THE JOURNAL OF NAVIGATION (2017), **70**, 1133–1152.
doi:10.1017/S0373463317000224

New Environmental Line Feature-based Vision Navigation: Design and Analysis

Zeyu Li, Jinling Wang, Kai Chen and Yu Sun

(*School of Civil and Environmental Engineering, UNSW Australia, Sydney, Australia*)
(E-mail: zeyu.li@student.unsw.edu.au)

Vision navigation using environmental features has been widely applied when satellite signals are not available. However, the matching performance of traditional environmental features such as keypoints degrades significantly in weakly textured areas, deteriorating navigation performance. Further, the user needs to evaluate and assure feature matching quality. In this paper, a new feature, named Line Segment Intersection Feature (LSIF), is proposed to solve the availability problem in weakly textured regions. Then a combined descriptor involving global structure and local gradient is designed for similarity comparison. To achieve reliable point-to-point matching, a coarse-to-fine matching algorithm is developed, which improves the performance of the point set matching algorithm. Finally, a framework of matching quality evaluation is proposed to assure matching performance. Through the comparison, it is demonstrated that the proposed new feature has superior overall performance especially on correctly matched numbers of keypoints and matching correctness. Also, using real image sets with weak texture, it is shown that the proposed LSIF can achieve improved navigation solutions with high continuity and accuracy.

KEYWORDS

1. Vision. 2. Indoor navigation. 3. Weak texture. 4. Line segment intersection feature.

Submitted: 19 May 2016. Accepted: 19 March 2017. First published online: 9 May 2017.

1. INTRODUCTION. The next generation of navigation systems aim to obtain position and orientation of a moving platform anywhere at any time. Among many emerging navigation techniques, vision-based navigation has become a promising and popular approach to achieve ubiquitous navigation as it is accurate, passive and low-cost (Jin et al., 2016; Liu et al., 2012; Xian et al., 2015). In this domain, natural environmental features including edges, corners and keypoints (points of interests or features) play an important role as they can be automatically detected, described and matched, which provides the capability of storing and transferring the geospatial information. As a type of natural environmental feature, keypoints attract interest from researchers due to their detector repeatability and descriptor distinctiveness. Repeatability is defined as robustness of keypoint location with regard to environment change such as translation, scaling, rotation and viewpoint change. Distinctiveness means the generated descriptor should be distinctive from others

https://doi.org/10.1017/S0373463317000224 Published online by Cambridge University Press

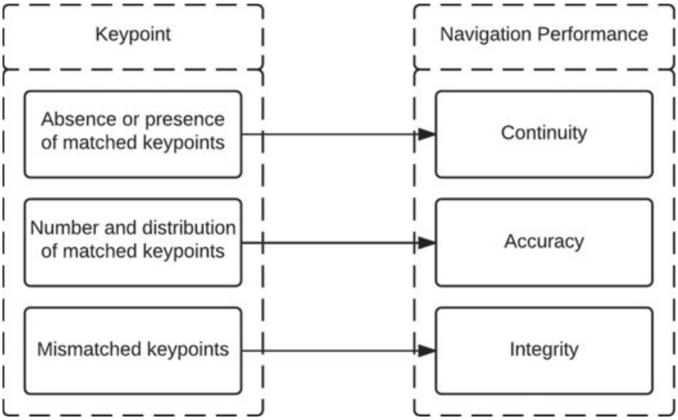

Keypoint
Navigation Performance
Absence or presence of matched keypoints
Continuity
Number and distribution of matched keypoints
Accuracy
Mismatched keypoints
Integrity

Figure 1. The relationship between keypoints and navigation performance.

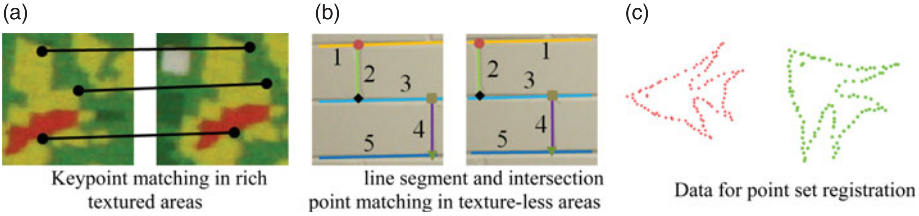

(a)
(b)
(c)
1
2
3
4
5
Keypoint matching in rich textured areas
line segment and intersection point matching in texture-less areas
Data for point set registration

Figure 2. Illustrative examples for keypoint matching, line segment matching and point set matching.

in the same image. The former characteristic enables keypoints to be tracked and the latter provides the possibility for keypoints to be distinguished and matched.

Keypoints are closely related to the performance of vision-based navigation, which is illustrated in Figure 1. The presence or absence of keypoints directly affects the continuity since geo-spatial information contained in the keypoints cannot be obtained if matched keypoints do not exist. The number and distribution of keypoints affect the accuracy of navigation solutions through error propagation (Li et al., 2016). If there are mismatched keypoints, the correctness of the navigation solution may be violated, degrading the integrity of the navigation system.

