3D Landmarks Extraction from a Range Imager Data for SLAM

Jianguo Jack Wang, Gibson Hu, Shoudong Huang and Gamini Dissanayake
ARC Centre of Excellence for Autonomous Systems
Faculty of Engineering and Information Technology
University of Technology, Sydney
PO Box 123 Broadway, Ultimo, NSW 2007, Australia
Email: jwang@eng.uts.edu.au

Abstract

This paper introduces a new 3D landmark extraction method using the range and intensity images captured by a single range camera. Speeded up robust features (SURF) detection and matching is used to extract and match features from the intensity images. The range image information is used to transfer the selected 2D features into 3D points. The range measurement bias and uncertainty of the range camera are analysed, and their models are developed for improving the range estimation. After outliers' detection and removal using random sampling consensus (RANSAC), reliable 3D points are obtained. 3D landmarks for simultaneous localisation and mapping (SLAM) are selected from the 3D points considering several factors, such as the uncertainty and geometry of their locations. Because of the availability of the SURF descriptor, the data association in SLAM has been performed using both the geometry and the descriptor information. The proposed method is tested in unstructured indoor environments, where the range camera moves in six degrees of freedom. Experimental results demonstrate the success of the proposed 3D landmark extraction method for SLAM.

1 Introduction

A fundamental issue in mobile robotics is navigation and mapping. Only a robot capable of navigating in a safe and reliable way can achieve true autonomy. State-of-the-art approaches for metric localization and mapping are probabilistic methods explicitly considering uncertainty information modelling sensor noise and imperfections in robot motion. The approaches can be categorized by their map representation, including feature, grid or raw data etc. In feature-based approaches it is assumed that the physical environment can be modelled by geometric features [Weingarten, 2006]. Due to the small quantity of data required to represent these features, the resulting maps are compact and the associated algorithms efficient. On the other hand, it requires a reliable feature extraction mechanism, which depends on the quality of a sensor data and chosen feature type.

Landmark based Simultaneous localisation and mapping (SLAM) is the process of building a map of landmarks in the environment while simultaneously uses the map to localize the robot. Apart from the estimation or optimization algorithm used in SLAM, landmark extraction, selection, and association are the key steps for a successful SLAM implementation. This paper introduces a novel 3D landmark selection, and association method for SLAM with data collected from a SwissRanger range camera for robotic search and rescue application.

1.1. Related Work

Laser scanners are one of the most popular sensors employed in 2D and 3D SLAM because of the accurate range information they can provide. The most common approach is to rotate 2D laser scanner to generate 3D scans [Cole and Newman, 2006] [Lingemann et al., 2005] [Nuchter *et. al.* 2007]. The 3D map is usually obtained by scan matching and trajectory optimization. However, the use of laser sensor only may exhibit a number of important limitations: (i) it is difficult provide an appropriate representation of the map, (ii) the number of laser points is large and the computation cost is high (iii) the data association is not easy especially when the loop is closed [Karlsson *et al.*, 2005].

Vision based SLAM has attracted more attention recently due to the low cost, light weight and low power requirements of the camera sensor [Konolige and Agrawal 2008]. In particular, approaches to efficiently represent the salient regions of an image such as Scale Invariant Feature Transformation (SIFT) [Lowe, 2004] and SURF [Bay *et al.*, 2008], provide reliable data association which is an essential component of SLAM. However, an image from a 2D camera can not provide depth information unless stereo vision systems are used [Se *et al.*, 2005] [Paz *et al.*, 2008]. Although there have been some interesting SLAM algorithms using a single camera, in general they are not robust or reliable enough to provide accurate depth information, particularly at longer distances.

Combining range sensors with cameras will certainly allow for the building of more accurate 3D maps. Recently, some research along this direction has been published. For example, [Cole and Newman, 2006] make use of a tilting 2D laser scanner and a camera for SLAM in outdoor environments. The camera is mainly used in the loop closure detection. [Ohno and Tadokoro, 2005] use a 3D scanner to generate the 3D map, and then use the dense texture images to recover the 3D shape.

Ellekilde *et al.* [2007] proposed a combination of sensors in the form of a digital passive camera with one low resolution range camera, CSEM SwissRanger SR-2 [Oggier *et al.*, 2004]. They concluded that the resolution of the SwissRanger is inadequate to be able to extract reliable geometric features of the environment directly from the intensity image with SIFT. Thus visually salient features extracted from a conventional camera are used to help in associating range images and representing the environment.

1.2. Contributions

SwissRanger is a time-of-flight range imaging device that delivers distances as well as grey-level (i.e. intensity) images. Therefore in principle, it is possible to extract 2D feature points from the intensity image as from common images. We demonstrate in this paper that reliable 2D features can be detected from SwissRanger's intensity images with SURF [Bay *et al.*, 2008]. Data collected by SwissRanger only can be processed to provide reliable 3D landmarks for the navigation of robots with six degrees of freedom. This approach not only reduces the complexity of robot navigation and mapping system by avoiding sensor fusion, but also satisfies the special requirement in some applications where no enough nature light is available and visible illumining is forbidden, such as victim search and rescue.

