6D Scan Registration using Depth-Interpolated Local Image Features

Henrik Andreasson *, Achim J. Lilienthal

Centre for Applied Autonomous Sensor Systems
Dept. of Technology, Örebro University
SE-70182 Örebro, Sweden

Abstract

This paper describes a novel registration approach that is based on a combination of visual and 3D range information. To identify correspondences, local visual features obtained from images of a standard color camera are compared and the depth of matching features (and their position covariance) is determined from the range measurements of a 3D laser scanner. The matched depth-interpolated image features allows to apply registration with known correspondences. We compare several ICP variants in this paper and suggest an extension that considers the spatial distance between matching features to eliminate false correspondences. Experimental results are presented in both outdoor and indoor environments. In addition to pair-wise registration, we also propose a global registration method that registers all scan poses simultaneously.

Key words: registration, vision, laser range finder, SLAM
PACS:

1. Introduction

Registration is the process of transforming data into a consistent coordinate system. Since registration of scans (in this paper, a scan is understood to be a data set recorded at a fixed position) is trivial if the change between the measurement locations is precisely known, the problem regresses to determining the change in pose between the scans. Scan registration is thus the process of estimating the relative pose between scans considering their specific features. It is a core component of many Simultaneous Localization and Mapping (SLAM) algorithms.

In contrast to a substantial amount of previous publications, which consider input from a 2D laser scanner and 2D poses, making the assumption of planar motion, we address the general case of 6-dimensional poses and consider input data, which consist of 3D range data *and* visual information.

Since visual information is particularly suited to solve the correspondence problem (data association), vision-based systems have been applied as an addition to laser scanning based SLAM approaches for detecting loop closing. This general approach was implemented for systems based on a 2D laser scanner [1] and a 3D laser scanner [2]. When using registration methods that rely on a relatively weak criterion for correspondence, for example point to point distance as in [3], a good initial estimate is very important for the robustness of the system. Here, we are using instead the strong correspondences visual features can provide and thus the initial estimate can be very poor. Indeed, it can be argued for areas of reasonable size that an initial estimate is not necessary at all [2]. Another advantage of considering visual information is that

* Corresponding author.
Email addresses: henrik.andreasson@oru.se (Henrik Andreasson), achim.lilienthal@oru.se (Achim J. Lilienthal).

vision can enable solutions in highly cluttered environments where pure laser range scanner based methods fail [4]. The benefits of using vision come at almost no extra cost since a camera is much less expensive than a 3D laser scanner.

In this paper, we present a registration method that takes input from a perspective camera and a 3D laser scanner (which was realized by a 2D laser scanner mounted on nodding pan/tilt unit). The key idea is to combine the strong visual correspondences with the depth accuracy obtained from the laser scanner. The remainder of this paper is organized as follows. First, related work is presented in Sec. 2, followed by a description of the proposed methods for pairwise (Sec. 3) and global registration (Sec. 4), and the experimental setup (Sec. 5). Sec. 6 presents experimental results and followed by conclusions and suggestions for future work in Sec. 7.

2. Related Work

The discriminative power of local visual features was combined with a 3D laser scanner based SLAM approach in Newman et al. [2]. In their work, SIFT features [5] were used to detect loop closure events and to obtain an initial estimate of the relative pose between the two images, which correspond to the loop closure, by determining the essential matrix. Our approach also relies on local visual features and we also implemented them using the SIFT algorithm in this paper. However, using the data from the laser scanner, our method associates depth values with those SIFT features for which matching features were found and carries out registration using only these visually salient 3D points. Thus, in contrast to the work by Newman et al., we do not consider the full point cloud for registration.

While work that combines local image features with 3D range data is sparse, there are several approaches to registration and SLAM that use either vision or range data alone. Registration methods that utilize 3D laser data are commonly based on the ICP algorithm [6,7]. Another 3D laser based approach is the 3D-NDT method by Magnusson et al. [8]. Common to all 3D laser based registration methods is that they require a sufficiently accurate initial estimate.

Research work in which visual information is incorporated directly into scan registration has, to the best of our knowledge, been restricted to extensions of the distance function that is minimized in ICP. To estimate the relative pose between stereo camera depth maps, Color ICP has been used [9], an ICP variant which also incorporates the color as part of the distance function. In [10], the standard ICP method is combined with a constraint based on the optical flow.

Registration using visual features alone does not work in a straight forward manner. A particular problem is to determine the scale [11]. One approach is to use a predefined pattern with known geometrical properties as part of the first image [12]. However, unless an object with known geometrical properties is shown again, such an approach would encounter problems with scale drift. Another commonly used approach is to have multiple cameras and to use triangulation to get a depth estimate for each visual feature as, for example, in [13]. However, generally speaking, the position of the features is not known precisely enough for accurate registration. To improve the estimate of each landmark position the landmarks can be tracked over a sequence of images [11], which however requires that the camera poses are known. The simultaneous estimation of the position of features and camera poses is directly related to the SLAM problem. Many approaches rely on initial pose estimates from odometry [14–17] and update the position estimates of visual features using Extended Kalman Filters (EKF) [12,16], or Rao-Blackwellised Particle Filters (RBPF) [18,15], for example. Alternatively, if only a limited number of successive frames is considered to search for corresponding features this is the problem of visual odometry [14].

