

GEOMETRY-IMAGE-INTENSITY COMBINED FEATURES FOR REGISTRATION OF TERRESTRIAL LASER SCANS

Shahar Barnea, Sagi Filin

Mapping and Geo-Information, Technion – Israel Institute of Technology, Haifa, 32000, Israel
(barneas,filin@technion.ac.il)

Commission III/2

KEY WORDS: Terrestrial Laser Scanning; Registration; Resolution; Scale; Laser scanning; Algorithms

ABSTRACT:

The complexity of natural scenes and amount of information acquired by terrestrial laser scanners turns their registration into a complex problem. This complexity becomes even more pronounced when considering the relatively low angular resolution and manmade surfaces monotonicity that makes the detection of corresponding objects difficult. Since modern scanners are accompanied by consumer cameras of relatively high-quality and acquire laser pulse intensity data, it stands to reason utilizing these data sources for the registration. An integrative scheme benefits from the large body of keypoint-matching related research, thereby providing an efficient means for registration of pairs and multiple scans simultaneously. This paper proposes an efficient registration strategy via integration of the data content provided by terrestrial laser scanners. For this purpose, a model that is based on perspective invariant features is developed. Feature detection utilizes the scanner intensity data, with geometric characteristics derived from the scan; descriptive attributes are extracted from the images content. In this regards, the paper demonstrates how different information sources can be used advantageously for registration of terrestrial scans. Following the presentation of the model, its application on multiple scans of an outdoor complex scene is demonstrated. Results show that accurate point driven approaches are possible for very wide baseline scenarios.

1. INTRODUCTION

Registration of terrestrial laser scans (TLS) into a common reference frame is a fundamental processing step, with common practices involves deployment of artificial targets in the scene as reference objects, typically in the form of spheres or reflectors. To avoid manual intervention in the registration process, a growing body of research addresses the problem of autonomous registration in relation to both range images and terrestrial laser scans. Range image registration received much attention in the computer vision literature, focusing on feature based schemes and computational techniques. Feature based registration engages in extraction of geometric feature detection followed by generation of designated 3D descriptors to enhance the registration and eliminate the need for initial transformation parameters. Huber (2002) introduces spin images, Gelfand et al. (2005) propose integral volume descriptors, Huang et al. (2006) extract surface characteristics in multi-scale, and curvature based descriptors have been described in Gal and Cohen-Or (2006). Making use of image content, Seo (2005) and Smith et al. (2008) derive image based descriptors using the range image information. Differing from range images, terrestrial laser scans usually depict cluttered surroundings that are characterized by strong scale variations, occlusions, and foreshortening. The predominant registration approaches are geometry driven, where Brenner and Dold (2007) propose planar segment based matching, Dold and Brenner (2008) study different score functions as a means to validate the matching results, von Hansen (2006) develops an exhaustive-search mechanism for matching planes from two datasets, and Bae and Lichti (2008) use curvature variations as a matching criterion on local points. The descriptive content requires normal vector and curvature estimation, both are noisy measurements. Focusing on point features, Barnea and Filin (2008) propose a combinatorial approach for matching interest points that were extracted from range panoramas, and von Hansen et al. (2008) outlines a line-matching-based registration procedure.

Use of external or additional information sources for laser scans registration has received growing attention, particularly due to their texture richness compared to range data. Ulirsch et al. (2003) set a framework for integrating images and scanning data. Al-Manasir and Fraser (2006) use relative orientation of images for scan registration via signalized, artificial, targets. Liu et al. (2006) consider a framework with no rigid attachment between camera and scanner but with the imposition of specific geometric constraints. Kang et al. (2007) use the Moravec operator and cross-correlation as a means to find point correspondence between images and use those for the registration phase. Dold and Brenner (2006) use cross correlation of two textured planner patches as a matching criterion. Barnea and Filin (2007) proposed the scale invariant feature transform (SIFT) operator (Lowe, 2004) to extract corresponding points among images from different scans position, with Böhm and Becker (2007) extracting SIFT based corresponding points between the intensity channels of two scans. Kang et al. (2009) extract SIFT feature from the intensity channel and match those points within the 3D-space in a coarse-to-fine scheme. Wang and Brenner (2008) extract SIFT features from the intensity channel and use local geometric information related to the keypoint in order to filter noisy points, thereby improving the matching inlayers/outliers percentage. Modern image-based feature extraction and matching schemes (e.g., Lowe, 2004; Mikolajczyk et al., 2005; Lingua et al., 2009) relieve the need for registration parameters approximation, and simplify mutual features association via descriptors. Benefiting from these properties, particularly while resolving the low texture content of the range data, the utilization of additional information sources is only natural. Color images, provide high spatial resolution and rich descriptive information (e.g., Al-Manasir and Fraser, 2006; Barnea and Filin, 2007), while the laser intensity-channel provides a co-aligned data source that reflects difference in material, shares resolution, and is almost unaffected by external illumination conditions. Notwithstanding, naïve application of extraction concepts ignores the wide

