International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XXXIX-B3, 2012
XXII ISPRS Congress, 25 August – 01 September 2012, Melbourne, Australia

A STEP TOWARDS DYNAMIC SCENE ANALYSIS WITH ACTIVE MULTI-VIEW RANGE IMAGING SYSTEMS

Martin Weinmann and Boris Jutzi

Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology (KIT)
Kaiserstr. 12, 76128 Karlsruhe, Germany
{martin.weinmann, boris.jutzi}@kit.edu

Commission III, WG III/5

KEY WORDS: LIDAR, Multisensor, Point Cloud, Imagery, Automation, Close Range, Dynamic

ABSTRACT:

Obtaining an appropriate 3D description of the local environment remains a challenging task in photogrammetric research. As terrestrial laser scanners (TLSs) perform a highly accurate, but time-dependent spatial scanning of the local environment, they are only suited for capturing static scenes. In contrast, new types of active sensors provide the possibility of simultaneously capturing range and intensity information by images with a single measurement, and the high frame rate also allows for capturing dynamic scenes. However, due to the limited field of view, one observation is not sufficient to obtain a full scene coverage and therefore, typically, multiple observations are collected from different locations. This can be achieved by either placing several fixed sensors at different known locations or by using a moving sensor. In the latter case, the relation between different observations has to be estimated by using information extracted from the captured data and then, a limited field of view may lead to problems if there are too many moving objects within it. Hence, a moving sensor platform with multiple and coupled sensor devices offers the advantages of an extended field of view which results in a stabilized pose estimation, an improved registration of the recorded point clouds and an improved reconstruction of the scene. In this paper, a new experimental setup for investigating the potentials of such multi-view range imaging systems is presented which consists of a moving cable car equipped with two synchronized range imaging devices. The presented setup allows for monitoring in low altitudes and it is suitable for getting dynamic observations which might arise from moving cars or from moving pedestrians. Relying on both 3D geometry and 2D imagery, a reliable and fully automatic approach for co-registration of captured point cloud data is presented which is essential for a high quality of all subsequent tasks. The approach involves using sparse point clouds as well as a new measure derived from the respective point quality. Additionally, an extension of this approach is presented for detecting special objects and, finally, decoupling sensor and object motion in order to improve the registration process. The results indicate that the proposed setup offers new possibilities for applications such as surveillance, scene reconstruction or scene interpretation.

1 INTRODUCTION

An appropriate 3D description of the local environment is represented in the form of point clouds consisting of a large number of measured 3D points and, optionally, different attributes for each point. Such point clouds can directly be acquired with different scanning devices such as terrestrial laser scanners (TLSs), time-of-flight (ToF) cameras or devices based on the use of structured light. However, a single scan often is not sufficient and hence, multiple scans have to be acquired from different locations in order to get a full scene coverage. As each captured point cloud represents 3D information about the local area only with respect to a local coordinate frame, a basic task for many applications consists of a point cloud registration. This process serves for estimating the transformation parameters between different point clouds and transforming all point clouds into a common coordinate frame. Existing techniques for point cloud registration rely on

- 3D geometry,
- 3D geometry and the respective 2D representation as range image and
- 3D geometry and the corresponding 2D representation of intensity values.

Standard approaches involving only the spatial 3D information for calculating the transformation parameters between two partially overlapping point clouds are based on the Iterative Closest Point (ICP) algorithm (Besl and McKay, 1992) and its variants (Rusinkiewicz and Levoy, 2001). Iteratively minimizing the difference between two point clouds however shows a high computational effort for large numbers of points. Hence, other registration approaches are based on information extracted from the point clouds. This information may for instance be derived from the distribution of the points within each point cloud by using the normal distributions transform (NDT) either on 2D scan slices (Brenner et al., 2008) or in 3D (Magnusson et al., 2007). If the presence of regular surfaces can be assumed in the local environment, various types of geometric features are likely to occur, e.g. planes, spheres and cylinders. These features can directly be extracted from the point clouds and strongly support the registration process (Brenner et al., 2008; Pathak et al., 2010; Rabbani et al., 2007). In cluttered scenes, descriptors representing local surface patches are more appropriate. Such descriptors may be derived from geometric curvature or normal vectors of the local surface (Bae and Lichti, 2008).

