Robust Wide Baseline Scene Alignment Based on 3D Viewpoint Normalization*

Michael Ying Yang[1], Yanpeng Cao[2], Wolfgang Förstner[1], and John McDonald[2]

[1] Department of Photogrammetry, University of Bonn, Bonn, Germany
[2] Department of Computer Science, National University of Ireland, Maynooth, Ireland

Abstract. This paper presents a novel scheme for automatically aligning two widely separated 3D scenes via the use of viewpoint invariant features. The key idea of the proposed method is following. First, a number of dominant planes are extracted in the SfM 3D point cloud using a novel method integrating RANSAC and MDL to describe the underlying 3D geometry in urban settings. With respect to the extracted 3D planes, the original camera viewing directions are rectified to form the front-parallel views of the scene. Viewpoint invariant features are extracted on the canonical views to provide a basis for further matching. Compared to the conventional 2D feature detectors (e.g. SIFT, MSER), the resulting features have following advantages: (1) they are very discriminative and robust to perspective distortions and viewpoint changes due to exploiting scene structure; (2) the features contain useful local patch information which allow for efficient feature matching. Using the novel viewpoint invariant features, wide-baseline 3D scenes are automatically aligned in terms of robust image matching. The performance of the proposed method is comprehensively evaluated in our experiments. It's demonstrated that 2D image feature matching can be significantly improved by considering 3D scene structure.

1 Introduction

Significant progress has recently been made in solving the problem of robust feature matching and automatic Structure from Motion (SfM). These advances allow us to recover the underlying 3D structure of a scene from a number of collected photographs [1], [2], [3]. However, the problem of automatically aligning two individual 3D models obtained at very different viewpoints still remains unresolved. Since the captured images are directly linked to the 3D point cloud in the SfM procedure, 3D points can be automatically related in terms of the matching of their associated 2D image appearances. Previously a number of successful techniques [4], [5], [6], [7], [8], [9] have been proposed for robust 2D image matching - a comprehensive review was given in [10]. 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

* The first two authors contributed equally to this paper.

G. Bebis et al. (Eds.): ISVC 2010, Part I, LNCS 6453, pp. 654–665, 2010.

produce reliable matching results of features extracted on wide baseline image pairs. In this paper our goal is to integrate 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.

In predominantly planar environments (urban scenes), fitting a scene with a piecewise planar model has become popular for urban reconstruction [11], [12]. In this paper, we proposed a novel approach to extract a number of dominant planes in the 3D point cloud by integrating RANSAC and MDL. The derived planar structures are used to represent the spatial layout of a urban scene. 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 rectified to form the front-parallel views of the scene. Viewpoint invariant features are then extracted on these canonical views for further matching. The key idea of the proposed method is schematically illustrated in Fig. 1. Knowing how everything looks like from a front-parallel view, it becomes easier to recognize the same surface from different viewpoints. Compared with some previous efforts on combining 2D feature with 3D geometry [13], [14], our method exploited the planar characteristics of man-made environment and extracted a number of dominant 3D planes to represent its 3D layout. Viewpoint normalization can be performed *w.r.t.* the planes to achieve better efficiency and robustness.

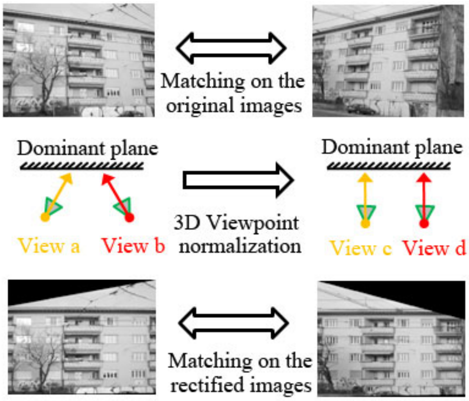

Matching on the original images
Dominant plane
View a
View b
3D Viewpoint normalization
Dominant plane
View c
View d
Matching on the rectified images

Fig. 1. The major procedure of generating and matching viewpoint invariant features

The remainder of the paper is organized as follow. Section 2 reviews some existing solutions for robust feature matching and 3D model alignment. The proposed method for 3D dominant plane extraction is presented in Section 3. 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. We finally conclude with a brief summary in Section 6.