Simultaneous registration of a set of scans (global registration) can be formulated as a graph problem where each node represents a robot scan pose and an edge represents a constraint between the nodes. SLAM methods which operate on this structure are called graph-based or simply graph-SLAM. There exists a variety of graph-based SLAM approaches. One of the first examples using 2D data is the work by Lu and Milios [19]. Olson et al. [20] use an approach based on stochastic gradient descent (SGD) to optimize the global poses. In the work by Olson, only 2D data were evaluated. Their approach cannot be directly applied in 3D since they make the assumption of linear angular subspaces, which does not hold in the case of 3D data.The problem of linear subspaces has been addressed by Grisetti et al. [21]. In their approach a variant of the gradient descent method and a tree parametrization is applied together with incremental spherical linear interpola-

tion (SLERP), to address the non-commutativity of rotations in 3D.

An important observation is that, in approaches which use vision only, typically, the uncertainty regarding the pose of visual features is high and that this prevents using the feature poses directly for accurate registration. By contrast the accuracy that laser range scanners provide makes it unnecessary to extract and track features in 3D laser data.

We propose in this paper a full 6D registration approach that does not require initial pose estimates. By using interpolated range values from the laser scanner to estimate the 3D position of local image features (and their covariance), our approach also avoids that image features have to be tracked over a sequence of frames. To the best of our knowledge, such a registration method has not been suggested before. We also present an extension of our method to the case of global registration.

Global registration is similar to graph-SLAM in that it optimizes all relative poses, which correspond to edges in the graph, simultaneously. However, graph-based SLAM approaches typically apply pair-wise scan matching and represent the result as an edge in the graph together with an assigned covariance estimate. In global registration, on the other hand, edges represent sets of matched features and the optimization is performed by simultaneous registration of *all* connected scans together. An approach to simultaneous registration of more than two scans was proposed by Biber et al. [22] for the case of 2D laser range data.

3. DIFT registration

The proposed registration approach is based on depth-interpolated image features (DIFT) and we therefore termed it DIFT registration. The visual feature descriptor and detector used in this paper is SIFT developed by Lowe [5] but other local image features could be used as well. The position and covariance in 3D for visual features are obtained from the laser range measurements surrounding the visual feature location. For example, if the detected feature is located on the planar surface of a poster, the feature's position covariance will be smaller (especially perpendicular to the surface) compared to a feature that is located at a thin branch of a tree.

As stated in the previous section, most current approaches to scan registration depend on fairly accurate initial pose estimates. In the proposed method, correspondences are solely determined using highly distinctive visual features and not from spatial distance alone. As a result, no initial pose estimates are required.

Shortly the registration procedure can be described as follows: first, SIFT features are computed in the planar images recorded with the current scan $\mathcal{S}_c$ (please remember that in this paper a scan denotes the 3D points from the laser range scanner and a set of planar images) and compared to the SIFT features found in the images belonging to another scan $\mathcal{S}_p$. Next, the depth values are estimated for all matching feature pairs in $\mathcal{S}_p$ and $\mathcal{S}_c$, using the closest projected 3D laser point as described in Sec. 3.2. The covariance of the visual features is computed using the neighboring 3D points (Sec. 3.3). Pairs of matching features are then used together with the feature position covariance to obtain the final relative pose estimate (see Sec. 3.6).

Please note that we do not model the errors introduced by calibration inaccuracy nor the sensor noise of the laser scanner, the wrist or the camera, assuming that all these sources create only negligible errors.

In Sec. 4 we present the global registration approach. This approach utilizes image similarity to determine the set of scans that are subsequently all registered simultaneously.

3.1. *Detecting Visual Correspondences*

Given two images I_a and I_b, local visual features are extracted using the SIFT algorithm [5] resulting in two sets of features F_a and F_b, corresponding to the two images. Each feature $f_i = \{[X, Y]_i, H_i\}$ in a feature set $F = \{f_i\}$ comprises the position $[X, Y]_i$ in pixel coordinates and a histogram H_i containing the SIFT descriptor.

The feature matching algorithm calculates the Euclidean distance between each feature in image I_a and all the features in image I_b. A potential match is found if the smallest distance is less than 60% of the second smallest distance. This criterion was found empirically and was also used in [23], for example. It reduces the risk of falsely declaring correspondence between SIFT features by excluding cases where several almost equally well matching alternatives exists. Please note that due to this relative matching criterion, feature matching is not a symmetric operation. It is possible that a feature $f_i \in F_a$ matches feature $f_j \in F_b$ but not the other

way around. It is also possible that several features in F_a match a certain feature in F_b. To handle this issue, the feature with the highest similarity is selected if more than one other matching candidate is found.

The feature matching step results in a set of feature pairs $P_{a,b}$, with a total number $M_{a,b} = |P_{a,b}|$ of matched pairs. Since the number of extracted features varies heavily depending on the image content, we normalize the number of matches to the average number of features in the two images and define a similarity measure $S_{a,b} \in [0, 1]$ as:

$$S_{a,b} = \frac{M_{a,b}}{\frac{1}{2}(n_{F_a} + n_{F_b})} \quad (1)$$

where $n_{F_a} = |F_a|$ and $n_{F_b} = |F_b|$ are the number of features in F_a and F_b respectively. An alternative to the normalization in Eq. 1 would be to normalize the number of matches to the maximum number of features in the images. We did, however, not experience problems with using the normalization in Eq. 1.

