SKYLINE-BASED REGISTRATION OF 3D LASER SCANS

Andreas Nüchter, Stanislav Gutev, Dorit Borrmann, and Jan Elseberg

Jacobs University Bremen gGmbH
School of Engineering and Science
Campus Ring 1, 28759 Bremen, Germany

a.nuechter@jacobs-university.de
http://www.nuechti.de

Commission V/4, III/2, III/4, IV/4, IV/8

KEY WORDS: LIDAR, point cloud processing, 3D city modeling, marker-free registration, place recognition

ABSTRACT:

Acquisition and registration of terrestrial 3D laser scans is a fundamental task in mapping and modeling of cities in three dimensions. To automate this task marker-free registration methods are required. Based on the existence of skyline features this paper proposes a novel method. The skyline features are extracted from panoramic 3D scans and encoded as strings enabling the use of string matching for merging the scans. Initial results of the proposed method in the old city center of Bremen are presented.

1 INTRODUCTION

3D city modeling is the process of creating digital representations of urban areas. A wide variety of applications benefit from the availability of such digital representation, e.g., driver assistance and intelligent transportation systems, urban planning and architecture, simulation of wide load trucks, etc. All these applications require up-to-date models and therefore automation of city modeling is of increasing interest.

Automatic marker-free registration of terrestrial 3D laser scans is a fundamental scientific issue in automating 3D city modeling. The availability of such methods drastically reduces the amount of time spent in the field and during post processing. The task can be formulated as follows: Given a series of acquired 3D laser scan data, i.e., three-dimensional point clouds, find initial point or feature correspondences for every two scan pairs, that allow to compute a relative orientation between the scans. Otherwise return 'not-matchable'. Having such an initial guess, the well-known iterative closest point algorithm (ICP) can then be used for a precise alignment.

Automation of mapping calls for robotic mapping, which has attracted a lot of attention in recent years. A key technique used by robotic systems is to build a map while at the same time navigating in an unknown environment. This problem is known as simultaneous localization and mapping (SLAM) in the robotics community. One of the main components of the algorithms developed to solve the SLAM problem is the so-called loop closing, i.e., a robot has to detect, when it moves to a position close to a position, where it has been before. If the sensor data of the systems consists of 3D laser scans, then this robotic problem corresponds to the task of relating 3D scans to each other, to find overlapping 3D scans, and to compute relative orientations. This process is also know as place recognition.

The method proposed here exploits the fact, that 3D scans in urban environments always contain a skyline, i.e., a border between buildings and the sky, and additionally that urban ground is mostly flat. The beam of a laser scanner is reflected by the buildings and yields valid measurements whereas invalid or max. range measurements are returned in case the beam goes into the sky. This border is unique for every place and shows only a little variation between connected poses; a fact we are going to exploit (cf. Figure 1). With the development of our method, we aim at presenting a computationally inexpensive and simple method for automatic registration.

The remainder of this paper is structured as follows: First, we summarize the state of the art. Next, section 3 describes the processing of 3D laser scans and the feature computation. Section 4 explains the applied string matching algorithm. Results are presented in section 5 followed by the conclusions and future work.

2 STATE OF THE ART

Many state-of-the-art registration methods rely on initial pose (position and orientation) guesses, acquired through odometry of a robot, global positioning systems (GPS), or local positioning using artificial landmarks or markers as reference. As pose information is hard to acquire and in many scenarios prone to errors or not available at all, registration without initial pose estimates and place recognition are highly active fields of research. Common appearance based place recognition approaches often rely on camera data and are not suitable for laser scans (Booij et al., 2007; Cummins and Newman, 2007, 2008; Konolige et al., 2009; Valgren and Lilienthal, 2010).

