Development and Evaluation of a Two-Staged 3D Keypoint Based Workflow for the Co-Registration of Unstructured Multi-Temporal and Multi-Modal 3D Point Clouds

Steffen Isfort[1], Melanie Elias[1], Hans-Gerd Maas[1]

[1] Institute of Photogrammetry and Remote Sensing, TUD Dresden University of Technology, Germany - (steffen.isfort, melanie.elias, hans-gerd.maas)@tu-dresden.de

Keywords: 3D keypoints, keypoint detectors, keypoint descriptors, rigid registration, deformation analysis

Abstract

Robust and automated point cloud registration methods are required in many geoscience applications using multi-temporal and multi-modal 3D point clouds. Therefore, a 3D keypoint-based coarse registration workflow has been implemented, utilizing the ISS keypoint detector and 3DSmoothNet descriptor. This paper contributes to keypoint-based registration research through variations of the standard workflow proposed in the literature, applying a two-staged strategy of global and local keypoint matching as well as prototypical keypoint projection and fine registration based on ICP. Further, by testing the utilized detector and descriptor on unstructured, multi-temporal and multi-source point clouds with variations in point cloud density, generalization ability is tested outside benchmark data. Therefore, data of the Bøverbreen glacier in Jotunheimen, Norway has been acquired in 2022 and 2023, deploying UAV-based image matching and terrestrial laser scanning. The results show good performance of the implemented robust matching algorithm PROSAC, requiring fewer iterations than the well-known RANSAC approach, but solving the rigid body transformation with TEASER++ is faster and more robust to outliers without demanding pre-knowledge of the data. Further, the results identify the keypoint detection as most limiting factor in speed and accuracy. Summarizing, keypoint-based coarse registration on low density point clouds, applying a global and local matching strategy and transformation estimation using TEASER++ is recommended. Keypoint projection shows potential, increasing number and precision in low density clouds, but has to be more robust. Further research needs to be carried out, focusing on identifying a fast and robust keypoint detector.

1. Introduction

3D Point clouds are used for spatio-temporal high-resolution change detection and deformation analysis in various applications such as in urban areas (Stilla and Xu, 2023), landslide monitoring (Esposito et al., 2017; Mayr et al., 2018) and glacier monitoring (Ulrich et al., 2021; Zahs et al., 2019; Schwalbe et al., 2008), making use of various techniques like image matching on the basis of unmanned aerial vehicles (UAVs) and terrestrial laser scanning (TLS). This requires the co-registration of multi-temporal point clouds in a shared coordinate reference system (CRS). This task is usually solved by georeferencing all epochs either directly by additional sensors like GNSS or indirectly by usage of measured ground control points (GCP).

3D point cloud registration without GCPs is in general defined as an optimization problem to find rigid transformation parameters minimizing the projection error between a pair of point clouds. In multi-temporal point clouds, robust transformation based on stable areas is required due to changes between both epochs. An overview on registration approaches and categorizations is given in several papers, e.g. by Stilla and Xu (2023), Dong et al. (2020), Zhang et al. (2020), Wujanz et al. (2016) and Pomerleau et al. (2015). According to the categorization by Stilla and Xu (2023), one registration approach is the feature-correspondence-based registration. It promises 3D point cloud registration without prior knowledge in any orientation. The basic idea is the calculation of corresponding features, followed by a transformation estimation based on those feature correspondences. While features can be embedded in global-information-based and end-to-end registration pipelines, this paper focuses on local-information-based registration. Hand-crafted local feature detectors such as the intrinsic shape signature (ISS) (Zhong, 2009) and descriptors like FPFH (Rusu et al.,

2009) have been around for many years, using information in a local neighborhood of points. Local learning-based detectors like USIP (Li and Lee, 2019) and descriptors like 3DSmoothNet (Gojcic et al., 2019), implemented in this paper, have more recently been developed.

Common to all learning-based methods is the dependency on sufficient training data, which is especially relevant for registration methods based on global information and end-to-end learning (Stilla and Xu, 2023; Zhang et al., 2020). Many of the methods developed originate from the field of computer vision and use indoor data sets for training and evaluation like 3DMatch (Zeng et al., 2017) or data sets designed for object recognition like Random Views set (Tombari et al., 2013). Frequently used outdoor data sets are ETH (ETH Zürich, n.d.), which is a TLS benchmark data set for viewpoint registration, and the KITTI odometry (Geiger et al., 2012) data set. However, there is a gap between algorithm development on benchmark data and testing the generalization ability on real outdoor applications with variations in data and sensors. Probably best investigated are hand-crafted local feature detectors and descriptors, e.g. Guo et al. (2016) and Hänsch et al. (2014). But, besides an evaluation on outdoor data by Pirotti et al. (2023), there is a lack on testing those methods on unstructured data captured in outdoor environments, as with the learning-based methods.

This paper investigates robust and automated registration methods for 4D glacier monitoring using multi-temporal and multi-source point clouds based on a 3D keypoint coarse registration approach described in, e.g., Gojcic et al. (2019). The influence of data source and point cloud density is analyzed and performance limiting influencing factors are identified. Further, a two-staged matching strategy is tested, as well as a prototyp-
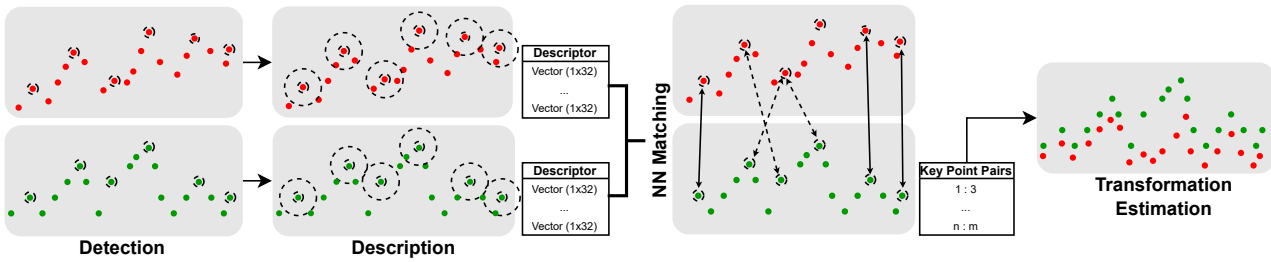
Figure 1. Implemented keypoint based registration workflow consisting of three major steps: keypoint detection using ISS, keypoint description using 3DSmoothNet and matching in feature space and transformation estimation using SVD/TEASER++.

ical keypoint projection and fine registration approach. Section 2 describes the implemented workflow divided in 3D keypoint detection, description, transformation estimation and workflow modifications. Section 3 describes the data sets used to evaluate the workflow, results are presented and discussed in section 4. Section 5 summarizes the key findings.