As the scans are acquired on a regular grid resulting from a cylindrical or spherical projection, the spatial 3D information can also be represented as range image. This range image provides additional features such as distinctive feature points which strongly support the registration process (Barnea and Filin, 2008; Steder et al., 2010).

Currently, most of the scanning devices can not only capture 3D information but also either co-registered camera images or panoramic reflectance images representing the respective energy of the backscattered laser light. The additional information typ-

ically is represented as intensity image. This intensity image might provide a higher level of distinctiveness than shape features (Seo et al., 2005) and thus information about the local environment which is not represented in the range measurements. Hence, the registration process can efficiently be supported by using reliable feature correspondences between the respective intensity images. Although different kinds of features can be used for this purpose, most of the current approaches are based on the use of feature points or keypoints as these tend to yield the most robust results for registration without assuming the presence of regular surfaces in the scene. Distinctive feature points simplify the detection of point correspondences and for this reason, SIFT features are commonly used. These features are extracted from the co-registered camera images (Al-Manasir and Fraser, 2006; Barnea and Filin, 2007) or from the reflectance images (Wang and Brenner, 2008; Kang et al., 2009). For all point correspondences, the respective 2D feature points are projected into 3D space using the spatial information. This yields a much smaller set of 3D points for the registration process and thus a much faster estimation of the transformation parameters between two point clouds. Furthermore, additional constraints considering the reliability of the point correspondences (Weinmann et al., 2011; Weinmann and Jutzi, 2011) allow for increasing the accuracy of the registration results.

Once 2D/2D correspondences are detected between images of different scans, the respective 3D/3D correspondences can be derived. Thus knowledge about the closest neighbor is available and the computationally expensive ICP algorithm can be replaced by a least squares adjustment. Least squares methods involving all points of a scan have been used for 3D surface matching (Gruen and Akca, 2005), but since a large overlap between the point clouds is required which can not always be assumed, typically sparse 3D point clouds consisting of a very small subset of points are derived from the original 3D point clouds (Al-Manasir and Fraser, 2006; Kang et al., 2009). To further exclude unreliable 3D/3D correspondences, filtering schemes based on the RANSAC algorithm (Fischler and Bolles, 1981) have been proposed in order to estimate the rigid transformation aligning two point clouds (Seo et al., 2005; Böhm and Becker, 2007; Barnea and Filin, 2007).

For dynamic environments, terrestrial laser scanners which perform a time-dependent spatial scanning of the scene are not suited. Furthermore, due to the background illumination, monitoring outdoor environments remains challenging with devices based on structured light such as the Microsoft Kinect device which uses random dot patterns of projected infrared points for getting reliable and dense close-range measurements in real-time. Hence, this paper is focused on airborne scene monitoring with range imaging devices mounted on a sensor platform. Although the captured point clouds are corrupted with noise and the field of view is very limited, a fast, but still reliable approach for point cloud registration is presented. The approach involves an initial camera calibration for increased accuracy of the respective 3D point clouds and the extraction of distinctive 2D features. The detection of 2D/2D correspondences between two successive frames and the subsequent projection of the respective 2D points into 3D space yields 3D/3D correspondences. Using such sparse point clouds significantly increases the performance of the registration process, but the influence of outliers has to be considered. Hence, a new weighting scheme derived from the respective point quality is introduced for adapting the influence of each 3D/3D correspondence on a weighted rigid transformation. Additionally, an extension of this approach is presented which is based on the already detected features and focuses on a decoupling of sensor and object motion.

The remainder of this paper is organized as follows. In Section 2, the proposed methodology for successive pairwise registration in dynamic environments is described as well as a simple extension for decoupling sensor and object motion. The configuration of the sensor platform is outlined in Section 3. Subsequently, the performance of the presented approach is tested in Section 4. The derived results are discussed in Section 5. Finally, in Section 6, the content of the entire paper is concluded and suggestions for future work are outlined.