2 Related Work

Automatic 3D scene alignment is a key step in many computer vision applications including large scale 3D modelling, augmented reality, and intelligent urban navigation. Given two sets of 3D points obtained at different viewpoints, the task is to estimate an optimal transformation between them. The most popular class of method for solving this problem is the Iterative closest point (ICP) based techniques [15], [16], [17]. They compute the alignment transformation by iteratively minimizing the sum of distances between closest points. However, the performances of ICP-based methods reply on a good estimation initialization and require good spatial configuration of 3D points. Recently many researchers proposed to enhance the performances of 3D point cloud alignment by referring to their associated 2D images. In [18], an effective method was presented for automatic 3D model alignment via 2D image matching. [19] presented 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. [20] presented an automated 3D range to 3D range registration method that relies on the matching of reflectance range image and camera image. In [21], 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 the small observation changes. To produce satisfactory registration results of 3D points clouds captured at significantly changed viewpoints, we need an effective image feature scheme to establish reliable correspondences between wide baseline image pairs.

A large number of papers have reported on robust 2D image feature extraction and matching, cf. [10] 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) [7] is the best scale-invariant feature scheme and the Maximally Stable Extremal Regions (MSER) [6] shows superior affine invariance. In [22], 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) [13], based on 3D normalized patches was proposed for 3D model matching and querying. In [14], 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 model 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 scenes), piece-wise planar 3D models are more robust, compact, and efficient for viewpoint normalization of cameras with wide baselines.

3 3D Dominant Plane Extraction

One of the most widely known methodologies for plane extraction is the RANdom SAmple Consensus (RANSAC) algorithm [23]. It has been proven to successfully detect planes in 2D as well as 3D. RANSAC is reliable even in the presence of a high proportion of outliers.

Based on the observation that RANSAC may find wrong planes if the data has a complex geometry, we introduces a plane extraction method by integrating RANSAC and minimum description length (MDL).

Here, we apply MDL for plane extraction, similar to the approach of [24]. 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), 6 planes and outliers (6P+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

$$\#bits(points \mid O) = n_0 \cdot (3lb(R/\epsilon)) \tag{1}$$

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

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

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 [25]. 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 $\mathbf{x} \sim N(\mu, \Sigma)$ which leads to the sixth term.

$\#bits(points \mid 1P+O)$, $\#bits(points \mid 3P+O)$, $\#bits(points \mid 4P+O)$, $\#bits(points \mid 5P+O)$, and $\#bits(points \mid 6P+O)$, and so on, can be deducted in a similar way. 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.

4 3D Viewpoint Invariant Features

In this step we perform normalization with respect to 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 an image of the 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 a homography function. Since we know the 3D positions of the points shown in the scene and their corresponding image, we can compute the homography relating the plane to its image from four (or more) 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 examples 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 [7]. Potential keypoints are identified by scanning local extreme in a series of Difference-of-Gaussian (DoG) images. For each detected keypoint $\mathbf{x}$, appropriate scale $\mathbf{s}$ and orientation θ are assigned to it and a 128-element SIFT descriptor $\mathbf{f}$ is created based upon image gradients of its local neighbourhood.

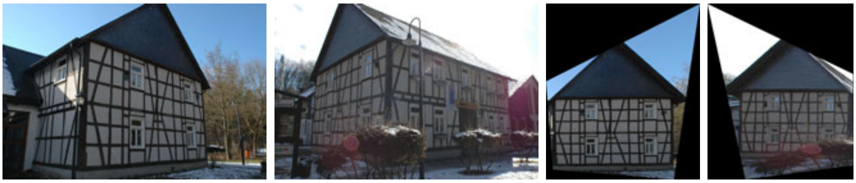


Fig. 2. Some examples of viewpoint normalization. *Left*: Original images; *Right*: Normalized front views. Note the perspective distortions are largely reduced in the warped front-parallel views of the building walls (e.g. a rectangular window in the 3D world will also appear rectangular in the normalized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