Traditional keypoints such as Scale-Invariant Feature Transform (SIFT) (Lowe, 1999) are generally good natural features in rich texture areas as illustrated in Figure 2(a). However, keypoint matching in ubiquitous weakly textured areas such as homogenous walls and man-made objects is challenging. Most traditional keypoint detection and description methods are based on intensity or colour variation (e.g. gradients). If the intensity or colour is stable, it will be difficult to detect keypoints. Moreover, although a number of keypoints may be detected, the descriptor of such a keypoint will tend to lose its distinctiveness as the descriptor comes from the neighbouring area of the detected keypoint, eventually affecting matching performance. In this scenario, the continuity of the vision-based navigation system will be seriously affected. Although some keypoints may be detected and matched correctly, they tend to be located in certain small rich texture areas. Besides, the number is significantly reduced, hence the geometry of vision-based navigation will deteriorate,

https://doi.org/10.1017/S0373463317000224 Published online by Cambridge University Press

finally affecting the accuracy of navigation solutions (Li et al., 2016). Also, the ratio of mismatches will be enlarged, threatening the integrity of the navigation system.

There have been studies aiming to improve navigation performance in texture-less areas. One aspect is to directly match the detected line segments to achieve vision-based navigation. As Figure 2(b) shows, line segments with the same number are correctly matched. The geo-spatial information contained in the line segments can be transferred through line segment matching. Then the position of the camera can be determined by space resection. Zhou et al. (2015) put forward Structural Simultaneous Localisation And Mapping (StructSLAM) that made use of the structured lines to reduce the error of navigation, and simultaneously generated maps made up of detected line segments. However, for line segment matching, it is difficult to assure its correctness in weakly textured areas as the distinctiveness of description depends on the neighbouring areas. Another aspect is to develop new keypoints. As illustrated in Figure 2(b), the intersection points with the same shape are correctly matched. Similarly, Kim and Lee (2012) employed the line segment pair to generate a Line Intersection Context Feature (LICF) that was invariant under perspective projection in weakly textured areas. The matched LICFs provided the clue for line segment matching, point-to-point matching and epipolar geometry reconstruction in weakly textured areas. However, original LICF employs the Normalised Cross-Correlation (NCC) method for matching tentative points. There is still room to improve its matching performance as NCC is not invariant to scale, rotation and shearing differences (Lewis, 1995).

Point set registration is related to keypoint matching. As shown in Figure 2(c), the main objective of point set registration is to find the correspondence between two point sets only using the positions of the point sets as the input, which is beneficial for keypoint matching in weakly textured areas. The transformation of the two point sets often includes not only rigid transformation such as rotation and scaling, but also non-rigid deformation, noise and outliers. However, one difference between the traditional point set and detected keypoints is that the mismatch ratio can be very high (e.g. larger than 50%) in the texture-less environment, which poses a challenge for the robustness of the traditional point set matching algorithm. Therefore the robustness of the point set matching algorithm needs to be enhanced through the appropriate design.

In this paper, we address the limitation of natural features in weakly textured areas in the indoor environment by the detection of a new feature – Line Segment Intersection Feature (LSIF). Considering the difficulty of matching in texture-less areas, its description is given and the matching algorithm is designed. Quality control is conducted to provide a measure of the trust for matching correctness and to further ensure matching correctness. The rest of this paper is structured as follows. Sections 2, 3, 4 and 5 propose the detailed design for the LSIF detection, description, matching and validation algorithm respectively. In Section 6, real images for weakly textured areas in indoor environment are employed to compare and analyse the matching and navigation performance of LSIF. Section 7 presents concluding remarks.

2. LINE SEGMENT INTERSECTION FEATURE DETECTION.

LSIFs originate from the intersections of line segment pairs. Here, a Line Segment Detector (LSD) (Von Gioi et al., 2008) with sub-pixel accuracy and controlled false detection rate is employed. It creates a level-line field, and then pixels with similar orientations are segmented to form line support regions. In each region, the principal inertial axis is used as the main rectangle

https://doi.org/10.1017/S0373463317000224 Published online by Cambridge University Press

