Robust Alignment of Wide Baseline Terrestrial Laser Scans via 3D Viewpoint Normalization*

Yanpeng Cao[1], Michael Ying Yang[2], John McDonald[1]
[1]Department of Computer Science, National University of Ireland, Maynooth, Ireland
[2]Department of Photogrammetry, University of Bonn, Bonn, Germany
ycao@cs.nuim.ie, michaelyangying@uni-bonn.de, johnmcd@cs.nuim.ie

Abstract

The complexity of natural scenes and the amount of information acquired by terrestrial laser scanners turn the registration among scans into a complex problem. This problem becomes even more challenging when two individual scans captured at significantly changed viewpoints (wide baseline). Since laser-scanning instruments nowadays are often equipped with an additional image sensor, it stands to reason making use of the image content to improve the registration process of 3D scanning data. In this paper, we present a novel improvement to the existing feature techniques to enable automatic alignment between two widely separated 3D scans. The key idea consists of extracting dominant planar structures from 3D point clouds and then utilizing the recovered 3D geometry to improve the performance of 2D image feature extraction and matching. The resulting features are very discriminative and robust to perspective distortions and viewpoint changes due to exploiting the underlying 3D structure. Using this novel viewpoint invariant feature, the corresponding 3D points are automatically linked in terms of wide baseline image matching. Initial experiments with real data demonstrate the potential of the proposed method for the challenging wide baseline 3D scanning data alignment tasks.

1. Introduction

Terrestrial laser scanners are frequently used for the collection of highly detailed 3D urban modelling [3], [4]. In most cases, several scanning data at different viewpoints are needed to obtain full scene coverage, and therefore requires registration of the individual scans into one global reference frame. For the registration, the common practice involves the manual deployment of artificial targets, which are easily distinguishable in the scene as tie objects. Since most modern laser scanners are accompanied with digital cameras, the registration of terrestrial laser scans can be aided by using the images that are simultaneously captured with the scanning data. Given images directly linked to the 3D point clouds, the focus of this paper is to automatically align two individual laser scans obtained at very different viewpoints in terms of the matching of their associated 2D image appearances. Previously a number of successful techniques [1], [23], [13], [5], [14] have been proposed for robust 2D image matching - a comprehensive review was given in [16]. However the performances of these techniques are limited in that they only consider the 2D image texture and ignore important cues related to the 3D geometry. These methods cannot produce reliable matching results of features extracted on wide baseline image pairs. In this paper, we employ a method, which integrates recent advances in 2D feature extraction with the concept of 3D viewpoint normalization, to improve the descriptive ability of local features for robust matching over largely separated views [25].

Our goal is to build a framework to automatically align two widely separated 3D scenes captured by laser scanner. The framework includes two major steps: (1) to connect the images captured by a hand-held camera to the 3D laser data; (2) to establish robust matches between two groups of widely separated images. For the first step, we obtain corresponding feature point between the camera images and the laser provided image, thus we can connect the image pixels to the 3D points (pixel-to-point correspondences). Since the photos are captured from similar viewpoints of the laser scanner, the standard SIFT matching is suitable for this task. For the second step, we present a novel scheme to establish robust feature correspondences between two widely separated 3D scenes based on the concept of 3D viewpoint normalization. A number of dominant planes in the 3D point cloud are extracted to represent the spatial layout of the environment. The 2D image features can be normalized with respect to these recovered planes to achieve viewpoints invariance. The individual patches on the original image, each corresponding to an identified 3D planar region, are rec-

[12]*The first two authors contributed equally to this paper.

978-1-4244-9497-2/10/$26.00 ©2010 IEEE

tified to form the front-parallel views of building facades. Viewpoint invariant features are then extracted on these rectified views to provide a basis for further matching. Knowing how everything looks like from a front-parallel view, it becomes easier to recognize the same surface from different viewpoints. The resulting features are very robust to the perspective distortions caused by large viewpoint changes, thus they are very suitable for wide baseline image matching. Also the viewpoint invariant features contain enough information to completely define a point-to-point mapping relation given a single correspondence. It leads to a much more efficient RANSAC-based matching scheme. Compared with some previous approaches on combining 2D feature with 3D geometry [24], [10], our method extracted a number of dominant 3D planes to represent the 3D layout of an urban setting. The resulting piece-wise planar 3D model offers more robustness to the errors occurred in 3D reconstruction. Moreover, feature extraction can be performed with respect to the extracted planes in a single pass to achieve better efficiency.