2 METHODOLOGY

The proposed methodology provides fast algorithms which are essential for time-critical surveillance applications and should be capable for a real-time implementation on graphic processors. After data acquisition (Section 2.1), a preprocessing has to be carried out in order to get the respective 3D point cloud (Section 2.2). However, the point cloud is corrupted with noise and hence, a quality measure is calculated for each point of the point cloud (Section 2.3). Subsequently extracting distinctive features from 2D images allows for detecting reliable 2D/2D correspondences between different frames (Section 2.4), and projecting the respective 2D points into 3D space yields 3D/3D correspondences of which each 3D point is assigned a value for the respective point quality (Section 2.5). The point cloud registration is then carried out by estimating the rigid transformation between two sparse point clouds where the weights of the 3D/3D correspondences are derived from the point quality of the respective 3D points (Section 2.6). Finally, a feature-based method for object detection and segmentation is introduced (Section 2.7) which can be applied for decoupling sensor and object motion.

2.1 Data Acquisition

In contrast to the classical stereo observation techniques with passive sensors, where data from at least two different viewpoints has to be captured, the monostatic sensor configuration of the PMD[vision] CamCube 2.0 preserves information without the need of a co-registration of the captured data. A PMD[vision] CamCube 2.0 simultaneously captures various types of data, i.e. geometric and radiometric information, by images with a single shot. The images have a size of 204×204 pixels which corresponds to a field of view of $40^\circ \times 40^\circ$. Thus, the device provides measurements with an angular resolution of approximately 0.2°. For each pixel, three features are measured, namely the respective range R, the active intensity I_a and the passive intensity 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 a single frame consisting of a range image $\mathbf{I}_R$, an active intensity image $\mathbf{I}_a$ and a passive intensity image $\mathbf{I}_p$ can be updated with high frame rates of more than 25 releases per second, this device is well-suited for capturing dynamic scenes.

2.2 Preprocessing

In a first step, the intensity information of each frame, i.e. $\mathbf{I}_a$ and $\mathbf{I}_p$, has to be adapted. This is achieved by applying a histogram normalization of the form

$$I_n = \frac{I - I_{min}}{I_{max} - I_{min}} \cdot 255 \tag{1}$$

which adapts the intensity information I of each pixel to the interval $[0, 255]$. The modified frames thus consist of a normalized active intensity image $\mathbf{I}_{n,a}$, a normalized passive intensity image $\mathbf{I}_{n,p}$ and the range image $\mathbf{I}_R$ which are illustrated in Figure 1.

For all subsequent tasks, it is essential to get the 3D information as accurate as possible. Due to radial lens distortion and decentring distortion, however, the image coordinates have to be adapted in order to be able to appropriately capture a scene. Hence, a camera calibration is carried out for the used devices. This yields a corrected grid of image coordinates with the principal point as origin of the new 2D coordinate frame. For each point $\mathbf{x} = (x, y)$ on the new grid, the respective 3D information in the local coordinate frame can then be derived from the measured range value R with

$$R = \sqrt{X^2 + Y^2 + Z^2} \tag{2}$$

and a substitution of X and Y with

$$X = \frac{x}{f_x} \cdot Z \quad \text{and} \quad Y = \frac{y}{f_y} \cdot Z \tag{3}$$

where f_x and f_y are the focal lengths in x- and y-direction. Solving for the depth Z along the optical axis yields

$$Z = \frac{R}{\sqrt{\left(\frac{x}{f_x}\right)^2 + \left(\frac{y}{f_y}\right)^2 + 1}} \tag{4}$$

and thus, the 3D point $\mathbf{X} = (X, Y, Z)$ corresponding to the 2D point $\mathbf{x} = (x, y)$ has been calculated. Consequently, the undistortion of the 2D grid and the projection of all points onto the new grid lead to the respective point cloud data.

Figure 1: Image representation of normalized active intensity, normalized passive intensity and range data.