baseline among scanner placements (much wider than common images baselines), as well as the panoramic data acquisition which is different in considerations than the narrow image field of view. As a result, the overlapping part among data from different scans may exhibit considerable difference in scale and perspective. Such characteristics reduce the level of success using common matching techniques (Mikolajczyk et al., 2005). Aiming at overcoming these limitations, this paper proposes exploiting knowledge about scene geometry for the registration process. Instead of a straightforward application of image-based concepts, a data- and scene-aware extraction methodology is proposed using integration of the available data sources, including: the range, intensity channel, and image based data. Such integration enables registration over wide baselines, offers robustness, and handles complex cluttered scenes, while accommodating the large data volume typical to TLS. The challenge is integrating the different cues that each channel provides into a general registration framework. As the paper shows, the proposed model enables high level of registration accuracy over wide-baselines, while being applied to complex scenes, and using no a priori knowledge of the scanners pose parameters.

2. METHODOLOGY

A successful registration requires deriving a rich set of localized keypoints, ones which are scale and perspective invariant. Data for that purpose include range data, intensity data, and images. Our proposed extraction scheme is based on keypoint detection in the intensity channel using geometric information from the range data. Matching is then performed using the rich image content, adapted to maintain perspective invariance. Range data availability, enables performing the actual registration and outlier detection in 3D object-space, offering an efficient computational scheme and robust transformation model.

2.1 Representation and Mutual alignment

The scanner and mounted camera, feature two reference frames which are co-aligned by a boresight transformation. The camera-scanner boresight relation can be encapsulated by a 3×4 projection matrix \mathbf{P} which represents the relation between an object space point and an image point:

$$x = \mathbf{P}X = \mathbf{K}(\mathbf{I} - t)X \quad (1)$$

where $X = [x \ y \ z \ 1]^T$ and $x = [u \ v \ 1]^T$, are object- and image-space points, respectively, in homogeneous representation; \mathbf{K} the calibration matrix, \mathbf{I} the identity matrix, and \mathbf{R} and t , the rotation matrix and translation vector, respectively. Radial and decentering lens distortions are calibrated and corrected for. For each scan, n images are acquired at predefined “stops” (every $360/n$ degrees). Assuming that, *i*) the camera is rigidly mounted to the scanner, *ii*) the intrinsic camera parameter are fixed and calibrated in advance, and *iii*) the acquisition position (of the “stop”) is fixed across all scanning positions, enable using the same projection matrices for all images of the same “stop” within different scans.

The scanned data (ranging and intensity), is represented in polar coordinates

$$(\rho \cos \varphi \cos \theta, \rho \cos \varphi \sin \theta, \rho \sin \varphi)^T = (x, y, z)^T \quad (2)$$

with φ and θ latitudinal and longitudinal coordinates of the firing direction and ρ the measured range. Polar coordinates offer lossless raster data representation as the angular spacing is fixed. Range and intensity values set pixels content.

2.2 Keypoints detection

Optimal detection should provide a sufficient amount of accurately localized keypoints with high repeatability rate and maintain high localization accuracy. Accuracy (or detection error) relates to the distance between two correctly matched points, and repeatability relates to the chance that a certain 3D point will be detected in all scans in which it appears. Foreshortening, and sampling resolution may affect the detection accuracy, and so are internal ranging and pointing accuracy.

Considering the information sources available for the detection, range data suffer from lack of surface texture and a consequent keypoint extraction in the object-to-background transitions zones. Different object appearance from different viewpoints will lead to low repeatability likelihood as well as poor localization of the detection across scans. Image-based keypoints enjoy higher-resolution and richer texture compared to the other two alternatives. However, viewpoint differences due to camera-to-scanner offsets, misalignment errors, and differences in acquired information between the scanned and imaged data, may lead to erroneous detections and affect repeatability and accuracy of the actual 3D points. Considering the shortcomings related to both data source, the detection is applied on the intensity channel, which offers richer texture than the range data, and is innately co-aligned to range measurements. Descriptive data is then extracted from the acquired images that accompany the scan and are texture richer. A major challenge in accurately localizing keypoints and maintaining high repeatability rates is accommodating variation in scale, particularly when considering the wide baselines over which data are acquired. An even keypoint coverage which compensates for variations in the object-to-scanner distance (objects closer to the scanner are recorded at higher spatial resolution than distant ones) is achieved via detection in a scale (distance) aware manner. Detection is performed adaptively, as a function of point range from the scanner. An adaptive smoothing is implemented as a family of convolution kernels characterized by different σ values where:

$$\sigma(\rho) = \frac{D}{8\rho \tan\left(\frac{\Delta}{2}\right)} \quad (3)$$

with D , the object-space size of the window, and Δ the angular sampling resolution. Keypoint extraction is then implemented using the Harris operator:

$$H = \det(\mathbf{A}) - k \cdot \text{tr}(\mathbf{A})^2 \quad (4)$$

with \mathbf{A} , the autocorrelation first-order partial derivatives matrix, and k , a tunable sensitivity parameter (Harris and Stephens, 1988). The proposed scheme leads to detection of points by introducing geometrically-aware scale sense, contrary to SIFT-like schemes that evaluate scale in a pattern-recognition-like manner.

2.3 Invariant Features Extraction

To extract descriptive features for the keypoints, the wide baseline is taken into consideration. Due to the strong perspective differences among scans, naïve image-patch extraction becomes insufficient as a means to obtain consistent features. Thus, local properties are extracted via a perspective invariant transformation.

Defining k as a keypoint whose 3D scan-related coordinates are X_k , and whose surface normal is defined by \vec{n} ; both keypoint

and corresponding normal define an independent reference frame whose origin is at X_k , and axes, \vec{u} , \vec{v} , \vec{w} , defined, by:

$$\vec{w} = \vec{n} ; \vec{u} = \vec{n} \times \mathbf{e}_3 ; \vec{v} = \vec{n} \times \vec{u} \quad (5)$$

in the scanner reference frame, with \vec{w} pointing at the normal direction, \vec{u} is arbitrarily determined, and is thus set to co-align with scanning system z vector, and \vec{v} completes a right-handed reference coordinate system, and \mathbf{e}_3 , the third vector in the \mathbb{R}^3 canonical basis. A point U in the local reference frame is transformed to the scanner frame via:

$$X = [\hat{u} \ \hat{v} \ \hat{w}] \cdot U + X_k = \begin{pmatrix} \hat{u} & \hat{v} & \hat{w} & X_k \\ 0 & 0 & 0 & 1 \end{pmatrix} U \quad (6)$$

Linking an image point, x , to its corresponding local reference frame point, U , is carried out via:

$$x = \mathbf{P} \begin{pmatrix} \hat{u} & \hat{v} & \hat{w} & X_k \\ 0 & 0 & 0 & 1 \end{pmatrix} U \quad (7)$$

when substituting Eq. (6) into Eq. (1). Perspective invariance is achieved by reprojecting the image data into an artificial image patch which is embedded in the uv plane (orthogonal to the keypoint normal). Defining u as an image coordinate in the artificial image patch; its relation to U is given by:

$$u = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u \\ v \\ w \\ 1 \end{pmatrix} = \mathbf{C} \cdot U \quad (8)$$

namely an orthogonal projection of the point U to the artificial image point. Eq. (8) defines a transformation from the local 3D reference frame, onto the artificial patch frame. When substituted into Eq. (7) (considering the plane at $w=0$), the following relation can be written:

$$x = \mathbf{P} \begin{pmatrix} \hat{u} & \hat{v} & \hat{w} & X_k \\ 0 & 0 & 0 & 1 \end{pmatrix} \mathbf{C}^T u \Rightarrow x = \mathbf{H}u \quad (9)$$

this relation holds, as multiplication by $\mathbf{C}^T \mathbf{C}$ sets w to zero, which is valid due to the localized nature of the point. These three transformations provide an explicit form of a homographic transformation, \mathbf{H} , between image patch pixels and their actual imaged points. It enables generating image patches which are independent of the actual viewing direction, providing an artificial front looking image of the detected keypoint. As such, the patch is not affected by the scanning position, but only by the local normal, thus becoming invariant to perspective viewing effects.

