

AUTOMATIC FEATURE-BASED POINT CLOUD REGISTRATION FOR A MOVING SENSOR PLATFORM

Martin Weinmann, André Dittrich, Stefan Hinz, and Boris Jutzi

Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology (KIT)
Englerstr. 7, 76131 Karlsruhe, Germany
{martin.weinmann, andre.dittrich, stefan.hinz, boris.jutzi}@kit.edu

Commission WG III/4

KEY WORDS: LIDAR, Point Cloud, Registration, Imagery, Dynamic, Close Range

ABSTRACT:

The automatic and accurate alignment of multiple point clouds is a basic requirement for an adequate digitization, reconstruction and interpretation of large 3D environments. Due to the recent technological advancements, modern devices are available which allow for simultaneously capturing intensity and range images with high update rates. Hence, such devices can even be used for dynamic scene analysis and for rapid mapping which is particularly required for environmental applications and disaster management, but unfortunately, they also reveal severe restrictions. Facing challenges with respect to noisy range measurements, a limited non-ambiguous range, a limited field of view and the occurrence of scene dynamics, the adequate alignment of captured point clouds has to satisfy additional constraints compared to the classical registration of terrestrial laser scanning (TLS) point clouds for describing static scenes. In this paper, we propose a new methodology for point cloud registration which considers such constraints while maintaining the fundamental properties of high accuracy and low computational effort without relying on a good initial alignment or human interaction. Exploiting 2D image features and 2D/2D correspondences, sparse point clouds of physically almost identical 3D points are derived. Subsequently, these point clouds are aligned with a fast procedure directly taking into account the reliability of the detected correspondences with respect to geometric and radiometric information. The proposed methodology is evaluated and its performance is demonstrated for data captured with a moving sensor platform which has been designed for monitoring from low altitudes. Due to the provided reliability and a fast processing scheme, the proposed methodology offers a high potential for dynamic scene capture and analysis.

1 INTRODUCTION

An adequate description of a 3D scene is typically derived in the form of point clouds consisting of a large number of measured 3D points and, optionally, different attributes for each point such as intensity or color. The sampling of observed object surfaces should be as dense and complete as possible. Due to occlusions resulting from objects in the scene or areas with low point density, typically multiple point clouds have to be captured from different locations in order to obtain complete objects and full scene coverage. However, as the spatial coordinates of each point cloud are only determined with respect to a local coordinate frame of the sensor, all captured point cloud data has to be transferred into a common coordinate frame which is commonly referred to as point cloud registration or 3D scan matching.

The approaches for point cloud registration can be categorized by considering the data they exploit. Standard approaches such as the Iterative Closest Point (ICP) algorithm (Besl and McKay, 1992) or Least Squares 3D Surface Matching (LS3D) (Gruen and Akca, 2005) only exploit spatial 3D information and minimize either the difference between point clouds or the distance between matched surfaces. The use of point distributions or geometric primitives such as planes has also been proposed in literature (e.g. (Magnusson et al., 2007; Brenner et al., 2008)) and belongs to this category. Considering that the captured scans typically represent data measured on a regular scan grid, the spatial 3D information can also be represented as range image. Exploiting visual features in this range image significantly alleviates the registration process. As most of the modern active 3D sensors provide intensity or color information in addition to the spatial 3D information, respective intensity or color images may also be available. The intensity images are typically derived from reflectance information representing the respective energy of the backscattered

laser light, whereas color information is usually obtained from co-registered camera images. Such intensity or color images provide a higher level of distinctiveness and allow for detecting reliable correspondences between visual features.

Nowadays, many approaches for point cloud registration exploit visual features derived from intensity or color images in order to obtain sparse point clouds. Detected feature correspondences between the respective images indicate corresponding 3D points. Hence, the registration of such sparse point clouds may for instance be based on a standard rigid transformation (Eggert et al., 1997) which is typically combined with the RANSAC algorithm for increased robustness in case of existing outlier correspondences (Seo et al., 2005; Boehm and Becker, 2007; Barnea and Filin, 2007). As a powerful alternative, the transfer to solving the Perspective- n -Point (PnP) problem has been proposed (Weinmann et al., 2011; Weinmann and Jutzi, 2011).