The proposed landmark extraction method combines several strategic approaches. First 2D features are extracted from intensity images and matched by SURF. Adequate number of matches is obtained by adjusting a threshold parameter. Then 2D constraints are applied to remove obvious mismatches. 3D feature points are then interpolated with the four closest range measurements of each feature. After that 3D rotation and translation RANSAC [Fischler and Bolles, 1981] is applied to exclude outliers. The number of points in each frame is controlled in the landmark selection processes. Considering the uncertainty and geometry of their location, good quality 3D points are further selected as landmarks. The SURF descriptor of all the landmarks are kept for robust data association and close loop detection. Each landmark's locating quality can be represented by the range measurement uncertainty. Field tests with ground truth have been conducted to collect data for performance evaluation.

1.3. Framework

In order to achieve robust 3D landmark extraction from SwissRanger data, it is necessary to design a proper data processing procedure, which is illustrated in Figure 1.

SURF features are extracted from the 2D intensity image and matched across those from previous images. After removing outliers, matched features are scored and selected for 3D point interpolation. The ranger images are rectified to remove camera lens distortion and range measure bias. Then the 3D position of selected features is computed from the corresponding pixels in the range image. After coordinate transformation and uncertainty estimation, RANSAC is applied to remove poor quality 3D points and outliers, and to compute the initial estimation of relative camera pose. If adequate 3D points have been found, a 3D landmark selection algorithm is applied based on the uncertainty and geometry of their location. Otherwise, all the points are registered as landmarks.
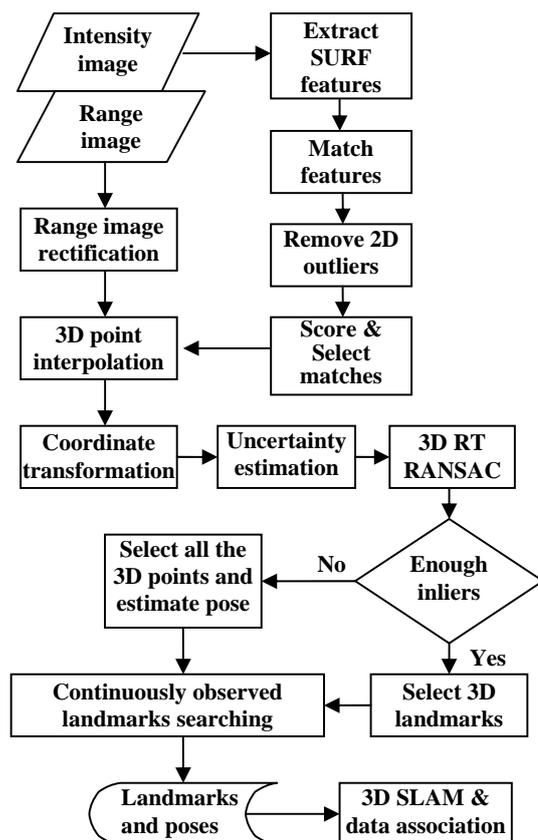

Figure 1. Flowchart of 3D landmark selection

Data association is conducted after the landmark selection. One challenge in data association is that how to efficiently and reliably search previously observed landmarks for finding the re-observed ones. The details of important steps in this flowchart are presented in following sections. The rest of the paper is organized as follows. Intensity image processing is described in Section 2, including feature detection, matching and selection. Range image processing consists of image rectification, measurement uncertainty estimation, 3D point interpolation, outlier detection and landmark selection. All these are introduced in Section 3. Section 4 addresses data association and Section 5 is for field test and results analysis, following by conclusion.

2 Intensity Image Processing

A SwissRanger camera consists of a two dimensional dedicated image sensor and a modulated near-infrared (NIR) light source. Every pixel on the sensor samples the amount of modulated light reflected by objects in a scene four times every period at equal intervals. Its range measurement is calculated from the four measurements, and its intensity image gets from the average of them. The intensity images captured by a SwissRanger camera are basically the same as images from common cameras, but have low resolution. Despite the low resolution, we found that some feature extraction methods developed for common images can still be effective for the intensity images.

2.1 SURF Feature Extraction

A wide variety of scale and rotation invariant feature extraction methods have already been proposed for finding correspondences between images. SIFT and SURF methods are the most popular ones. The main strength of them is to produce a feature descriptor that allows quick comparisons with other features and is rich enough to allow these comparisons to be highly discriminatory. The robust data association between their features is particularly effective. We tested both methods and find that SURF can detect more features and more suitable to our application.

The descriptor generated by SURF describes a distribution of Haar-wavelet responses within the interest point neighbourhood, with only 64 dimensions. This reduces the time for feature computation and matching, and increases simultaneously the robustness. Its indexing step is based on the sign of the Laplacian, which improves not only the matching speed, but also the robustness of the descriptor.