Given a number of features extracted on the canonical views, we applied the criterion described in [26] to generate the putative feature correspondences. Two features are considered matched if the cosine of the angle between their descriptors $\mathbf{f}_i$ and $\mathbf{f}_j$ is above some threshold δ as:

$$\cos(\mathbf{f}_i, \mathbf{f}_j) = \frac{\mathbf{f}_i \cdot \mathbf{f}_j}{\|\mathbf{f}_i\|_2 \, \|\mathbf{f}_j\|_2} > \delta \tag{3}$$

where $||\cdot||_2$ represents the $L2$-norm of a vector. This criterion establishes matches between features having similar descriptors and does not falsely reject potential correspondences extracted on the images of repetitive structures which are very common in man-made environments.

After obtaining a set of putative feature correspondences based on the matching of their local descriptors, we impose certain global geometric constraints to identify the true correspondences. The RANSAC technique [23] 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.

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

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

The geometrical model can be significantly simplified via the use of these novel features, and thus, lead to a more efficient matching method. Since the effects of perspective transformation are not compensated in the standard SIFT scheme, only the 2D image coordinates of SIFT features can be used to generate geometric constraints (F-Matrix or H-Matrix). Therefore, a number of SIFT matches are required to compute F-Matrix (7 correspondences) or H-matrix (4 correspondences). In comparison, the viewpoint invariant features are extracted on the front-parallel views of the same continuous flat building facade, taken at different distances and up to a camera translation and rotation around its optical axis. Every feature correspondence provides three constraints: scale (camera distance), 2D coordinates on the canonical view (camera translation), and dominant orientation (rotation around its optical axis). Therefore, a single feature correspondence is enough to completely define a point-to-point mapping relation between two canonical views. Consider a pair of matched features $(\mathbf{x}_1^m, \mathbf{s}_1^m, \theta_1^m)$ and $(\mathbf{x}_2^n, \mathbf{s}_2^n, \theta_2^n)$ both extracted on the normalized front-parallel views, a 2D similarity translation hypothesis is generated as follows:

$$\begin{bmatrix} x_1 - x_1^m \\ y_1 - y_1^m \\ 1 \end{bmatrix} = \begin{bmatrix} \Delta s & 0 & 0 \\ 0 & \Delta s & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\Delta\theta & -\sin\Delta\theta & 0 \\ \sin\Delta\theta & \cos\Delta\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_2 - x_2^m \\ y_2 - y_2^m \\ 1 \end{bmatrix} \tag{4}$$

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. Our experimental evaluations in Section 5.2 show that for all ground true correspondences the scale ratios and orientation differences are equal up to a very small offset. It means that the information of patch scale and dominant orientation associated with the viewpoint invariant features are robust enough to generate geometrical hypothesis, which is impossible in the SIFT scheme. Using this simplified geometric model, a much smaller number of samples are needed to guarantee the generation of the correct hypothesis (c.f. Table 1 for comparison). The correspondences consistent with each generated hypothesis (e.g. the symmetric transfer error is less than a threshold) are defined as its inliers. The hypothesis with the most supports is chosen and its corresponding inliers are defined as true matches.

5 Experimental Results

We conducted experiments to evaluate the performance of the proposed method on urban scenes, with focus on the building facade images.

5.1 Point Cloud Sets Generation

We have taken 15 pairs of images over largely separated views with a calibrated camera. Each pair consists of 10 images, of which 5 images represent left view, the other 5 images represent right view. Only one pair is exceptional, as shown in Fig. 7 (a), of which 10 images represent left view, the other 10 images represent right view. We intend to use this pair for further comparison *w.r.t.* multi-view image numbers. Then, we applied orientation software AURELO [27] to achieve full automatic relative orientation of these multi-view images. And we used the public domain software PMVS (patch-based multi-view stereo) [28] 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. The algorithm starts by detecting features in each image, matches them across multiple images to form an initial set of patches, and uses an expansion procedure to obtain a denser set of patches, before using visibility constraints to filter away false matches. An example for a point cloud derived with this software is given in Fig. 3 *Middle*. Finally, 5 dominant planes were extracted from each point cloud, while the rest planes were removed. One example demonstrating dominant planes extraction is shown in Fig. 3 *Right*.

