image content. Such problem can be approached in two ways: i) solving the image (relative) pose parameters and then computing the scanner pose parameters using a boresight transformation, see e.g., Al-Manasir and Fraser (2006), and ii) using the boresight computation between scanner and images to find the local 3D point coordinates and computed directly the scanner pose parameters using a rigid body transformation.

While the first approach offer slight advantages in terms of the quality of the matched entities (therefore, leading to better registration accuracy) it leads to a more complex framework involving the simultaneous orientation of multiple images. In contrast, the second approach that estimates a rigid body transformation, involves only a single transformation per scan, one that is relatively easier to compute.

2.1 Camera to scanner registration

The camera mounted on top of the scanner can be linked to the scanner body by finding the transformation between the two frames shown in Figure 2. Such relation involves three offset parameters and three angular parameters. This relation can also be formulated via the projection matrix P. With P a 3x4 matrix that represents the relation between world 3D point (X) and image 2D point (x) in homogeneous coordinates. Compared to the six standard boresighting pose parameters, the added parameters (five in all) will account to interior orientation parameters. The projection matrix can be formulated as follows:

$$x = KR[I \mid -t]X = PX \qquad (1)$$

with

$$K = \begin{bmatrix} f_x & s & x_0 \\ & f_y & y_0 \\ & & 1 \end{bmatrix}$$

f_x and f_y are the focal lengths in the x and y directions respectively, s is the skew value, x_0 and y_0 are the perspective offset across the two image axes. R is the rotation matrix between the scanner and the camera reference frames (the red and the blue coordinate systems in the figure respectively) and t the translation vector (Hartley and Zisserman, 2003).

For the estimation of the relative pose offset between the scanner and the camera image, points for which well-defined 3D laser points exist are selected. Using the laser points as control information allows computing the projection matrix directly and linearly. In this regard, we point that the calibration of the lens distortion parameters (radial and decentring) will provide an even better accuracy. At each scanning position, n images are acquired in predefined “stops” along the scan (e.g., every 360/n degrees). For each image, the projection matrix, P, represents the relation between the image and the scan. The proposed model assumes that, i) the camera is rigidly mounted to the scanner, ii) the interior camera parameter are fixed and known, and iii) the acquisition position is fixed across all scanning positions. These standard assumptions enable using the same projection matrices for all images in the same “stop” in different scans.

2.2 Detection of corresponding points

Finding an image points correspondence has been an active research for several decades. Mikolajczyk and Schmid (2004) present a comparative review of the modern methods, and note that they are composed of two fundamental steps: extraction, and matching. The goal of the extraction phase is to detect keypoints (sometimes terms interest points) in a repeatable manner. The challenge in this stage is to yield high repeatability rate even under extreme viewpoint, resolution, and exposure changes (e.g., brightness and contrast). The goal of the matching phase is to find correspondence among the keypoints that were extracted from the different images. For this purpose, descriptors that provide distinctive characterization of the keypoint are used. Following the generation of a descriptor for each detected keypoint, the matching is performed by searching for similar descriptors in different images and upon finding them, recording them as candidate tie-points. The challenge in the matching phase is to design a descriptor that offers unique and descriptive features while being insensitive to small detection errors and perspective deformation. Following the generation of proposed correspondences phase, some correct and some not, comes the computation of the transformation between the images. This will usually be driven by the Random Sampling Consensus (RANSAC) algorithm (Fishler and Bolles, 1981). An important aspect in the application of the RANSAC algorithm is the minimal number of points required to compute the hypothesis transformation in each iteration. This number affects the number of required iterations and thus, the chances to finally converge to the correct solution. In this regard, one should prefer a geometric model with a small set of points to calculate the hypothesis transformation.

For the extraction of keypoints and their descriptors, we make use of the Scale Invariant Feature Transform (SIFT) that was proposed in Lowe (2004), and was applied in photogrammetry in Shragai et al. (2005), and Läbe and Förster (2006).