2.2 Feature Matching

There are two popular strategies to match feature descriptors. One is called nearest- similarity-ratio (NSR) matching strategy, where the difference of the best and second best matches must be less than a certain ratio to guarantee reliable matching. The other one is named as similarity-threshold-based (STB) matching strategy. This technique is better suited to represent the distribution of the descriptor in its feature space [Bay et al., 2008].

The feature matching algorithm provided in the SURF package is based on NSR [Torr, 2002]. It cannot avoid multi-matching, and the threshold in the algorithm needs to be adaptively adjusted as a proper number of matches are expected. As both NSR and STB matching strategies have pros and cons, it is anticipated to get better matching results with a wise combination of them.

A new feature matching and selection algorithm is proposed in this paper by combing NSR and STB matching strategies. It can avoid multi-matching and provide required number of matched feature pairs. First the threshold parameter of NSR is set to reasonable high to allow enough matches to be detected. Then some constraints based on statistical value of the features' 2D displacement are implemented to remove evidential mismatches. The similarity scores of the matches are sorted and the best ones of desired number are selected for further process.

Figure 2 shows an example of SURF features matching in a pair of SwissRanger images. The right two images are the pair of intensity images and the left two are range images. There are 53 matched pairs after NSR, showing in the right figures. It is clearly shown that some mismatched pairs appear. After the proposed process, the prefer number of best matches (here is 30) are selected as shown in the left images. Most of them (if not all of them) are inliers.

In order to investigate how the SURF feature matching score affect the reliability of the selected features, we conducted field tests to find the relation between the matching score and the reliability of the SURF features. Test results show that the features with high reliability also have high feature matching scores. This indicates that our approach that combines NSR and STB to select reliable features as landmarks is feasible.
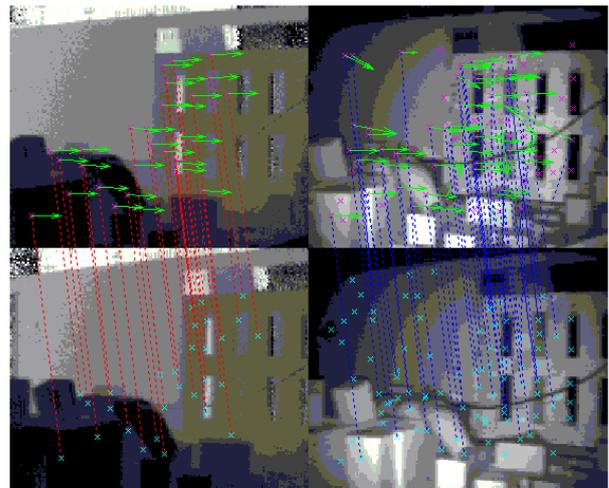

Figure 2. Example of SURF feature matching

It should be noted that sufficient overlapping between images is required for a significant number of features to be matched. Note also that the set of SURF feature descriptors are saved for each landmark. This is very important for data association, which is detailed in Section 4; and one of the major advantages of the proposed method using SURF features.

3 Range Image Processing

The SwissRanger cameras are based on the principle of emitting modulated NIR light with a selected frequency and then measuring the phase shift of the reflection [CSEM, 2006]. The distance measurements are encoded into 16 bit integers. Ideally the mapping should be such that 0 corresponds to 0 m and 65536 to 7.5 m, the known non-ambiguity range of the sensor. The cameras have been calibrated to compensate offset and scaling errors.

In addition to range images, intensity images are also captured, which is related to the reflected signal's strength. The S/N at each pixel has been manifested to exhibit a direct correspondence to the strength of the reflected signal [Oggier et al., 2004]. It is thus reasonable to assume that noise on acquired distance measurement is primarily related to the variance of the signal's strength. Robbins *et al.* [2009] analysed this relation by conducting a test. We have investigated the uncertainty and bias of range measurement and proposed a few models for their estimation.

3.1 Range Image Rectification

In order to calibrate the SwissRanger camera, a two-stage procedure was adopted [Weingarten, 2006]. As the sensor has a standard optical lens refracting the reflected light onto the imaging chip, distortions and misalignments can be compensated using common video camera calibration techniques [Bouguet, 2008].

The relation between range resolution and distance was provided in the product manual [CSEM, 2006]. Weingarten [2006] proposed an empirical relation to calibrate range measurement, which is only depended on the range measurement itself and missing some other factors. It has been demonstrated that for a large range of illumination levels the range uncertainty is essentially

only limited by the noise of the available light [Oggier et al., 2004]. This allows reliably predicting the obtainable range uncertainty if the S/N can be measured. In other words, for every range measurement there is a reliable prediction of its uncertainty, which can be exploited by robot navigation algorithms.