5.2 Performance Evaluations

After extracting 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

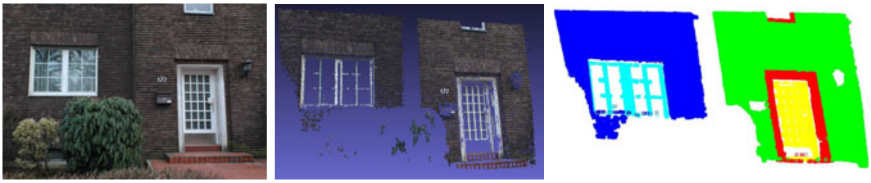


Fig. 3. *Left*: One of three images taken for a building facade scene. *Middle*: A snap-shot image of corresponding 3D point cloud generated by PMVS. *Right*: The five dominant planes automatically extracted from the point cloud.

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 [22] 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 [22]. We adjusted the threshold value to vary the number of resulting feature correspondences.

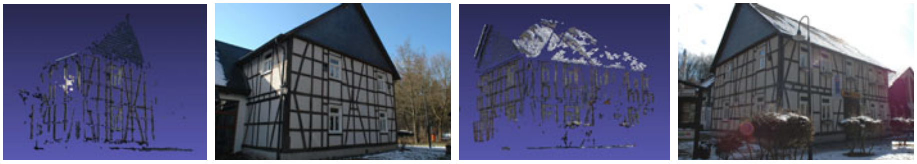


Fig. 4. Two 3D point clouds and their associated images captured at widely separated views

Our goal is to evaluate 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 *Left*. It's noted that the descriptors of corresponding features extracted on the front-parallel views become very similar. It's because 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 *Right*, respectively. 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. It means that the information of patch scale and dominant orientation associated with the

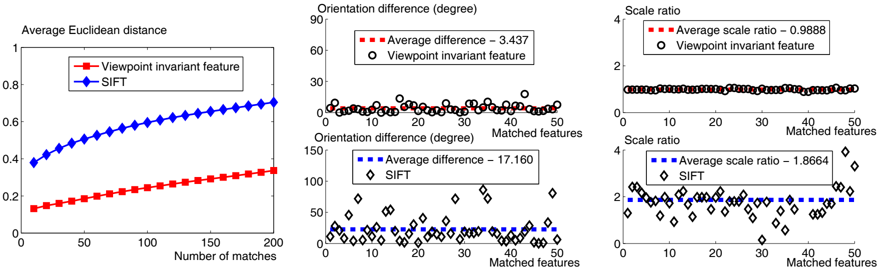


Fig. 5. Performance comparison between SIFT and Viewpoint invariant features. *Left*: The average Euclidean distances between the descriptors of matched features; *Middle*: The orientation differences between matched features; *Right*: The scale ratios between matched feature. The matched feature extracted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

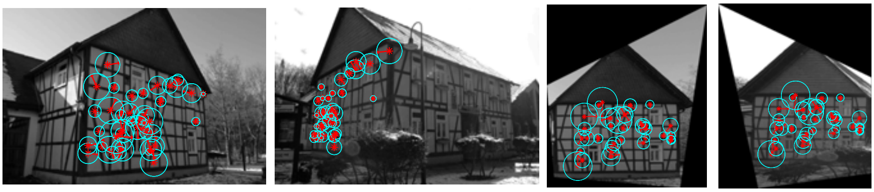
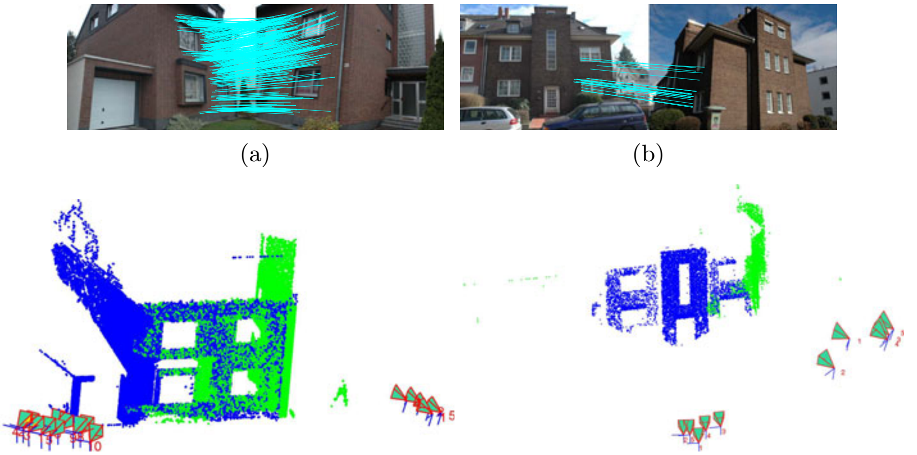