3.2. *Estimating Visual Feature Depth*

The image data consist of a set of image pixels $\mathcal{P}_j = (X_j, Y_j, C_j)$, where X_j, Y_j are the pixel coordinates and $C_j = (C_j^1, C_j^2, C_j^3)$ is a three-channel color value. By projecting 3D point $p_i = [x_i, y_i, z_i]$ obtained from the laser scanner with the range r_i, onto the image plane, a projected laser range reading $\mathbf{R}_i = (X_i, Y_i, r_i, (C_i^1, C_i^2, C_i^3))$ is obtained, which associates a range value r_i with the coordinates and the color of an image pixel.

The visual feature f_i is located in the image at a sub-pixel position $[X_i, Y_i]$ and we now want to assign a depth estimate r_i^* to the feature f_i. This is the vision-based interpolation problem that is addressed in [24]. In this paper, we apply the simple Nearest Range Reading method and assign the laser range reading r_i to the feature f_i, which corresponds to the projected laser range reading $\mathbf{R}_i$ that is closest to $[X_i, Y_i]$. From the estimated range r_i^* and the pixel coordinates $[X_i, Y_i]$, we finally obtain the 3D position $\mu_{f_i} = (x, y, z)$ of the feature f_i.

3.3. *Estimating Visual Feature Position Covariance*

To obtain the position covariance of each visual feature point C_f, the closest projected laser point p_0 relative to the visual feature f in the image plane

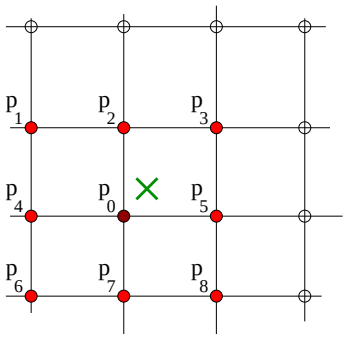

Fig. 1. Laser points used to estimate the position covariance of an image feature (indicated by × in the figure). Circles represent range readings. Filled dots $p_{0..M}$ represent range readings used to compute the covariance estimate. The filled dot p_0 represents the laser point from which the depth of the visual feature was determined. The horizontal lines indicate 2D laser range scanning planes and the vertical lines the tilt movement of the wrist.

is used together with M surrounding laser points $p_{1..M}$, see Fig. 1. The covariance C_f is then calculated as

$$C_f = \frac{1}{M-1} \sum_{i=0}^{M} (p_i - \mu)^2, \quad (2)$$

where $\mu = \frac{1}{M} \sum_{i=0}^{M} p_i$. In our experimental evaluation we used $M = 8$, see Fig. 1.

The motivation for selecting the points to compute the covariance estimate in image space is that the bearing of visual features is typically more accurate than their depth. In the image space, range readings from the laser scanner that have a similar bearing can be easily found. If the selection of the points was done using the 3D points alone, issues with depth discontinuities would have to be handled.

3.4. *ICP*

The iterative closest points (ICP) algorithm [3,25], finds a rigid body transformation $(\mathbb{R}, \mathbf{t})$ between two scan poses $\mathbf{x}_p$ and $\mathbf{x}_c$ by minimizing the following function

$$J(\mathbb{R}, \mathbf{t}) = \sum_{i=1}^{N} ||p_i^c - \mathbb{R}p_i^p - \mathbf{t}||^2, \quad (3)$$

where p_i^p and p_i^c are corresponding points from scans $\mathcal{S}_p$ and $\mathcal{S}_c$. Corresponding point pairs are determined by searching for closest points using a distance metric that varies for different ICP variants. Searching for closest points is the most time consuming part of the algorithm. A common approach

to decrease the search time is to use a k-d tree. If the correspondences are known, there exist various closed-form solutions to obtain the rigid transformation that minimizes Eq. 3. We have adopted the singular value decomposition method proposed by Arun et al. [26]. In our approach, correspondences are detected using visual features, thus an exhaustive search in the spatial domain is not required. As for all ICP methods we make the assumption that the measurement noise is Gaussian and independent and identically distributed.

3.5. *Generalized Total Least Squares ICP*

Generalized Total Least Squares ICP (GTLS-ICP) has been proposed by San-Jose et al. [27] as an extension of ICP. This method is similar to standard ICP but considers the covariance of each point. Instead of Eq. 3, GTLS-ICP minimizes the following function:

$$J(\mathbb{R},\mathbf{t}) = \sum_{i=1}^{N}(q_i - p_i^c)^T C_{q_i}^{-1}(q_i - p_i^c) + \sum_{i=1}^{N}(p_i^c - q_i)^T C_{p_i^c}^{-1}(p_i^c - q_i), \quad (4)$$

where $q_i = \mathbb{R}p_i^p + \mathbf{t}$. The covariance matrix C_{q_i} is obtained by rotating the eigen vectors of the covariance matrix $C_{p_i^p}$, Eq. 2, with the rotation matrix $\mathbb{R}$. However, there is no closed-form solution to minimize this function and therefore the GTLS-ICP function is minimized using an iterative optimization method.

3.6. *Trimmed ICP Extension*

Since visual features are used to establish corresponding points, no further means of data association, (such as searching for closest data points in ICP) is necessary. Although the SIFT features were found to be very discriminative (see for example [28]), there is of course still a risk that some of the correspondences are not correct. To further decrease the possibility of erroneous point associations, only a set fraction of the correspondences with the smallest spatial distance between the corresponding points is used for registration. In the experiments presented in this paper the fraction was set to 70%. Because the relative pose estimate affects the spatial distance between corresponding points, relative pose updates

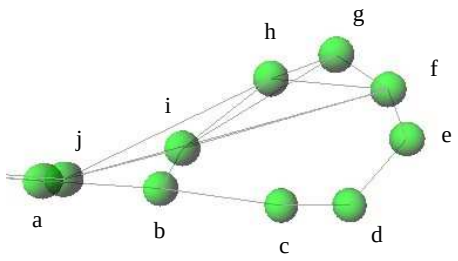

Fig. 2. An example of a pose graph in 3D, seen from above. Each sphere represents a scan pose (node) $\mathbf{x}$ and each line represents an edge e.

are calculated repeatedly until a stopping criterion is met. Any initial pose estimate can be used. In our implementation we always start with an initially identical pose. For the stopping criterion, we consider the change of the sum of the squared distance between the corresponding points compared to the previous iteration. The optimization was stopped if the difference was less than 10^{-6} m^2.