The remainder of the paper is organized as follows. Section 2 reviews some existing solutions for 3D model alignment and robust feature matching. In Section 3, we present the preprocessing step where the hand-held camera photos are linked to the 3D Laser scans. In Section 4, we explain the procedures of 3D viewpoint normalization and propose an effective scheme to match the resulting viewpoint invariant features. In Section 5, the performance of the proposed method is comprehensively evaluated. Finally, the conclusion is given in Section 6.

2. Related Work

Terrestrial laser scanning has been proven effective in 3D urban reconstruction of architectural details and building facades. However, one of the biggest problems encountered while processing the scans is 3D point cloud registration. Given two sets of 3D points captured at different viewpoints, the task to obtain tie points and to estimate an optimal transformation between them. Commercial software typically requires users to manually deploy artificial targets in the scene as corresponding points. Until present, several matching algorithms have been proposed to avoid the manual intervention. The most popular class of methods is the Iterative closest point (ICP) based techniques [2], [28], [19]. They compute the alignment transformation by iteratively minimizing the sum of distances between closest points. However, the performances of ICP-based methods rely on a good estimation initialization and require good spatial configuration of 3D points. Since laser-scanning instruments nowadays are often equipped with an additional image sensor, many researchers proposed to enhance the performances of 3D point cloud alignment by referring to their associated 2D images. In [21], an effective method is presented for automatic 3D model alignment via 2D image matching. [12] present a general framework to align 3D points from SfM with range data. Images are linked to the 3D model to produce common points between range data. Ikeuchi [9] presents an automated 3D range to 3D range registration method that relies on the matching of reflectance range image and camera image. In [8], a flexible approach was presented for the automatic co-registration of terrestrial laser scanners and digital cameras by matching the camera images against the range image. These techniques work well for frames with small observation changes (e.g continuous videos). To produce satisfactory registration results of 3D points clouds captured at significantly changed viewpoints, we need an effective image feature scheme which is capable of establishing robust correspondences between wide baseline image pairs.

A large number of papers have reported on robust 2D image feature extraction and matching, cf. [16] for a detailed review. The underlying principle for achieving invariance is to normalize the extracted regions of interest so that the appearances of a region will produce the same descriptors (in an ideal situation) under the changes of illumination, scale, rotation, and viewpoint. Among them the Scale-invariant feature transform (SIFT) [13] is the best scale-invariant feature scheme and the Maximally Stable Extremal Regions (MSER) [5] shows superior affine invariance. In [15], the authors conducted a comprehensive evaluation of various feature descriptors and concluded that the 128-element SIFT descriptor outperforms other descriptor schemes. Robust 2D feature extraction techniques have been successfully applied to various computer vision tasks such as object recognition, 3D modelling, and pose estimation. However, the existing schemes cannot produce satisfactory feature matching over largely separated views because perspective effects will add severe distortions to the resulting descriptors. Recently, many researchers have considered the use of 3D geometry as an additional cue to improve 2D feature detection. A novel feature detection scheme, Viewpoint Invariant Patches (VIP), based on 3D normalized patches was proposed for 3D model matching and querying [24]. In [22], the authors developed a physical scale space for detecting keypoints, which extends a 2D image-based detection and description framework to 3D using an image back-projected onto a range scan. In [10], both texture and depth information were exploited for computing a normal view onto the surface. In this way they kept the descriptiveness of similarity invariant features (e.g. SIFT) while achieving extra invariance against perspective distortions. However these methods directly make use of the preliminary 3D point clouds from SfM. Viewpoint normalization with respect to the local computed tangent planes are prone to errors occurred in the process of 3D reconstruction. For predominantly planar scenes (urban environment),

a piece-wise planar 3D model is more robust, compact, and efficient for viewpoint normalization of cameras with wide baselines.

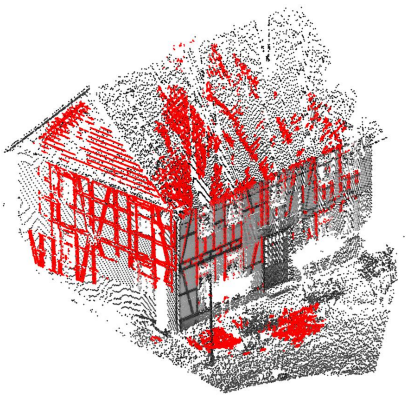
3. Preprocessing

We plan to make use of the image contents simultaneously acquired with the 3D scanning data to enable autonomous terrestrial laser registration. The captured image is directly linked to the 3D point cloud if the camera is directly integrated to the laser scanner. However, fixing the relative position between the 3D range and 2D image sensors has two major limitations. First, the rigid setting sacrifices the flexibility of 2D image capture. The acquisition of the images and range scans has to occur at the same time and from the same viewpoint. This will cause constraints on the deployment of the camera. A good position for 3D range scanning might not be suitable for 2D image capturing. Second, the images may need to be captured at different times, particularly if there were poor lighting conditions at the time that the range scans were acquired. To overcome above drawbacks, we applied the technique described in [12] to link a group of independently captured images to the 3D range data.