2.3 Scale Invariant Feature Transform

The Scale Invariant Feature Transform - SIFT (Lowe, 2004) is a methodology for finding corresponding points in a set of images. The method designed to be invariant to scale, rotation, and illumination. The methodology consists of the following four steps:

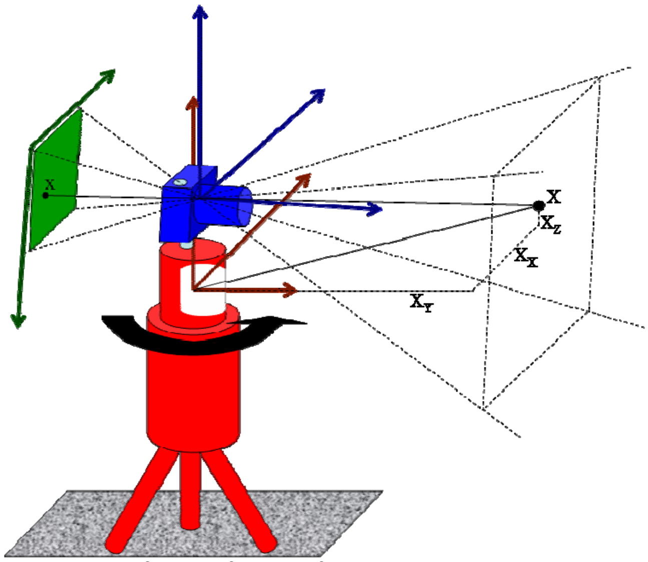

Figure 2. Reference frames of the scanning system with a mounted camera.

1. Scale-space extrema detection – using the difference of Gaussian (DoG), potential interest points are detected.
2. Localization – detected candidate points are being probed further. Keypoints are evaluated by fitting an analytical model (mostly in the form of parabola) to determine their location and scale, and are then tested by a set of conditions. Most of them aim guaranteeing the stability of the selected points.
3. Orientation assignment – orientation is assigned to each keypoint based on the image local gradient. To ensure scale and orientation invariance, a transformation (in the form of rotation and scale) is applied on the image keypoint area.
4. Keypoint descriptor – for each detected keypoint a descriptor, which is invariant to scale, rotation and changes in illumination, is generated. The descriptor is based on orientation histograms in the appropriate scale. Each descriptor consists of 128 values.

With the completion of the keypoint detection (in which descriptors are created), the matching process between images begins. Matching is carried out between the descriptors, so the original image content is not considered here. Generally, for a given keypoint, matching can be carried with respect to all the extracted keypoints from all images. A minimum Euclidian distance between descriptors will then lead to finding the correspondence. However, matching in this exhaustive manner can be computationally expensive (i.e., $\mathbf{O}(N^2)$ with N the number of keypoints). Common indexing schemes cannot be applied to improve the search here because of the descriptors dimensionality. However, an indexing paradigm, called Best Bin First (BBF) can be applied (Lowe, 2004). The BBF algorithm reduces the search to a limited number of the most significant descriptors values and then tries locating the closest neighbor with high probability. Compared to the exhaustive matching, this approach improves the performance by up to two orders of magnitude, while difference between the amounts of matched points is small. Our proposed solution follows Brown and Lowe (2003) where all key points from all images are organized in one K-d tree. Once a set of matching points has been generated, another filtering process is applied.

Figure 3 shows the keypoints extracted in a scene that mixes structured and unstructured objects, the squares around each keypoint illustrates the scale in which it was detected and the small vector, its orientation.

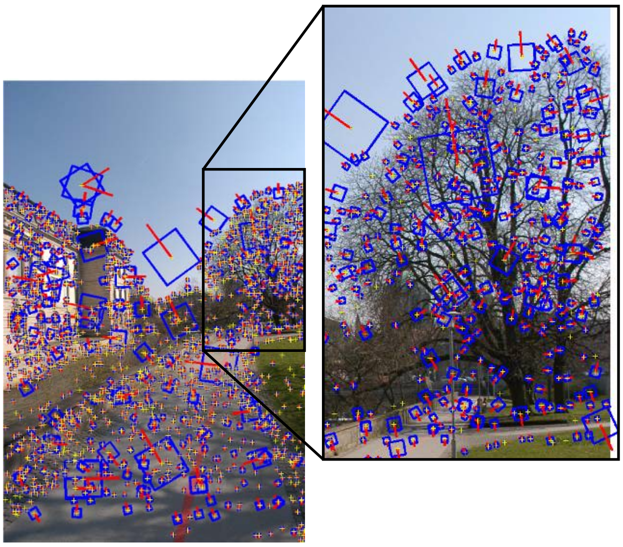

Figure 3. SIFT keypoints with orientation and scale.

2.4. Linking the laser scans and the image information

Since the registration scheme is based on a rigid body transformation, the extraction of keypoints in image space should now be transferred into the local 3D object space. Generally, this transfer requires tracing the ray into object space. However, we apply here a back projection of the 3D point cloud onto the image using the boresight parameters that were derived in the calibration phase (see Section 2.1). We then assign the 3D coordinates of the relevant laser point to the keypoints. The result of the back-projection of the laser point cloud into the imaging system reference-frame is demonstrated in Figure 4. Notice that vegetation expression in the range image compared to intensity one.