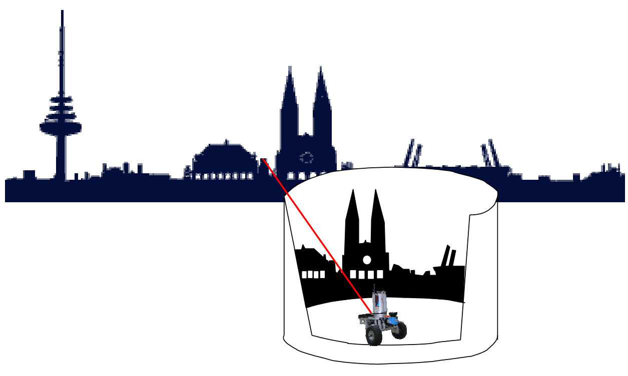

Figure 1: The skyline of Bremen projected in cylindrical coordinates is observed with a 3D laser scanner.

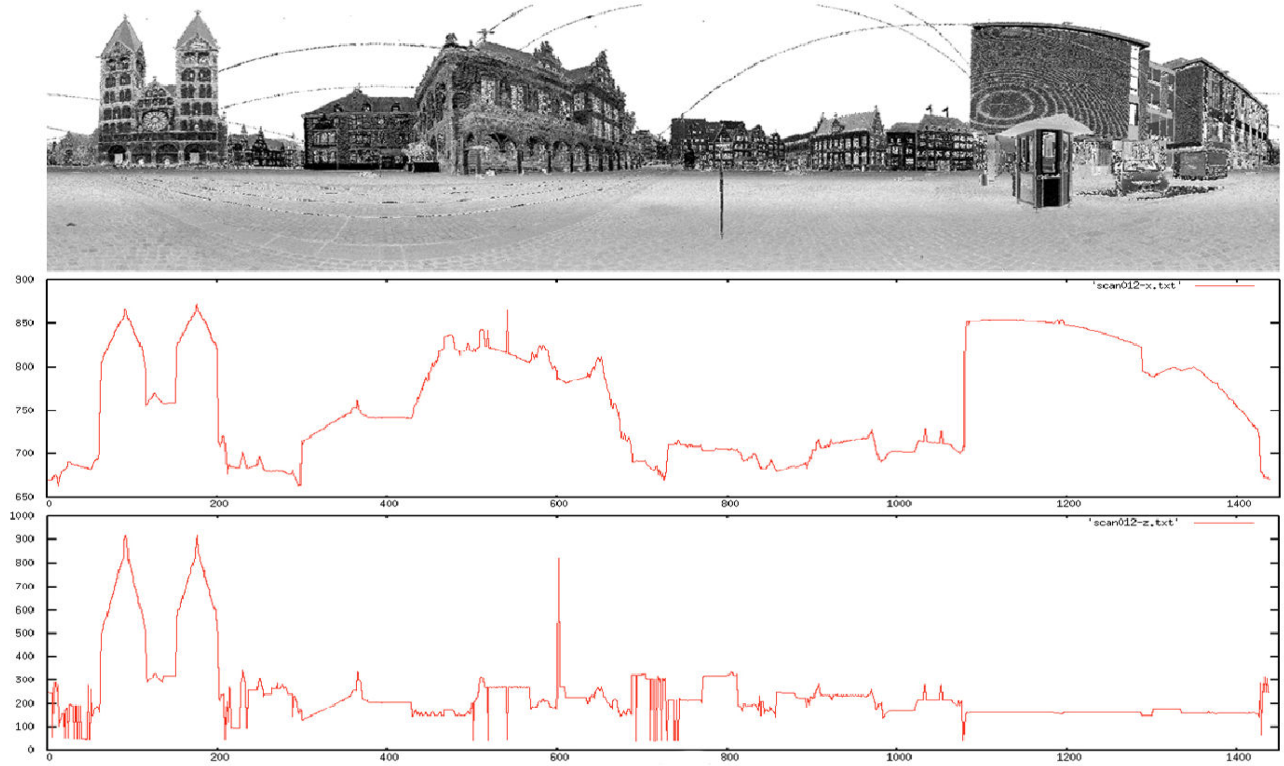

Figure 2: Range image in spherical coordinates with reflectance values, extracted skyline, and skyline in cylindrical coordinates