2.3 Point Quality Assessment

For further calculations, it is feasible to derive a measure which describes the quality of each 3D point. Those points which arise from objects in the scene will probably provide a smooth surface, whereas points corresponding to the sky or points along edges of the objects might be very noisy. Hence, for each point on the regular 2D grid, the standard deviation σ of all range values within a 3×3 neighborhood is calculated and used as a measure describing the reliability of the range information of the center point. This yields a 2D confidence map according to which the influence of a special point on subsequent tasks can be weighted. For the example depicted in Figure 1, the corresponding confidence map is shown in Figure 2.

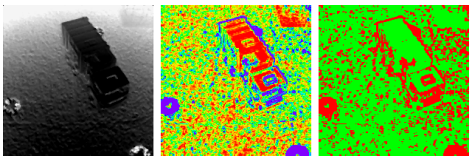

Figure 2: Range image, confidence map (pseudo-color representation where reliable points are marked in red and unreliable ones in blue) and thresholded confidence map (green: $\sigma \leq 0.05$ m).

2.4 2D Feature Extraction

As each frame consists of range and image data acquired on a regular grid, the alignment of two point clouds is based on using both kinds of information. However, instead of using the whole 3D information available which results in a high computational effort, the intensity information is used to derive a much smaller set of 3D points. Hence, distinctive 2D features are extracted from the intensity information which later have to be projected into 3D space. For this purpose, the Scale Invariant Feature Transform (SIFT) (Lowe, 2004) is carried out on the normalized active intensity image as well as on the normalized passive intensity image. This yields distinctive keypoints and 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. The vector representation of these descriptors allows for deriving correspondences between different images by considering the ratio

$$r = \frac{d(N_1)}{d(N_2)} \tag{5}$$

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. This ratio $r \in [0, 1]$ describes the distinctiveness of a keypoint. Distinctive keypoints arise from low values and hence, the ratio r has to be below a certain threshold t_{des}. Typical values for this threshold are between 0.6 and 0.8. This procedure yields n_a correspondences between the normalized active intensity images of the two frames and n_p correspondences between the normalized passive intensity images. For the registration process, it is not necessary to distinguish between the two types of correspondences as only the spatial relations are of interest. Hence, a total number of $n = n_a + n_p$ correspondences is utilized for subsequent tasks.

2.5 Point Projection

In contrast to the measured range and intensity data which are only available on a regular grid, the location of SIFT features is determined with subpixel accuracy. Hence, an interpolation has to be carried out in order to obtain the respective 3D information as well as the respective range reliability. For this purpose, a bilinear interpolation is used. Assuming a total number of m SIFT features extracted from an image, this yields a set of samples s_i with $i = 1, \ldots, m$ which are described by a 2D location $\mathbf{x}_i$, a 3D location $\mathbf{X}_i$ and a quality measure σ_i. Compared to the original point cloud, the derived 3D points $\mathbf{X}_i$ represent a much smaller point cloud where each point is assigned a quality measure σ_i.

Extending this on two frames with m_1 and m_2 SIFT features, between which $n \leq \min\{m_1, m_2\}$ correspondences have been detected, yields additional constraints. From the set of all n correspondences, it is now possible to derive subsets of

- 2D/2D correspondences $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ which can be used for image-based techniques, e.g. using the fundamental matrix (Hartley and Zisserman, 2008),
- 3D/3D correspondences $\mathbf{X}_i \leftrightarrow \mathbf{X}'_i$ which can be used for techniques based on the 3D geometry such as the ICP algorithm (Besl and McKay, 1992) and approaches estimating a rigid or non-rigid transformation, or
- 3D/2D correspondences $\mathbf{X}_i \leftrightarrow \mathbf{x}'_i$ which can be used for hybrid techniques such as the methods presented in (Weinmann et al., 2011) and (Weinmann and Jutzi, 2011) which involve the EPnP algorithm (Moreno-Noguer et al., 2007).

The additional parameters σ_i can also be included for weighting the influence of each correspondence on any of the algorithms described above.