The 3D coordinate assignment is not immediate, however. Keypoints are defined by their position and scale (window size), therefore, for each keypoint, candidate 3D coordinates are collected from the scale dependent corresponding window (see Figure 3). Generally, the coordinate assignment problem can be partitioned into two cases the first is when the point falls on a solid object; the second is when the point falls between surfaces. In the first case, we assign the nearest 3D coordinate in terms of angular distance between the keypoint direction and laser point direction, while in the second we assign the 3D coordinates of the point closest to the imaging system. The motivation for this is as follows, for solid objects the keypoint location is well defined and, therefore, the nearest 3D point will have the smallest bias among all candidates (we note that some refinement to the ray direction can be applied, but this is negligible). For the other case, with lack of any other information we opt toward assigning the closest distance within the candidate 3D points under the realization that it is the foreground object, which is likeliest to do with the detection of the point as keypoint. Differentiation between the two cases is achieved by computing the *std*. of the 3D points' depth.

2.5 Registration between scans

With the candidate matches, the registration of the laser scan becomes an estimation problem of the rigid body transformation,

$$X = X_0 + (I + S)^{-1}(I - S)x \qquad (2)$$

where I is a 3x3 identity matrix, and S is an skew-symmetric matrix, defined as:

$$S = \begin{bmatrix} 0 & c & -b \\ -c & 0 & a \\ b & -a & 0 \end{bmatrix}$$

The transformation can estimated linearly using such methods as the one proposed in Horn et al. (1988). Since some of the proposed matches are outliers, a RANSAC solution guides the parameter estimation. One of the appealing properties of the registration based on the rigid body transformation is that only three points are needed to generate a hypothesis. Therefore, even if a small fraction of inliers is assumed, the number of trials will be controllable and very efficient. Choosing the relative orientation option and using, for example, the well-known eight-point algorithm to estimate the fundamental matrix (Hartley and Zisserman, 2003) will obviously have a much higher cost under a small fraction of inliers assumption.

Figure 4. Depth image calculated to fit the original image, left: the depth image, right: the original image. Because the spatial resolution of the laser point cloud is much sparser than the image resolution (0.12^{o} compared to 0.03^{o} here) filling of depth image was applied for demonstration purposes only.

3. RESULTS

To demonstrate our approach we test the proposed algorithm on three scans acquired in a row by Riegl 360. The image sequences of the three scans are presented in Figure 5. The distance between the scanners is 8.15, and 22.28 [m] respectively, and the maximal scanning range ~100 [m]. Six mega-pixel size images acquired by the Nikon-D100 were processed in full resolution. For each image SIFT keypoint were extracted with 4,000-11,000 keypoints per image evaluated for the matching. Figure 3 shows a typical set of keypoints (with some pruning for visual clarity). Matches are then evaluated between each image in a scan to all seven images in the counterpart scan (for multiple scans a similar procedure will apply). Tables 1, 2 list the number of matches (descriptor wise) between each image in one scan and the images in the other. Even though Table 1 has a dominant diagonal, the structure of the match matrix is arbitrary and depends on similarity between the images in the scans. Figure 5 clearly shows why the first set is diagonal dominant. Figures 6 (top and center) shows the matched keypoints between the pair of sixth images in set 1-2. Generally most matches are correct, but some outliers can be

seen, e.g., point 134 (encircled) that has no counterpart. Figures 6(bottom) shows the matched keypoints between image 7 of scan 2 and 3 of scan 3. One can see that the number and quality of the matches is relatively poor compared to the first pair.
Overall, 1256 matched points (sum of all values in the table) were found all scans in set 1-2, and 123 points between 2 and 3. For each matched keypoint, 3D coordinates are assigned (see Section 2.3). Image pairs with less than four matched points are overlooked due to the realization that such a small number is most likely the result of lack of overlap between the images with only accidental matches found (this was also validated by manual inspection). This further pruning reduces the number of matched keypoints to 1219 and 68 matches respectively. Following the assignment of the 3D coordinates to the matched keypoint comes the RANSAC guided 3D rigid body transformation.