2. Workflow

Figure 1 illustrates the major workflow steps of keypoint detection, description and transformation estimation.

2.1 Keypoint Detection

In the detection step, for both point clouds P and Q a subset of keypoints $p \in P$ and $q \in Q$ is subsampled. Since corresponding keypoints are to be found between p and q, the detected keypoints should match as closely as possible in terms of the represented terrain point for a number of reasons: First, the description of a pair of best corresponding keypoints will vary if the detected points don't represent identical terrain, making matching by descriptors more difficult. Second, the achievable transformation accuracy is lower, if the two points are not identical. Third, as the computation time depends on the number of keypoints, detection of a small number of high quality keypoints is preferred to reduce computational costs. But since the sensor and the chosen viewpoints influence the spatial sampling and occlusions, as well as the quality of the measured 3D points, 3D keypoint detection is limited in its capability to detect best possible matching points.

A simple and fast approach is to randomly filter a fixed number n of points from both point clouds, like applied by Gojcic et al. (2019), without the need for point features like normals or eigenvalues. But since a random subsampling is limited to a random chance of identical point detection, keypoint detector algorithms have been developed. Algorithms such as Harris3D (Sipiran and Bustos, 2011), SIFT-3D (Rusu and Cousins, 2011) and ISS (Zhong, 2009) sample points with high saliency and rich geometry, making them good to describe and to find correspondences. In this paper, the ISS algorithm is used. It has shown good repeatability in studies by Guo et al. (2016). Points with locally maximal and sufficiently varying eigenvalues are defined as keypoints. The eigenvalues are calculated from the scatter matrix of the surrounding points in a local radius. For comparison, a random keypoint subsampling scheme is tested as well.

2.2 Keypoint Description and Matching

The second workflow step is keypoint description. For each detected keypoint $p_i \in p$ and $q_i \in q$, a feature vector $f_{p_i} \in F_p$

and $f_{q_i} \in F_q$ is calculated, where F_p and F_q denotes the global n-dimensional feature spaces for both sets of keypoints. Such a feature vector should be invariant to 3D affine transformations and should describe the point as uniquely as possible to enable high matching rates and thus enable fast transformation estimation. Handcrafted feature vectors such as FPFH (Rusu et al., 2009), USC (Tombari et al., 2010) and SHOT (Salti et al., 2014) first estimate a local reference frame to reach rotation invariance. Depending on the algorithm, feature vectors describe variations of point density or normal vector orientations in a local neighborhood. Since learning-based descriptors outperform those handcrafted descriptors in tests on benchmark data, only the learning-based descriptor 3DSmoothNet (Gojcic et al., 2019) is used. It is rotation invariant and although trained on 3DMatch RGB-D indoor data, it showed good generalization across sensors and scenes on ETH outdoor data. In this paper, the 32-dimensional feature vector is used.

Matching is done by a nearest neighbor (NN) search in global feature space using an octree structure, resulting in a set of keypoint correspondences C. Since there are mismatches, a filter is applied to consider high quality matches in the following transformation estimation. Following Gojcic et al. (2019), only mutual correspondences defined as

$$C = \{\{p_i \in p, q_j \in q\} \mid f_{q_j} = nn(f_{p_i}, F_q) \wedge \\ f_{p_i} = nn(f_{q_j}, F_p)\} \quad (1)$$

are retained, where $nn()$ denotes the nearest neighbor search applying the $L2$ norm.

Besides the best match, the second nearest match is considered for each pair of corresponding keypoints, in the following denoted as NN_{1st} and NN_{2nd}. For simplification, searching is only done in one direction. Calculating the distance ratio R between the first and second match as

$$R = 1 - \frac{NN_{1st}}{NN_{2nd}} \quad \{0 \leq R \leq 1\} \quad (2)$$

provides a quality estimation for each pair of matched keypoints, used in the following transformation estimation algorithm.

2.3 Transformation Estimation

The resulting set of keypoint pairs is used in a transformation estimation pipeline making use of the Progressive Sample Consensus (PROSAC) algorithm (Chum and Matas, 2005) and singular value decomposition (SVD) (Umeyama, 1991). Since the matches are expected to have a high amount of outliers,
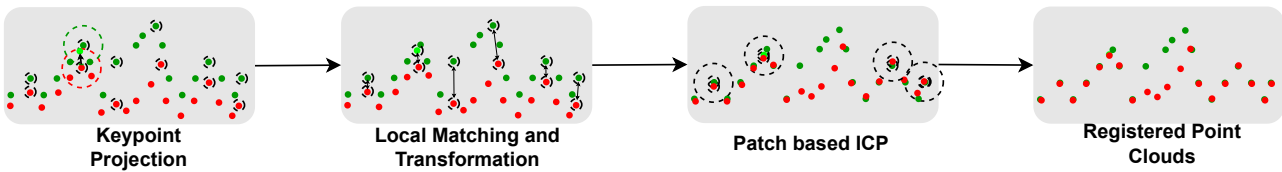
Figure 2. Workflow modifications after coarse registration: Improving keypoint identity by keypoint projection (light green), followed by a local matching strategy and a patch based ICP.