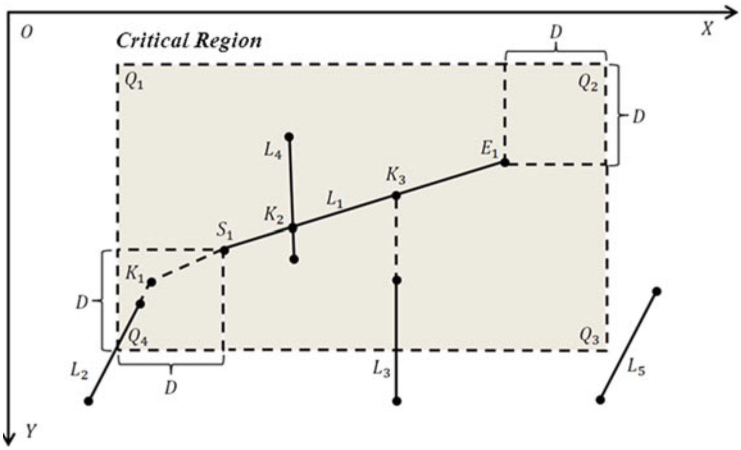


Figure 3. The critical region $\prod$ of L_1

direction. Pixels whose level-line angles are within a certain tolerance are identified as aligned points. After the validation process based on *a contrario* approach using the number of aligned points, the detected line segments are extracted with two end points.

However, the simple intersection for the detected line segments will not contribute to the following matching performance since there are several chaotic intersection points from unnecessary line segment pairs. Normally these line segments have small lengths. The main line segments tend to be vertical with others in a man-made indoor environment, which is a beneficial characteristic for eliminating unnecessary line segments. Therefore, a filter based on the line segments' lengths and intersection angles with others is applied. That is, only the line segments with lengths larger than a predefined threshold (e.g. ten pixels), and at the same time, intersection angles with others lying in a range such as [80° 100°], are preserved. The retained line segments can be represented by $L_1, L_2, \ldots, L_k$.

As shown in Figure 3, a neighbouring area alongside L_1 is named as the critical region $\prod$ within the image coordinate system, which can be defined as a rectangle $Q_1 Q_2 Q_3 Q_4$ by extending two squares with the length being equal to D from the two ending points respectively. The squares always lie on different sides of L_1. The rule for detecting LSIFs from L_1 is that if intersections between L_1 or L_1's extensions and other line segments or their extensions lie in $\prod$, they are detected as LSIFs. For example, K_1, K_2 and K_3 are detected LSIFs as their corresponding line segments or extensions intersect with L_1 in $\prod$. The intersection between L_5's extension and L_1's extension does not lie in $\prod$, therefore its intersection is discarded.

The coordinates of the detected LSIFs can be represented by:

$$\mathrm{C_{LSIF}} = \left\{ \mathrm{INEPT}(\mathrm{L_i}, \mathrm{L_j}) \mid \mathrm{INEPT}(\mathrm{L_i}, \mathrm{L_j}) \in \prod \right\} \tag{1}$$

where INEPT represents the intersection's coordinates of the two detected line segments or their corresponding extension lines. The main differences between the proposed LSIF and LICF proposed by Kim and Lee (2012) lie in two aspects: firstly, the critical region of LSIF is a square for simplified calculation, while the boundary of determining LICF involves the calculation of two circles. Secondly, the filter based on the length and intersection angle excludes the unnecessary line segments which contribute little for structure description.

3. LINE SEGMENT INTERSECTION FEATURE DESCRIPTION.

The aim of LSIF description is to generate descriptors according to gradients and structure information (e.g. geometric relationship), and provide tentative correspondences by similarity comparison.

Although the number and distribution of detected LSIFs vary for each image, the global structure of the detected LSIFs, which can be referred to as the general 2D geometric relationship between one keypoint and others, is comparatively stable. For example, if one detected $\mathrm{LSIF_a}$ in image I correctly matched with another $\mathrm{LSIF_b}$ in image II, significant similarities will exist between $\mathrm{LSIF_a}$'s geometric relationship (e.g. distance and relative position) with other LSIFs in image I, and $\mathrm{LSIF_b}$'s geometric relationship with other LSIFs in image II, and vice versa. The benefit of this is that the stable global structure description purely uses LSIFs' image coordinates without the extraction of gradients or colour information from texture-less areas. Therefore, global structure should be the main clue for LSIF matching.

Compared with pure point set matching, the image itself also brings benefits for LSIF description. Although the uniqueness of neighbouring gradient information (e.g. variation of intensity) for the detected LSIFs from a texture-less area is reduced, it can provide the secondary clues for LSIF matching if the weighting factor is appropriately controlled (Section 4.1). Besides, a few detected LSIFs may exist in the small areas with rich texture. Therefore, the description of local gradient contributes to reducing the matching ambiguity.

Considering these two aspects, this paper has designed a combined descriptor that utilises local gradient and global structure of the detected LSIF and is illustrated as follows:

$$D_C = (D_S, D_I) \tag{2}$$

where I and S are local gradient descriptor and global structure descriptor respectively.

4. LINE SEGMENT INTERSECTION FEATURE MATCHING.

The matching correctness is important as it will affect the integrity of the vision-based navigation system. This section designs a coarse to fine matching strategy to assure correctness.