Fig. 6. A number of matched features are shown. *Left*: on the original images; *Right*: on the front parallel views. Their scales and orientations are annotated. The feature matches on the viewpoint normalized views have very similar orientations and consistent scale ratios.

viewpoint invariant features are robust enough to determine camera distance and camera rotation around it optical axis, respectively. To qualitatively demonstrate the improvements, a number of matched features are shown on the original images (cf. Fig. 6 *Left*) and on the normalized images (Fig. 6 *Right*).

5.3 Wide Baseline Alignment

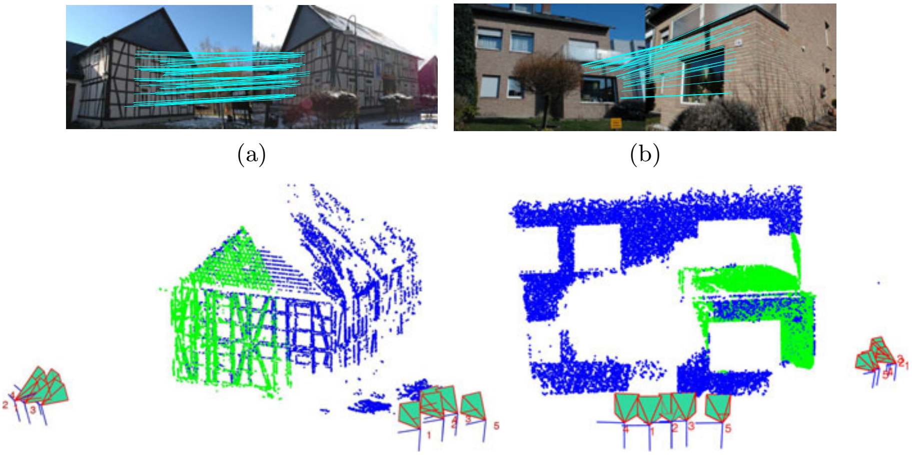
Next we demonstrate the advantages of the proposed feature matching scheme by applying it to some very difficult wide baseline alignment tasks. First, we extracted a number of viewpoint invariant features and establish putative correspondences according to Eq. 3 (the threshold δ was set at 0.9). Then, we applied the RANSAC algorithm impose the global geometric constraint (Eq. 4) to identify inliers. The number of inlier correspondences and correct ones were counted manually. For comparison, we applied SIFT and MSER for the same task. A set of putative matches were firstly established, among them the inlier correspondences were selected by imposing the homography constraint. In many cases, SIFT and MSER cannot generate enough correctly matched features to



(a)

(b)

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



(a)

(b)

Fig. 8. Some other example results of wide baseline 3D scene matching. Our technique successfully aligned 3D scenes with very small overlap.

compute the correct H-matrix for identifying inlier correspondences due to the large viewpoint changes. Some matching results are shown in Fig. 7 and Fig. 8 with the quantitative comparisons provided in Tab. 2.

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.)

Scene	SIFT			MSER			Our method		
	T	N	I	T	N	I	T	N	I
7a	70	89	7117	303	550	4511	420	421	8165
7b	0	13	704	3	13	512	23	23	658
8a	19	28	901	7	16	690	79	80	901
8b	0	10	640	4	19	412	41	41	804

6 Conclusions

We have proposed an intuitive scheme for aligning two widely separated 3D scenes via the use of viewpoint invariant features. To achieve this, we extracted viewpoint invariant features on the normalized front-parallel views *w.r.t.* 3D dominant planes derived from point cloud of a scene. This enables 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 to some difficult wide baseline alignment tasks of a variety of urban scenes. Our evaluation demonstrates that viewpoint invariant features are an improvement on current methods for robust and accurate 3D wide baseline scene alignment.