Recently, an experimental setup for surveillance applications in indoor and outdoor environments has been presented which is suited for simulating airborne scene monitoring from low altitudes fairly realistically (Hinz et al., 2011). The first results presented in (Weinmann and Jutzi, 2012) indicate a high potential for dynamic scene capture and analysis, but they also reveal that additional effort is required to obtain satisfying results. For dynamic scene capture, highly accurate 3D measurements as provided by a terrestrial laser scanner cannot be assumed, but the modern scanning devices offer a simultaneous image-based acquisition of intensity and range information with high update rates. Hence, the registration process has to provide high-quality estimates of the transformation parameters, low computational effort and robustness with respect to noisy range measurements. Furthermore, the limited field of view and the limited non-ambiguous range have to be taken into account. An important and meanwhile commonly

used first step is the reduction of captured point cloud data to sparse point clouds by using visual features in 2D imagery. A reliable feature matching yields 2D/2D correspondences and the respective 3D/3D correspondences. In contrast to previous work (Weinmann and Jutzi, 2011; Weinmann and Jutzi, 2012), the proposed methodology involves an improved scheme for outlier rejection and estimation of inlier reliability. This scheme exploits information derived from the reliability of range measurements as well as the reliability of feature correspondences which is based on intensity measurements. With the further consideration of the plausibility of corresponding 3D points, a straightforward approach for aligning point clouds can be applied which would not be suitable without a reliable outlier removal. The contribution of this paper is a new methodology for fast and accurate point cloud registration which

- exploits sparse point clouds with additional point attributes in form of quality measures based on geometric and radiometric information,
- introduces an improved weighting scheme considering the derived quality measures, and
- involves a plausibility check taking into account the detected 3D/3D correspondences.

After presenting the methodology for successive pairwise registration in Section 2, the configuration of the sensor platform is described in Section 3. Subsequently, in Section 4, an evaluation is carried out which demonstrates the performance of the new approach for a realistic test scenario. In Section 5, the derived results are discussed with respect to basic requirements and other approaches. Finally, the content of the entire paper is concluded in Section 6 and suggestions for future work are outlined.

2 METHODOLOGY

The proposed methodology focuses on airborne scene monitoring with a moving sensor platform. After data acquisition (Section 2.1), a preprocessing is carried out in order to get normalized intensity images and the respective 3D point cloud (Section 2.2). As the captured point clouds are corrupted with noise, a quality measure is derived for each 3D point (Section 2.3). Subsequently, distinctive features are extracted from 2D intensity images (Section 2.4). A comparison of these features yields reliable 2D/2D correspondences between different frames as well as a quality measure taking into account the distinctiveness of matched features. Additionally, the projection of the respective 2D points into 3D space yields 3D/3D correspondences. As the influence of each 3D/3D correspondence on the registration process should rely on a respective quality measure, a weighting scheme considering geometric and radiometric information is introduced (Section 2.5). Finally, the point cloud registration is carried out by estimating the rigid transformation between two sparse point clouds with a weighted least squares alignment (Section 2.6).

2.1 Data Acquisition

The proposed concept focuses on the use of range imaging devices which are also referred to as range cameras, i.e. devices which provide 2D image representations of captured range and intensity/color. These devices should additionally provide a high update rate for capturing dynamic scenes or for rapid mapping.

2.2 Preprocessing

The first step consists of adapting the recorded data according to (Weinmann and Jutzi, 2012), where a histogram normalization is carried out which maps the captured intensity information to the

interval $[0, 255]$. For color images, a conversion to gray-valued images could be applied in order to obtain intensity images. Furthermore, the lens distortion has to be taken into account which involves an initial camera calibration and the respective correction of the captured 3D information.

2.3 Point Quality Assessment

As the range measurements might be corrupted with noise, it is suitable to add a quality measure as attribute for each measured 3D point. Considering the 2D representation of the measured range information, the variation of the range values within small and local image neighborhoods has a strong influence on the reliability of measured 3D points (Weinmann and Jutzi, 2011; Weinmann and Jutzi, 2012). Hence, for each point on the regular 2D grid, the reliability of the respective range information is described with the standard deviation $\sigma \in \mathbb{R}$ of all range values within a 3×3 neighborhood. Low values σ indicate a 3D point on a smooth surface and are assumed to be reliable, whereas high values indicate noisy and unreliable range measurements. Resulting from this, a confidence map M_C is available. In addition to this, the quality measure could further exploit the active intensity measurements representing the energy of the backscattered laser light if these are available, e.g. for range imaging devices such as PMD[vision] CamCube 2.0 or MESA Imaging SR4000.