4.1. *Coarse Matching using Combined Descriptors.*

Compared with point set registration where the coordinates of all the keypoints are directly used as the description, the coarse matching preserves the keypoints with high possibility to be correctly matched. The coarse matching procedure for two sub-descriptors should consider their own characteristics. In this paper, Shape Context (Belongie et al., 2002) and SIFT descriptor (Lowe, 1999) are chosen as the global structure descriptor and local gradient descriptor, respectively. Therefore, the similarity between the sub-descriptors is measured by a χ^2 test statistic and L_2 norm respectively. Assume there are m and n keypoints in the two images. For each sub-descriptor, a $m \times n$ cost matrix can be generated. Lower cost means that the two corresponding descriptors have a higher possibility to be matched.

The cost matrix can be generated as follows: originally the similarity measure for Shape Context is constrained between 0 and 1 by χ^2 test statistic. Similarly, to combine with the Shape Context, the similarity measure of the SIFT descriptor needs to be normalised between 0 and 1. Assume the cost matrix generated by S is C_S and the one generated by I is

C_I. Since the distinctiveness of local gradient descriptors is weakened in texture-less areas, more weight should be assigned to a global structure descriptor. Hence w_S should be larger than w_I (e.g. $w_S = 0{\cdot}6$ and $w_I = 0{\cdot}4$). Therefore, the final cost matrix C_C of the combined descriptor D_C can be denoted as:

$$C_C = w_S C_S + w_I C_I (w_S > w_I) \tag{3}$$

The establishment of point-to-point correspondences is done by the Hungarian method (Kuhn, 1955). Other algorithms for finding the optimal path of the cost matrix can still be applied. After that, epipolar geometry-based Random Sample Consensus (RANSAC) (Hartley and Zisserman, 2003) is employed to further reduce mismatches. Then the initial point-to-point matching is generated.

As expected, the initial matching result purely from the coarse matching stage is not completely reliable. The ratio of mismatches depends on the distinctiveness of keypoints' global structure and local gradient.

4.2. *Original Affine Coherent Point Drift Algorithm.* For completeness, affine Coherent Point Drift (CPD) matching (Myronenko and Song, 2010) is briefly described in this section. The affine CPD algorithm models one point sets as the centroids from the Gaussian Mixture Model (GMM), and the other set is then generated by GMM. If one data point is correctly matched, its GMM *posterior* probability is maximised. In the matching process, the topological structure of the point is preserved as the centroids move coherently as a group.

Assume $X_{N\times 2}$ and $Y_{N\times 2}$ are the coordinates of the tentative matches in two images respectively after coarse matching. $X_{N\times 2}$ is the data point set and $Y_{N\times 2}$ is the set of GMM centroids. An additional uniform distribution represents the noise and mismatches. The mixture model is constructed as:

$$\mathrm{p(x)} = \mathrm{w}\frac{1}{N} + (1-w)\sum_{n=1}^{N}\frac{1}{N}p(x \mid n) \tag{4}$$

where $\mathrm{p(x|m)} = \frac{1}{2\pi\sigma^2}e^{-\frac{\|x-y_n\|}{2\sigma^2}}$, and $\mathrm{w}(0 < w < 1)$ is the weight for the uniform distribution. The location of the GMM centroid can be determined by the affine parameter B and t by minimising the negative log-likelihood function in Equation (5).

$$\mathrm{E(B,t,}\sigma^2) = -\sum_{n=1}^{N} log \sum_{n=1}^{N+1} p(n)p(x_n|n) \tag{5}$$

An Expectation Maximisation (EM) algorithm is employed to find B and t. In the Expectation step, the objective function shown in Equation (6) is the upper bound of the negative log-likelihood function in Equation (5).

$$\mathrm{Q(B,t,}\sigma^2) = \frac{1}{2\sigma^2}\sum_{n=1}^{N}\sum_{n=1}^{N} P^{old}(n|x_n)\|x_n - By_n - t\|^2 + N_p log\sigma^2 \tag{6}$$

where $\mathrm{N_p} = \sum_{n=1}^{N}\sum_{n=1}^{N} p^{old}(n|x_n)$. According to Bayes' theorem, the components of the *posteriori* probability matrix $\mathrm{p^{old}}$ equals:

$$\mathrm{p^{old}(n|x_n)} = P(i,j) = \frac{exp(-\frac{1}{2\sigma^2}\|x_i - (By_j + t)\|^2)}{\sum_{n=1}^{N} exp(-\frac{1}{2\sigma^2}\|x_i - (By_j + t)\|^2) + 2\pi\sigma^2\frac{w}{1-w}} \tag{7}$$

https://doi.org/10.1017/S0373463317000224 Published online by Cambridge University Press

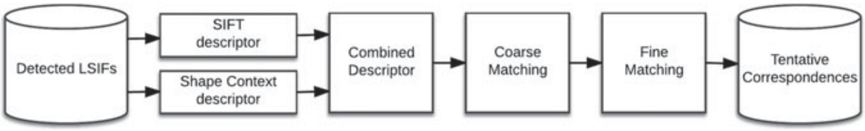


Figure 4. LSIF description and matching process.

where σ^2 equals $\frac{1}{2N^2}\sum_k^N\sum_k^N\|x_k-y_k\|$. As the number of the two tentatively matched LSIFs is the same, a $N\times N$ *posterior* probability matrix P can be constructed as the indicator for matching correctness.