Table 1: number of matches, scans 1 & 2 –baseline 8.15 [m]

		scan 2						
	img #	1	2	3	4	5	6	7
scan 1	1	4	6	4	0	1	3	0
	2	4	1	11	3	4	3	3
	3	0	5	16	5	0	3	0
	4	0	1	2	35	36	2	4
	5	4	1	1	0	347	115	6
	6	2	3	4	0	55	414	38
	7	5	2	1	0	2	4	96

Table 2: number of matches, scans 2 & 3 –baseline 22.28 [m]

		scan 3						
	img #	1	2	3	4	5	6	7
scan 2	1	3	2	1	1	2	0	2
	2	0	2	1	1	2	2	2
	3	0	3	0	6	2	4	1
	4	3	5	0	0	3	5	0
	5	0	12	4	8	1	2	2
	6	1	6	7	4	0	2	0
	7	0	2	13	3	2	1	0

Out of 1219 proposed matches, 979 were found correct (amount to 80.3% of the proposed matches) for set 1-2. In contrast, out of the initial 68 candidates in the 2-3 scan, 18 proposed correspondences were found (amounts to 26.5%). The differences in correct correspondences reflects the change in the baseline between the scan pair (8 compared to 22 [m]). The comparison of the estimated parameters to manual calculation, considered as ground truth, shows that the translation error on the scanning position is on order of 0.65 [m] for the first pair and 1.15 [m] for the second one; the angular error was (0.12, 0.3, 0.01) [^{o}] for ω,φ,κ angles respectively and (0.18,0.07,1.09) for the second. Those offsets can related to errors that are accumulated in the course of the process (calibration errors, image to range data conversion errors and matching accuracy errors). However, these values are good enough to launch an ICP procedure between the point clouds, which is advisable to perform for tuning the registration.

4. CONCLUSIONS

The registration results of the two scans show the great potential of registration via images. As the paper has demonstrated when considering the image-based registration problem between scans as a platform for an eventual rigid body transformation, the rich image-based information (extracted keypoints) allows using homologous registration candidates which wouldn't have been naturally detected using any of the range data registration methods one finds. The rigid body transformation also allows using small subsets of points for the RANSAC hypothesis generation, thereby allowing greater flexibility in the feature extraction phase.

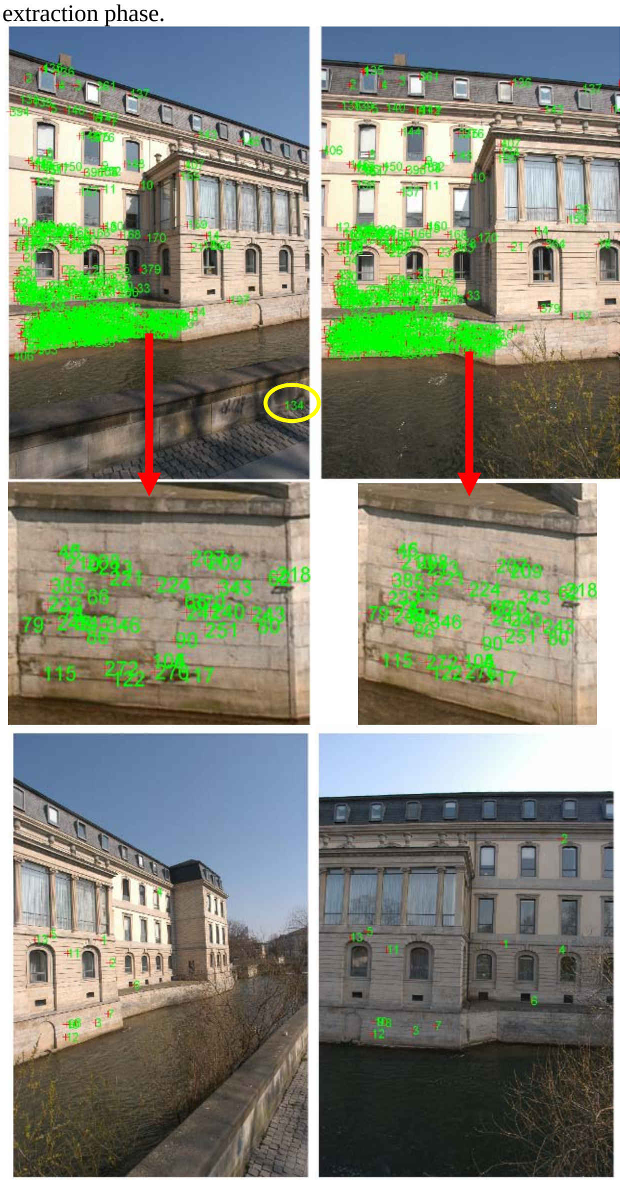

Figure 6. Matched keypoints between images pairs, up) from scans 1-2, center) blowup showing the quality of the matches, bottom) matches from scan 2-3 the different viewing geometry dropped the number of matches.

5. ACKNOWLEDGEMENT

The authors would like to thank Dr. Claus Brenner for making the data used for our tests available.