either by changed topography, false detections or false matches, a fast and robust algorithm has to be used. In contrast to the Random Sample Consensus (RANSAC) algorithm, where each pair of correspondences is selected with equal probability, the PROSAC algorithm sort the pairs according to a quality function, in this case the distance ratio R. Starting from a minimum subset of three pairs to solve the transformation, the set from which the points are drawn is extended by the next best pair of points. This continues until all pairs of points have been used and the PROSAC algorithm continues like the RANSAC algorithm. In consequence, high quality pairs are more likely to be selected, solving the transformation with a smaller number of iterations. Its success depends on the quality function used. As with the RANSAC algorithm, the number of inliers is estimated in each iteration by checking the distances of the keypoint pairs for exceeding a threshold value τ in Euclidean space. Like RANSAC, the PROSAC algorithm terminates when a predefined number of inliers is reached or a predefined number of iterations has been tested, so it requires prior knowledge of the data. As an alternative, TEASER++ (Yang et al., 2021) can be applied for rigid body transformation estimation. It promises a very fast and robust transformation estimation without prior knowledge and is tested for comparison.

2.4 Modifications

Assuming a coarse registration is possible through the proposed workflow, suggestions are made in the following to increase the number and accuracy of valid keypoint matches (see figure 2).

2.4.1 Local Matching. To improve the descriptor matching with a given coarse registration, the global feature matching task in F_p and F_q (see eq. 1) can be reduced to a feature matching task in a local feature space. For a pair of correspondences $\{p_i \in p, q_j \in q\}$, the subsets of local features in F_p and F_q are defined as:

$$
\begin{aligned}
F_q(local) &= \{f \in F_q \mid L2_{Norm}(q_f, p_i) \leq r\} \\
F_p(local) &= \{f \in F_p \mid L2_{Norm}(p_f, q_j) \leq r\}
\end{aligned}
\quad (3)
$$

where p_f and q_f denote the keypoints corresponding to a feature f.

2.4.2 Keypoint Projection. To overcome the problem of missing keypoint identity between two point clouds, for a pair of corresponding keypoints $\{p_i \in p, q_j \in q\}$, keypoint p_i shall be projected in point cloud Q. Therefore, in both point clouds a patch of k points is selected in a local sphere with radius r around the keypoints.

$$
\begin{aligned}
Patch_P &= \{k \in P \mid L2_{Norm}(k, p_i) \leq r\} \\
Patch_Q &= \{k \in Q \mid L2_{Norm}(k, q_j) \leq r\}
\end{aligned}
\quad (4)
$$

Since both clouds are already roughly registered, ICP should be applicable in case of a correct correspondence. Thus, both patches are registered by a point-to-point ICP algorithm, where the Euclidean distance between both corresponding keypoints $\{p_i, q_j\}$ is used as initial transformation estimation of both patches. By applying the resulting transformation T to the first keypoint, it is projected to the second cloud, where a new, originally not measured point is created with a better correspondence than the first matched keypoint. The list of corresponding keypoints is also updated. To reduce the computational costs, the keypoint projection is applied only to correspondences with a high estimated quality.

2.4.3 Fine Registration. Despite the optimizations, the keypoint registration is still based on point pairs which don't correspond in identity. So, a fine registration algorithm is needed, making use of the rich point cloud geometry. Therefore, all points in a local sphere of radius r around all final inlier keypoints $\{p_{final}, q_{final}\}$ are searched. Practically, these sets are local patches (LP) in the point clouds.

$$
\begin{aligned}
LP_P &= \{k \in P \mid \exists p \in p_{final}, L2_{Norm}(k, p) \leq r\} \\
LP_Q &= \{k \in Q \mid \exists q \in q_{final}, L2_{Norm}(k, q) \leq r\}
\end{aligned}
\quad (5)
$$

Based on those local patches around the keypoint pairs, an ICP fine registration is performed. This simple approach follows the assumption of stable areas in a local neighborhood of inlier keypoints.

3. Evaluation Data

For evaluation, a multi-temporal and multi-sensor measurement of the front area of the Bøverbreen glacier in Jotunheimen, Norway has been carried out in mid September 2022 and late August 2023 with UAV imaging and TLS in both epochs. Figure 3 shows the study area with an illustration of the glacier outlines and targets used for georeferencing.

UAV Image Data. In total, 385 aerial images were taken in 2022 and 881 images in 2023 by nadir cross-grid flights with a few additional 45° off-nadir (oblique) images applying the UAV DJI Phantom 4 RTK. Four GCPs measured with RTK-GNSS as well as the UAV RTK-GNSS were used for integrated georeferencing in WGS84/UTM32N and ellipsoidal heights. Some remaining signalized ground points were used as check points. Image data processing was done by Structure-from-Motion (SfM) and Multi-View Stereo (MVS) with Agisoft Metashape version 2.0.2, resulting in two 3D point clouds for 2022 and 2023. More detail on the data is provided in Elias et al. (2024). Using Metashape, both point clouds were slightly filtered by a minimum of three images used for 3D point reconstruction and subsampled to four clouds with a pointspacing of 2 cm, 5 cm,
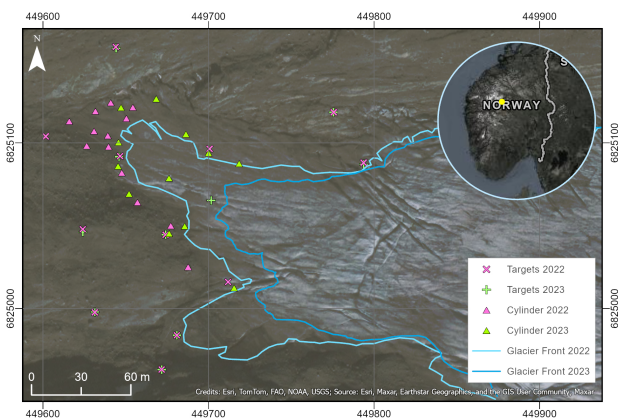
Figure 3. Study area: Bøverbreen glacier in Jotunheimen, Norway. Four UAV GCPs were used for georeferencing, others as check points. Cylinder targets were used for TLS georeferencing. Map: WGS84/UTM32N.