The affine parameter B and t can be directly deduced by the partial derivatives from Equation (6) in the Maximisation step of EM process. The EM process will iterate Expectation and Maximisation steps until convergence. More details can be seen in the work by Myronenko and Song (2010).

4.3. *Modified Coherent Point Drift Algorithm.* As mentioned above, the original affine CPD employs the initial values of affine parameters to estimate the probability of matching correctness (E step). However, the constraints from global structure similarity can be beneficial in assuring the matching correctness in the EM process.

Shape Context is applied as the descriptor for the global structure similarity comparison. The cost matrix of Shape Context for each initial keypoint pair can be transferred to the prior matching probability matrix, which imposes constraints on the original matching probability matrix in Equation (7).

The construction of the prior probability matrix P_{Prior} is as follows. The weighting factor for the 'correct' correspondence in the coarse matching stage is set as W_{Prior} (e.g. 0.8). In the first iteration, the prior probability matrix is a $N\times N$ diagonal matrix with diagonal elements equalling to W_{Prior}, and remaining elements equals to $\frac{1-W_{Prior}}{N-1}$. For the following iterations, the elements on the optimal path of the cost matrix from the Hungarian method are set as W_{Prior}, and similarly, the other elements are set as $\frac{1-W_{Prior}}{N-1}$. Therefore the new matching probability matrix is set as the multiplication of $P(n,n)$ and $P_{Prior}(n,n)$ as shown in Equation (8).

$$P_{New}(n,n) = P(n,n)P_{Prior}(n,n) \tag{8}$$

By using Equation (8), the correct correspondences are preserved, while the weights of mismatches are controlled. In summary, the description and matching process is summarised in Figure 4.

5. MATCHING VALIDATION.

After the above three steps, a few mismatches may still exist in the pairs. Moreover, if the matching quality does not meet the requirement for correctness, the algorithm should have the ability to warn users. Therefore matching validation and quality evaluation are necessary to further detect and exclude the mismatches as well as to assess the matching quality. Assume $X_{N\times 2}$ and $Y_{N\times 2}$ are tentative matching pairs. With the obtained affine transformation with B and t, each keypoints' coordinates in $X_{N\times 2}$ can be transferred to $\hat{Y}_{N\times 2}$. If they are correctly matched, the Euclidean distance between estimated coordinates $\hat{Y}_{N\times 2}$ and $\mathrm{Y}_{\mathrm{N}\times 2}$ will be small. If mismatches exist, their corresponding distances will be distant from the majority. Therefore, mismatch elimination is transferred to a one-dimensional outlier detection problem. Based on this, a Minimum Covariance Determinant (MCD) is applied in this paper to detect and exclude mismatches.

https://doi.org/10.1017/S0373463317000224 Published online by Cambridge University Press

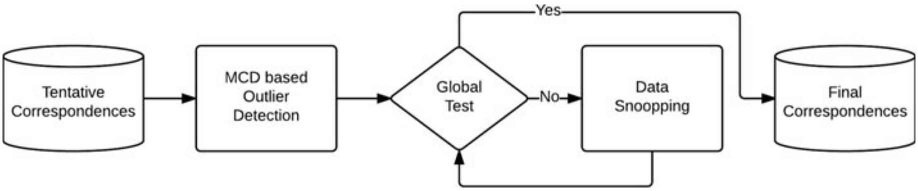


Figure 5. The validation processes.

The details of MCD can be found in Humenberger et al. (2010). Other outlier detection algorithms can still be applied.

After mismatch detection and exclusion based on MCD, all the obtained matching pairs $Pair_i = \{(x_i, y_i) \mid i = 1, 2, \ldots, N$ can be modelled in a linear form based on affine transformation as the functional model, and its corresponding stochastic model as illustrated below:

$$Ax = l + v \tag{9}$$

$$\Sigma = \sigma_o^2 \mathrm{Q} = \sigma_o^2 \mathrm{P}^{-1} \tag{10}$$

where x represents unknowns in the affine transformation matrix, l is the $2N \times 1$ observation vector composed of y_i, v is the residual vector and A is the $2N \times 6$ design matrix coming from x_i. For the stochastic model, σ_o^2 is the *a priori* variance factor, Q is the $2N \times 2N$ cofactor matrix, and P is the $2N \times 2N$ weight matrix.

To evaluate the quality of correspondences, a two-step outlier detection procedure including global test and data snooping is conducted as shown in Figure 5.