In order to improve convergence, the initial estimate used in the trimmed GTLS-ICP method (Tr. GTLS-ICP) was obtained from the trimmed ICP method (Tr. ICP) to increase the convergence rate. We did not observe convergence problems using this approach.

4. Global Registration

In this section we present an extension of our method that matches a set of multiple scans simultaneously. Visual features are used to determine which scans are included in the registration process by simply thresholding the similarity measure $S_{a,b}$, Eq. 1.

The proposed approach, which is related to graph-based SLAM methods, can be described as follows: Given a set of n scans $\mathcal{S}_{1..n}$ and their estimated poses $\mathbf{x}_{1..n}$ together with a set of extracted features $F_{\mathcal{S}_{1..n}}$, an initial estimate of all poses can be calculated by performing sequential registration, i.e. registering each scan $\mathcal{S}_i$ with the previous one $\mathcal{S}_{i-1}$. With this approach, however, the errors will accumulate. If the current scan $\mathcal{S}_i$ can be registered to a previous scan $\mathcal{S}_j$ that is not its direct predecessor (i.e. $j < i-1$), the uncertainty of the pose estimates can be bounded. After each loop closing an additional edge is therefore added to the edges between subsequent scans so that we end up with n scan poses

and m edges where $m > n$, i.e. we obtain an overestimated equation system. By adding more edges or constraints, the pose of each node can be determined more accurately since more measurements are incorporated (given that the certainty of the constraint is estimated correctly). A pose graph containing both scan poses $\mathbf{x} = [a, b, ..., i]$ and edges e can be seen in Fig. 2. An edge that connects two nodes which previously were separated by many edges generally provides more information in terms of pose error reduction than an edge which connects two nodes that are separated with few edges. For example, the edge $e_{i,b}$ in Fig. 2 reduces the number of edges between a to i to two, compared to nine if only successive edges were used.

The problem of determining all the scan poses simultaneously given the edge constraints can now be defined as minimizing:

$$L(\mathbf{x}_{1..n}) = \sum_{i=0}^{n} \sum_{j=i+1}^{n} V(i,j) J'(\mathbb{R}_{\mathbf{x}_i}, \mathbf{t}_{\mathbf{x}_i}, \mathbb{R}_{\mathbf{x}_j}, \mathbf{t}_{\mathbf{x}_j}). \quad (5)$$

$V(i,j)$ is a binary variable that decides whether the similarity measure $S_{i,j}$ is above a preselected threshold and J' is the extension of the function J (either Eq. 3 or Eq. 4) where each point cloud has its own rotation and translation. The total number of summations needed in Eq. 5 is the number of edges m. For the optimization of Eq. 5, we apply the Fletcher-Reeves conjugate gradient optimization method [29]. For this method the Hessian has to be computed which was done numerically. All six dimensions were optimized, three for translation and three for rotation. Euler angles were used to represent the rotation.

5. Experimental Setup

5.1. *Hardware*

For the experiments presented in this paper we used the ActivMedia P3-AT robot "Tjorven" shown in Fig. 3, equipped with a 2D laser ranger scanner (SICK LMS 200) and a 1-MegaPixel (1280x960) color CCD camera. The CCD camera and the laser scanner are both mounted on a pan-tilt unit from Amtec with a displacement between the optical axes of approx. 0.2 m. The angular resolution of the laser scanner was set to 0.25 degrees.

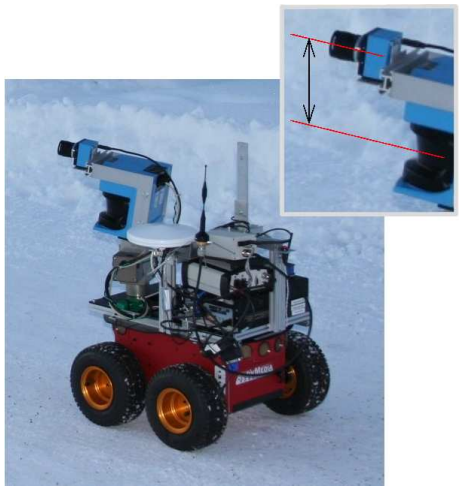

Fig. 3. Our mobile robot platform "Tjorven" equipped with the sensors used in this paper: the SICK LMS 200 laser range scanner and a color CCD camera both mounted on an Amtec pan tilt unit. The close-up shows the displacement between the camera and the laser which causes parallax errors.

5.2. *Data Collection*

For each scan pose, 3D range and image data were collected as follows: First, three sweeps are carried out with the laser scanner at -60, 0 and 60 degrees relative to the robot orientation (horizontally). This results in three separate sets of 3D points that can be straightforwardly merged using the known positions of the laser scanner during the three sweeps. During each of these sweeps, the tilt of the laser scanner is continuously shifted from -40 degrees (looking up) to 30 degrees (looking down). After the three range scan sweeps, seven camera images were recorded at -90, -60, -30, 0, 30, 60, and 90 degrees relative to the robot orientation (horizontally) and at a fixed tilt angle of -5 degrees (looking up). A full data set acquired at a single scan pose is visualized in Fig. 4.