10 cm (see figure 4) and 50 cm for performance reasons as well as workflow evaluation on varying densities. Ground truth point cloud registration of both epochs was done using the georeferencing as basis, followed by a fine registration applying an ICP in manually selected stable areas.

TLS. The TLS data was collected with a RIEGL VZ-400i using 21 scan positions in 2022 and 12 in 2023, distributed in the glacier's front area. Georeferencing was done using 13 respectively 12 temporary cylindrical targets in 2022 and 2023, measured with RTK-GNSS in WGS84/UTM32N and ellipsoidal heights. Registration of single scan positions and both epochs was done using RiSCAN PRO version 2.8. First, all scan positions from 2022 were registered using RiSCAN's automatic registration algorithm and multi station adjustment. Second, all scan positions from 2023 were aligned to the 2022 point cloud with fixed registered scan positions, again using the automatic registration workflow and the multi station adjustment. For the registration of scan positions, the maximum residuals at the cylinders are 1.7 cm for the scan positions in 2022 and 1.5 cm for the scan positions in 2023. Both are considered within the expectable accuracy range. In 2022, the maximum uncertainty in plane features and cylinders is in the southeastern area, because of a scan position arrangement without loop closure. In 2023, most uncertainty is in the glacier front area from north to south, since there are rather large changes in topography. But no systematic error due to changes between the two epochs can be identified. Finally, the project was georeferenced using the cylindrical targets from both epochs. Besides some manual outlier filtering, point clouds from both epochs have been downsampled applying a voxel filtering with a resolution of 2 cm, 10 cm and 50 cm, reducing the original size from about 200 M/130 M points to 20 M/15 M, 4 M/4 M and 0.4 M/0.4 M for 2022/2023, respectively. Figure 5 shows the resulting point cloud for 2022. The laser scans are limited in terms of completeness on the sheer ice of the glacier, due to the scanner's wavelength in near infrared. Cross-source ground truth registration between UAV and TLS point clouds was done like UAV point cloud registration, using the georeferencing as basis, followed by a fine registration applying an ICP in manually selected stable areas.

The proposed keypoint based pairwise registration workflow has been tested on nine pairs of source and target clouds, using inter-source (TLS:TLS, UAV:UAV) and cross-source
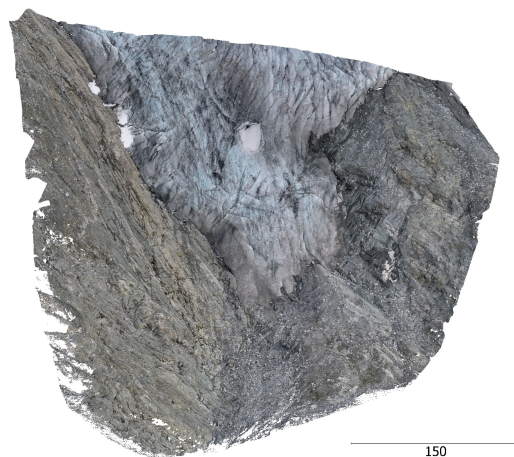

Figure 4. UAV point cloud from 2023, subsampled with a voxel filter of 10 cm to approx. 16 M points. Scale in meters.
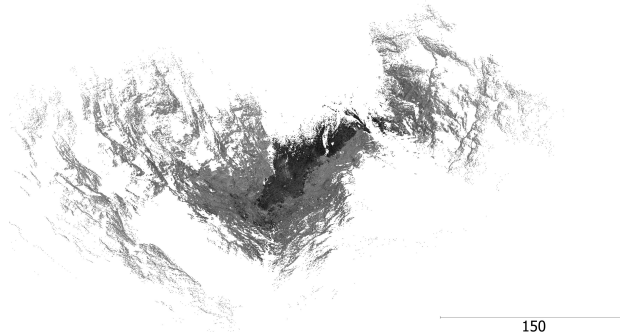

Figure 5. TLS point cloud from 2022, subsampled with a voxel filter of 2 cm to approx. 20 M points. Scale in meters.

(TLS:UAV) clouds with varying point densities of 2 cm, 10 cm and 50 cm. Please note that due to the large amount of data, the UAV:UAV registration could only be calculated after a voxel filtering of 5 cm. Since both point clouds are initially correctly registered, the source cloud is transformed using an arbitrary transformation matrix T. Thus, the transformation searched for is its inverse T^{-1}.

4. Results and Discussion

For comparability, results will be discussed grouped by subtasks and not fully consistent with the workflow sequence presented.