First, a 3D laser scan is acquired and a sequence of 2D images is independently gathered using a hand-held camera from various positions that do not necessarily coincide with the viewpoint of the range scanner. A point cloud can be reconstructed from these multiple-view images by using the structure-from-motion (SfM) algorithm [18]. Then, the SIFT features are extracted on the camera images and the image captured by the laser for correspondence matching. A number of putative matches are found using local appearance descriptors, and then the RANSAC algorithm is used to eliminate false correspondences by imposing a plane-to-plane mapping homography. Since the difference of viewpoints between camera and laser is not significant, the standard SIFT technique is capable of producing robust image matches as tie points between the 3D SfM point cloud and the 3D laser scan. Finally, we compute the transformation that aligns the 3D models gathered via range sensing and computed via structure from motion, thus the complete set of 2D images is automatically linked to the 3D point cloud. More abundant appearance information will be associated with the 3D ranging data after the preprocessing step. Fig. 1 shows a result of such preprocessing.

4. 3D viewpoint invariant features

In this step, we apply an effective method to extract a number of dominant 3D planes in the 3D points, as described in [26]. The RANSAC algorithm [6] is applied for plane hypothesis generation and minimum description length (MDL) principle [17] is used to evaluate several competing hypothesis. The method can avoid detecting wrong planes due to the complex geometry of the 3D data. The detected 3D planes will be used to represent the spatial layout of the environments. The 2D image features are normalized with respect to these recovered planes to achieve viewpoints invariance. Viewpoint invariant features are extracted on the rectified front-parallel views to provide a basis for further matching.



(a) SfM point cloud mapping on laser scan.

(b) Corresponding five images.

Figure 1. SfM point cloud mapping on laser scan. The red colour points refer to 3D SfM from images. The density of laser scan is 1/100 of original data by resampling.

4.1. Dominate planes extraction

MDL is applied for plane extraction, similar to the approach of [17]. Given a set of points, we assume several competing hypothesis, here namely, outliers (O), 1 plane and outliers (1P+O), 2 planes and outliers (2P+O), 3 planes and outliers (3P+O), 4 planes and outliers (4P+O), 5 planes and outliers (5P+O), ect..

Let n_0 points x_i, y_i, z_i be given in a 3D coordinate and the coordinates be given up to a resolution of ϵ and be within range R. The description length for the n_0 points, when assuming outliers (O), therefore is

$$\#bits(points \mid O) = n_0 \cdot (3lb(R/\epsilon))$$

where $lb(R/\epsilon)$ bits are necessary to describe one coordinate.

If we now assume n_1 points to sit on a plane, n_2 points to sit on the second plane, and the other $\bar{n} = n_0 - n_1 - n_2$

points to be outliers, we need

$$\#bits(points \mid 2P+O) = n_0 + \bar{n} \cdot 3lb(R/\epsilon) + 6lb(R/\epsilon) + n_1 \cdot 2lb(R/\epsilon) + n_2 \cdot 2lb(R/\epsilon) + \left[\sum_{i=1}^{n_1+n_2} \left\{ \frac{1}{2ln2} \cdot (\boldsymbol{v}_i)^T \Sigma^{-1} (\boldsymbol{v}_i) + \frac{1}{2} lb(|\Sigma| / \epsilon^6) + \frac{k}{2} lb2\pi \right\} \right]$$

where the first term represents the n_0 bits for specifying whether a point is good or bad, the second term is the number of bits to describe the bad points, the third term is the number of bits to describe the parameters of two planes, which is the number of bits to describe the model complexity, a variation of [20]. We assumed the n_1 good points to randomly sit on one plane which leads to the fourth term, and the n_2 good points to randomly sit on the other plane which leads to the fifth term, and to have Gaussian distribution $\boldsymbol{x} \sim N(\boldsymbol{\mu}, \Sigma)$ which leads to the sixth term.

$\#bits(points \mid 1P+O)$, $\#bits(points \mid 3P+O)$, $\#bits(points \mid 4P+O)$, and $\#bits(points \mid 5P+O)$, and so on, can be deducted in a similar way.

Incremental RANSAC is applied to extract planes in the point cloud. The MDL principle, deducted above, for interpreting a set of points in 3D space,is employed to decide which hypothesis is the best one. This method of integrating RANSAC and MDL has been shown to avoid detecting wrong planes [26]. One example demonstrating dominant plane extraction is shown in Fig. 3.