5.3. *Calibration*

In our setup the displacement between the laser scanner and the camera is fixed. It is necessary to determine six external calibration parameters (three for rotation and three for translation) once. This is done by simultaneously optimizing the calibration parameters for several calibration scans. The method we apply requires a special calibration board, see Fig. 5, which is also used to determine the internal calibration parameters of the camera. The calibration board was framed with reflective

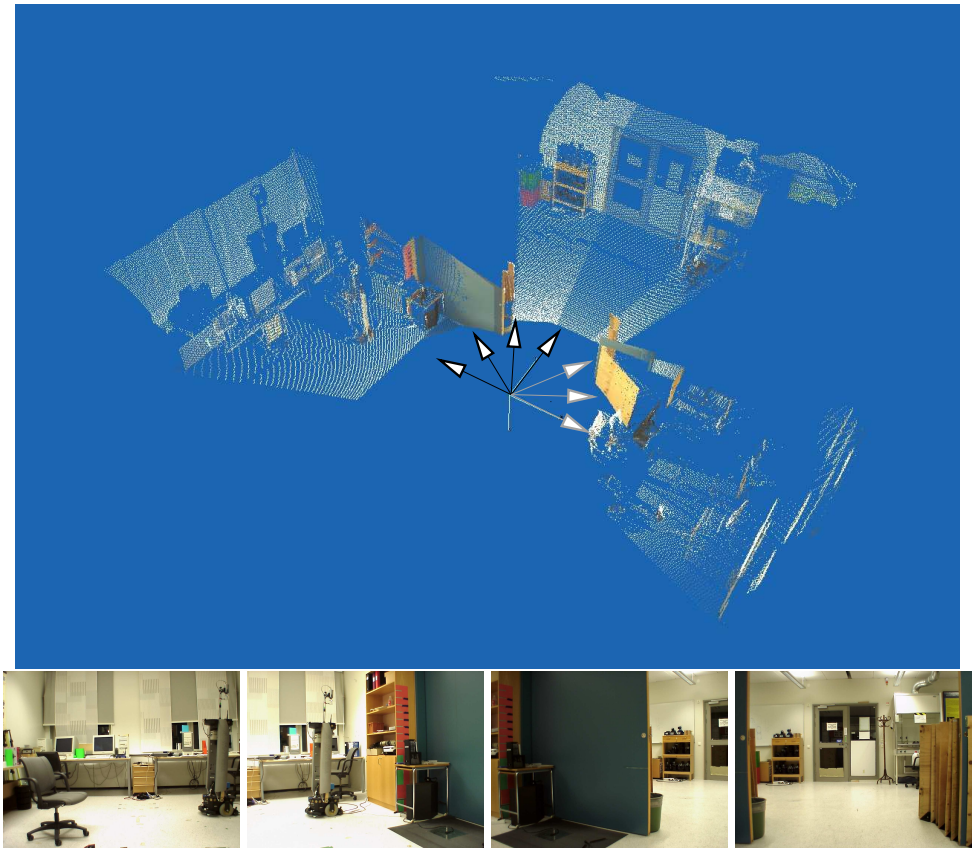

Fig. 4. Top: Full data set acquired for a single scan pose comprising three sweeps with the laser scanner fused with color information from seven camera images. The first four images (marked with dark arrows in the top figure) are shown at the bottom.

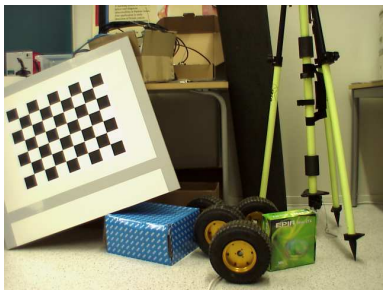

Fig. 5. Calibration board used to determine the calibration parameters of the camera, with a chess board texture and reflective tape (gray border) to locate the board using remission / intensity values from the laser scanner.

tape enabling to use the reflective (remission) values from the laser scanner to automatically estimate the 3D position of the chess board corners detected in the image. The external parameters for the camera are obtained by minimizing the sum of squared distances (SSD) between the chess board corners found in the image and the 3D position of the chess board corners derived from the laser range readings.

6. Experimental Results

6.1. *Indoor Experiment*

A data set consisting of 22 scan poses, containing 66 laser scanner sweeps and 154 camera images was collected as described in Sec. 5.2 in an indoor lab environment. The first scan pose and the last scan pose were recorded at a similar position. An example of the final registration result can be seen in Fig. 6.