Keypoint Detection. Keypoints were detected using ISS with a radius of 1.0 m for 2 cm (5 cm) and 10 cm point spacing and 2.0 m for 50 cm point spacing, controlling the local neighborhood. Local roughness parameters (λ_1, λ_2) were set to 0.9. Figure 9 shows example results for ISS keypoint detection in the UAV point cloud from 2023. For comparison, a similar number of random keypoints has been sampled. It is in the nature of the detection method, that results for randomly detected keypoints may vary with each iteration, but tests confirmed that they remain quite stable in their illustrated magnitudes. Applying the ground truth transformation, detected keypoints were matched by a mutual NN search in Euclidean space, providing information about theoretically possible true matches. Figure 6 summarizes the results at certain distance thresholds. First of
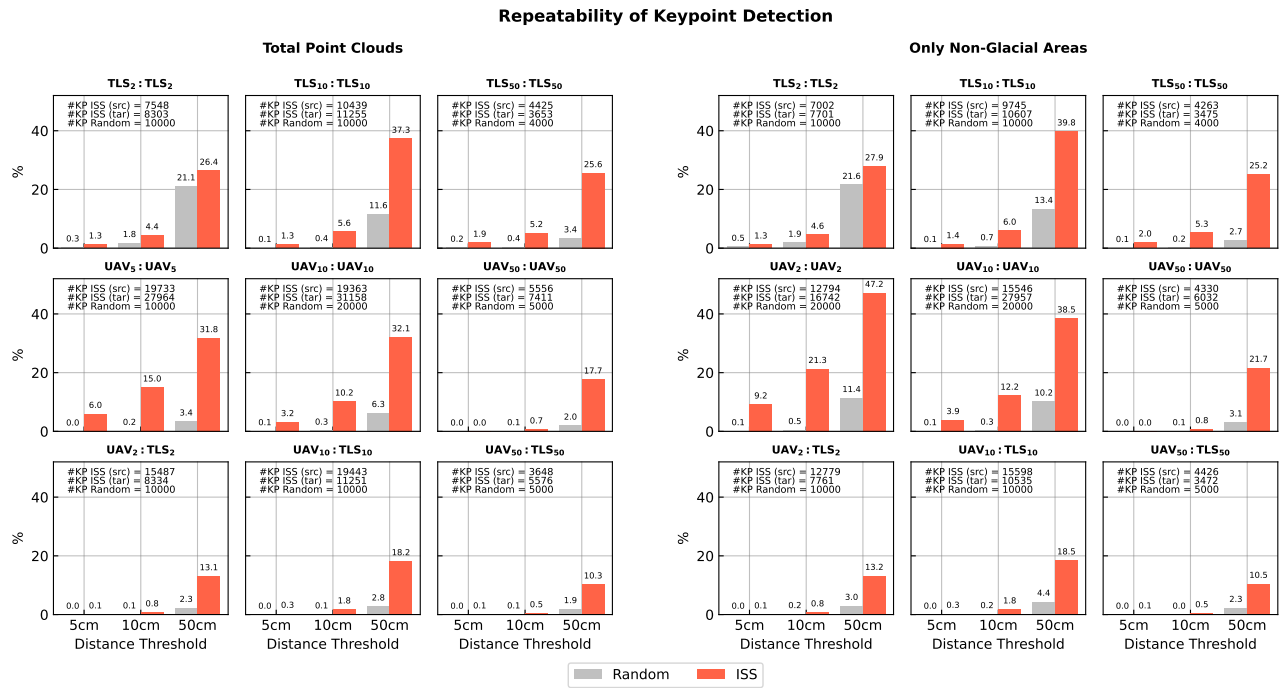
Figure 6. Repeatability of keypoint detection between two epochs as number of mutual nearest neighbor (NN) matches between the source and target point clouds in relation to the smaller total number of detected keypoints (%). Thresholds on x-axes denote to the maximum NN distance under ground truth transformation. Left: Tested on total point clouds, Right: non-glacial parts of the clouds. Subplot title: Tested source to target (src : tar) registration for inter-source and cross-source data. Subscription denotes for the downsampling voxel filter resolution in cm.

all, it can be stated that the random detection of keypoints (expectably) has a worse repeatability of about 80 % lower in average than the detection of keypoints with a dedicated detection algorithm. Further, repeatability in TLS:TLS and UAV:TLS registration increases with a lower point density and remains quite stable in UAV:UAV registration. Probably, filtering the TLS data has the effect of a low-pass filter and reduces the complexity to the essential geometry, which is even detectable in lower densities or with gaps in data. Overall, the repeatability is rated as too low. This depends in particular on the desired accuracy and therefore the threshold value. However, considering the high information density and quality of the point clouds, a higher repeatability should be aimed for. Even with a threshold value of 50 cm, about 60 % of the detected keypoints are still not repeatable, which considerably limits the efficiency of the entire workflow. The lack of repeatability is not essentially caused by the deformations between the two epochs. Detection in only non-glacial areas, which can be regarded as largely stable in the context of the applied threshold values, brought only minor improvements. In the following analysis, the ISS keypoints were used for matching and transformation estimation, since they show the better results compared to randomly chosen ones.

Keypoint Description and Matching. Feature description has been done in a local radius of 3.0 m in high and medium densities and 6.0 m in low densities. Global and local matching results are shown in figure 7, categorized according to a distance threshold under ground truth transformation and in relation to the respective possible matches from the NN search.

Matching results show, that in general, global feature matching delivers higher success rates in lower point densities. Second, the matching can be significantly improved by a regional matching strategy applying a coarse registration, even with a conservative radius of 5 m set in this investigation. However, the absolute number of valid matches is quite low in relation to the absolute number of detected keypoints. For example 205 keypoints have been globally matched below 10 cm in TLS:TLS with a
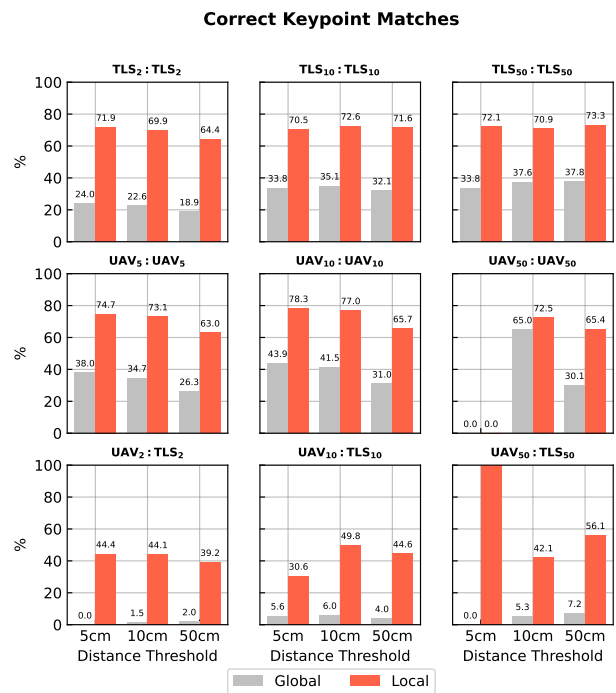

Figure 7. Recall of global and local mutual NN-based keypoint matching in feature space as percentage of possible mutual nearest neighbors in Euclidean space under ground truth transformation and different thresholds. Features computed using 3DSmoothNet.

medium density, which is 35.1 % of possible valid matches, but just about 2 % of the total number of keypoints. The problem is particularly evident in global cross-source registration, where the total number of possible matches is already low. In addition, the performance of the descriptor is significantly worse here. As a result, only a few individual points are matched correctly.