2.6 Point Cloud Registration

The spatial relation between two point clouds with n 3D/3D correspondences $\mathbf{X}_i \leftrightarrow \mathbf{X}'_i$ with $\mathbf{X}_i, \mathbf{X}'_i \in \mathbb{R}^3$ can formally be described as

$$\mathbf{X}'_i = \mathbf{R}\mathbf{X}_i + \mathbf{t} \tag{6}$$

where $\mathbf{R} \in \mathbb{R}^{3\times3}$ represents a rotation matrix and $\mathbf{t} \in \mathbb{R}^3$ represents a translation vector. A fully automatic estimation of the transformation parameters can be derived from minimizing the error between the point clouds. Including a weighting $w_i \in \mathbb{R}$ for each 3D/3D correspondence $\mathbf{X}_i \leftrightarrow \mathbf{X}'_i$ yields an energy function E with

$$E = \sum_{i=1}^{n} w_i \|\mathbf{X}'_i - (\mathbf{R}\mathbf{X}_i + \mathbf{t})\|^2 \tag{7}$$

for the registration process. For minimizing this energy function E, the registration is carried out by estimating the rigid transformation from all 3D/3D correspondences and the weigths are derived from a histogram-based approach. This approach is initialized by dividing the interval [0m, 1m] into $n_b = 100$ bins of equal size. For all detected correspondences, the calculated quality measures σ_i and σ'_i assigned to the 3D points $\mathbf{X}_i$ and $\mathbf{X}'_i$ are mapped to the respective bins b_j and b'_j. Points with standard deviations greater than 1 m are mapped to the last bin. The occurrence of mappings to the different bins is stored in histograms $\mathbf{h} = [h_j]_{j=1,\ldots,100}$ and $\mathbf{h}' = [h'_j]_{j=1,\ldots,100}$. Subsequently, cumulative histograms

$$\mathbf{h}_c = \left[\sum_{j=1}^{i} h_j\right]_{i=1,\ldots,100} \quad \text{and} \quad \mathbf{h}'_c = \left[\sum_{j=1}^{i} h'_j\right]_{i=1,\ldots,100}$$

are derived. The entries of the cumulative histograms reach from 0 to the number n of detected correspondences. As points with a low standard deviation are more reliable, they should be assigned a higher weight. For this reason, the inverse cumulative histograms

$$\mathbf{h}_{c,inv} = \left[n - \sum_{j=1}^{i} h_j\right]_{i=1,\ldots,100} \tag{8}$$

and

$$\mathbf{h}'_{c,inv} = \left[n - \sum_{j=1}^{i} h'_j\right]_{i=1,\ldots,100} \tag{9}$$

are calculated. Finally, the weight w_i of a 3D/3D correspondence $\mathbf{X}_i \leftrightarrow \mathbf{X}'_i$ is set to

$$w_i = \min\{\mathbf{h}_{c,inv}(\sigma_i), \mathbf{h}'_{c,inv}(\sigma'_i)\} \tag{10}$$

where σ_i and σ'_i are considered as quality measures for the respective 3D points $\mathbf{X}_i$ and $\mathbf{X}'_i$. Estimating the transformation parameters can thus be carried out for both range imaging devices separately. However, as the relative orientation between the devices is already known from a priori measurements and both devices are running synchronized, the rigid transformation can be estimated from the respective correspondences detected by both devices between successive frames. Combining information from both devices corresponds to extending the field of view and this yields more reliable results for the registration process. The extension can be expressed by transforming the projected 3D points $\mathbf{X}_i$ which are related to the respective camera coordinate frame (superscript c) into the body frame (superscript b) of the sensor platform according to

$$\mathbf{X}_i^b = \mathbf{R}_c^b \cdot \mathbf{X}_i^c + \mathbf{t}_c^b \tag{11}$$

where $\mathbf{R}_c^b$ describes the rotation and $\mathbf{t}_c^b$ denotes the translation between the respective coordinate frames. For this, it is assumed that the origin of the body frame is in the center between both range imaging devices.