Figure 3 shows the range and intensity images of a calibration board. It shows clearly in the range image (left) that the noise of range measurement corresponding to the black squares in the intensity image (right) is much higher than the noise in white squares.
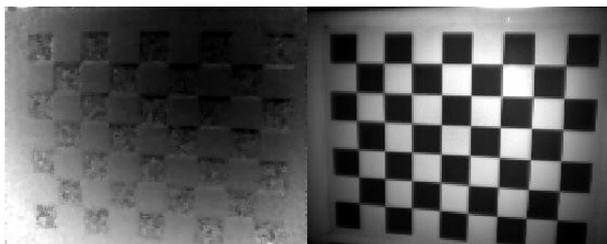
Figure 3. Range and intensity image of a calibration board

We found that actually the S/N not only affects uncertainty of range measurement, but also the range bias. Previous research also found this and suggested a formula for uncertainty estimation [May *et al.*, 2006]. However, no quantitative model for bias has been proposed.

Two models are proposed here for estimating range measurement uncertainty. One directly calculates the standard deviation (STD) of neighbourhood range measurements. This is true when an object surface is flat and perpendicular to the direction of the range. However it is inaccurate in other situations.

The relation of the neighbourhood STD and intensity is investigated. Some of the results are listed in Table 1. It is clearly shown that neighbourhood STD is highly correlated with the mean intensity value.

Table 1. Relation between STD and intensity.

STD (mm)	30.43	11.82	8.53	5.26	3.32	2.20
Intensity	328	428	824	2056	4940	6368

The model we proposed to calculate the range uncertainty according to both range and intensity measurement is expressed as the following equation. This is based on the fact that S/N decides the range measurement uncertainty, which can be estimated in the case of constant background light condition. Positioning the SwissRanger camera at a known distance from the planar shown in Figure 3, a set of images were acquired and the noise at each pixel can be statistically determined. A function can then be fitted to the points obtained by plotting the uncertainty as a function of the intensity and range.

$$\delta r = r \cdot (a + b / (I - c)^k)$$

where I is the intensity measurement; a is the minimum range measurement noise; and c the minimum intensity to get range measurement, which is decided by both the sensor and background noise. b and k are the other two parameters in the equation. a, b, c and k can be estimated

by range image calibration, as -19.1882, 3000, 300 and 0.5 respectively.

Figure 4 is the plot of the test data showing the relationship between the intensity of features and their range measurement uncertainty. This is similar to previous test results [Ellekilde et al., 2007].
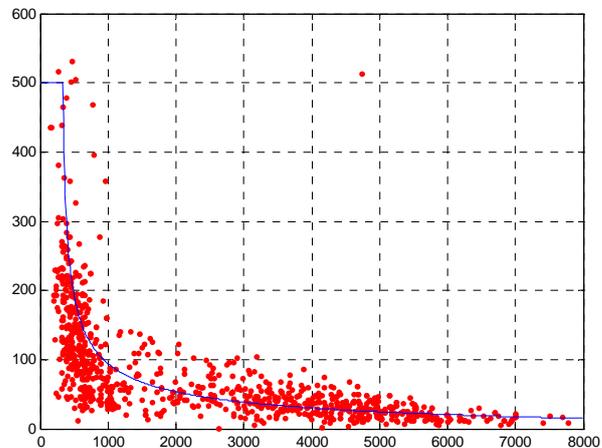

Figure 4. Relationship between intensity and uncertainty

Although the cameras have been calibrated to compensate range offset and scaling errors, they still have range bias caused by the intensity variation. Figure 5 is an example of the plot showing the relationship between SwissRanger image intensity and the range bias. The blue dots are the samples of range measurements with corresponding intensity, the red stars are the median of the samples and the red line is the true distance.
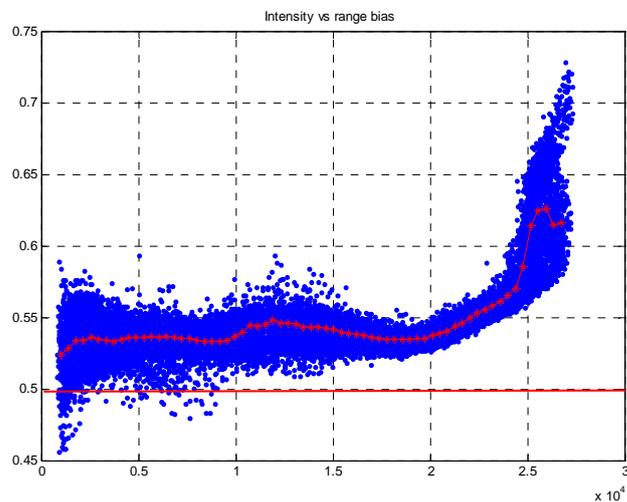

Figure 5. The model of relative range bias and intensity

The range bias can also be empirically modelled by analysing the range image calibration results. Table 2 and 3 are the proposed lookup tables for range bias estimation according to intensity and range respectively.

Table 2. Range bias vs. intensity.