Keypoint Projection. After applying the coarse transformation estimated from the global matches, local matches with a quality indicating feature distance ratio of 0.6 and higher were projected in patches with a local radius of 0.5 m in high and medium densities and 2.0 m in low densities from the source cloud to the target cloud. Figure 8 shows the results as an increase or decrease in the number of valid keypoint matches. In most registrations, the number of valid matches could be increased slightly, especially in UAV:UAV registrations. The largest increase can be seen in the low density TLS:TLS registration. However, in some cases the number of valid matches was lower. This is likely due to the influence of point density on the ICP, which is particularly relevant for TLS:TLS and mixed registration and requires further investigation. As a standard point-to-point ICP algorithm is used, the computational time for projection is high, especially in high-density clouds. However, at low densities it may have the potential to significantly increase the accuracy of matched keypoint pairs.
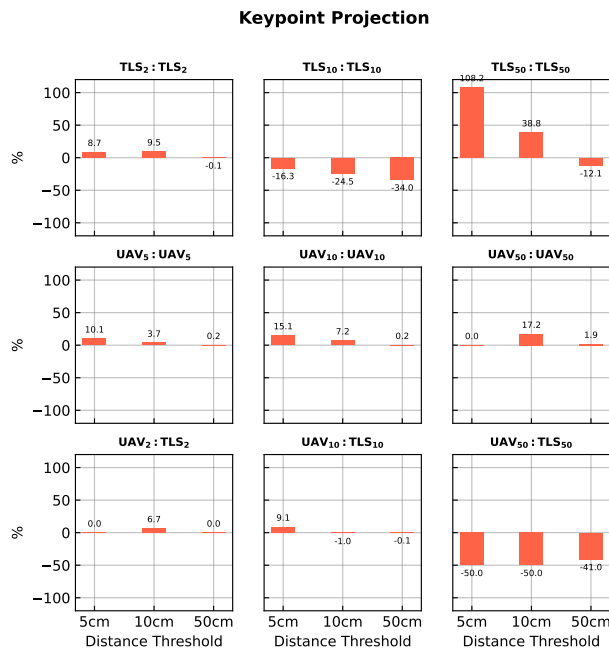

Figure 8. Increase or decrease of true keypoint matches after keypoint projection under coarse transformation.

Transformation Estimation. For transformation estimation, PROSAC in combination with SVD and for comparison TEASER++ is tested in global and local transformation estimation. The PROSAC inlier distance is set to 20 cm in global and 5 cm in local transformation. Figure 10 shows an example of PROSAC inliers in the UAV point cloud from 2023. Following Pomerleau et al. (2015), results are evaluated by comparing the estimated transformation matrix and the ground truth transformation, considering the errors in translation e_t and rotation e_r:

$$\Delta T = \begin{bmatrix} \Delta R & \Delta t \\ \mathbf{0} & 1 \end{bmatrix} = T_r T_g^{-1} \tag{6}$$
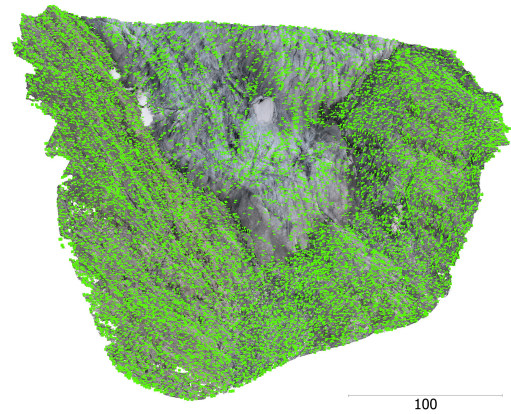

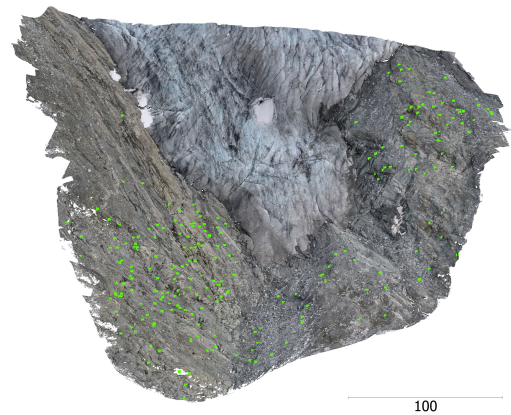Figure 9. ISS detected 31158 keypoints in UAV 2023 point cloud (10 cm). Scale in meters.