4.2. Viewpoint invariant feature generation

In this step, we perform normalization with respect to the extracted dominant 3D planes to achieve viewpoint invariance. Given a perspective image of a world plane, the goal is to generate the front-parallel view of the plane. This is equivalent to obtaining the image of a world plane where the camera viewing direction is parallel to the plane normal. It's well known that the mapping between a 3D world plane and its perspective image is defined as a 3×3 homography. Since we know the 3D positions of the points on the building facade and their corresponding image coordinates, we can compute the homography relating the facade plane to its image given at least four correspondences. The computed homography H enables us to warp the original image to a normalized front-parallel view where the perspective distortion is removed. Fig. 2 shows some results of such viewpoint normalization.

Within the normalized front-parallel views of the scene, the viewpoint invariant features are computed in the same manner as the SIFT scheme [13]. Given a number of extracted dominant 3D planes, features extraction can be efficiently performed in a single pass *w.r.t.* the planes. Potential keypoints are identified by scanning local extreme in a series of Difference-of-Gaussian (DoG) images. For each

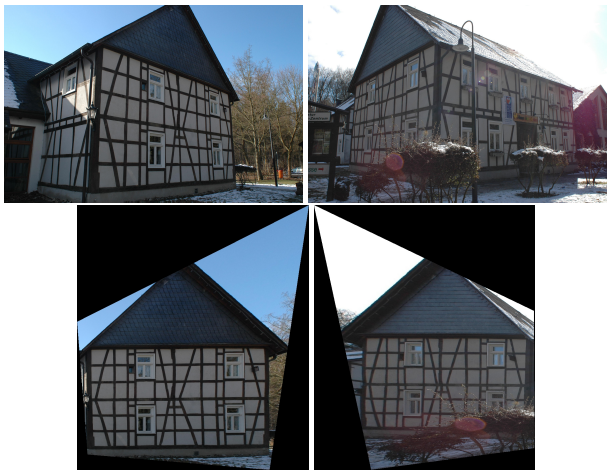


Figure 2. Some examples of viewpoint normalization. *Top*: Original images; *Bottom*: Normalized front views. Note the perspective distortions are largely reduced in the warped front-parallel views of the building walls (e.g. a rectangular window in the 3D world will also appear rectangular in the normalized images)

detected keypoint, appropriate scale and orientation are assigned to it and a 128-element SIFT descriptor is created based upon image gradients of its local neighbourhood. A complete viewpoint invariant feature consists of the following components: (1) $\boldsymbol{X}$ is its 3D position in the space; (2) $\boldsymbol{x}$ is its 2D coordinates in the normalized front-parallel view; (3) s is its corresponding spatial patch scale; (4) θ is the dominant gradient orientation of the normalized patch; (5) $\boldsymbol{f}$ is the 128-element descriptor; and (6) $\boldsymbol{n}$ is the normal of the plane it belongs to.

4.3. Viewpoint invariant feature matching

In [13] a pair of SIFT features are considered matched if the ratio between distances to the closest match and to the second closest is below some predefined threshold. The ratio check scheme is justified because the correct match for a discriminative keypiont is often significantly better (closer in the descriptor space) than the incorrect ones [13]. However, in urban environments where many repetitive structures (e.g. windows) exist, this criterion will falsely reject many correct matches since a feature cannot find a unique distinctive match. We applied the criterion described in [27] to generate the putative correspondences. We consider two features matched if the cosine of the angle between their associated descriptors $\boldsymbol{f}_i$ and $\boldsymbol{f}_j$ is above some threshold δ as:

$$\cos(\boldsymbol{f}_i, \boldsymbol{f}_j) = \frac{\boldsymbol{f}_i \cdot \boldsymbol{f}_j}{\|\boldsymbol{f}_i\|_2 \left\|\boldsymbol{f}_j\right\|_2} > \delta \qquad (1)$$

where $\|\cdot\|_2$ represents the $L2$-norm of a vector. In case multiple matches pass the criteria, we keep the top 5 correspondences for further RANSAC matching. This criterion establishes matches between features having similar descrip-

Outlier ratio	40%	50%	60%	70%	80%
Ours (1 p)	4	5	6	9	14
H-matrix (4 p)	22	47	116	369	1871
F-matrix (7 p)	106	382	1827	13696	234041

Table 1. The theoretical number of samples (M) required for RANSAC to ensure 95% confidence (ρ) that one outlier free sample is obtained for estimation of geometrical constraint. The actual required number is around an order of magnitude more.

tors and does not falsely reject potential correspondences extracted on the images of repetitive structures which are common in urban environments.