Bias (mm)	-40.0	-20.0	0.0	0.0	28.0	92.00
Intensity	1	300	1500	20000	24300	25600

Table 3. Range bias vs. range.

Bias (mm)	30.0	30.0	120.0	140.0	140.0	110.0	320.0
Range(m)	0.1	1.0	1.5	2.0	2.5	3.0	7.5

3.2   3D Feature Point Interpolation

As matched pairs of SURF features are obtained from 2D intensity images, it is necessary to find corresponding 3D position of the features for SLAM implementation. The range images are rectified first in both image and range aspects according to camera and range calibration results. Then 3D interpolation is applied to calculate the range of each feature. After that the 3D position of the point features are transformed to Cartesian coordinate system. It should be noted that the intensity images do not need to be rectified because all their pixels exactly match the ones in corresponding range images.

There is an option to change the order of range image processing showing in Figure 1 if only rectified features, instead of all the range data, are required. Range image rectification block can be put after the 3D feature interpolation so only the selected 3D feature points are rectified. This will reduce computation load further.

3.3   3D Rotation and Translation RANSAC

The initial matches between two sets of 3D points in two frames are obtained from SURF matching, as described earlier. In fact the matching process is not unambiguous and can produce mismatched pairings. In order to make the landmark selection more robust and estimate the relationship between two camera poses, 3D rotation and translation RANSAC [Fischler and Bolles, 1981] is applied to refine the matches.

This algorithm uses the typical RANSAC based approach to find the inliers of the matched 3D feature points. N pairs of matched points are randomly selected from the data set first. Least square 3D point fitting is applied to compute the rotation and translation that best fit the N pairs of matched feature points. This transformation is then applied to all the points within the data set to find the total number of pairs that are consistent with this particular transformation. The random selection of N pairs of points and the above check is repeated M times to find the transformation with the maximal number of consistent pairs. This entire subset of matches regarded as correct can then be used to calculate the camera pose using the least square fitting of all of them.

Two methods are used for the least square 3D point fitting to improve the robustness. One is the least square 3D point fitting techniques proposed by Arun et al. [1987], where no initial value for the transformation is required but the uncertainty of the point position cannot be taken into account. Another is the iterated linearized least square approach where the uncertainty of the point position can be considered but an initial value of the transformation is needed. One way to get the initial value is using the result from Arun et al. [1987].

As a large number of matched 3D points between two frames are usually detected, a subset selection is needed to limit the number of landmarks. A quality measure has been chosen to avoid features being cluttered within a small area out of those with the smallest range measurement error.

4   Data Association

One of the key tasks for SLAM is to find correct data association. A robot has to determine whether a detected landmark corresponds to a previously seen landmark or to a new one. Vision-based SLAM approaches often use the similarity of features' descriptors from SIFT or SURF etc. algorithms, as a similarity measurement. If the similarity between both descriptors is above a certain threshold, the landmarks are considered to be the same. This technique can only provide good correspondences when the feature has been observed from similar viewing angles [Gil *et al.*, 2006].

Since SURF or SIFT feature descriptors are only partially invariant to affine projection, the descriptor of the same feature may be significantly different when observing it from different viewpoints. For example, a robot may be unable to make the correct data association when moving through the same corridor but from different directions. The descriptor also changes as the scale factor or brightness changes, or image blur happens, as analysed by [Bay *et al.*, 2008].

[Gil *et al.*, 2006] proposed a method to deal with the data association in the context of SIFT features from a pattern classification point of view. The key idea of their method is to track visual landmarks during several consecutive frames and select only those features that stay comparably stable under different viewing angles. This reduces the number of landmarks in the resulting map representation. However, this method requires a feature is tracked in several consecutive frames before it is selected as landmark, which is hard to be met when robot is turning or fast moving.

In this work we proposed a different approach to deal with the data association in the context of SURF features. For each landmark, a set of SURF feature descriptors are associated with it. The selection procedure of the descriptors is as follows. As soon as a landmark is selected the first SURF feature descriptor is associated with it. Then the following descriptors of the landmark are abandoned unless they are different with the first descriptor and the other selected ones in certain degree. At the same time the estimated camera poses number associated with the selected descriptors are also saved for improving data association.

Data association can be done either before the SLAM or during SLAM. Performing data association during SLAM allows the more accurate landmark and poses information to be used, and thus used in this approach. Figure 6 shows the flowchart of the proposed data association process. The aim of this data association algorithm is to be robust as well as computationally efficient.

First geometry of possible matches is estimated. SURF descriptor matching is then applied with an adaptive matching ratio to guarantee enough match quantity without losing match quality. This approach greatly increases data association accuracy.