2.4 Robust normal assessment

Considering \mathbf{P} and \mathbf{C} being constant, the only varying parameter in Eq. (8) is \vec{n} , the local surface normal. \vec{n} is readily derived from the range panorama representation, $\rho(\varphi, \theta)$ via the cross product of the gradients at each point:

$$\vec{n}_{i,j} = \nabla \rho(\theta_{i,j}, \varphi_{i,j}) = \frac{\partial \rho_{i,j}}{\partial \varphi} \times \frac{\partial \rho_{i,j}}{\partial \theta} = \begin{pmatrix} X\varphi_{i,j} \\ Y\varphi_{i,j} \\ Z\varphi_{i,j} \end{pmatrix} \times \begin{pmatrix} X\theta_{i,j} \\ Y\theta_{i,j} \\ Z\theta_{i,j} \end{pmatrix} \quad (10)$$

with

$$\begin{bmatrix} X\varphi_{i,j} \\ Y\varphi_{i,j} \\ Z\varphi_{i,j} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} X_{i+1,j} - X_{i-1,j} \\ Y_{i+1,j} - Y_{i-1,j} \\ Z_{i+1,j} - Z_{i-1,j} \end{bmatrix} ; \begin{bmatrix} X\theta_{i,j} \\ Y\theta_{i,j} \\ Z\theta_{i,j} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} X_{i,j+1} - X_{i,j-1} \\ Y_{i,j+1} - Y_{i,j-1} \\ Z_{i,j+1} - Z_{i,j-1} \end{bmatrix} \quad (11)$$

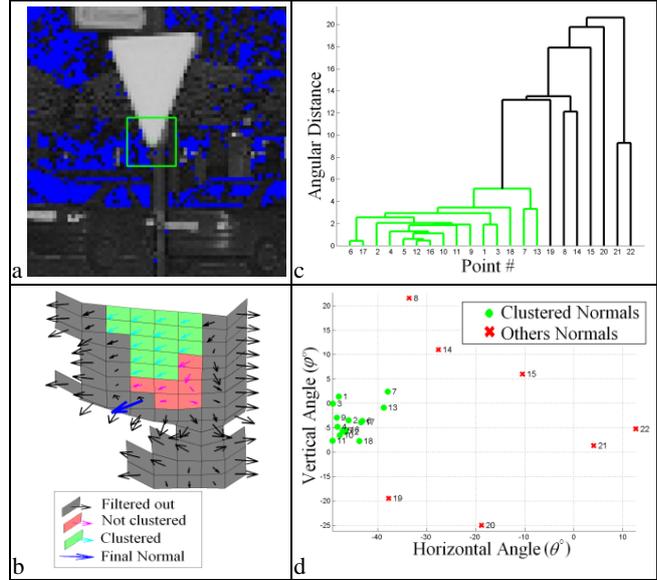


Figure 1: Robust normal assessment, a) detected keypoint as extracted from the intensity channel (the frame surrounding the keypoint is the window derived by Eq. (3)), b) filtering normals that lie around edges, c-d) remaining normals are clustered via dendrogram analysis, where the keypoint normal is calculated by averaging the cluster elements (blue arrow in 'b'); notice the significant difference between the final resulting normal and the one that was computed for that particular location.

The local differential tangent plane at the point is well-defined as long as the keypoint is lying on planar surface. Nonetheless, in most cases detected points may only partially lie on such surfaces, and in other cases the normal of points surrounding the keypoint may vary considerably (keypoints of the latter kind are the more likely to appear). To relax the surface planarity demand for the keypoint surroundings (where the normal determination is well-defined mathematically), an alternative approach that seeks the dominant normal direction is applied. Normal dominance is determined by evaluating all normals within a close neighborhood to the point. Neighborhood size is set as a function of the range, with points that are closer to the scanner having larger windows (Eq. (3)). Focusing on the actual object on which the keypoint lies, we analyze only normals of points that fall within a distance bounds (in reference to the keypoint) around the keypoint. Similarly, normals relating to points located on the object boundaries are discarded as they are unreliable due to discontinuity (or mathematically undefined). Such points are easily filtered using edge analysis on the range map. Even after filtering out potentially outlying normals, not all the remaining ones may lie on the dominant surface. The dominant one is selected by analyzing the distribution of the remaining normals within the window and selecting the most significant cluster among them. The sought cluster should be characterized by a dense set of samples. Differing from standard clustering related problems where the aim is finding different groups (mostly predefined number of groups) within a given datasets, the concern here is finding the most dominant set of instances and separating it from the rest of the data. Thus, the clustering is considered as a density based problem and a hierarchical/agglomerative approach is applied via dendrogram formation and analysis. The dendrogram is formed via single linkage (nearest neighbor) using the smallest distance between objects in the two clusters (Johnson and Wichern, 2002; Miller and Han, 2001) namely,

$$d(r, s) = \min \left(\text{dist}(x_r, x_s) \right), \quad i \in (1, \dots, n_r), \quad j \in (1, \dots, n_s) \quad (12)$$

Clustering is performed according to the horizontal and vertical angle of the normals, with a cutoff value defined by an angular distance (cf. Figure 1).