After obtaining a set of putative feature matches based on the matching of local descriptors, we impose certain global geometric constraints to identify the true correspondences. The RANSAC technique [6] is applied for this task. The number of samples M required to guarantee a confidence ρ that at least one sample is outlier free is given in Table 1. When the fraction of outliers is significant and the geometric model is complex, RANSAC needs a large number of samples and becomes prohibitively expansive.

The geometrical model can be greatly simplified via the use of these novel features, and thus, lead to a more efficient matching method. For the conventional 2D feature techniques, only the 2D image coordinates of extracted features can be used to generate geometric constraints (F-Matrix or H-Matrix). Therefore, a number of feature matches are required to compute F-Matrix (7 correspondences) or H-matrix (4 correspondences). In comparison, the viewpoint invariant features contain enough information to completely define a point-to-point mapping relation given a single feature correspondence. Consider a pair of matched features $(\boldsymbol{x}_1^m, s_1^m, \theta_1^m)$ and $(\boldsymbol{x}_2^n, s_2^n, \theta_2^n)$ both extracted on the normalized front-parallel views, a 2D similarity translation hypothesis is generated as follows:

$$\begin{bmatrix} x_1 - x_1^m \\ y_1 - y_1^m \\ 1 \end{bmatrix} = \begin{bmatrix} \Delta s & 0 & 0 \\ 0 & \Delta s & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\Delta\theta & -\sin\Delta\theta & 0 \\ \sin\Delta\theta & \cos\Delta\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_2 - x_2^m \\ y_2 - y_2^m \\ 1 \end{bmatrix} \quad (2)$$

where $\Delta s = s_1^m / s_2^n$ is the scale ratio and $\Delta\theta = \theta_1^m - \theta_2^n$ is the orientation difference. Using this simple geometric model, a much smaller number of samples are needed to guarantee the generation of the correct hypothesis (c.f. Table 1 for comparison). Moreover, using the viewpoint invariant features the RANSAC can successfully return the true correspondences from a putative feature set of high outlier percentage. This is particularly advantageous for image matching in urban environments. The man-made buildings usually contain lots of respective structures (e.g. windows, doors, bricks). Setting a strict matching criteria (ratio check) will falsely reject the true correspondences. Using the viewpoint invariant features, we can set a relatively loose criteria (Eq. 1) to establish a large number of putative matches (lots of outliers contained) and then apply the one-point RANSAC algorithm to identify correct ones. This can't be achieved using the standard SIFT features.

5. Experimental Results

In this section, we conducted experiments to evaluate the performance of the proposed viewpoint invariant features and demonstrated their applications for automatic alignment of wide baseline terrestrial Laser scans, with focus in the urban environments.

5.1. Data generation and preprocessing

We have taken two groups of laser scanning data using Leica HDS6000. Each group contains two individual 3D point clouds of a same building captured at largely separated viewpoints. The laser-equipped camera simultaneously captured an intensity image which provides pixel-to-point correspondences to the 3D point cloud. For each laser scan, we also took 5 images using a hand-held camera at similar viewpoints. We applied the orientation software AURELO [11] to achieve full automatic relative orientation of these multi-view images. And we used the public domain software PMVS (patch-based multi-view stereo) [7] for deriving a dense point cloud for each view of image pairs. It provides a set of 3D points with normals at those positions where there is enough texture in the images. Then we applied the standard SIFT matching scheme to establish correspondences between the camera images and the laser provided image. The pixel-to-pixel tie points allow us to link the 3D point cloud from SfM to the laser scanning (cf. an example shown in Fig. 1). Finally, a number of dominant planes were extracted from each point cloud, while the rest 3D points were removed. Fig. 3 shows an example of the captured laser scan and the extracted dominant 3D planes.

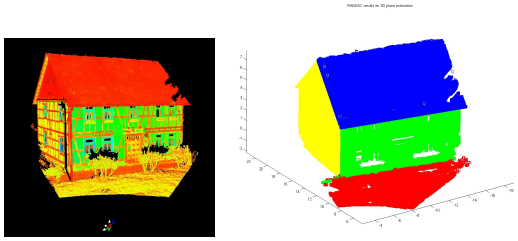


Figure 3. *Left*: A snap-shot image of 3D point cloud taken by laser scanner. *Right*: The four dominant planes automatically extracted from the point cloud.

5.2. Performance evaluations

Given the extracted dominant planes, we perform normalization w.r.t. these planes to achieve viewpoint invariance. After viewpoint normalization, corresponding scene elements will have more similar appearances. The resulting features will suffer less from the perspective distortions and show better descriptiveness. We tested our method on two wide baseline 3D point clouds, as shown in Fig. 4, to demonstrate such improvements. It's noted that both 3D point clouds covered a same dominant planar structure which can be easily related through a homography. A number of SIFT and viewpoint invariant features were extracted on the original images and on the normalized front-parallel views, respectively. Then we followed the method described in [15] to define a set of ground truth matches. The extracted features in the first image were projected onto the second one using the homography relating the images (we manually selected 4 well conditioned correspondences to calculate the homography). A pair of features is considered matched if the overlap error of their corresponding regions is minimal and less than a threshold [15]. We adjusted the threshold value to vary the number of resulting feature correspondences.