Aside from range values laser scanners record the intensity of the reflected light. These intensities provide additional information for the registration process. Böhm and Becker (2007) suggest to use SIFT features for automatic registration and present an example of a successful registration on a 3D scan with a small field-of-view. Wang and Brenner (2008) have extended this work by using additional geometry features to reduce the number of matching outliers in panoramic outdoor laser scans. Kang et al. (2009) proposed a similar technique for indoor and outdoor environments. Flint et al. (2007) use a key point detector called THRIFT, to detect repeated 3D structure in range data of building facades.

Other approaches purely rely on the 3D structure. Brenner et al. (2008) use 3D planar patches and the normal distribution transform (NDT) on several 2D scan slices respectively for a coarse registration. Similarily, Pathak et al. (2010) evaluate the use of planar patches and found that it is mostly usable. A solution using the NDT in 3D has been given by Magnusson et al. (2009). While this approach computes global features of the scan several groups use features that describe small regions of the scan for place recognition and registration (Huber, 2002; Steder et al., 2010; Barnea and Filin, 2008)

In addition to coarse registration, many authors use the well-known iterative closest point (ICP) algorithm for fine registration (Besl and McKay, 1992). ICP requires no computation of features. Instead, it matches raw point clouds by selecting point correspondences on the basis of smallest distances and minimizing the resulting Euclidean error. This iterative algorithm converges to a local minimum. Good start estimates improve the matching results drastically, i.e., ensures that ICP converges to a correct minimum.

Up to our knowledge, skyline features have not been employed for coarse registration or 3D city model construction. However, in the robotics and automation community the characteristics of the skyline have been investigated for several algorithms. Ramadingam et al. (2009) uses an upward pointing catadioptric, i.e., fisheye lens, camera and 3D urban models for precise localization without GPS. Stein and Medioni (1995) also solve the robot localization problem based on the horizon. Bazin et al. (2009) presents a dynamic programming approach that uses the skyline. This paper investigates on the effectiveness of a very simple dynamic programming approach, i.e., performing string matching, for coarse registration of panoramic laser scans.

3 SKYLINE STRING COMPUTATION

3.1 Range Image Computation

After the acquisition of a terrestrial 360 degree 3D laser scan, range images are computed. Common representations for these images use spherical coordinates, i.e., the horizontal image axis represents the rotation of the scanner around the yaw-axis and the vertical axis represents the mirror rotation, i.e., the pitch rotation. Either depth values or reflectance values are plotted. Figure 2 (top) shows an example of such an image with reflectance values.

The edge between buildings and the sky forms the skyline. To reliably extract it, street lights, power lines of trams or other artifacts have to be removed. This is achieved by traversing the range image on vertical image lines from bottom to the top. Figure 2 (middle) shows the extracted skyline in the range image.

The last step of our range image computation is the transformation of the image from spherical into cylindrical coordinates, i.e., we plot the scanner rotation around the yaw-axis on the horizontal axis, and the (scaled) height z on the vertical axis. Figure 2 (bottom) shows the resulting 2D curve, which is the input of all subsequent steps.