2.7 Object Detection and Segmentation

As 2D SIFT features have already been calculated for the registration process, they can also be utilized for detecting special objects in the scene. This allows for calculating the coarse area of an object and for automatically selecting features which should not be included in the registration process as they arise from objects which are likely to be dynamic. These features have to be treated in a different way as the static background being relevant for registration. For this purpose, image representations of several objects have to be stored in a database before starting the surveillance application. One of these images contains a template for the object present in the scene, but from a different measurement campaign at a different place and at a different season. Due to a similar altitude, the active intensity images show a very similar appearance. Comparing the detected SIFT features of the normalized active intensity image to the object templates in the database during the flight yields a maximum similarity to the correct template. Defining a spatial transformation based on the SIFT locations as control points, the template is transformed. The respective area of the transformed template is then assumed to cover the detected object. This procedure allows for detecting both static and moving objects in the scene as well as for decoupling sensor and object motion. Hence, the presented approach for registration also remains reliable in case of dynamic environments if representative objects are already known.

3 ACTIVE MULTI-VIEW RANGE IMAGING SYSTEMS

The proposed concept focuses on airborne scene monitoring with range imaging devices. For simulating a future operational system involving such range imaging devices fairly realistically, a scaled test scenario has been set up. However, due to the large payload of several kilograms for the whole system, mounting the required components for data acquisition and data storage on an unmanned aerial vehicle (UAV) still is impracticable. Hence, in order to investigate the potentials of active multi-view range imaging systems, a cable car moving along a rope is used as sensor platform which is shown in Figure 3. The components mounted on this platform consist of

- two range imaging devices (PMD[vision] CamCube 2.0) for recording the data,
- a notebook with a solid state hard disk for efficiently storing the recorded data and
- a 12 V battery with 6.5 Ah for independent power supply.

As the relative orientation of the two range imaging devices can easily be changed, the system allows for variable multi-view options with respect to parallel, convergent or divergent data acquisition geometries.

However, due to the relatively large influence of noise effects arising from the large amount of ambient radiation in comparison to the emitted radiation as well as from multipath scattering, the utilized devices only have a limited absolute range accuracy of a few centimeters and noisy point clouds can be expected. Furthermore, due to the measurement principle of such time-of-flight cameras, the non-ambiguous range R_n with

$$R_n = \frac{1}{2}\,\frac{c}{f_m} \tag{12}$$

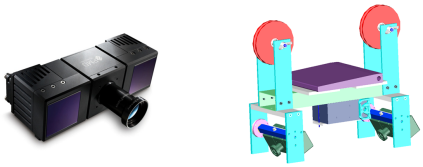

Figure 3: PMD[vision] CamCube 2.0 and model of a cable car equipped with two range imaging devices.

depends on the modulation frequency f_m, where c_0 denotes the speed of light. A modulation frequency of 20 MHz thus corresponds to a non-ambiguous range of 7.5 m. In order to overcome this range measurement restriction, image- or hardware-based unwrapping procedures have been introduced (Jutzi, 2009; Jutzi, 2012). When dealing with multiple range imaging devices, it also has to be taken into account that these may influence each other and that interferences are likely to occur. This can be overcome by choosing different modulation frequencies.

4 EXPERIMENTAL RESULTS

The estimation of the flight trajectory of a sensor platform requires the definition of a global world coordinate frame. This world coordinate frame is assumed to equal the local coordinate frame of the sensor platform at the beginning. The local coordinate frame has a fixed orientation with respect to the sensor platform. It is oriented with the X-direction in forward direction tangential to the rope, the Y-direction to the right and the Z-direction downwards. For evaluating the proposed methodology, a successive pairwise registration is performed. The threshold for the matching of 2D features is selected as $t_{des} = 0.7$. The resulting 2D/2D correspondences are projected into 3D space which yields 3D/3D correspondences. Including the weights in the estimation of the rigid transformation yields position estimates and, finally, an estimated trajectory which is shown in Figure 4 in nadir view and in Figure 5 from the side. The green and blue points describe thinned point clouds captured with both range imaging devices and transformed to the global world coordinate frame.