2.5 Matching and Registration

Following the formation of the artificial image patch (a fixed sized patch that matches the size of the object-space window, D), the generation of a descriptor follows. The plane homography provides a perspective invariance, up to rotation around the normal as the \vec{u} , \vec{v} components are arbitrarily set (Eq. (5)). The patch is derived from the image in which it appears closest to the center. In order to generate a descriptor and simultaneously overcome the rotation ambiguities we use SIFT descriptors which are rotation invariant. As a first step in the descriptor generation phase, the strongest bin out of 36 orientation histogram is set as a reference orientation. Following the orientation assignment, the actual descriptor is computed by dividing the patch into 4×4 regions. An eight bins orientation histogram is created for each one of the 16 regions and lead to 128 elements descriptor. The quantization of gradient locations and orientations makes the descriptor robust to small geometric distortions and small errors in the region detection.

Having a set of descriptors for each scan, the matching procedure is carried out. Euclidian distance between descriptors is used as a proximity measure for finding correspondence. Uniqueness of the matching is established by evaluating the quotient of the first and second best matches, requiring it to be greater than a preset value. Matching in an exhaustive manner can be computationally expensive (i.e., $O(n^2)$ with n the number of keypoints); following Agarwal et al. (2009) and Brown and Lowe (2003) all keypoints from all relevant scans are organized in a single kD tree. For each scan, matching is performed by traversing the tree and seeking the closest descriptor (and therefore keypoint) in the other scans. Computational complexity is reduced then to $O(n \log(n))$, and other than the improvement in the computational efficiency, the necessity to follow a pairwise framework is eliminated.

The matched descriptors provide a set of candidate points from which the transformation among scans is evaluated. Notice the distinction between the keypoint coordinate that is obtained directly from the laser scan and the descriptor that was derived from the rich image content. This way the actual 3D keypoint coordinates are not affected by registration errors between the scanner and the mounted camera, while at the same time rectification of the image patch is computed by the actual normal direction of the point. Once a set of matching points has been generated, the registration of the laser scan becomes an of the rigid body transformation estimation problem,

$$\mathbf{X} = \mathbf{X}_0 + (\mathbf{I} + \mathbf{S})^{-1} (\mathbf{I} - \mathbf{S}) \mathbf{x} \quad (13)$$

where \mathbf{I} is a 3×3 identity matrix, and \mathbf{S} , a skew-symmetric matrix. The transformation can be estimated linearly (e.g., Horn et al., 1988). Since some of the proposed matches are outliers, a RANSAC solution guides the parameter estimation (Fischler and Bolles, 1981). One of the appealing properties of a registration which is 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 is controllable and efficient.

3. RESULTS

The application of the proposed model is tested on scan set containing 16 scans. The scans were acquired by the Riegl LMS Z360i laser scanner (angular resolution of $\Delta\theta = \Delta\varphi = 0.12^\circ$) with scan contains 2.25 million ranges (creating a 750×3000 pixels image) spanning 360° horizontally and 90° vertically with

maximum ranges on order of 80m. Seven, six-mega-pixel size images were acquired per scan acquired by a Nikon-D100 camera with a 14 mm lens, and processed in full resolution. Overlap between images was $\sim 13.5\%$, equivalent to ~ 270 pixels. The data covers a courtyard-like square and offers a typical urban environment that is dominated by manmade objects. The environment features a significant amount of clutter.

Results summarizing the application of the model are presented in Table 1 and show that the registration reaches a centimeter level of positional accuracy and a hundredth of a degree level in angular terms. Accuracy is measured in terms of offsets between the computed parameters and manual registration of the scans, which act as ground truth. These offsets are well within the manual registration level of accuracy. Comparing the results to alternative point-based registration schemes show a considerable level of improvement. Using keypoint-based detection on the range data, accuracy was on order of ~ 0.5 - 1.5 m and angular accuracy was on order of $\sim 1^\circ$. These results enabled solving the coarse registration and to initiate a refinement, ICP-driven phase. A slightly better accuracy level was achieved when applying the SIFT on the acquired images. The level of accuracy reached here improves both settings dramatically. obtaining it, can be attributed to two contributors: *i*) the varied texture of the intensity data that allows homing on more distinct, and therefore accurate, keypoints than those that can be extracted in the range data, and *ii*) accurate localization – that relates to both the quality of detection process that extract points in their actual scale, and avoidance of alignment errors as well as viewing related artifacts between the images and the scanner. Table 1 also shows that accuracy is hardly influenced by the distance between scanners.