Acknowledgment

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.

References

1. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge (2003)
2. Snavely, N., Seitz, S.M., Szeliski, R.: Modeling the world from Internet photo collections. IJCV 80(2), 189–210 (2008)
3. Pollefeys, M., Van Gool, L., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J., Koch, R.: Visual modeling with a hand-held camera. IJCV 59(3), 207–232 (2004)
4. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-up robust features (SURF). Comput. Vis. Image Underst. 110(3), 346–359 (2008)
5. Tuytelaars, T., Van Gool, L.: Matching widely separated views based on affine invariant regions. IJCV 59(1), 61–85 (2004)
6. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide baseline stereo from maximally stable extremal regions. In: BMVC (2002)

7. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. IJCV 60(2), 91–110 (2004)
8. Donoser, M., Bischof, H.: Efficient maximally stable extremal region (MSER) tracking. In: CVPR, pp. 553–560 (2006)
9. Mikolajczyk, K., Schmid, C.: Scale and affine invariant interest point detectors. IJCV 60(1), 63–86 (2004)
10. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Van Gool, L.: A comparison of affine region detectors. IJCV 65(1-2), 43–72 (2005)
11. Sinha, S., Steedly, D., Szeliski, R.: Piecewise planar stereo for image-based rendering. In: ICCV, pp. 1881–1888 (2009)
12. Furukawa, Y., Curless, B., Seitz, S., Szeliski, R.: Manhattan-world stereo. In: CVPR, pp. 1422–1429 (2009)
13. Wu, C., Clipp, B., Li, X., Frahm, J., Pollefeys, M.: 3d model matching with viewpoint-invariant patches (VIP). In: CVPR, pp. 1–8 (2008)
14. Koeser, K., Koch, R.: Perspectively invariant normal features. In: ICCV, pp. 14–21 (2007)
15. Besl, P., McKay, N.: A method for registration of 3-d shapes. PAMI 14(2), 239–256 (1992)
16. Zhao, W., Nister, D., Hsu, S.: Alignment of continuous video onto 3d point clouds. PAMI 27(8), 1305–1318 (2005)
17. Pottmann, H., Huang, Q., Yang, Y., Hu, S.: Geometry and convergence analysis of algorithms for registration of 3d shapes. IJCV 67(3), 277–296 (2006)
18. Seo, J., Sharp, G., Lee, S.: Range data registration using photometric features. In: CVPR II, pp. 1140–1145 (2005)
19. Liu, L., Stamos, I., Yu, G., Zokai, S.: Multiview geometry for texture mapping 2d images onto 3d range data. In: CVPR II, pp. 2293–2300 (2006)
20. Ikeuchi, K., Oishi, T., Takamatsu, J., Sagawa, R., Nakazawa, A., Kurazume, R., Nishino, K., Kamakura, M., Okamoto, Y.: The great buddha project: Digitally archiving, restoring, and analyzing cultural heritage objects. IJCV 75(1), 189–208 (2007)
21. Gonzalez Aguilera, D., Rodriguez Gonzalvez, P., Gomez Lahoz, J.: 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)
22. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. PAMI 27(10), 1615–1630 (2005)
23. Fischler, M., Bolles, R.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Comm. of the ACM 24(6), 381–395 (1981)
24. Pan, H.: Two-level global optimization for image segmentation. ISPRS Journal of Photogrammetry and Remote Sensing 49, 21–32 (1994)
25. Rissanen, J.: Modelling by shortest data description. Automatica 14, 465–471 (1978)
26. Zhang, W., Košecká, J.: Hierarchical building recognition. Image Vision Comput. 25(5), 704–716 (2007)
27. Läbe, T., Förstner, W.: Automatic relative orientation of images. In: Proceedings of the 5th Turkish-German Joint Geodetic Days (2006)
28. Furukawa, Y., Ponce, J.: Accurate, dense, and robust multi-view stereopsis. PAMI (2009)