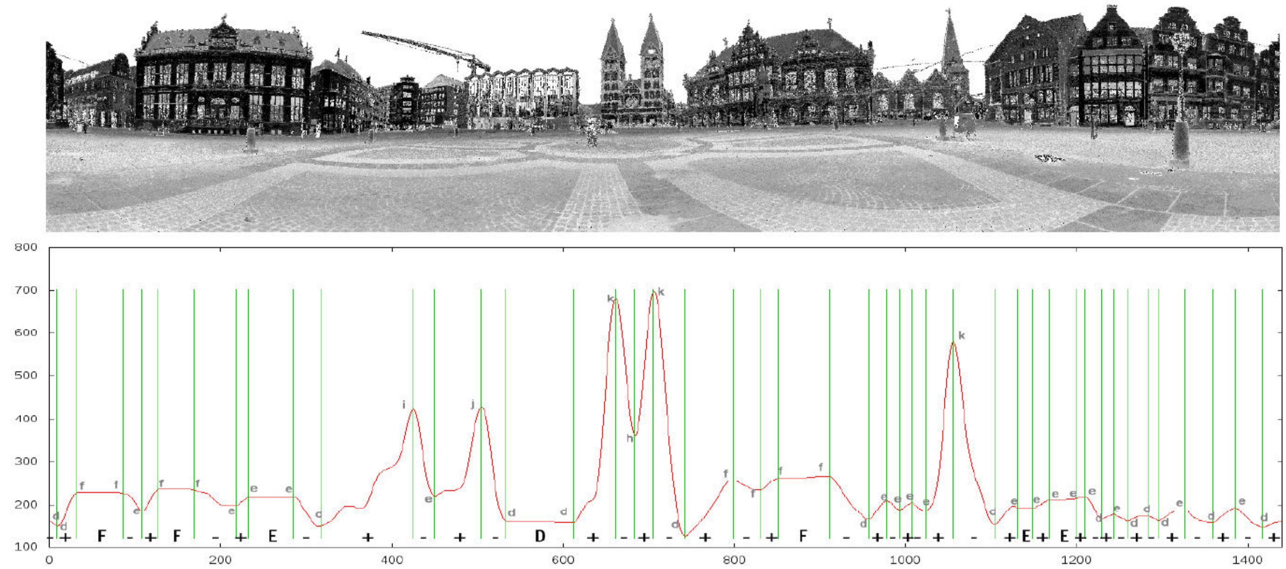

Figure 3: Reflectance image and skyline with extracted features (of Scan I). In addition to the feature descriptors as letters rising and falling edges are classified with $+$ and $-$.

3.2 Feature Detection

We aim at computing simple recognizable features from the extracted skyline. After smoothing the skyline curve with a median filter and a Gaussian kernel, we compute the derivative to extract minima and maxima as well as regions with no change of slope. Maxima correspond to pitched roofs, while flat roofs show no slope. Extrema and flat regions serve as features in our approach.

3.3 String Encoding

Features are encoded as characters. For all 3D scans we create a unique string. Local maxima and minima are labeled with lower case letters, while flat regions get upper case letters. The letter is defined by the height of the object. The step-width is approximately 10 m, i.e., we change the letter accordingly. Thus, in Figure 3 the 90 m high church towers are labelled with the character k.

4 STRING MATCHING

Given a sequence of 3D scans encoded into strings, we aim at finding overlapping 3D scans by string comparison. Our method assumes that it is possible to match 3D scans based on the encoded skyline, if they share enough common features, i.e., share many characters in theirs strings.

Given are two 3D scans as strings, i.e, sequences of characters $X = \langle x_1, x_2, \ldots, x_m \rangle$ and $Z = \langle z_1, z_2, \ldots, z_k \rangle$. Z is called a subsequence of X if there is a strictly increasing sequence of k indices $\langle i_1, i_2, \ldots i_k \rangle$, i.e., $(i_1 < i_2 < \ldots < i_k)$ such that $Z = \langle x_{i_1}, x_{i_2}, \ldots x_{i_k} \rangle$. The longest common subsequence of two strings X and $Y = \langle y_1, y_2, \ldots, y_m \rangle$ is the longest sequence Z which is both a subsequence of X and Y.

The longest common subsequence (LCS) problem is a standard problem in computer science of finding the longest subsequence common to all sequences in a set of sequences. The simple brute-force solution to the problem would be to try all possible subsequences from one string, and search for matches in the other string. As there is an exponential number of possible subsequences the runtime grows exponentially with the length of the string. However, for a fixed number of sequences the problem can be solved efficiently by dynamic programming. Therefore, we propose to compare always two encodings of scans at a time.