A detailed analysis of the model application focuses on two scans (6-7), whose baseline is 26.5 m, exhibiting strong perspective variations. Fig. 2 presents the intensity channel for both, showing that the captured scene is cluttered, dominated by a set of building facades that surround an inner court and is obstructed by trees and parking vehicles. Mutual objects appearing in both scans differ dramatically by scale and perspective. For the two scans, 26 correct matches were found, 20 of them are distinctive, namely separated in distance from one another. As an illustration, the blowup in Fig. 2 shows a point that appear in 26.8 and 5.3 m from the scanner respectively, which is front looking in scan '6' (upper), it is strongly distorted perceptively in scan '7'. Despite the strong differences between the two views of the same point following the rectification both patches look alike both in terms of scale and view. As can be noticed, using image data to generate the descriptor, benefits from high resolution and richer texture than the intensity channel. It increases the chances to establish a match. In order to evaluate the contribution of the proposed approach to alternative SIFT driven ones (e.g., Böhm and Backer, 2007; and Wang and Brenner, 2008) the matching is performed based on an image-to-image basis (similar to Barnea and Filin, 2007) and then on the intensity channel directly (Lowe's SIFT implementation was used here). As for the image-to-image based matching, all 7×7 images matching were preformed, but to only six correct distinctive matches total.

As an example, Fig. 3 shows matching results of one image pair and compares it to the results obtained by our approach. Notice that all the SIFT matches that were obtained are outlying, contrasting the six correct matches that were obtained using the proposed method. The difference between the number of extracted points is relevant not only in terms of quantity but is echoed in the reliability of identifying the actual matches (6 vs. 20), and in terms of the quality of the parameter estimation as points are being spread more widely across the scan. Applying the SIFT directly on the intensity channel of this particularly

dataset (images not shown) yielded almost no actual inliers (when applying image enhancements procedures, only one correct match was found).

L Scan	R Scan	dist [m]	distance error [cm]	Max. angular error. [(1/100)°]	# correct points
1	2	10.7	2.5	7.46	19
2	3	19	1.1	0.65	37
3	4	11.3	2.9	0.64	198
3	5	24.6	2.5	0.59	21
4	5	14.7	1.8	0.7	80
4	6	17.2	2.7	1.17	14
5	6	15.1	1.5	2.17	15
5	7	27.8	8.2	6.93	25
6	7	26.4	4.7	6.72	23
7	8	25	7.1	3.95	37
8	10	28.5	5.2	7.96	15
8	11	27.5	7.0	8.05	11
8	9	21.9	4.4	4.93	71
9	10	12.6	3.4	0.5	142
9	11	12.7	4.1	3.01	26
10	11	1.41	1.2	0.51	856
10	12	17.5	5.5	4.74	26
11	12	17	4.2	1.25	20
12	13	17	0.8	0.67	156
13	14	18.1	1.6	0.69	93
14	15	19.8	2.3	0.97	45
15	16	27.4	9.8	7.68	32

Table 1. Registration accuracy and number of correct tie-points obtained by applying the proposed method.

To demonstrate the model properties, selected matched keypoints are presented in Fig. 2. Fig. 2a demonstrates the more standard application of the proposed model of a strong perspective difference of a (largely) planar object. It shows a corresponding point that was detected on window corner. Despite the difference between the two patches due to perspective projection and scale, following the homographic projection and rescaling, the two patches appear similar. Fig. 4b shows an example of matched keypoints, significantly varying in scale. The left patch is at a distance of 16.5 m from the scanner, and the patch size is 128 pixels while the right keypoint is at a distance of 43.8 m generating a 44 pixels window. Notice that the computation has managed identifying the actual scale correctly, bringing the two, different patches into the same size when rescaled. Fig. 4c features a more complex case in which the actual 3D keypoint location falls outside of the background plane. The projective distortion between the two image patches is much more distinct than in the previous examples. Seeking the dominant normal direction, the clustering has managed to detect the actual plane and projecting both patches correctly. Fig. 4d shows an example of an object with different background under large perspective and scale deformation.

It is also noted that the descriptor generation enabled a natural application of a multi-scans registration scheme. As noted is Sec. 2.5, using a kD-tree based approach enables extending the matching from a pairwise framework into multiple scans. Fig. 3 demonstrates the extraction of keypoints that were detected as common to three scans (3, 4, 5) whose baselines are 3, 4: 11.3 m, 3, 5: 24.6 m, and 4, 5: 14.7 m respectively. Six matched triplets were identified among the three scans. Both detection over wide base lines and the ability to identify points in multiple scans allude to the repeatability of the proposed scheme and its potential use for autonomous efficient laser scans registration.