To evaluate the registration performance, we register scans sequentially and add up the relative poses. Then we perform a final registration of the last scan with the first one and add the corresponding pose as well. Finally, we compare the resulting total relative pose estimate with the ground truth that is $\mathbf{t} = (0, 0, 0)$ and $\mathbb{R} = I_{3\times3}$ by construction. This method works if the robot was driven along a

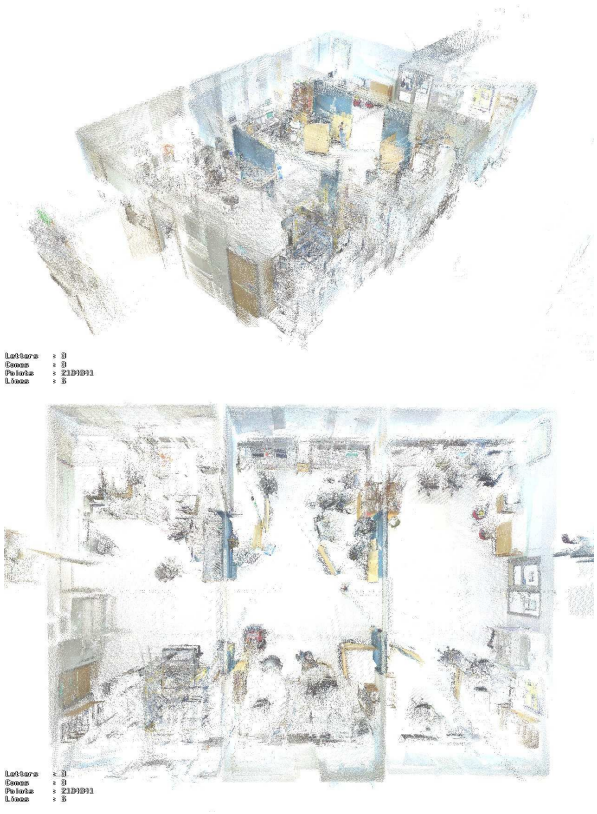

Fig. 6. Result of sequential registration of 22 scan poses. The visualized data comprise of 3 × 22 registered scans and the corresponding colors from 7 × 22 camera images.

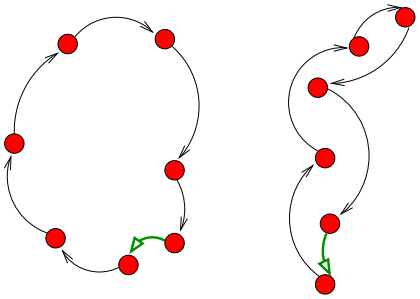

Fig. 7. Construction to obtain the ground truth for sequential registration of scans (see description in the text). The thicker arrow indicates the final registration with the first scan. Left: case in which the robot path formed a loop. Right: case in which the robot path does not form a loop.

round trip but not if the robot followed a straight path and did not return to the initial pose. In the latter case we create a virtual loop by "moving forward" selecting every second scan and moving backwards using the remaining scans, see Fig. 7. The results of this evaluation show sensitively the accuracy of the pairwise registration since even small registration errors can heavily influence the estimate of the final pose.

Table 1 presents a sequence of results in which the number of correspondences was limited to a certain number $\mathcal{N}$. The $\mathcal{N}$ correspondences used for registration were selected randomly from the set of matching image features and each registration experiment was repeated 20 times. The table shows the Euclidean pose error d (in meters) and the rotational error α (in radians). These results show that the performance of Tr. GTLS-ICP is better compared to Tr. ICP, especially when there are fewer corresponding matches and $\mathcal{N}$ is low.

Average execution times of Tr. ICP and Tr. GTLS-ICP using the indoor data set and $\mathcal{N} = 30$ are shown in Table 2. One can clearly see that the closed form solution provides a superior performance. The results in Table 2 were obtained with un-optimized source code on a Pentium 4 with 2GHz and 512 MB of RAM.

If standard ICP with Euclidean distance is used with all the 3D laser points, the sequential registration results in a distance error of 7.10 meters and an angular error of 2.17 radians. This result was obtained using the same initial estimate (that the two scans are located at the same pose) as in the other registration experiments presented in this paper. It is clear that due to the low quality of the initial pose estimate the standard ICP method performs very bad.

Table 1
Registration results given in meters and radians using the trimmed registration versions using the indoor data set

	Tr. ICP		Tr. GTLS-ICP	
$\mathcal{N}$	$d \pm \sigma_d$	$\alpha \pm \sigma_\alpha$	$d \pm \sigma_d$	$\alpha \pm \sigma_\alpha$
10	4.60±5.16	1.52±1.86	3.59±4.87	1.11±1.32
15	0.83±1.04	0.34±0.74	0.71±1.16	0.32±0.79
20	1.09±1.91	0.57±1.32	0.54±0.89	0.28±0.84
30	0.32±0.26	0.07±0.04	0.20±0.10	0.05±0.02
40	0.23±0.13	0.04±0.02	0.20±0.09	0.03±0.02
60	0.16±0.08	0.03±0.02	0.16±0.06	0.02±0.01

Table 2
Execution time for the registration alone with pre-calculated correspondences (without calculating the feature descriptors, covariances etc.) using the indoor data set with $\mathcal{N} = 30$.

method	time
Tr. ICP	0.0022 s±0.0005 s
Tr. GTLS-ICP	19.0 s±13.13 s

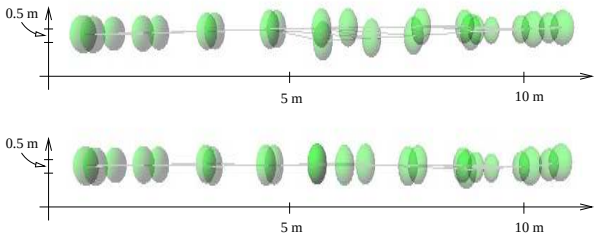

Fig. 8. Comparison of the pose edge graphs seen from the side (see also Table 3). Since the robot was driven indoor on a flat surface, the nodes should appear on a straight line in a side view. Please note the scale of the z-axis (height) that was stretched to emphasize the differences between the presented results. Top: sequential pair-wise registration. Bottom: global registration.