Figure 10. 225 PROSAC inlier keypoints after local matching in UAV 2023 point cloud (10 cm). Scale in meters.
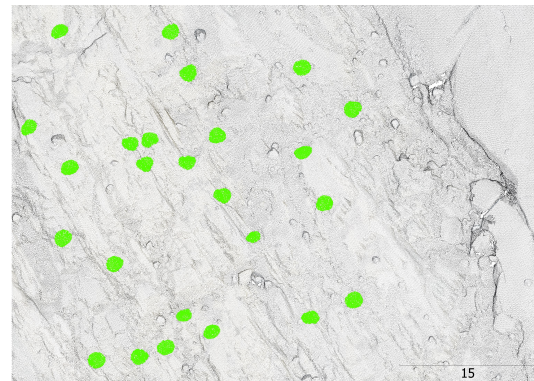

Figure 11. Patches around inlier keypoints used for ICP fine registration. Scale in meters.

$$e_t = \|\Delta t\| = \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2} \tag{7}$$

$$e_r = arccos\left(\frac{trace(\Delta R) - 1}{2}\right) \tag{8}$$

Figure 12 shows the resulting errors for global and local transformation estimation. In general, the registration estimation is more precise in inter-source registration tasks. While showing precise registration estimations over all point densities in TLS:TLS registration, the precision is better with higher point
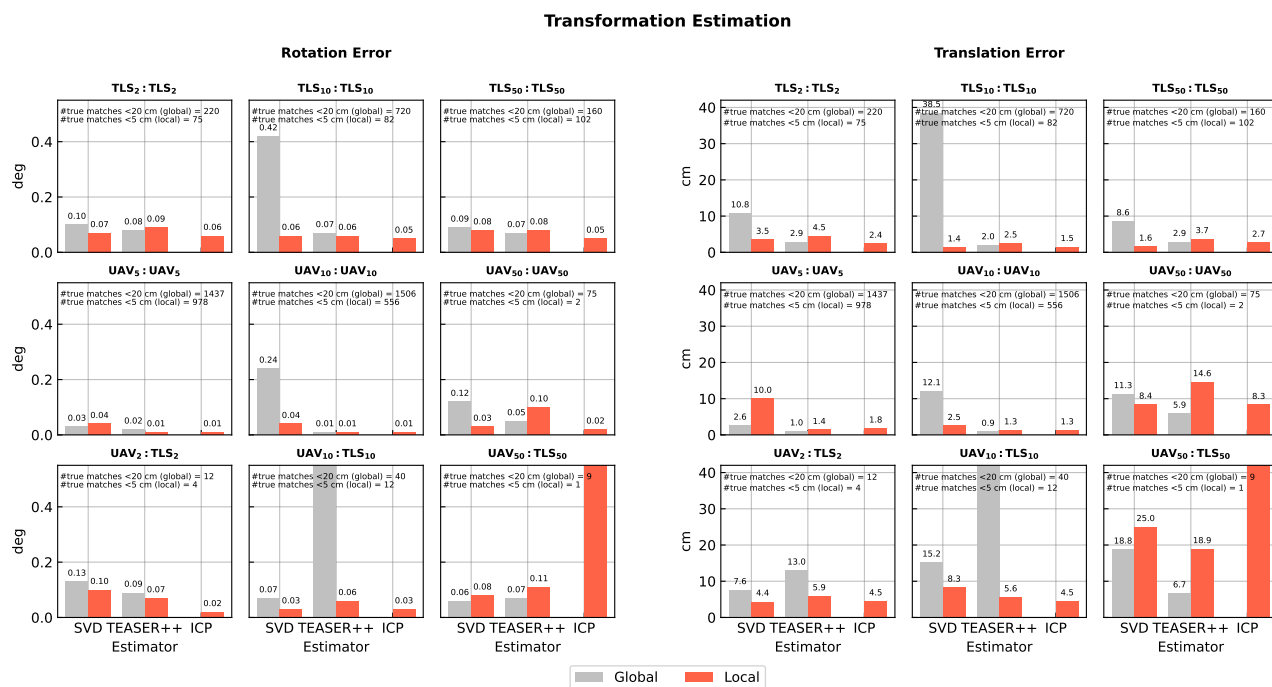
Figure 12. Global and local transformation estimation with error in rotation (e_r) and translation (e_t) with respect to the ground truth transformation matrix.

densities in UAV:UAV and cross-source registration. A correlation between the repeatability of the keypoint recognition and the precision of the registration can be recognised, resulting in more precise registration results with better repeatability and a higher absolute number of possible matches. Further, the registration with TEASER++ is more precise in most cases and shows less outliers while at the same time being significantly faster than PROSAC and SVD. But, the implemented PROSAC showed good results in comparison to the standard RANSAC, solving the transformation in a few iterations, provided a realistic amount of inliers. The following ICP registration is based on local patches around PROSAC inlier keypoints (see figure 11). It improves registration results in most cases, but shows outliers in low density cross-source registration tasks. In most inter-source registration tasks, global registration shows compatible results to local registration with more outliers applying PROSAC and SVD. However, this might be due to a lower inlier threshold set to PROSAC search in global registration. In conclusion, precise registration is possible with global registration at low densities and can be increased if a larger number of correct matches is available. However, a control mechanism should be implemented to detect erroneous estimates.

5. Conclusion

The results demonstrate that keypoint-based coarse registration is possible with an accuracy in the low cm-range. Keypoint detection is clearly identified as the most limiting factor, restricting accuracy by constraining registration to a relatively small number of point pairs with precise repeatability. In addition, it limits the computation time. Since only a small number of repeatable keypoints is detected, a large number of keypoints must be recognised to ensure enough true pairs in the final registration, which renders most of the detection and following computation redundant since most keypoints are outliers. Therefore, there is a need for an algorithm for keypo-

int detection that is characterized by high repeatability and robustness against varying point cloud densities, allowing cross source registration and reducing the total number of keypoints to be detected. Keypoint description most likely showing a lack in cross source description. Matching can be significantly improved applying a local matching strategy. Hence, a two-stage registration pipeline is recommended using global matching in the first instance, followed by a local matching strategy. Despite the good results of the implemented PROSAC algorithm, TEASER++ is recommended as it is significantly faster. Working on low density point clouds seams to be sufficient, especially when considering the much lower computational costs. The benefits and robustness of keypoint projection has to be investigated further, but might improve the precision in low densities. In future work, besides the influence of different parameterizations, learning-based keypoint detectors and in addition, other promising approaches such as image-to-geometry registration (e.g. Elias et al., 2019) based on rendered images of the 3D point clouds shall be tested.

References

Chum, O., Matas, J., 2005. Matching with prosac - progressive sample consensus. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 1, 220–226 vol. 1.