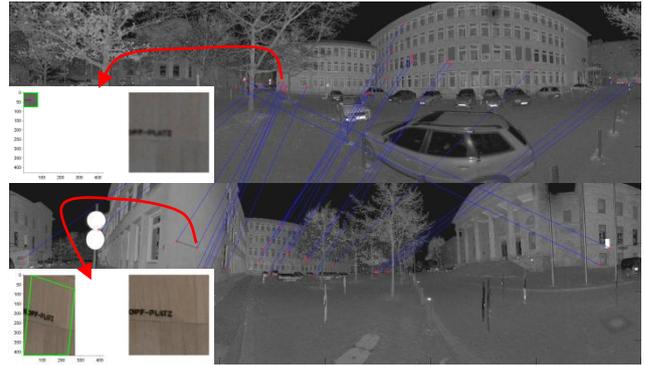


Figure 2: correctly matched points on scans 6-7.

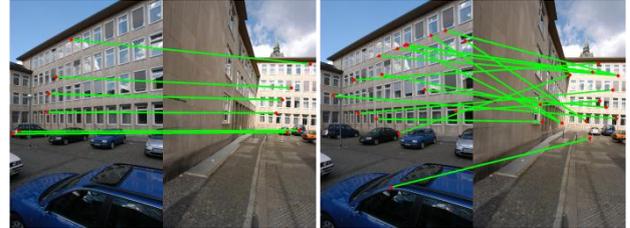


Figure 3: matched points on two images from scans 6 and 7; left) the proposed method, right) image-to-image SIFT results.

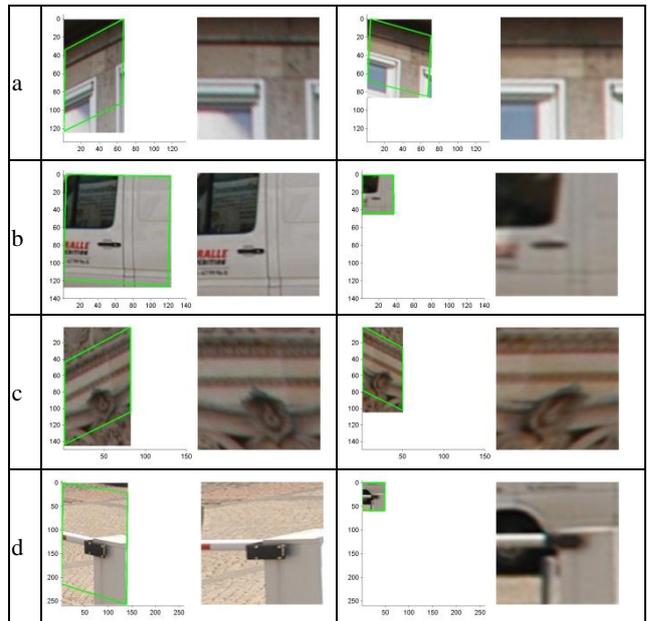


Figure 4: Features that matched using the proposed method, showing the difference between the original image patches as cropped from the image (left hand side of each feature) and the similar appearance of the object following the orthogonal projection (right). All patches are taken within a neighborhood of 0.5 m, and transformed into a 128x128 pixels size patch.

4. SUMMARY AND CONCLUSIONS

The registration results show how integration of the data captured by the scanner facilitates accurate registration, one equivalent to applying ICP methods. Aware design of the keypoint detector, based on the scanner intensity channel which is co-aligned with the range while providing richer textural content, enabled obtaining large amount of localized points that are also scale attentive. Via perspective invariance and relaxation of the planarity demand, the derived features provided suitable information for the subsequent matching phase. As the paper has demonstrated when considering the intensity channel

and image-driven data as a platform for an eventual rigid body transformation, sufficient amount of information has been extracted. Such information would not have been naturally detected using any of the range data registration methods. This information enabled accommodating the wide based-line scanner placements and the clutter that characterizes natural scenes. The rigid body transformation also allows using small subsets of points for the RANSAC hypothesis generation, thereby allowing greater robustness in the registration process. Application of the proposed framework may be extended into, e.g., 3D object querying, or image-to-image matching when DTM is available.

ACKNOWLEDGEMENTS

Funding for this research was provided in part by the Volkswagen-Niedersachsen Foundation.

REFERENCES

Agarwal S., Snavely N., Simon I., Seitz S. M., Szeliski R., 2009. Building Rome in a day. 12th IEEE International Conference on Computer Vision, Kyoto, Japan.

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

Bae, K. H. and Lichti D.D., 2008. A method for automated registration of unorganised point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, 63(1), 36-54.