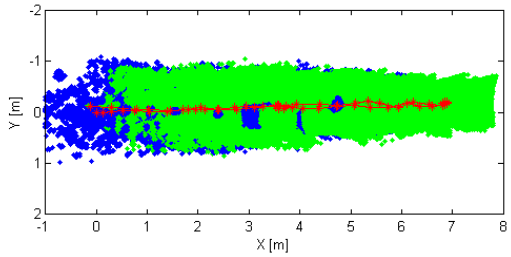

Figure 4: Projection of the estimated trajectory and thinned point cloud data onto the XY-plane.

A limitation of the experimental setup seems to be the fact that no reference values are available for checking the deviation of the position estimates from the real positions. However, due to the relative orientation of the sensor platform to the rope, the projection of the real trajectory onto the XY-plane should approximately be a straight line. Additionally, the length of the real trajectory projected onto the ground plane can be estimated from aerial images or simply be measured. Here, the distance Δ_{ground} between the projections of the end points onto the ground plane has been measured as well as the difference $\Delta_{altitude}$ between maximum and minimum altitude. From the measured values of $\Delta_{ground} = 7$ m and $\Delta_{altitude} = 1.25$ m, a total distance of

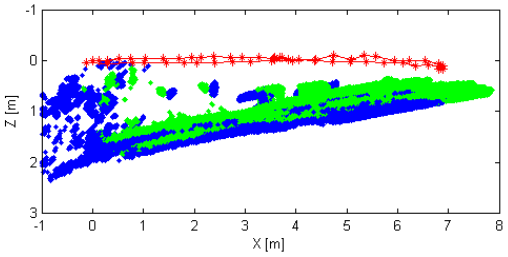

Figure 5: Projection of the estimated trajectory and thinned point cloud data onto the XZ-plane.

approximately 7.11 m can be assumed. A comparison between the start position and the point with the maximum distance on the estimated trajectory results in a distance of 6.90 m. As a consequence, the estimated trajectory can be assumed to be of relatively high quality. The results for a subsequent object detection and segmentation is illustrated for an example frame in Figure 6.

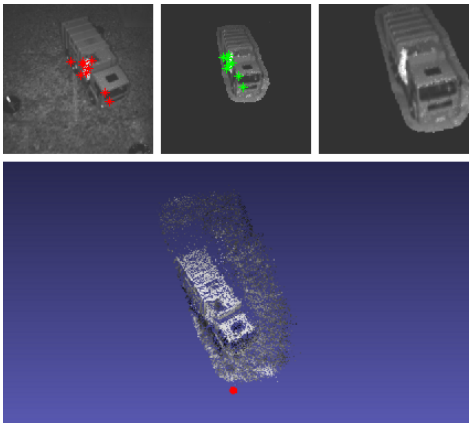

Figure 6: SIFT-based object detection and segmentation: normalized active intensity image, template and transformed template (upper row, from left to right). The corresponding point cloud for the area of the transformed template and the sensor position (red dot) are shown below.

5 DISCUSSION

The presented methodology is well-suited for dynamic environments. Instead of considering the whole point clouds, the problem of registration is reduced on sparse point clouds of physically almost identical 3D points. Due to this fact and the non-iterative processing scheme, the proposed algorithm for point cloud registration is very fast which is required for monitoring in such demanding environments. Although the current Matlab implementation is not fully optimized with respect to parallelization of tasks, a total time of approximately 1.63 s is required for preprocessing, point quality assessment, feature extraction and point projection. Further 0.46 s are required for feature matching, calculation of weights and point cloud registration. This can significantly be reduced with a GPU-implementation of SIFT, as the calculation of SIFT features already takes approximately 1.54 s.

Furthermore, the simple estimation of a rigid transformation is not sufficient, as used 3D/3D correspondences have the same weight, even if the uncertainty of the respective 3D points is very high or if outlier correspondences not fitting to the transformation have been detected. Hence, a quality measure for 3D/3D correspondences has been introduced which is based on the quality of the respective 3D points. This quality measure is used for

weighting the influence of each 3D/3D correspondence on the estimation of the rigid transformation. As most of the 3D points of a frame are assigned a higher quality, the introduced weights of 3D/3D correspondences with low quality are approximately 0. Consequently, the presented approach shows similar characteristics as a RANSAC-based approach, but it is faster and a deterministic solution for the transformation parameters is calculated.