Dong, Z., Liang, F., Yang, B., Xu, Y., Zang, Y., Li, J., Wang, Y., Dai, W., Fan, H., Hyyppä, J., Stilla, U., 2020. Registration of large-scale terrestrial laser scanner point clouds: A review and benchmark. *ISPRS Journal of Photogrammetry and Remote Sensing*, 163, 327–342.

Elias, M., Isfort, S., Eltner, A., Maas, H.-G., 2024. UAS Photogrammetry for Precise Digital Elevation Models of Complex Topography: A Strategy Guide. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* [accepted].

Elias, M., Kehl, C., Schneider, D., 2019. Photogrammetric water level determination using smartphone technology. *The Photogrammetric Record*, 34(166), 198-223.

Esposito, G., Salvini, R., Matano, F., Sacchi, M., Danzi, M., Somma, R., Troise, C., 2017. Multitemporal monitoring of a coastal landslide through SfM-derived point cloud comparison. *The Photogrammetric Record*, 32(160), 459–479.

ETH Zürich, n.d. Eth prs tls data set. https://prs.igp.ethz.ch/research/completed_projects/automatic_registration_of_point_clouds.html (9 February 2024).

Geiger, A., Lenz, P., Urtasun, R., 2012. Are we ready for autonomous driving? the kitti vision benchmark suite. *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 3354–3361.

Gojcic, Z., Zhou, C., Wegner, J. D., Wieser, A., 2019. The perfect match: 3d point cloud matching with smoothed densities. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5540–5549.

Guo, Y., Bennamoun, M., Sohel, F., Lu, M., Wan, J., Kwok, N. M., 2016. A Comprehensive Performance Evaluation of 3D Local Feature Descriptors. *International Journal of Computer Vision*, 116(1), 66–89.

Hänsch, R., Weber, T., Hellwich, O., 2014. Comparison of 3D interest point detectors and descriptors for point cloud fusion. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, II-3, 57–64.

Li, J., Lee, G. H., 2019. Usip: Unsupervised stable interest point detection from 3d point clouds. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE, 361–370.

Mayr, A., Rutzinger, M., Geitner, C., 2018. Multitemporal Analysis of Objects in 3D Point Clouds for Landslide Monitoring. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2, 691–697.

Pirotti, F., Guarnieri, A., Chiodini, S., Bettanini, C., 2023. Automatic Coarse Co-Registration of Point Clouds from Diverse Scan Geometries: A Test of Detectors and Descriptors. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, X-1/W1-2023, 581–587.

Pomerleau, F., Colas, F., Siegwart, R., 2015. A Review of Point Cloud Registration Algorithms for Mobile Robotics. *Foundations and Trends in Robotics*, 4(1), 1–104.

Rusu, R. B., Blodow, N., Beetz, M., 2009. Fast point feature histograms (fpfh) for 3d registration. *2009 IEEE International Conference on Robotics and Automation*, IEEE, 3212–3217.

Rusu, R. B., Cousins, S., 2011. 3d is here: Point cloud library (pcl). *2011 IEEE International Conference on Robotics and Automation*, IEEE, 1–4.

Salti, S., Tombari, F., Di Stefano, L., 2014. SHOT: Unique signatures of histograms for surface and texture description. *Computer Vision and Image Understanding*, 125, 251–264.

Schwalbe, E., Maas, H.-G., Dietrich, R., Ewert, H., 2008. Glacier Velocity Determination from Mulit Temporal Terrestrial Long Range Laser Scanner Point Clouds. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVII/B5, 457–462.

Sipiran, I., Bustos, B., 2011. Harris 3D: a robust extension of the Harris operator for interest point detection on 3D meshes. *The Visual Computer*, 27(11), 963–976.

Stilla, U., Xu, Y., 2023. Change detection of urban objects using 3D point clouds: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 197, 228–255.

Tombari, F., Salti, S., Di Stefano, L., 2010. Unique shape context for 3d data description. *Proceedings of the ACM Workshop on 3D Object Retrieval*, 3DOR '10, Association for Computing Machinery, New York, NY, USA, 57–62.

Tombari, F., Salti, S., Di Stefano, L., 2013. Performance Evaluation of 3D Keypoint Detectors. *International Journal of Computer Vision*, 102(1-3), 198–220.

Ulrich, V., Williams, J. G., Zahs, V., Anders, K., Hecht, S., Höfle, B., 2021. Measurement of rock glacier surface change over different timescales using terrestrial laser scanning point clouds. *Earth Surface Dynamics*, 9(1), 19–28.

Umeyama, S., 1991. Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(4), 376–380.

Wujanz, D., Krueger, D., Neitzel, F., 2016. Identification of Stable Areas in Unreferenced Laser Scans for Deformation Measurement. *The Photogrammetric Record*, 31(155), 261–280.

Yang, H., Shi, J., Carlone, L., 2021. TEASER: Fast and Certifiable Point Cloud Registration. *IEEE Transactions on Robotics*, 37(2), 314–333.

Zahs, V., Hämmerle, M., Anders, K., Hecht, S., Sailer, R., Rutzinger, M., Williams, J. G., Höfle, B., 2019. Multi-temporal 3D point cloud-based quantification and analysis of geomorphological activity at an alpine rock glacier using airborne and terrestrial LiDAR. *Permafrost and Periglacial Processes*, 30(3), 222-238. https://onlinelibrary.wiley.com/doi/abs/10.1002/ppp.2004.

Zeng, A., Song, S., Nießner, M., Fisher, M., Xiao, J., Funkhouser, T., 2017. 3dmatch: Learning local geometric descriptors from rgb-d reconstructions. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 199–208.

Zhang, Z., Dai, Y., Sun, J., 2020. Deep learning based point cloud registration: an overview. *Virtual Reality & Intelligent Hardware*, 2(3), 222–246.

Zhong, Y., 2009. Intrinsic shape signatures: A shape descriptor for 3d object recognition. *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*, IEEE, 689–696.