6. REFERENCES

Al-Manasir K., Fraser C., 2006. Registration of terrestrial laser scanner data using imagery. Photogr.. Rec.. 21 (115), 255–268.

Besl P., McKay N., 1992. A method for Registration of 3-D Shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 14 (2), 239–256.

Brown M., Lowe G. D., 2003. Recognising Panoramas. International Conference on Computer Vision. Nice, France, 1218-1227

Figure 5. Image sequences of the three scans

Chen Y., Medioni G., 1992. Object Modeling by Registration of Multiple Range Images. Image Vision Comp. 10(3), 145-155.

Dalley G., Flynn P., 2002. Pair-wise range image registration: a study in outlier classification. Computer Vision and Image Understanding archive, 87 (1-3), 104– 115.

Dold C., Brenner C., 2006. Registration of Terrestrial Laser Scanning Data using Planar Patches and Image Data. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 36 (Part 5), 25–27.

Fischler M.A., Bolles R.C., 1981. Random Sample Consensus: A paradigm for model fitting with application to image analysis and Automated Cartography. *Communication Association and Computing Machine*, 24(6), 381-395.

Gelfand, N., Mitra, N. J., Guibas, L. J., Pottmann, H. 2005. Robust global registration. Proceedings of the Third Eurographics Symposium on Geometry Processing, Vienna, Austria 4-6 July, 2005, pp. 197-206.

Gruen A., Akca D., 2005. Least squares 3D surface and curve matching. ISPRS J. of Phot. and Rem. Sens., 59(3), 151–174.

Hartley R., Zisserman A., 2003. Multiple View Geometry in Computer Vision. Cambridge University Press, Second Edition.

Horn B., Hilden H.M., Negahdaripour S., 1988. Closed-form solution of absolute orientation using orthonormal matrices. Journal of the Optical Society of America 5 (7), 1127–1638.

Huang Q.X., Flvry S., Gelfand N., Hofer M., Pottmann H. 2006. Reassembling Fractured Objects by Geometric Matching. ACM Trans. on Graphics 25 (3), 569–578.

Huber D., 2002. Automatic three-dimensional modeling from reality. PhD thesis, Carnegie Mellon University.

Kang Z., Zlatanova S., Gorte B., 2007. Automatic Registration of Terrestrial Scanning Data Based on Registered Imagery. FIG Working Week 2007, Hong Kong SAR, China, 13-17 May.

Läbe, T., Förstner, W., 2006. Automatic relative orientation of images. Proceedings of the 5th Turkish-German Joint Geodetic Days, March 29th - 31st, 2006, Berlin,

Liu L., Stamos I., Yu G., Wolberg G., Zokai S., (2006) Multiview Geometry for Texture Mapping 2D Images Onto 3D Range Data, IEEE Computer Society Conference on CVPR, New York, 17-22 June, Vol. 2 pp. 2293–2300.

Lowe D. G., 2004. Distinctive Image Features from Scale-Invariant Keypoints. IJCV 60(2), 91-110.

Mikolajczyk K., Schmid C., 2004. Scale & Affine Invariant Interest Point Detectors. IJCV 60 (1), 63–86.

Mitra N. J., Gelfand N., Pottmann H., and Guibas L., 2004. Registration of point cloud data from a geometric optimization perspective. Proceedings of the 2004 Eurographics/ACM SIGGRAPH Symposium on Geometry Processing,. 23-32.

Pottmann H., Huang Q. X., Yang Y. L., Hu S. M., 2006. Geometry and convergence analysis of algorithms for registration of 3D shapes. IJCV 67 (3), 277–296.

Rusinkiewicz S., Levoy M., 2001. Efficient variants of the ICP algorithm. In: Proceedings 3rd International Conference on 3D Digital Imaging and Modeling, pp. 145–152.

Seo J. K., Sharp G. C., Lee S. W., (2005) Range Data Registration Using Photometric Features IEEE Computer Society Conference on CVPR, 20-25 June, Vol. 2. 1140–1145.

Shragai Z., Barnea S., Filin S., Zalmanson G., Doytsher Y., 2005. Automatic Image Sequence Registration Based on a Linear Solution and Scale Invariant Keypoint Matching. International Archives of Photogrammetry and Remote Sensing. 36(3/W36): 5-11.

Ullrich, A., Schwarz, R., H. Kager, 2003. Using Hybrid Multi-Station Adjustment for an Integrated Camera Laser-Scanner System. in Proc. of Optical 3D Meas. Tech. VI, Zurich, Switzerland, 298-304.