2.4 2D Feature Extraction and Projection to 3D

Once the measured information has been assigned additional attributes, the registration process can rely on both range and intensity information, and a confidence map providing the respective quality measure. For detecting corresponding information, the Scale Invariant Feature Transform (SIFT) (Lowe, 2004) is applied on the intensity images. This yields distinctive keypoints at 2D image locations $\mathbf{x}_i \in \mathbb{R}^2$ as well as the respective local descriptors which are invariant to image scaling and image rotation, and robust with respect to image noise, changes in illumination and small changes in viewpoint. These properties of the descriptors allow a reliable feature matching relying on the ratio $\rho \in \mathbb{R}$ with

$$\rho = \frac{d(N_1)}{d(N_2)} \quad (1)$$

where $d(N_i)$ with $i = 1, 2$ denotes the Euclidean distance of a descriptor belonging to a keypoint in one image to the i -th nearest neighbor in the other image. A low value of ρ indicates a high similarity to only one of the derived descriptors belonging to the other image. Thus, the ratio $\rho \in [0, 1]$ describes the distinctiveness of the occurring features. Meaningful feature correspondences arise from a greater difference between $d(N_1)$ and $d(N_2)$ and hence, the ratio ρ has to satisfy the constraint $\rho \leq t_{des}$, where t_{des} is a certain threshold typically chosen within the interval $[0.6, 0.8]$. As the SIFT features are localized with subpixel accuracy, the assigned information has to be interpolated from the information available for the regular and discrete 2D grid, e.g. by applying a bilinear interpolation. Subsequently, the 2D/2D correspondences $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ between these visual features are used to reduce the captured point cloud data to sparse point clouds of physically almost identical 3D points $\mathbf{X}_i \leftrightarrow \mathbf{X}'_i$ with $\mathbf{X}_i, \mathbf{X}'_i \in \mathbb{R}^3$. Including the assigned attributes, each correspondence can be described with two samples of corresponding information according to

$$s_i = (\mathbf{x}_i, \mathbf{X}_i, \sigma_i, \rho_i) \leftrightarrow s'_i = (\mathbf{x}'_i, \mathbf{X}'_i, \sigma'_i, \rho'_i) \quad (2)$$

where σ_i and σ'_i indicate the quality of the derived 3D points with respect to measured range information, and $\rho_i^* = \rho_i = \rho'_i$ are the assigned quality measures with respect to the distinctiveness of the used intensity information.

2.5 Weight Calculation

For weighting the influence of each 3D/3D correspondences on the estimated transformation, a weight parameter has to be derived for each 3D/3D correspondence. Given the calculated values $\sigma_i, \sigma'_i \in [0, \infty)$ and $\rho_i^* \in [0, t_{des}]$, which are considered as quality measures for the respective 3D points \mathbf{X}_i and \mathbf{X}'_i , the influence of the i -th 3D/3D correspondence $\mathbf{X}_i \leftrightarrow \mathbf{X}'_i$ on the registration process can be weighted by applying a histogram-based approach (Weinmann and Jutzi, 2012). First, the interval $[0m, 1m]$ is divided into $n_b = 100$ bins of equal size, and the values σ_i and σ'_i are mapped to the respective bin. Those values above the upper boundary of 1m are mapped to the last bin. The occurrence of mappings to the different bins is stored in histograms $\mathbf{h} = [h_j]_{j=1, \dots, 100}$ and $\mathbf{h}' = [h'_j]_{j=1, \dots, 100}$. From these, the cumulative histograms \mathbf{h}_c and \mathbf{h}'_c are subsequently derived, where the entries reach from 0 to the number n of detected 3D/3D correspondences. For assigning those 3D points with a lower σ a higher weight, as these are more reliable according to the definition, the inverse cumulative histograms (ICHS)

$$\mathbf{h}_{c,inv} = \left[n - \sum_{j=1}^i h_j \right]_{i=1, \dots, 100} \quad (3)$$

and

$$\mathbf{h}'_{c,inv} = \left[n - \sum_{j=1}^i h'_j \right]_{i=1, \dots, 100} \quad (4)$$

are established. The quality measure ρ_i^* encapsulating the reliability of the matching process is also used for such a weighting scheme based on $n_b = 100$ bins of equal size covering the interval $[0, 1]$. Resulting from this, a histogram $\mathbf{h}^* = [h_j^*]_{j=1, \dots, 100}$ and a cumulative histogram \mathbf{h}^*_c are available as well as an inverse cumulative histogram $\mathbf{h}^*_{c,inv}$ with