Global Registration

The same indoor data set was used to evaluate the global registration method presented in Sec. 4. For the global registration evaluation Tr. ICP-GTLS was used.

Qualitative results are obtained by calculating the planarity of the estimated poses $\mathbf{x}_{1..22}$. Since the data were collected indoors the ground truth assumption is that all poses should be lying in a plane, see Fig. 8. This plane P is obtained from the two largest eigen vectors $\lambda_{1,2}$ calculated from the covariance matrix C of all poses. Using the plane P, we compute the mean squared error (MSE) from the distance of each pose estimate $\mathbf{x}$ to the plane P and the MSE of the angle between the plane normal and the yaw rotation axis of each pose estimate. The results in Table 3 show the expected improved accuracy of global registration.

Table 3
Comparison of the planarity and angle difference between the plane normal and the yaw axis of the estimated poses (smaller MSE is better) between sequential pair-wise registration and global registration (d - distance error and α - angular error).

	successive registration	global registration
$MSE\ d$	$1.172 \cdot 10^{-3}$ m	$0.187 \cdot 10^{-3}$ m
$MSE\ \alpha$	$1.370 \cdot 10^{-3}$ rad	$0.366 \cdot 10^{-3}$ rad

6.2. *Outdoor Experiment*

An outdoor data set consisting of 32 scan poses was collected close to a building. The ground truth was obtained as explained in the previous section (see Fig. 7, right). Results are shown in Table 4 and in Figs. 9 and 10.

Looking at Fig. 10, for example, the left wall appears much clearer when using Tr. GTLS-ICP indicating that the registration results are better. Also the lamp post appears to be duplicated when using Tr. ICP but not with Tr. GTLS-ICP. If the depth variance is high, the Tr. GTLS-ICP method predominantly uses the bearing of the feature rather than the actually estimated feature position. An overall error of 0.63 meters on average as it is obtained for $\mathcal{N} = 90$ is a very good result considering the challenging outdoor environment and the fact that subsequent scans were taken quite far apart (approx. 3 m).

Table 4
Registration results given in meters and radians using the trimmed registration versions on the outdoor data set.

	Tr. ICP		Tr. GTLS-ICP	
$\mathcal{N}$	$d \pm \sigma_d$	$\alpha \pm \sigma_\alpha$	$d \pm \sigma_d$	$\alpha \pm \sigma_\alpha$
10	26.06±26.76	0.98±0.78	15.81±10.97	0.53±0.29
15	10.91±7.24	0.35±0.35	6.39±4.53	0.20±0.10
20	6.40±2.32	0.24±0.13	4.26±3.19	0.12±0.09
30	4.86±2.48	0.17±0.10	2.66±1.60	0.09±0.04
40	4.48±3.24	0.17±0.18	1.84±1.32	0.06±0.03
60	3.67±1.87	0.17±0.05	1.01±0.68	0.04±0.02
90	3.42±1.24	0.17±0.07	0.63±0.24	0.04±0.02

7. Conclusions

In this paper we have proposed a registration method that uses visual features to handle the correspondence problem. The method integrates both vision and 3D range data from a laser scanner and does not rely on any initial estimate for registration. The range data are used to obtain depth estimates and a covariance estimate for the extracted visual features. A global registration method was presented which illustrate the usefulness of combining visual features with depth estimates from a laser range scanner. The evaluation results show the general applicability of the proposed approach, which allows for registration without an initial pose estimate. Our results demonstrate further the importance of including the position covariance estimate into the registration process, especially if few correspondences are available or under outdoor conditions where the uncertainty of each feature position is typically larger than indoors.

Fig. 9. An outdoor registration result using Tr. GTLS-ICP and unlimited $\mathcal{N}$, visualized with 1.5 million colored laser range readings.

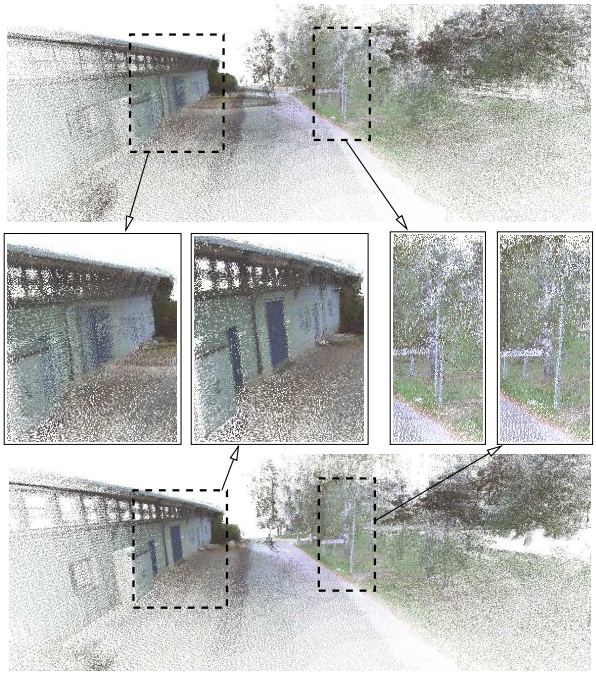

Fig. 10. Outdoor registration results using Tr. ICP (top) and Tr. GTLS-ICP (bottom) without using any limits of corresponding points $\mathcal{N}$. It can be seen that the building wall is more accurately constructed using Tr. GTLS-ICP. By the same token, please note that the highlighted lamp post appears duplicated when using the Tr. ICP method.