For all possible pairs of prefixes one computes the longest common subsequence. A prefix of a sequence is an initial string of values, e.g., the prefix $X_i = \langle x_1, x_2, \ldots, x_i \rangle$ contains the first i characters of X. Let $c[i, j]$ denote the length of the longest common subsequence of X_i and Y_j. We are interested in $c[m, n]$ since this will be the LCS of the two entire strings. The idea is to compute $c[i, j]$ assuming that we already know the values of $c[i', j']$ for $i' \leq i$ and $j' \leq j$, but not both equal. A dynamic programming algorithm that computed the maximal value of the LCS is realized by filling a table with values of c. The final LCS is extracted by using additionally stored back pointers. Details can be found in (Goeman and Clausen, 2002), where a more efficient procedure is presented as well, that uses a set of matches, chains (LCS) and so-called anti-chains.

Usually, there exists not only one unique LCS for two strings but rather the LCS forms a set. Since one can compute an estimate of the rotation and translation based on 3 matches, a RANSAC algorithm must follow the LCS computation. The matches can be selected from any element of the LCS set.

5 RESULTS AND DISCUSSION

Before we present our experiments and results we give some thoughts about the expected behavior of the matching method. In case of a pure rotation, the skyline string is ring-shifted and in the worst case the overlap is at most 50 %, due to the string representation as a simple data structure instead of a circular one. In case of a pure translation of the scanner, new letters will appear at a single point of the skyline. On the opposite side, letters will disappear. The effects to the string associated with the skyline is as follows: There exists a location in the string, where new letters appear and where existing letters are pushed to *both* sides. In addition, there exists one location where letters disappear and these locations are separated by 180 deg., i.e., approximately by half of the string length. Furthermore, the ring property holds, i.e., letters that are pushed out of the string array at the start or the end,

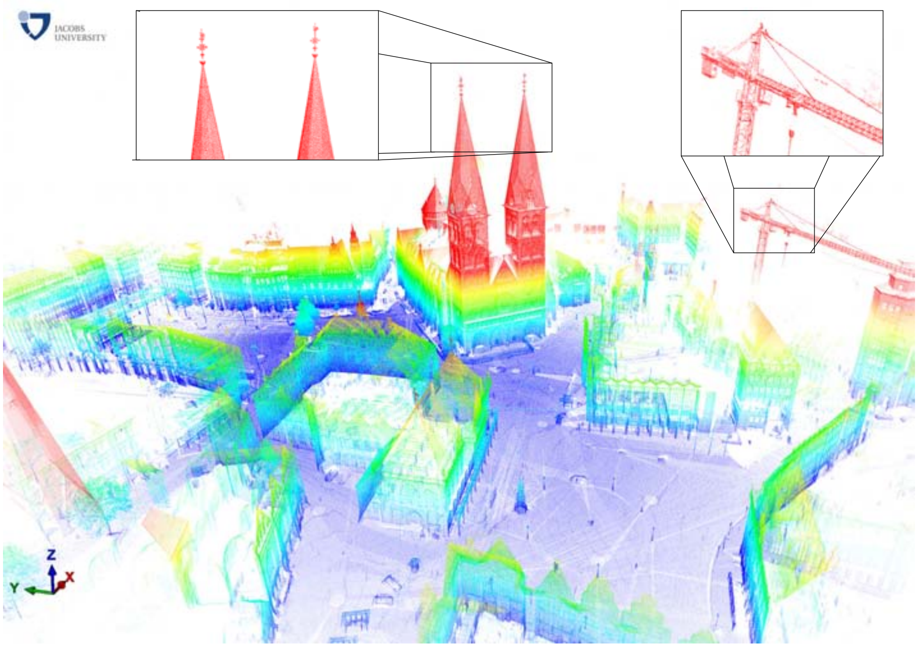

Figure 4: 3D view of the reconstructed scene using the scans in Figure 5.