$$\mathbf{h}^*_{c,inv} = \left[n - \sum_{j=1}^i h_j^* \right]_{i=1, \dots, 100} \quad (5)$$

from which additional weight parameters are derived. Thus, the new weighting scheme yields three weight parameters $\mathbf{h}_{c,inv}(\sigma_i)$, $\mathbf{h}'_{c,inv}(\sigma'_i)$ and $\mathbf{h}^*_{c,inv}(\rho_i^*)$ for each 3D/3D correspondence. From these, the respective weight w_i is finally determined with

$$w_i = \min \{ \mathbf{h}_{c,inv}(\sigma_i), \mathbf{h}'_{c,inv}(\sigma'_i), \mathbf{h}^*_{c,inv}(\rho_i^*) \} \quad (6)$$

as the minimum of these values.

2.6 Point Cloud Registration

Introducing a rotation matrix $\mathbf{R} = [r_{pq}] \in \mathbb{R}^{3 \times 3}$ and a translation vector $\mathbf{t} = [t_p] \in \mathbb{R}^3$, the spatial relation between two points $\mathbf{X}_i, \mathbf{X}'_i \in \mathbb{R}^3$ representing a 3D/3D correspondence $\mathbf{X}_i \leftrightarrow \mathbf{X}'_i$ can formally be described as

$$\mathbf{X}'_i = \mathbf{R}\mathbf{X}_i + \mathbf{t} \quad (7)$$

or more detailed as

$$X'_i = r_{11}X_i + r_{12}Y_i + r_{13}Z_i + t_1 \quad (8)$$

$$Y'_i = r_{21}X_i + r_{22}Y_i + r_{23}Z_i + t_2 \quad (9)$$

$$Z'_i = r_{31}X_i + r_{32}Y_i + r_{33}Z_i + t_3 \quad (10)$$

where a perfect mapping is achieved in the ideal case. However, the 3D/3D correspondences typically do not fit perfectly and therefore, a fully automatic estimation of the transformation parameters can be derived by minimizing the error between the

point clouds in the least squares sense. For this purpose, the three equations resulting from each of the n 3D/3D correspondences are concatenated. Subsequently introducing vector-matrix notation and separating the vector $\mathbf{u} \in \mathbb{R}^{12}$ containing the unknown parameters according to

$$\mathbf{u} = [r_{11}, r_{12}, r_{13}, r_{21}, r_{22}, r_{23}, r_{31}, r_{32}, r_{33}, t_1, t_2, t_3]^T \quad (11)$$

yields a linear equation system of the form

$$\mathbf{l} = \mathbf{A}\mathbf{u} \quad (12)$$

with $\mathbf{l} \in \mathbb{R}^{3n}$ and $\mathbf{A} \in \mathbb{R}^{3n \times 12}$. For solving this linear equation system, the least squares estimate can be derived as

$$\hat{\mathbf{u}} = \left(\mathbf{A}^T \mathbf{P} \mathbf{A} \right)^{-1} \mathbf{A}^T \mathbf{P} \mathbf{l} \quad (13)$$

where the matrix $\mathbf{P} \in \mathbb{R}^{3n \times 3n}$ is used for weighting the importance of the respective observations. As the observations are assumed to be independent, the weight matrix \mathbf{P} is considered as diagonal matrix. Exploiting the weights w_i defined in Section 2.5, which are stored in the vector $\mathbf{w} \in \mathbb{R}^n$, an initialization of the weight matrix \mathbf{P} according to

$$\mathbf{P} = \mathbf{P}_0 = \text{diag} \left\{ \left(\frac{w_i^*}{\|\mathbf{w}^*\|} \right)^2, \left(\frac{d_i}{\|\mathbf{d}\|} \right)^2 \right\}_{i=1, \dots, 3n} \quad (14)$$

is introduced, where $\mathbf{w}^* \in \mathbb{R}^{3n}$ contains the weights for each coordinate. The parameters d_i describe the difference between the respective change of a coordinate value and the mean change of the coordinate over all correspondences. The introduction of the parameters d_i is only possible for very small rotations of the sensor platform, but they ensure a weighting according to the plausibility of the respective 3D/3D correspondences which is derived from considering the major trend of all correspondences. Additionally, those correspondences whose 3D coordinates do not even fit to the confidence interval within at least two standard deviations of the mean coordinates, i.e. a confidence level of 95%, are removed. By considering the estimated improvement $\hat{\mathbf{v}}$ according to $\hat{\mathbf{v}} = \mathbf{A}\hat{\mathbf{u}} - \mathbf{l}$ with $\hat{\mathbf{v}} \in \mathbb{R}^{3n}$, an iterative update

$$\mathbf{P} = \mathbf{P}_k = \text{diag} \left\{ \frac{1}{\left(\frac{\hat{v}_i}{\|\hat{\mathbf{v}}\|} \right)^2} \right\}_{i=1, \dots, 3n} \quad (15)$$

is applied until the estimated transformation converges to changes below a certain threshold or until a maximum number of k_{max} iterations is reached.