6 CONCLUSIONS AND FUTURE WORK

In this paper, an experimental setup involving a moving sensor platform with multiple and coupled sensor devices for monitoring in low altitudes has been presented. For successive pairwise registration of the measured point clouds, a fast and reliable image-based approach has been presented which can also cope with dynamic environments. The concept is based on the extraction of distinctive 2D features from the image representation of measured intensity information and the projection into 3D space with respect to the measured range information. Detected 2D/2D correspondences between two frames, which have a high reliability, thus yield sparse 3D point clouds of 3D/3D correspondences. For increased robustness, the influence of each 3D/3D correspondence is weighted with a new measure derived from the quality of the respective 3D points. Finally, the point cloud registration is carried out by estimating the rigid transformation between two sparse point clouds which involves the calculated weights. As demonstrated, this approach can easily be extended towards using the already detected features for object detection and, even further, decoupling sensor and object motion which significantly improves the registration process in dynamic environments. The results indicate that the presented concept of active multi-view range imaging strongly supports navigation, point cloud registration and scene analysis.

The presented methodology can further be extended towards the detection, the segmentation and the recognition of multiple static or moving objects. Furthermore, a tracking method for estimating the trajectory of a moving object could be introduced as well as a model for further stabilizing the estimated trajectory of the sensor platform. Hence, active multi-view range imaging systems have a high potential for future research on dynamic scene analysis.

ACKNOWLEDGEMENTS

The authors would like to thank Michael Weinmann (Institute of Computer Science II, University of Bonn) for helpful discussions and Peter Runge (Geodetic Institute, Karlsruhe Institute of Technology) for constructing the sensor platform. Further thanks go to André Dittrich and Annette Schmitt for assistance during the measurement campaign.

REFERENCES

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

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

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.

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

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.

Böhm, 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.

Fischler, M. A. and Bolles, R. C., 1981. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* 24(6), pp. 381–395.

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.

Hartley, R. I. and Zisserman, A., 2008. *Multiple view geometry in computer vision*. University Press, Cambridge.

Jutzi, B., 2009. Investigations on ambiguity unwrapping of range images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 38 (Part 3 / W8), pp. 265–270.

Jutzi, B., 2012. Extending the range measurement capabilities of modulated range imaging devices by time-frequency multiplexing. *AVN - Allgemeine Vermessungs-Nachrichten* 2 / 2012.

Kang, Z., Li, J., Zhang, L., Zhao, Q. and Zlatanova, S., 2009. Automatic registration of terrestrial laser scanning point clouds using panoramic reflectance images. *Sensors* 9(4), pp. 2621–2646.

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.

Moreno-Noguer, F., Lepetit, V. and Fua, P., 2007. Accurate non-iterative O(n) solution to the PnP problem. *IEEE 11th International Conference on Computer Vision*, pp. 1–8.

Pathak, K., Birk, A., Vaskevicius, N. and Poppinga, J., 2010. Fast registration based on noisy planes with unknown correspondences for 3-D mapping. *IEEE Transactions on Robotics* 26(3), pp. 424–441.

Rabbani, T., Dijkman, S., van den Heuvel, F. and Vosselman, G., 2007. An integrated approach for modelling and global registration of point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing* 61(6), pp. 355–370.

Rusinkiewicz, S. and Levoy, M., 2001. Efficient variants of the ICP algorithm. *Proceedings of the Third International Conference on 3D Digital Imaging and Modeling*, pp. 145–152.

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.

Steder, B., Grisetti, G. and Burgard, W., 2010. Robust place recognition for 3D range data based on point features. *IEEE International Conference on Robotics and Automation*, pp. 1400–1405.

Wang, Z. and Brenner, C., 2008. Point based registration of terrestrial laser data using intensity and geometry features. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 37 (Part B5), pp. 583–589.

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).

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.