The landmarks can be further selected according to observation frequency. A step interval is setup to check all landmarks in the global map for landmarks consistency. Only key landmarks are kept with decision based on the state of the camera movement when observation is made and repetitiveness, which can reduce

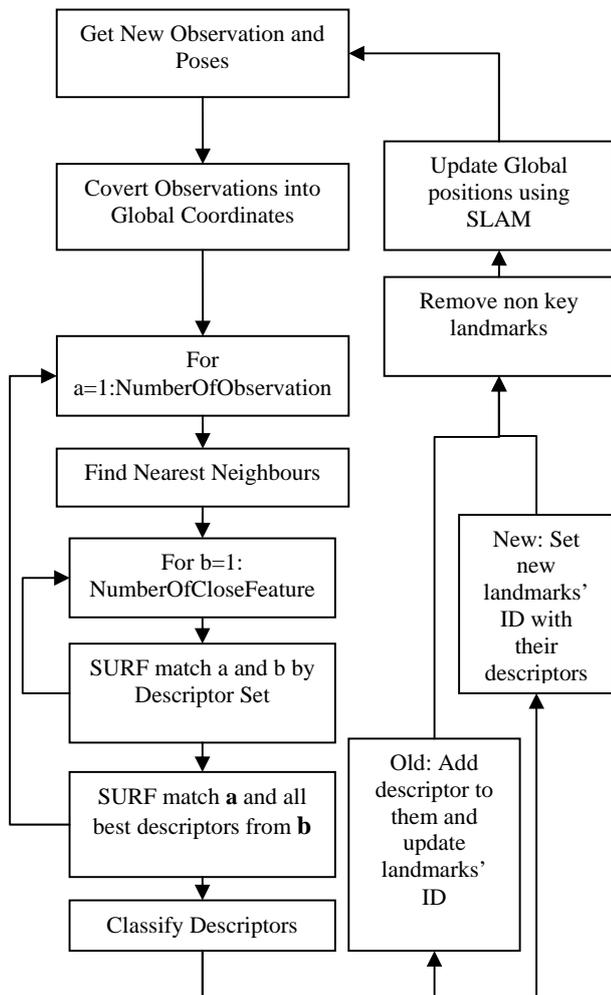the landmarks size and allow SLAM to perform much more efficiently.

Figure 6. Data association flow chart

To achieve a good estimate of global position of positions and odometry, our SLAM optimisation is performed at the end of each non key landmarks removal step. Assuming there is not a large degree of uncertainty in the trajectory, a good estimate of possible loop closures can be calculated.

5 Tests for SLAM

In order to verify the performance of proposed 3D landmark extraction method, field test data with ground truth were collected.

5.1 The experiment setup

The experiment is setup such that it mimics a real search and rescue environment. A SwissRanger camera is mounted on a push trolley, exactly 0.52m above the ground, which we negotiate through a set path marked on the ground. The path also has elevation in the z axis so that it can represent a more practical 3D SLAM situation. Features such as, blocks, crates, mannequins and blankets are scattered randomly throughout this environment.

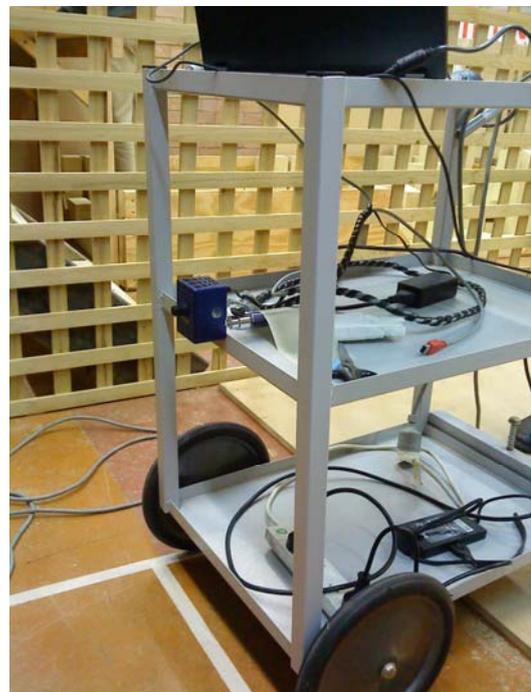Figure 7a. The test field with ground truth

Figure 7b. The test equipment settings

Data is collected by sampling the camera frame at a set rate. The intensity and the range images are stored while the trolley is pushed around the circuit. The trajectory is created in a way that similar features can be re-observed while also closing a loop. Figure 7a shows an example of the environment, and Figure 7b shows the test equipment settings.

5.2 The SLAM algorithm

The SLAM algorithm implemented is a recent developed efficient SLAM algorithm called 3D I-SLSJF: Iterative sparse local sub-map joining algorithm [Hu *et al.*, 2009]. This approach is computationally more efficient than the typical maximum likelihood (ML) method and also shows much better accuracy compared with 3D EKF (Extended Kalman Filter) SLAM. The method essentially builds several local maps with high local feature accuracy and

then joins them to form a global map both using the least squares approach. The algorithm itself exploits the sparseness of the map joining process making it computationally more efficient. Since the algorithm itself is incremental in nature, it can be easily implemented onto a real time system.