3 ACTIVE MULTI-VIEW RANGE IMAGING SYSTEM

For demonstrating the performance of the proposed methodology, a sensor platform is used which allows for monitoring from low altitudes (Hinz et al., 2011). This platform is shown in Figure 1, and it is equipped with

- two range imaging devices (PMD[vision] CamCube 2.0) for data acquisition,
- a notebook with a solid state drive for efficient data storage, and
- a 12V battery with 6.5Ah for independent power supply.

As the relative orientation of the two range imaging devices can easily be changed, the system allows for different multi-view op-



Figure 1: Sensor platform.

tions with respect to parallel, convergent or divergent data acquisition geometries. Due to the large payload of several kilograms, mounting the components for data acquisition and data storage on an unmanned aerial vehicle (UAV) is still impracticable. Hence, for scene capture, the sensor platform is moved along a rope. Considering typical surveillance applications, the combination of the sensor platform with a scaled test scenario allows a fairly realistic simulation of a future operational system.

A closer consideration of the devices used for data acquisition reveals the potential as well as the challenges of the proposed system. A PMD[vision] CamCube 2.0 simultaneously captures geometric and radiometric information in form of images with a single shot, and hence, the captured information can be organized in frames. Each frame consists of a range image I_R , an active intensity image I_a and a passive intensity image I_p . The active intensity depends on the illumination emitted by the sensor, whereas the passive intensity depends on the background illumination arising from the sun or other external light sources. As the captured images have a size of 204×204 pixels which corresponds to a field of view of $40^\circ \times 40^\circ$, measurements with an angular resolution of approximately 0.2° are provided. The frame can be updated with high frame rates of more than 25 releases per second. Hence, the device is well-suited for capturing dynamic scenes. A great advantage of the chosen range imaging device is that it can also be used within outdoor environments, but due to the relatively large influence of noise effects arising from multipath scattering as well as the large amount of ambient radiation in comparison to the amount of emitted radiation, a limited absolute range accuracy of a few centimeters and thus noisy point clouds can be expected. A visualization of captured data is shown in Figure 2.

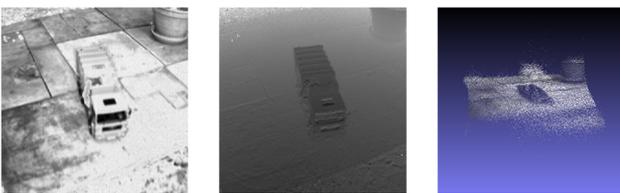


Figure 2: Visualization of captured data: normalized passive intensity image (left), range image (center) and textured 3D point cloud (right).

As the whole system involves multiple range imaging devices for extending the field of view, it has to be considered that these may influence each other and that interferences are likely to occur. This is overcome by choosing different modulation frequencies. Furthermore, a synchronization of both range imaging devices is required in order to obtain corresponding frames with respect to a temporal reference. For this purpose, a software-based trigger is introduced. Optionally, the range measurement restriction can

also be resolved with a hardware-based unwrapping procedure (Jutzi, 2012), which requires the use of different modulation frequencies for each of the two range imaging devices.

4 EXPERIMENTAL RESULTS

The experiments refer to the scene depicted in Figure 3. First, a local coordinate frame is defined in the center between both range imaging devices with fixed orientation with respect to the sensor platform. This coordinate frame is referred to as body frame (superscript b). The X^b -direction is oriented to the forward direction tangential to the rope, the Y^b -direction to the right and the Z^b -direction downwards. Subsequently, a global reference frame (superscript g) is defined which coincides with the initial position and orientation of the sensor platform. As the relative orientation between the devices and the platform is already known from a priori measurements, the projected 3D points \mathbf{X}_i^c which are related to the respective local coordinate frame of a range imaging device (superscript c) can directly be transformed into the body frame (superscript b) of the sensor platform according to

$$\mathbf{X}_i^b = \mathbf{R}_c^b \mathbf{X}_i^c + \mathbf{t}_c^b \quad (16)$$

where \mathbf{R}_c^b and \mathbf{t}_c^b describe the rotation and translation between the respective coordinate frames.