Figure 4. Two 3D point clouds and their associated images captured at widely separated views.

Our goal is to demonstrate that the performances of 2D image feature matching can be significantly improved by taking into account the underlying 3D geometry. We quantitatively measured how well two actually matched features relate with each other in terms of the Euclidean distance between their corresponding descriptors, their scale ratio, and their orientation difference. Given a number of matched features, we calculated the average Euclidean distance between their descriptors. The quantitative results are shown in Fig. 5 *Top*. The procedure of viewpoint normalization will compensate the effects of perspective distortion, thus the resulting descriptors are more robust to the viewpoint changes. For each pair of matched features, we also computed the difference between their dominant orientations and the ratio between their patch scales. The results are shown in Fig. 5 *Middle* and Fig. 5 *Bottom*, respectively.

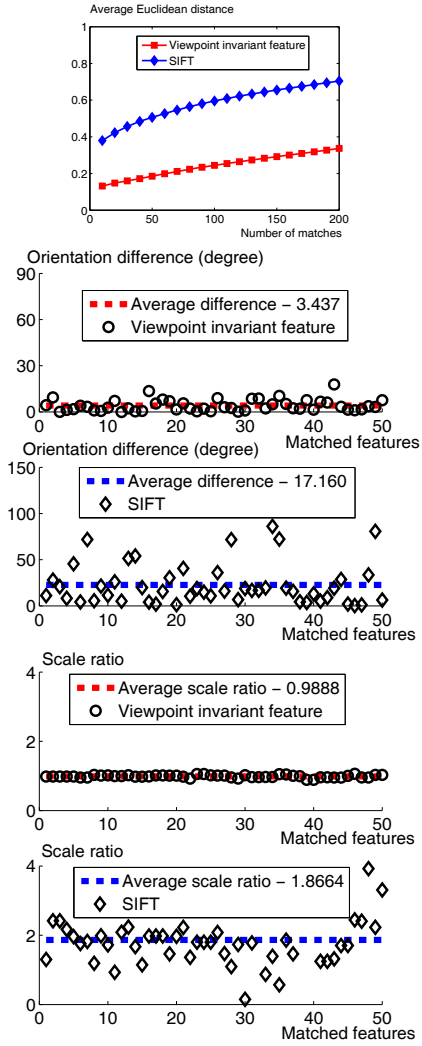


Figure 5. Performance comparison between SIFT and Viewpoint invariant features. *Top*: The average Euclidean distances between the descriptors of matched features; *Middle*: The orientation differences between matched features; *Bottom*: The scale ratios between matched feature. The matched feature extracted on the normalized front-parallel views show better robustness to viewpoint changes.

On the normalized front-parallel views, the viewing direction is normal to the extracted 3D plane. The matched features extracted on such normalized views have similar dominant orientations and consistent scale ratio. To qualitatively demonstrate the improvements, a number of matched features are shown on the original images (cf. Fig. 6 *Top*) and on the normalized images (Fig. 6 *Bottom*). Their corresponding scales and orientations are also annotated. On the normalized front-parallel views, the matched features have very similar orientations. Also their scale ratios show better consistency.

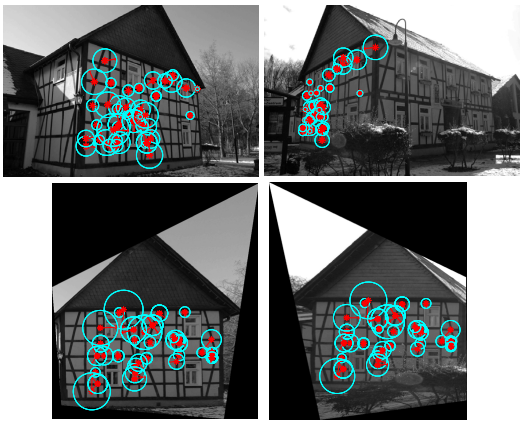


Figure 6. A number of matched features are shown. *Top*: on the original images; *Bottom*: on the front parallel views. Their scales and orientations are annotated. The feature matches on the viewpoint normalized views have very similar orientations and consistent scale ratios.