Figure 5: The market place in Bremen downtown (Scan 1, 2, and 5). Between the top and middle 3D scan the scanner was moved approximately 10 m, while the distance of the scan positions of the first and third scan was roughly 50 m.

Table 1: Results for the string matching. Colored in red is one possible matching. Top: Matches of the strings extracted from the skylines of 3D scans presented in Figure 5. Bottom: Matching result between the scan from Figure 3 and of the first scan of Figure 5.

Scan 1 Scan 2	fDdedkdEededededFfefefeEedheiDdckikdee FfDdedkddeeeddDdeDdedFfeFfeFfdEeiffffjDdkikdfefeF
Matching	fDdedkdedddededFfefefEeiDdkikdee
Scan 1 Scan 5	fDdedkdEededededFfefefeEedheiDdckikdee cFfefeFfdjdgeEefEegdfefFfcEebiDdDdeekhkDd
Matching	feeddeeedfeffEedDdkkd

Scan I Scan 1	ddFfeFfeEediejDdkhkdffFfdeeeekdEeEeededddeded fFfdddefdeeeDdEeedFfeFfFfDdEeeidjDdckikDdfefe
Matching	fFfdddfdeeedEeeededdee

appear on the other side. In this case, the overlap depends on the translation distance between the two scans. The last case concerns the combination of a rotation and translation. Again there exists a point in the string where new letters appear and one where the letters disappear. These locations are now unrelated, i.e., they depend on the applied sensor rotation and translation and are located somewhere in the string. Figure 6 illustrates these three cases.

Experiments for the proposed method have been carried out in the pedestrian area of Bremen downtown. The most common buildings in this area are in Renaissance style. A 3D view of the data set is given in Figure 4. Figure 5 presents three 3D scans of the market place. Between the different scans the scanner was only translated. In Figure 7 we show the matches between two scans. Visual inspection manyfests that nearly all of the found matches are correct. This can easily be verified by applying a RANSAC procedure Fischler and Bolles (1981). The number of correct matches with LCS suffices for computing an initial starting guess for fine registration with ICP. Table 1 (top) presents details of the string matching.

Additionally, we tried to match scans with our proposed method that originated from two different scanning campaigns, that was carried out with of roughly half a year in between. For example, the skyline from the first scanning campaign that is given in Figure 3 is matchable to the scans from the second campaign given in Figure 5. The results show that the skyline is an environment feature that remains mostly stable for long time periods with larger changes appearing only slowly over time. In our case, the crane configuration and the scaffold around the cathedral are the only prominent changes between the two campaigns. For the skyline the effects are minor. Table 1 (bottom) presents details of the string matching. The scans from the first campaign is labelled with the Roman I, while the scan from the second campaign is labelled 1.

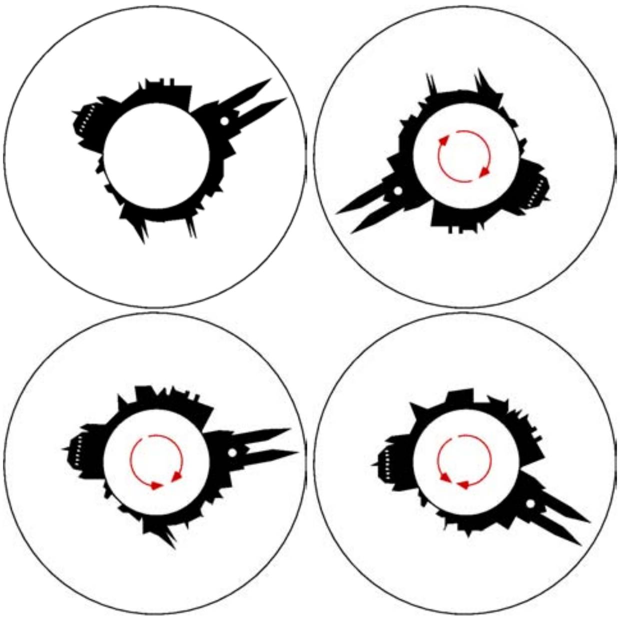

Figure 6: The cylinder presented in Figure 1 has been unfolded. From left to right: (1) sketch of the scene and its skyline. (2) pure rotation. (3) pure translation. (4) rotation and translation.