Figure 3: Illustration of the observed scene.

During the whole movement of the sensor platform, a total number of 116 frames is captured. Each frame contains simultaneous measurements of the two range imaging devices. Although there are no measured reference values for each single position in order to check the deviation of the position estimates from the real positions, a validation of the proposed methodology is still possible. As the sensor platform moves along a rope, the projection of the real trajectory onto the $X^g Y^g$ -plane represents a straight line. Furthermore, a loop closure constraint may be considered, where start position and end position of the sensor platform are identical. This is reached by repeating the first frame at the end of the movement. Additional criteria may involve a visual impression of several registered point clouds which yields insights about the scene and the quality of captured data. Hence, the evaluation of the proposed methodology involves

- the deviation σ_Y from the straight line in the $X^g Y^g$ -plane with $X^g = 0\text{m}$,
- the absolute error e_{loop} occurring when assuming an identical start and end position of the sensor platform, and
- a visual inspection of the registered point clouds.

For feature matching, a threshold of $t_{des} = 0.8$ is selected. The estimated trajectory obtained via successive pairwise registration is shown in Figure 4 in nadir view. The standard deviation σ_Y

of the position estimates projected into the X^gY^g -plane from the straight line with $X^g = 0m$ is $\sigma_Y = 0.0378m$, and the absolute error when considering a closed loop is $e_{loop} = 0.0967m$. A visualization of registered point clouds in a common coordinate frame is illustrated in Figure 5. The figure also shows that the raw point clouds contain many noisy 3D measurements which have to be removed for a subsequent scene analysis.

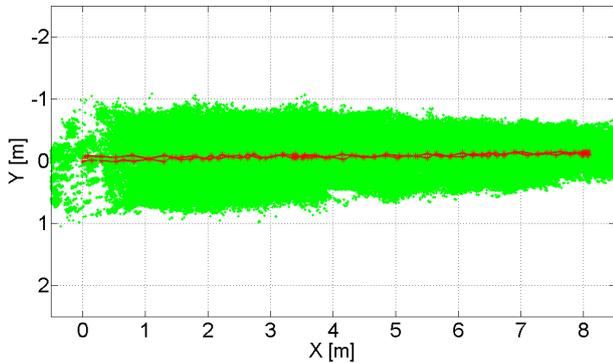


Figure 4: Projection of the estimated trajectory onto the X^gY^g -plane (red) and visualization of the distribution of measured 3D scene points (green).

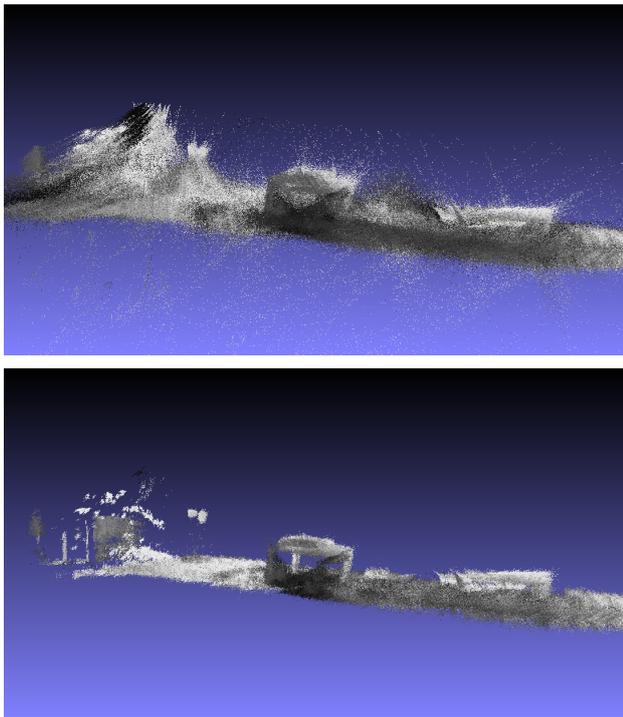


Figure 5: Visualization of registered point clouds: raw point cloud data (top) and those 3D points \mathbf{X}_i satisfying $\sigma_i \leq 0.03m$ (bottom).