The global test will verify the global consistency between the observation and the model including functional model and stochastic model. The test statistic can be formulated as (Knight et al., 2010):

$$\frac{f\,\hat{\sigma}_0^2}{\sigma_0^2} = \frac{v^T P v}{\sigma_0^2} \sim \chi^2_{1-\alpha,f} \tag{11}$$

where f is a number for redundancy which equals to $2N - 6$, $\hat{\sigma}_0^2$ is the *posteriori* variance factor, σ_0^2 is the *a priori* variance factor, v is the residual vector and P is the weight matrix. The local test (data snooping) shown in Equation (12) is to further find the faulty observations according to outlier statistics:

$$w_i = \frac{e_i^T P v}{\sigma_0 \sqrt{e_i^T P Q_v P e_i}} \sim N(0, 1) \tag{12}$$

where e_i is a $n \times 1$ vector containing zeros but one is in the corresponding position for the assumed outlier. If the test statistic is larger than the critical value determined by the confidence level, the observations may be faulty.

6. EXPERIMENTS AND ANALYSIS.

A pre-calibrated Digital Single Lens Reflex (DSLR) camera Canon 450D was used to capture the images. Three data sets, which

https://doi.org/10.1017/S0373463317000224 Published online by Cambridge University Press

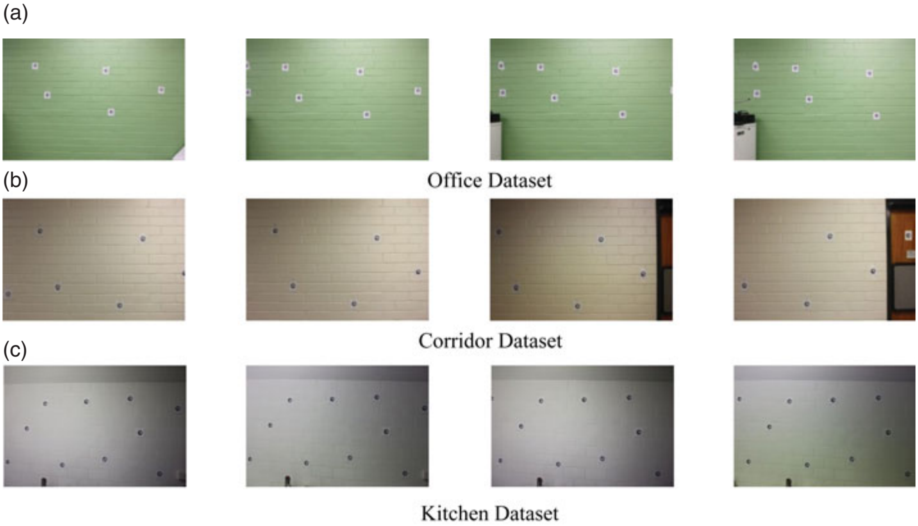


Figure 6. Samples of collected datasets

were named as Office (94 images), Corridor (66 images) and Kitchen (27 images) respectively, were collected in large weakly textured regions inside the UNSW Civil Engineering Building, which were shown in Figure 6.

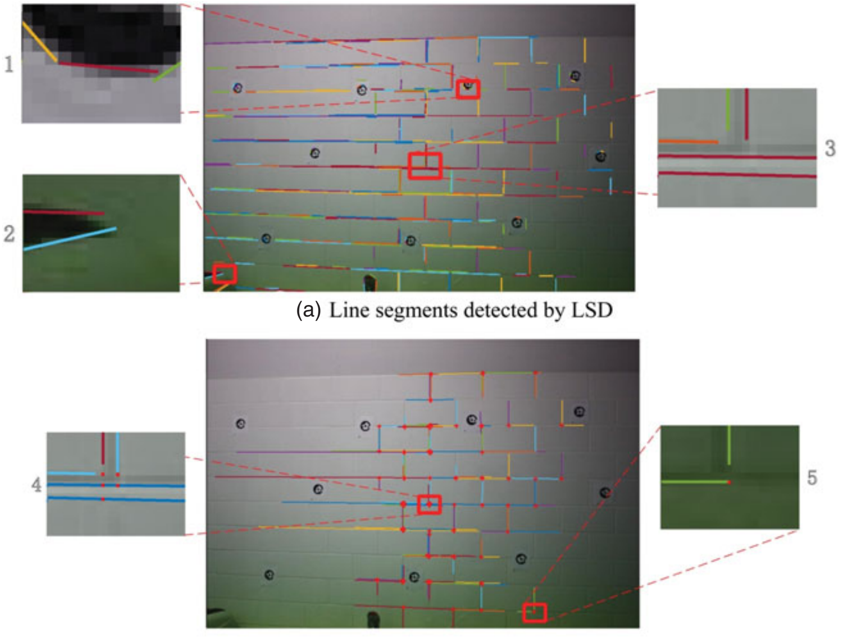
Each image contained a set of targets (e.g. Ground Control Points - GCPs) whose image coordinates and world coordinates were precisely known, which could be applied as the input to generate ground truth such as affine parameters and navigation solutions. It should be noted that in the mapping stage, the GCPs are used only to transfer the geo-information to the matched LSIFs by bundle adjustment. In the navigation stage (Sections 6.3 and 6.4), the LSIFs detected on these coded targets are removed. Therefore LSIFs are employed as natural features for navigation.