5.3   Test Results Comparison

The selected 3D landmarks and camera poses from the output of the 3D rotation and translation RANSAC are used as the input in the 3D I-SLSJF SLAM algorithm, as shown in Figure 1. The initial estimation of the camera poses is given a very high uncertainty covariance, so that the model will be heavily dependent on observations.

Figure 8a shows the results based on odometry only (the odometry is obtained from the relative poses computed from the least square fitting between consecutive frames), in which all the landmarks in each frames are treated as new landmarks. Comparing with the ground truth trajectory, the green line shown in the figure, it can be seen that the error is quite large.
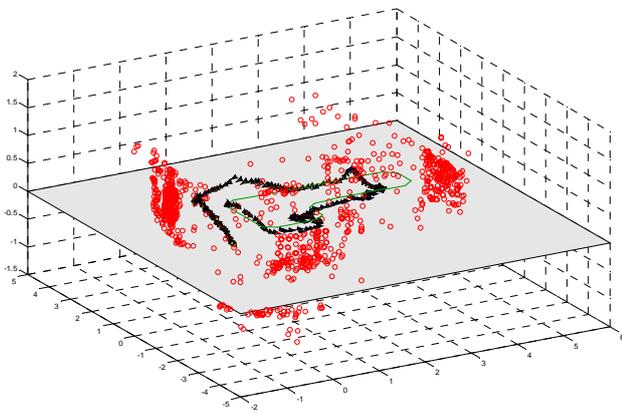

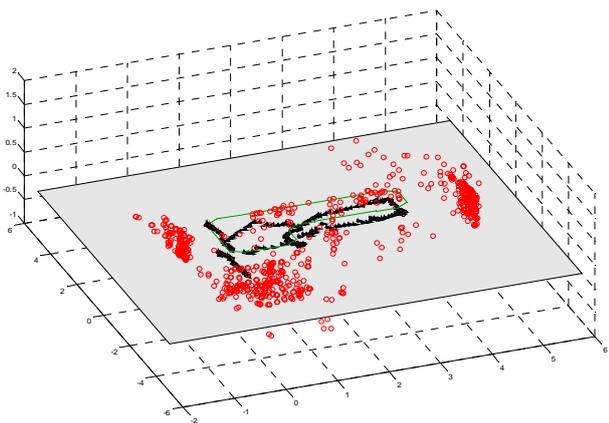Figure 8a. Navigation result with odometry estimation.

Figure 8b. SLAM result without data association.

SLAM results without applying the proposed data association for the 3D landmarks are shown in Figure 8b, where the continuously observed landmarks are identified. The results show that I-SLSJF SLAM can greatly reduce the error produced from feature matching. It improves the pure odometry estimation to some extent.
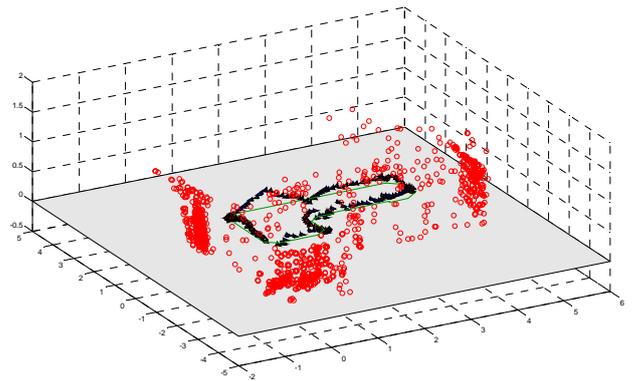

Figure 8c. SLAM results with the close loop information.

Figure 8c shows the SLAM results with loop closure using proposed data association for the 3D landmark data. The navigation result is improved significantly if a closed loop situation is experienced. Like all SLAM methods, The SLAM result is still not very satisfactory presumedly due to the range measurement bias and some other facts of imperfection. Accurate data association and landmark identification is very important as to avoid convergence towards the wrong solution.

6   Conclusion

This paper demonstrated that it is possible to extract and match 3D point features using SwissRanger data only for successful SLAM where the camera moves in six degrees of-freedom without odometry information. Although the quality of intensity image obtained from the SwissRanger is not very good, SURF feature detection can be used to successfully extract and match features in the image plane. The 3D position of the matched feature can be further obtained by the range image. After proper outlier removal and data association steps, the extracted 3D landmarks are used in the SLAM to obtain an estimate of the camera poses and landmark positions.

This work shows that it is sufficient to use a single SwissRanger to perform SLAM without the need of combination with other sensors. However, there are still some limitations of the current research. For example, the ambiguity of the range measurement need to be further analysed. The 3D I-SLSJF can still be slow if too many landmarks are selected for SLAM. The success rate of this method largely depends on data association associated with the feature matching algorithm.

Further research is necessary to make the feature selection and matching more efficient and reliable such that it can be used in real time SLAM.

Acknowledgements

help in SwissRanger camera configuration and calibration.