For data captured at two successive time steps during the movement of the sensor platform, the general appearance of the calculated histograms, cumulative histograms and inverse cumulative histograms is visualized in Figure 6, Figure 7 and Figure 8. The example is based on $n_1 = 480$ detected SIFT correspondences between the normalized active intensity images and $n_2 = 508$ detected SIFT correspondences between the normalized passive intensity images, i.e. a total number of $n = 928$ SIFT corre-

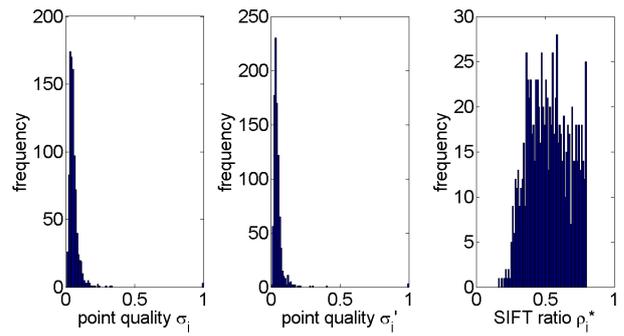


Figure 6: Histograms depicting the occurrence of range reliabilities σ_i and σ_i' (left and center), and the occurrence of SIFT correspondences with a SIFT ratio ρ_i^* (right).

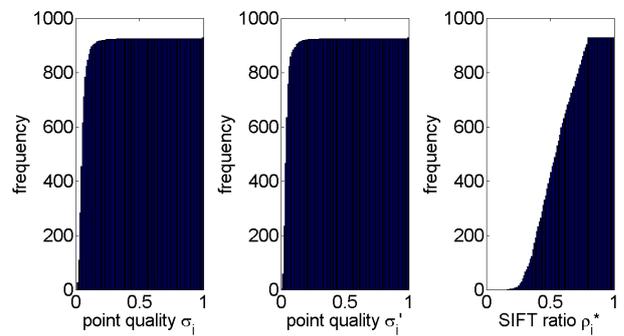


Figure 7: The respective cumulative histograms for the histograms depicted in Figure 6.

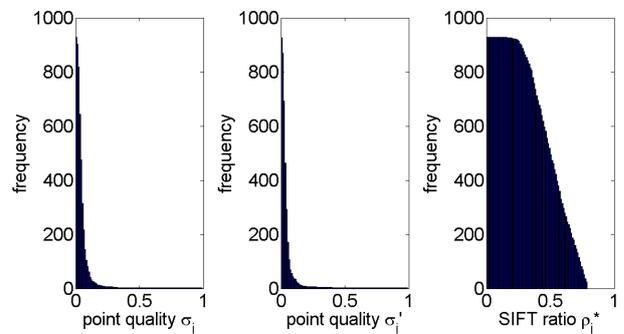


Figure 8: The respective inverse cumulative histograms (ICHs) for the histograms depicted in Figure 6.

spondences resulting in the same number of 3D/3D correspondences after the respective projection into 3D space. Finally, the required time effort for processing each frame and the relations between two frames has to be considered. The methodology has been implemented in Matlab and tested on a standard notebook (2.3GHz, 4GB RAM). For each frame, the average time required for preprocessing, point quality assessment, feature extraction and point projection is 1.6408s (Table 1). For obtaining the relations between two frames, the average time required for feature matching, calculation of weights and point cloud registration is 0.4821s (Table 2). The separate consideration of the average time required for the single tasks reveals that feature extraction and feature matching contribute to more than 90% of the whole time effort. This can significantly be reduced by applying a GPU-based implementation of SIFT or more efficient feature detectors and descriptors.

| Task | Time |
|--------------------------|---------|
| Preprocessing | 0.0766s |
| Point Quality Assessment | 0.0250s |
| Feature Extraction | 1.5295s |
| Point Projection | 0.0097s |
| Σ | 1.6408s |

Table 1: Average time effort for the processing of a single frame.

| Task | Time |
|----------------------|---------|
| Feature Matching | 0.4706s |
| Weighting Scheme | 0.0022s |
| Registration Process | 0.0092s |
| Σ | 0.4821s |

Table 2: Average time effort for aligning two point clouds.