6.1. *Illustrative example for LSIF Keypoint Matching Algorithm.* This section aims to provide an insight into the proposed LSIF detection, description, matching and validation algorithm with an example.

As shown in Figure 7(a), in total 230 line segments were detected. Depending on the LSD's performance, some vertical line segments laid on the left half of the images were not detected, potentially reducing the number of detected LSIFs. However, the detected line segments still capture the main structure. It was observed that there still existed small and chaotic line segments that did not contribute much in capturing the main structure. Therefore the line segments, whose lengths were less than $0{\cdot}02H$ and intersection angles were out of the range between 80 and 100 degrees, were excluded, where H was the height of image in pixels. For example, the line segments detected in sub-figure 1 of Figure 7(a) were removed as their length were less than $0{\cdot}02H$. Sub-figure 2 was also excluded due to the intersection angle being less than 80°. Only the line segments that met the aforementioned conditions in Section 2 were preserved. Sub-figure 3 was an example. Sub-figures 4 and 5 of Figure 7(b) were typical detected LSIF. Finally 102 LSIFs from 98 preserved line segments were detected.

With the designed detector, 102 and 120 LSIFs were detected for the two images respectively, as shown in Figure 8. In Figure 8(a), with the designed descriptor and coarse

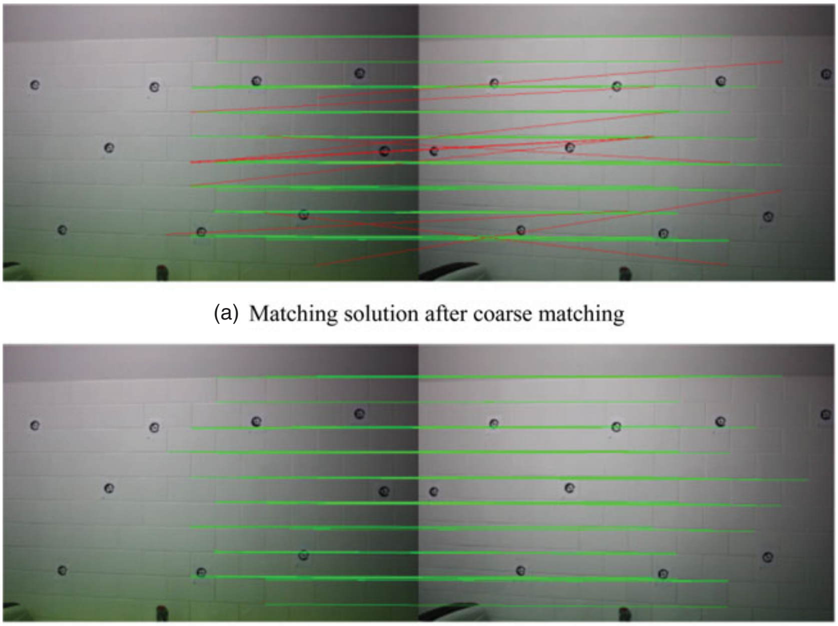
https://doi.org/10.1017/S0373463317000224 Published online by Cambridge University Press



(a) Line segments detected by LSD

(b) Filtered line segments and detected LSIFs

Figure 7. Line segment detection, filtering line segments and detected LSIFs



(a) Matching solution after coarse matching

(b) Matching solution after fine matching

Figure 8. Matching solution after coarse matching and fine matching.

https://doi.org/10.1017/S0373463317000224 Published online by Cambridge University Press

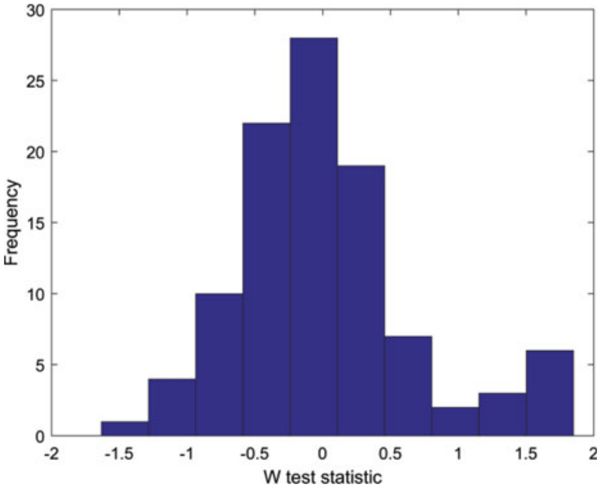


Figure 9. The histogram of W test statistics in the validation stage.