6 CONCLUSIONS AND FUTURE WORK

Beginning in the early days of photography, the skyline has become a fascinating unique identifier of a city. This work has aimed at bringing the skyline to the attention of our research community by exploring its capabilities to serve as a unique feature for coarse registration. We exploit the skyline, i.e., the edge between buildings and the sky as a feature and code this feature in a simple string representation. The data association problem is solved by simple string matching. The determined correspondences are the input for a novel coarse registration method for terrestrial 3D laser scans acquired in urban environments.

Needless to say, a lot of work remains to be done. Next, we will integrate the presented approach into our open-source registration toolkit 6D SLAM (`http://slam6d.sourceforce.net`) and investigate the robustness of the method. In addition, we will focus on approximate string matching method for replacing the LCS algorithm to handle cases, where the scanner is not exactly levelled or its height changes. Furthermore, we plan to use this computationally inexpensive algorithm in robotic mapping applications, where a large number of 3D scans must be registered, acquired for instance by the Velodyne laser scanner at the high frame rate of 10 Hz.

ACKNOWLEDGEMENTS

The data for the experiments was collected using a RIEGL laser scanner. We acknowledge Andreas Birk for fruitful discussions about feature based scan matching methods.

REFERENCES

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

Bazin, J.-C., Kweon, C., Demonceaux, C. and Vasseur, P., 2009. Dynamic programming and skyline extraction in catadioptric infrared images. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '09).