References

[1] K. L. Ho, P. Newman, Loop closure detection in SLAM by combining visual and spatial appearance, Robotics and Autonomous Systems 54 (9) (2006) 740–749.

[2] P. M. Newman, D. M. Cole, K. L. Ho, Outdoor SLAM using visual appearance and laser ranging, in: Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA), 2006, pp. 1180–1187.

[3] P. J. Besl, N. D. McKay, A method for registration of 3-d shapes, IEEE Trans. Pattern Analysis and Machine Intelligence 14 (2) (1992) 239–256.

[4] D. C. A. S. M. Abdallah, J. S. Zelek, Towards benchmarks for vision SLAM algorithms, in: Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA), 2006, pp. 1542–1547.

[5] D. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision 60 (2) (2004) 91–110.

[6] A. Nüchter, K. Lingemann, J. Hertzberg, H. Surmann, Heuristic-based laser scan matching for outdoor 6d slam, in: Proc. KI: Advances in Artificial Intelligence. 28th Annual German Conference on AI, 2005, pp. 304–319.

[7] R. Triebel, P. Pfaff, W. Burgard, Multi-level surface maps for outdoor terrain mapping and loop closing, in: Proc. of the IEEE Int. Conf. on Intelligent Robots & Systems (IROS), 2006.

[8] M. Magnusson, A. Lilienthal, T. Duckett, Scan registration for autonomous mining vehicles using 3D-NDT, Journal of Field Robotics 24 (10) (2007) 803–827.

[9] A. Johnson, S. B. Kang, Registration and integration of textured 3-d data, in: InternationalConference on Recent Advances in 3-D Digital Imaging and Modeling (3DIM '97), 1997, pp. 234 – 241.

[10] L.-P. Morency, T. Darrell, Stereo tracking using icp and normal flow constraint, in: Proc. of the Int. Conf. on Pattern Recognition (ICPR), Vol. 4, 2002, pp. 367–372.

[11] R. I. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, 2nd Edition, Cambridge University Press, ISBN: 0521540518, 2004.

[12] A. Davison, Real-time simultaneous localisation and mapping with a single camera, in: Proc. of the IEEE Int. Conf. on Computer Vision (ICCV), 2003, pp. 1403–1410.

[13] S. Se, D. Lowe, J. Little, Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks, International Journal of Robotics Research 21 (8) (2002) 735–758.

[14] M. Maimone, Y. Cheng, L. Matthies, Two years of visual odometry on the mars exploration rovers: Field reports, J. Field Robot. 24 (3) (2007) 169–186.

[15] T. Barfoot, Online visual motion estimation using FastSLAM with SIFT features, in: Proc. of the IEEE Int. Conf. on Intelligent Robots & Systems (IROS), 2005, pp. 579–585.

[16] P. Jensfelt, D. Kragic, J. Folkesson, M. Björkman, A

framework for vision based bearing only 3D SLAM, in: Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA), 2006, pp. 1944–1950.
[17] N. Karlsson, E. D. Bernardo, J. Ostrowski, L. Goncalves, P. Pirjanian, M. E. Munich, The vSLAM algorithm for robust localization and mapping, in: Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA), 2005, pp. 24–29.
[18] P. Elinas, R. Sim, J. Little, σSLAM: Stereo vision SLAM using the Rao-Blackwellised particle filter and a novel mixture proposal distribution, in: Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA), 2006, pp. 1564–1570.
[19] F. Lu, E. Milios, Globally consistent range scan alignment for environment mapping, Autonomous Robots 4 (4) (1997) 333–349.
[20] E. Olson, J. Leonard, S. Teller, Fast iterative optimization of pose graphs with poor initial estimates, in: Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA), 2006, pp. 2262–2269.
[21] G. Grisetti, S. Grzonka, C. Stachniss, P. Pfaff, W. Burgard, Efficient estimation of accurate maximum likelihood maps in 3d, in: Proc. of the IEEE Int. Conf. on Intelligent Robots & Systems (IROS), 2007.
[22] P. Biber, W. Straßer, nScan-matching: Simultaneous matching of multiple scans and application to SLAM, in: Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA), 2006, pp. 2270 – 2276.
[23] J. Gonzalez-Barbosa, S. Lacroix, Rover localization in natural environments by indexing panoramic images, in: Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA), IEEE, 2002, pp. 1365–1370.
[24] H. Andreasson, R. Triebel, A. Lilienthal, Vision-based interpolation of 3d laser scans, in: Proc. of the 2006 IEEE Int. Conf. on Autonomous Robots and Agents (ICARA), IEEE, 2006.
[25] Y. Chen, G. Medioni, Object Modelling by Registration of Multiple Range Images, Image and Vision Computing 10 (3) (1992) 145–155.
[26] K. S. Arun, T. S. Huang, S. D. Blostein, Least-squares fitting of two 3-d point sets, IEEE Trans. Pattern Analysis and Machine Intelligence 9 (5) (1987) 698–700.
[27] R. San-Jose, A. Brun, C.-F. Westin, Robust generalized total least squares iterative closest point registration, in: Seventh Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Lecture Notes in Computer Science, 2004.
[28] K. Mikolajczyk, C. Schmid, A Performance Evaluation of Local Descriptors, IEEE Trans. Pattern Analysis and Machine Intelligence 27 (10) (2005) 1615–1630.
[29] R. Fletcher, C. M. Reeves, Function minimization by conjugate gradients, The Computer Journal 7 (1964) 149–153.