5 DISCUSSION

In contrast to standard approaches such as the ICP algorithm, the presented approach does not require a good a priori alignment of the scans. Furthermore, it can be applied for very general scenes without assuming the presence of regular surfaces such as planes and can even cope with noisy measurements arising from monitoring within outdoor environments. The experimental results show that the proposed methodology is suited to recover the transformation parameters fast and accurately by exploiting synergies arising from the combined use of geometric and radiometric information. The proposed methodology directly takes into account the reliability of the 3D information captured with the two range cameras. Thus, it is even possible to strongly rely on the measurements of one range imaging device if the second range imaging device captures spatial information with more noise. Additionally considering how good the corresponding 3D/3D points fit together with respect to the gradient information of the local neighborhood in the 2D intensity representations strengthens the reliability with a further and complementary quality measure assigned to each 3D/3D correspondence. Subsequently, in the registration process, the plausibility of the respective 3D/3D correspondences and the iterative improvement of the geometric alignment of each coordinate are involved. If the estimated improvement \hat{v}_i is relatively large for a coordinate, the coordinate is considered as rather unreliable and the respective influence on the estimated transformation is assigned a lower weight. Due to all these considerations, a high reliability of the estimated flight trajectory can be expected. The only assumption of the registration process is the existence of structured 2D intensity representations in order to derive corresponding points via local image features. This, however, is a common assumption for all image-based approaches exploiting distinctive points or lines for point cloud registration.

6 CONCLUSIONS AND FUTURE WORK

In this paper, a new concept for data acquisition with a moving active multi-view range imaging system and a successive registration of the captured point clouds has been presented. Thus, the system is able to face the challenges arising from noisy range measurements, a limited non-ambiguous range and a limited field of view in dynamic environments. The proposed methodology focuses on the use of sparse point clouds and additional attributes from which a common quality measure considering geometric and radiometric information is derived via inverse cumulative histograms. This quality measure allows for weighting the influence of each 3D/3D correspondence on the estimated transformation

according to its reliability. Further applying a plausibility check for the detected 3D/3D correspondences, the registration process can be carried out with a weighted least squares adjustment. For future work, it would be desirable to detect relevant objects in the scene and describe their behavior by estimating the respective motion trajectories. This can also be achieved by exploiting the combined use of geometric and radiometric information as well as the respective consideration of 3D point cloud data and 2D image representations. Promising results can be expected.

REFERENCES

- Barnea, S. and Filin, S., 2007. Registration of terrestrial laser scans via image based features. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 36 (Part 3), pp. 32–37.
- Besl, P. J. and McKay, N. D., 1992. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14(2), pp. 239–256.
- Boehm, J. and Becker, S., 2007. Automatic marker-free registration of terrestrial laser scans using reflectance features. *Optical 3-D Measurement Techniques VIII*, pp. 338–344.
- Brenner, C., Dold, C. and Ripperda, N., 2008. Coarse orientation of terrestrial laser scans in urban environments. *ISPRS Journal of Photogrammetry and Remote Sensing* 63(1), pp. 4–18.
- Eggert, D. W., Lorusso, A. and Fisher, R. B., 1997. Estimating 3-D rigid body transformations: a comparison of four major algorithms. *Machine Vision and Applications* 9(5-6), pp. 272–290.
- Gruen, A. and Akca, D., 2005. Least squares 3D surface and curve matching. *ISPRS Journal of Photogrammetry and Remote Sensing* 59(3), pp. 151–174.
- Hinz, S., Weinmann, Ma., Runge, P. and Jutzi, B., 2011. Potentials of image based active ranging to capture dynamic scenes. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 38 (Part 4 / W19), pp. 143–147.
- Jutzi, B., 2012. Extending the range measurement capabilities of modulated range imaging devices by time-frequency multiplexing. *AVN - Allgemeine Vermessungs-Nachrichten* 02 / 2012, pp. 54–62.
- Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60(2), pp. 91–110.
- Magnusson, M., Lilienthal, A. and Duckett, T., 2007. Scan registration for autonomous mining vehicles using 3D-NDT. *Journal of Field Robotics* 24(10), pp. 803–827.
- Seo, J. K., Sharp, G. C. and Lee, S. W., 2005. Range data registration using photometric features. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2, pp. 1140–1145.
- Weinmann, Ma. and Jutzi, B., 2011. Fully automatic image-based registration of unorganized TLS data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 38 (Part 5 / W12), pp. 55–60.
- Weinmann, Ma. and Jutzi, B., 2012. A step towards dynamic scene analysis with active multi-view range imaging systems. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 39 (Part B3), pp. 433–438.
- Weinmann, Ma., Weinmann, Mi., Hinz, S. and Jutzi, B., 2011. Fast and automatic image-based registration of TLS data. *ISPRS Journal of Photogrammetry and Remote Sensing* 66(6), pp. S62–S70.