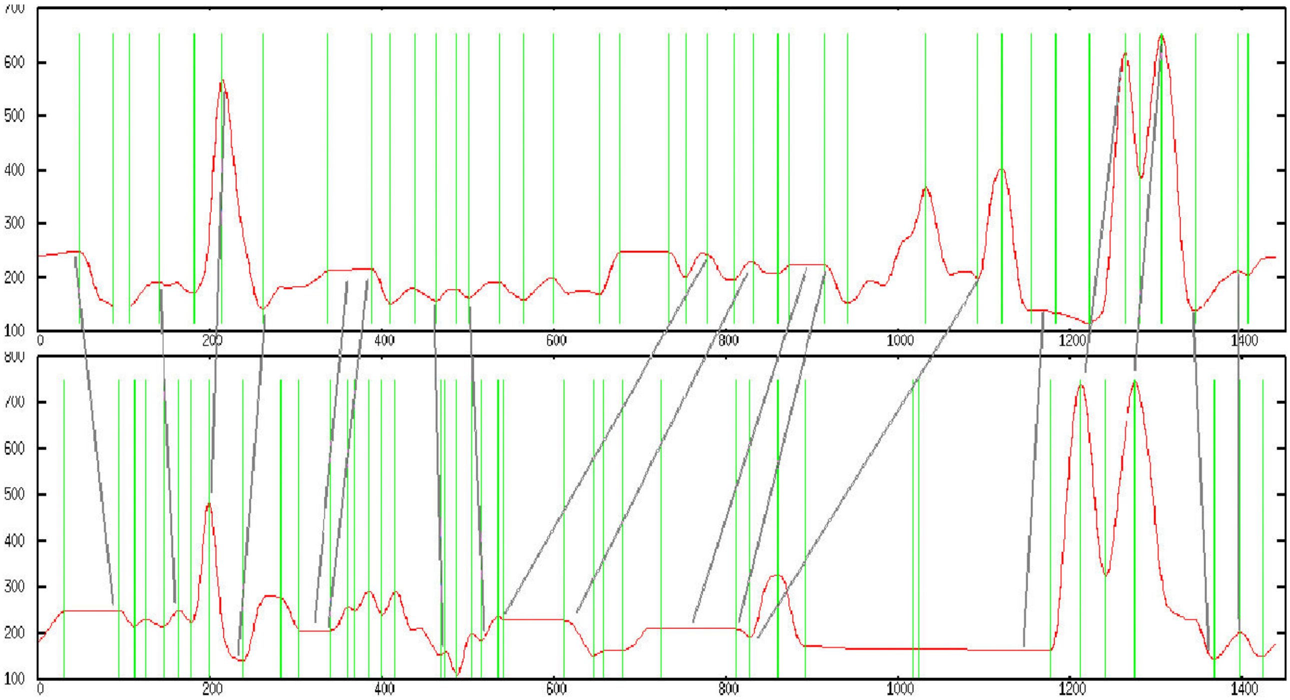

Figure 7: Successful feature matching using LCS.

Besl, P. and McKay, N., 1992. A Method for Registration of 3–D Shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 14(2), pp. 239 – 256.

Böhm, J. and Becker, S., 2007. Automatic marker-free registration of terrestrial laser scans using reflectance features. In: Proceedings of 8th Conference on Optical 3D Measurement Techniques, Zurich, Switzerland, pp. 338–344.

Booij, O., Terwijn, B., Z. and Krose, B., 2007. Navigation using an appearance based topological map. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '07), Rome, Italy, pp. 3927–3932.

Brenner, C., Dold, C. and Ripperda, N., 2008. Coarse orientation of terrestrial laser scans in urban environments. ISPRS Journal of Photogrammetry & Remote Sensing 63(1), pp. 4–18.

Cummins, M. and Newman, P., 2007. Probabilistic appearance based navigation and loop closing. In: In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '07), Rome, Italy, pp. 1828–1833.

Cummins, M. and Newman, P., 2008. Accelerated appearance-only slam. In: In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '08), Pasadena, CA, USA, pp. 1828–1833.

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.

Flint, A., Dick, A. and van den Hengel, A. J., 2007. Thrift: Local 3d structure recognition. In: Proceedings of the 9th Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications (DICTA '07).

Goeman, H. and Clausen, M., 2002. A new practical linear space algorithm for the longest common subsequence problem. Kybernetika 38(1), pp. 45–66.

Huber, D., 2002. Automatic Three-dimensional Modeling from Reality. PhD thesis, Carnegie Mellon University.

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 (4), pp. 2621–2646.

Konolige, K., Bowman, J., Chen, J. D., Mihelich, P., Calonder, M., Lepetit, V. and Fua, P., 2009. View-based maps. In: Robotics: Science and Systems (RSS '09), Seattle, USA.

Magnusson, M., Andreasson, H., Nüchter, A. and Lilienthal, A. J., 2009. Automatic appearance-based loop detection from 3d laser data using the normal distributions transform. Journal of Field Robotics (JFR), Special Issue on Three-Dimensional Mapping 26(11–12), pp. 892–914.

Pathak, K., Borrmann, D., Elseberg, J., Vaskevicius, N., Birk, A. and Nüchter., A., 2010. Evaluation of the robustness of planar-patches based 3d-registration using marker-based ground-truth in an outdoor urban scenario. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '10), Taipei, Taiwan.

Ramadingam, S., Bouaziz, S., Sturm, P. and Brand, M., 2009. SKYLINE2GPS: Localization in Urban Canyons using Omni-Skylines. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '10).

Steder, B., Grisetti, G. and Burgard, W., 2010. Robust place recognition for 3D range data based on point features. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '10).

Stein, F. and Medioni, G., 1995. Map-based localizatin using the panoramic horizon. IEEE Transaction on Robotics and Automation.

Valgren, C. and Lilienthal, A. J., 2010. Sift, surf & seasons: Appearance-based long-term localization in outdoor environments. Journal Robotics and Autonomous Systems (JRAS) 58(2), pp. 157–165.

Wang, Z. and Brenner, C., 2008. Point based registration of terrestrial laser data using intensity and geometry features. In: ISPRS Congress ('08).