5.3. Wide baseline laser scan alignment

In this section, we apply the proposed framework to automatically align laser scans captured at widely separated viewpoints. For the preprocessing step, we applied the standard SIFT scheme [13] to match the 2D camera images to the laser-provided image, as described in Section 3. Since the viewpoint change between camera and laser is not significant, the SIFT technique can produce robust image matches as tie points between the camera images and the 3D laser scan. Then, we applied the proposed viewpoint invariant features for the difficult wide baseline matching tasks. A set of putative matches were firstly established, among them the inlier correspondences were selected by imposing the geometrical constraint (Eq. 2). For comparison, we applied the scale-invariant feature scheme SIFT and the affine-invariant feature scheme MSER for the same task. The matching results are shown in Fig. 7 with the quantitative comparisons provided in Tab. 2. It's noted that the viewpoint invariant features can handle the large viewpoint changes (the view angles changed more than 90 degrees), for which SIFT and MSER do not work well. Finally, we computed the 3D transform matrix relating two individual laser scans given a number of matched 3D points. The laser alignment results are shown in Fig. 7.

6. Conclusions

Nowadays most laser-scanning equipments are accompanied with an additional image camera. In this paper we have proposed an automatic framework for aligning two widely separated 3D laser scans via the use of provided image contents. To achieve this, we brought in the concept of 3D viewpoint normalization and extracted features on the normalized front-parallel views w.r.t. 3D dominant planes derived from the point cloud of a scene. The resulting viewpoint invariant features enable us to link the corresponding 3D points automatically in terms of wide baseline image matching. We evaluated the proposed feature matching scheme against the conventional 2D feature detectors, and applied it to realistic wide baseline laser scanning data of a variety of urban scenes. The experimental results demonstrate the potential of viewpoint invariant features for robust and automatic wide baseline laser scan registration.

Scene	SIFT			MSER			Ours		
	T	N	I	T	N	I	T	N	I
Scene a	19	28	901	7	16	690	79	80	901
Scene b	0	13	704	3	13	512	23	23	658

Table 2. The quantitative results of wide baseline 3D scene matching. (I - the number of initial correspondences by matching descriptors, N - the number of inliers correspondences returned by the RANSAC technique, T - the number of correct ones.)

Acknowledgement

The work was funded by a Strategic Research Cluster grant (07/SRC/I1168) by Science Foundation Ireland under the National Development Plan, and Deutsche Forschungsgemeinschaft (German Research Foundation) FO 180/14-1 (PAK 274). The authors gratefully acknowledge these supports. The authors would also thank Dipl.-Ing. Martin Blome for providing laser scan data.

References

[1] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features (SURF). *CVIU*, 110(3):346–359, 2008.

[2] P. Besl and N. McKay. A method for registration of 3-d shapes. *PAMI*, 14(2):239–256, 1992.

[3] J. Böhm. Terrestrial laser scanning - a supplementary approach for 3d documentation and animation. In *Photogrammetric Week 2005*, pages 263–271, 2005.

[4] N. Cornelis, B. Leibe, K. Cornelis, and L. Gool. 3d urban scene modeling integrating recognition and reconstruction. *IJCV*, 78(2-3):121–141, 2008.

[5] M. Donoser and H. Bischof. Efficient maximally stable extremal region (MSER) tracking. *CVPR*, 2006.

[6] M. Fischler and R. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Comm. Of the ACM*, 24(6):381–395, June 1981.

[7] Y. Furukawa and J. Ponce. Accurate, dense, and robust multiview stereopsis. *PAMI*, 2009.

[8] D. Gonzalez Aguilera, P. Rodriguez Gonzalvez, and J. Gomez Lahoz. An automatic procedure for co-registration of terrestrial laser scanners and digital cameras. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(3):308–316, 2009.

[9] K. Ikeuchi, T. Oishi, J. Takamatsu, R. Sagawa, A. Nakazawa, R. Kurazume, K. Nishino, M. Kamakura, and Y. Okamoto. The great buddha project: Digitally archiving, restoring, and analyzing cultural heritage objects. *IJCV*, 75(1):189–208, 2007.

[10] K. Koeser and R. Koch. Perspectively invariant normal features. *ICCV*, 2007.

[11] T. Läbe and W. Förstner. Automatic relative orientation of images. *Proceedings of the 5th Turkish-German Joint Geodetic Days*, 2006.

[12] L. Liu, I. Stamos, G. Yu, G. Wolberg, and S. Zokai. Multiview geometry for texture mapping 2d images onto 3d range data. In *CVPR*, pages 2293–2300, 2006.

[13] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[14] K. Mikolajczyk and C. Schmid. Scale and affine invariant interest point detectors. *IJCV*, 60(1):63–86, 2004.

[15] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *PAMI*, 27(10):1615–1630, 2005.

[16] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A comparison of affine region detectors. *IJCV*, 65(1-2):43–72, 2005.