References

[Bay *et al.*, 2008] Bay, H., A. Ess, T. Tuytelaars, and L. V. Gool. SURF: Speeded Up Robust Features, *Computer Vision and Image Understanding (CVIU)*, *110*(3), 346--359.

[Bouguet, 2008] Bouguet, J.-Y. Camera Calibration Toolbox for Matlab.

[Cole and Newman, 2006] Cole, D. M., and P. Newman. Using laser range data for 3D SLAM in outdoor environments, in *Proceedings of IEEE International Conference on Robotics and Automation*.

[CSEM, 2006] CSEM. SwissRanger SR-3000 Manual, Version 1.02.

[Ellekilde *et al.*, 2007] Ellekilde, L.-P., S. Huang, J. Valls Miro, and G. Dissanayake. Dense 3D Map Construction for Indoor Search and Rescue, *Journal of Field Robotics*, *24*(1/2), 71-89.

[Fischler and Bolles, 1981] Fischler, M., and R. Bolles. RANdom SAmpling Consensus: a paradigm for model fitting with application to image analysis and automated cartography., *Communications of the Association for Computing Machinery*, *24*, 381–395.

[Gil *et al.*, 2006] Gil, A., O. Reinoso, O. M. Mozos, C. Stachniss, and W. Burgard. Improving Data Association in Vision-based SLAM, in *Proceedings of 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Beijing, China, October 9 - 15.

[Hu *et al.*, 2009] Hu, G., S. Huang, and G. Dissanayake. 3D I-SLSJF: A Consistent Sparse Local Submap Joining Algorithm for Building Large-Scale 3D Maps, in *48th IEEE Conference on Decision and Control*, Shanghai, China.

[Karlsson *et al.*, 2005] Karlsson, N., E. Di Bernardo, J. Ostrowski, L. Goncalves, P. Pirjanian, and M. E. Munich. The vSLAM algorithm for robust localization and mapping, in *Proceedings of IEEE International Conference on Robotics and Automation*, Barcelona, Spain.

[Konolige and Agrawal 2008] Konolige K., and M. Agrawal FrameSLAM: from Bundle Adjustment to Realtime Visual Mappping IEEE Transactions on Robotics,, Vol. 24, No. 5. (2008), pp. 1066-1077.

[Lingemann *et al.*, 2005] Lingemann, K., A. Nüchter, J. Hertzberg, and H. Surmann. High-speed laser localization for mobile robots, *Robotics and Autonomous Systems*, *51*, 275–296.

[Lowe, 2004] Lowe, D. G. Distinctive image features from scaleinvariant keypoints, *International Journal of Computer Vision*, *60*(2), 91–110.

[May *et al.*, 2006] May, S., G. Werner, H. Surmann, and K. Pervölz. 3D time-of-flight cameras for mobile robotics, in *Proceedings of the IEEE International Conference on Intelligent Robot and Systems*, Beijing, China.

[Nuchter *et. al.* 2007] Nuchter, A, K. Lingemann, J. Hertzberg, and H. Surmann, 6D SLAM - 3D Mapping Outdoor Environments *Journal of Field Robotics (JFR)*, Volume 24, Issue 8-9, pages 699 - 722, 2007.

[Oggier *et al.*, 2004] Oggier, T., M. Lehmann, R. Kaufmann, M. Schweizer, M. Richter, P. Metzler, G. Lang, F. Lustenberger, and N. Blanc. An all-solid-state optical range camera for 3D real-time imaging with sub-centimeter depth resolution (SwissRanger), in *Proceedings of SPIE The International Society of Optical Engnieering*.

[Ohno and Tadokoro, 2005] Ohno, K., and S. Tadokoro. Dense 3D map building based on LRF data and color image fusion, in *Proceedings of the IEEE International Conference on Intelligent Robot and Systems*.

[Paz *et al.*, 2008] L. M. Paz, P. Pinies, J. D. Tardos, and J. Neira, Large-Scale 6-DOF SLAM With Stereo-in-Hand, IEEE Transactions on Robotics,, Vol. 24, No. 5. (2008), pp. 946-957.

[Robbins et al. 2009] Robbins, S, B. Murawski, and. B. Schroeder, Photogrammetric calibration and colorization of the SwissRanger SR-3100 3-D range imaging sensor, Optical Engineering, 48(5), 053603,_May 2009.

[Saez *et al.*, 2005] Saez, J. M., F. Escolano, and A. Peñalver. First steps towards stereo-based 6DOF SLAM for the visually impaired, in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition—Workshops*, San Diego, CA, USA.

[Se *et al.*, 2005] Se, S., D. G. Lowe, and J. Little. Vision-based global localization and mapping for mobile robots, *IEEE Transactions on Robotics*, *21*(3), 364–375.

[Torr, 2002] Torr, P. H. S. A Structure and Motion Toolkit in Matlab: Interactive Adventures in S and M.

[Weingarten, 2006] Weingarten, J. *Feature-based 3D SLAM*, Swiss Federal Institute of Technology in Lausanne.