Barnea S., Filin S., 2007. Registration of Terrestrial Laser Scans via Image Based Features. *International Archives of Photogrammetry and Remote Sensing*. 36(3/W52): 26-31.

Barnea S., Filin S., 2008. Keypoint based autonomous registration of terrestrial laser point-clouds *ISPRS Journal of Photogrammetry & Remote Sensing*, 63(1): 19–35.

Böhm J., Becker S., 2007. Automatic Marker-Free Registration of Terrestrial Laser Scans using Reflectance Features. 8th Conference on Optical 3D Measurement Techniques, Zurich, Switzerland. 338-344.

Brenner C., Dold C., 2007. Automatic relative orientation of terrestrial laser scans using planar structures and angle constraints. *International Archives of Photogrammetry XXXVI-3/W52*, Espoo, Finland pp. 84-89.

Brown M., Lowe G. D., 2003. Recognizing Panoramas. 9th IEEE International Conference on Computer Vision, Nice, France.

Dold, C., Brenner, C., 2006. Registration of terrestrial laser scanner data using planar patches and image data. *International Archives of Photogrammetry and Remote Sensing*, XXXVI(5). Dresden, Germany, pp. 78-83.

Dold, C., Brenner, C., 2008. Analysis of Score Functions for the Automatic Registration of Terrestrial Laser Scans. *International Archives of Photogrammetry XXXVII*, Beijing, China, pp. 417-22.

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.

Gal R., Cohen-Or D., 2006. Salient geometric features for partial shape matching and similarity. *ACM Trans. Graph.* 25(1): 130–150.

Gelfand N., Mitra N. J., Guibas L. J., Pottmann H., 2005. Robust global registration the 3rd Eurographics Symposium on Geometry Processing, Vienna, Austria. 197–206.

Harris, C., Stephens, M.J., 1988. A combined corner and edge detector. *The 4th Alvey Vision Conference*. 147–152.

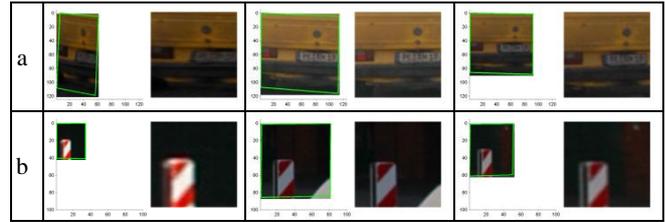


Figure 5: Features from three scans extracted using the proposed method.

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.

Johnson, R.A., Wichern, D.W., 2002. *Applied Multivariate Statistical Analysis*, 5th ed., Prentice Hall, NJ.

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.

Kang Z., Zlatanova S., 2009. Automatic Registration of Terrestrial Scan Data based on Matching Corresponding points from Reflectivity Images. *Urban Rem. Sens. Joint event*, 1-7.

Lingua A., Marenchino D., Nex F., 2009. Performance Analysis of the SIFT operator for automatic feature extraction and matching in photogrammetric Applications. *Sensors* 9(5): 3745-66.

Liu L., Stamos I., Yu G., Wolberg G., Zokai S., 2006. Multiview Geometry for Texture Mapping 2D Images onto 3D Range Data. *CVPR*, New York City, 2: 2293–2300.

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

Mikolajczyk K., Tuytelaars T., Schmid C., Zisserman A., Matas J., Schaffalitzky F., Kadir T., and Van Gool L., 2005. A comparison of affine region detectors. *IJCV*, 65(1/2):43–72.

Miller H.J., Han J., 2001. *Geographic data mining and knowledge discovery: an overview*. London, New York : Taylor & Francis.

Seo J. K., Sharp G. C., Lee S. W., 2005. Range Data Registration Using Photometric Features. *CVPR*, San Diego, 2: 1140–1145.

Smith E. R., King B. J., Stewart V. C., Radke R. J., 2008. Registration of Combined Range-Intensity Scans: Initialization Through Verification *CVIU*, 110(2): 226-244.

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

von Hansen, W., 2006. Robust Automatic Marker-free Registration of Terrestrial Scan Data, *International Archives of Photogrammetry XXXVI part 3*, Bonn, Germany, pp. 105-110.

von Hansen W., Gross H., Thoennessen U., 2008. Line-Based Registration of Terrestrial and Airborne Lidar Data. *International Archives of Photogrammetry XXXVI (B3a)*, Beijing, China, pp. 161-167.

Wang Z., Brenner C., 2008. Point Based Registration of Terrestrial Laser Data Using Intensity and Geometry Features. *International Archives of Photogrammetry XXXVII*, Beijing, China, pp. 583-589.