[17] H. Pan. Two-level global optimization for image segmentation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 49:21–32, 1994.

[18] M. Pollefeys, L. Van Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops, and R. Koch. Visual modeling with a hand-held camera. *IJCV*, 59(3):207–232, 2004.

[19] H. Pottmann, Q. Huang, Y. Yang, and S. Hu. Geometry and convergence analysis of algorithms for registration of 3d shapes. *IJCV*, 67(3):277–296, 2006.

[20] J. Rissanen. Modelling by shortest data description. In *Automatica*, volume 14, pages 465–471, 1978.

[21] J. Seo, G. Sharp, and S. Lee. Range data registration using photometric features. *CVPR*, pages II: 1140–1145, 2005.

[22] E. Smith, R. Radke, and C. Stewart. Physical scale intensity-based range keypoints. In *3DPVT*, 2010.

[23] T. Tuytelaars and L. Van Gool. Matching widely separated views based on affine invariant regions. *IJCV*, 59(1):61–85, 2004.

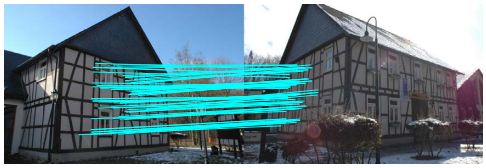
[24] C. Wu, B. Clipp, X. Li, J. Frahm, and M. Pollefeys. 3d model matching with viewpoint-invariant patches (VIP). *CVPR*, 2008.

[25] M. Y. Yang, Y. Cao, W. Förstner, and J. McDonald. Robust wide baseline scene alignment based on 3d viewpoint normalization. In *International Symposium on Visual Computing*, pages 654–665, 2010.

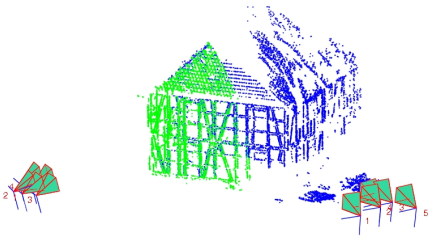
[26] M. Y. Yang and W. Förstner. Plane detection in point cloud data. Technical Report TR-IGG-P-2010-01, Department of Photogrammetry, University of Bonn, 2010.

[27] W. Zhang and J. Košecká. Hierarchical building recognition. *Image Vision Comput.*, 25(5):704–716, 2007.

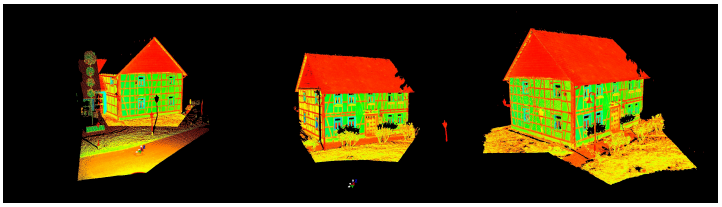
[28] W. Zhao, D. Nister, and S. Hsu. Alignment of continuous video onto 3d point clouds. *PAMI*, 27(8):1305–1318, 2005.



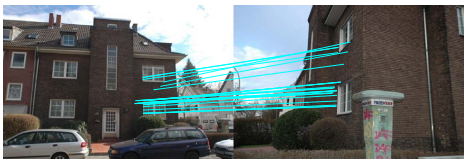
(a) Scene a: image matching pairs.



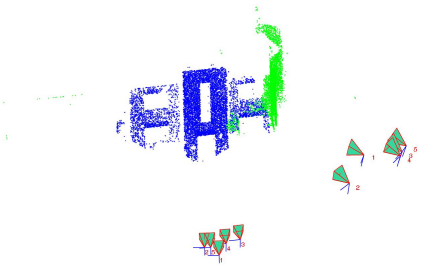
(b) Scene a: corresponding SfM.



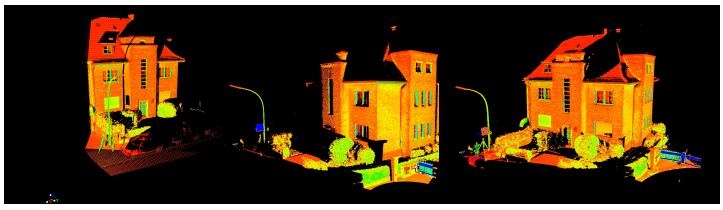
(c) Two different views of laser scan and corresponding alignment result.



(d) Scene b: image matching pairs.



(e) Scene b: corresponding SfM.



(f) Two different views of laser scan and corresponding alignment result.

Figure 7. Two example results of wide baseline 3D scene matching. Significant viewpoint changes can be observed on the associated image pairs.