matching algorithm, 56 of 66 detected LSIFs were correctly matched. But there were still ten mismatches. The coordinates of tentatively matched keypoints were then employed as the input for the modified CPD-based matching approach. Using Shape Context and LSIF validation, the mismatches could be completely eliminated as illustrated in Figure 8(b). The degree of freedom after fine matching was 96 according to Equation (11). If the confidence level was set as 99·5%, the critical value should be 135·433, since the calculated global test statistic (42.275) was less than the critical value. The global test showed that the correspondences were consistent with the affine transformation model. W tests were also conducted for each matched keypoint pair, and the histogram of W test statistics is illustrated in Figure 9. The critical value was set as 3·29 if the confidence level was 99·9%. Since W test statistics for each correspondence were less than 3·29, it was validated that there were no mismatches in the generated correspondences.

6.2. *Matching Performance Comparison and Analysis.* To compare the performance of LSIF with LICF and SIFT, 30 pairs of images from the aforementioned datasets were employed as the testing data. The algorithms shown in Table 1 included three parts: LICF-based approaches (Algorithms 1–4), LSIF-based approaches (Algorithms 5–8) and SIFT-based approaches (Algorithm 9–12), which all applied different description and matching strategies. All the tentatively matched keypoints were refined by RANSAC algorithm based on epipolar constraints except Algorithms 4, 8 and 12. Algorithms 3, 7 and 11 directly used the detected keypoints' coordinates as the input for original affine CPD matching approach. In the following paragraphs, all the algorithms are represented by Detector/Descriptor for better readability. For example, Algorithm 8 is represented by LSIF/PCD.

The matching performances are compared in five aspects. NoF represents the number of matched keypoints after RANSAC or corresponding validation approaches. NoM represents the number of mismatches. Matching accuracy is quantified by Average Transformation Error (ATE) illustrated in Equation (13):

$$E_{Trans} = \frac{1}{2N}\sum_{i=1}^{N}(d(x_i', Fx_i) + d(x_i, F^T x_i')) \tag{13}$$

https://doi.org/10.1017/S0373463317000224 Published online by Cambridge University Press

Table 1. Matching performance comparison for LICF based approaches, LSIF based approaches and SIFT based approaches.

Algorithm No.	Detector	Descriptor	Matching Algorithm	NoF	NoM	Continuity	ATE (SD) (Pixel)	Processing Time (s)
1	LICF	Shape Context (SC)	χ^2 test statistic	2179	532	96·67%	0·193 (0·194)	221·109
2	LICF	SIFT	L_2 norm	1520	68	100%	0·182 (0·181)	201·773
3	LICF	Coordinates (CO)	CPD matching	1241	52	70·00%	0·470 (1·459)	209·773
4	LICF	Proposed Combined Descriptor (PCD)	Proposed Matching algorithm	2043	25	100%	0·156 (0·142)	254·851
5	LSIF	Shape Context (SC)	χ^2 test statistic	3118	488	100%	0·201 (0·330)	57·320
6	LSIF	SIFT	L_2 norm	2347	29	100%	0·196 (0·207)	71·957
7	LSIF	Coordinates (CO)	CPD matching	3831	134	76·67%	0·274 (0·399)	42·540
8	LSIF	Proposed Combined Descriptor (PCD)	Proposed Matching algorithm	3138	0	100%	0·159 (0·145)	117·174
9	SIFT	Shape Context (SC)	χ^2 test statistic	2396	1160	86·67%	0·449 (0·304)	783·881
10	SIFT	SIFT	L_2 norm	1238	1213	26·67%	0·215 (0·250)	49·397
11	SIFT	Coordinates (CO)	CPD matching	1941	793	66·67%	0·688 (0·638)	125·248
12	SIFT	Proposed Combined Descriptor (PCD)	Proposed Matching algorithm	3036	303	96·67%	0·331 (0·426)	981·275

https://doi.org/10.1017/S0373463317000224 Published online by Cambridge University Press

where N is the number of correctly matched keypoint pairs and x_i and x_i' are image coordinates of correctly matched keypoints respectively. F is the estimated fundamental matrix from RANSAC using the correctly matched keypoints. $d(x_i', Fx_i)$ represents the epipolar distance from the keypoint x_i' to epipolar line. Continuity is defined as the ratio between the number of image pairs where at least four keypoints are finally matched and the number of all the image pairs. The threshold for continuity is set as three because if the number of matched keypoints is equal or less than three, the solution for the camera's position cannot be unique (Thompson, 1966). The processing time is also compared.

LICF/PCD shows superior performances with the largest NoF and the smallest NoM. However, merely using a global structure descriptor such as Shape Context (LICF/SC) does not contribute to improving continuity and eliminating mismatches. Because LICF detection involves the calculation of two curves as the boundary, which takes more time, the processing time for LICF-based approaches are larger than for LSIF-based approaches.

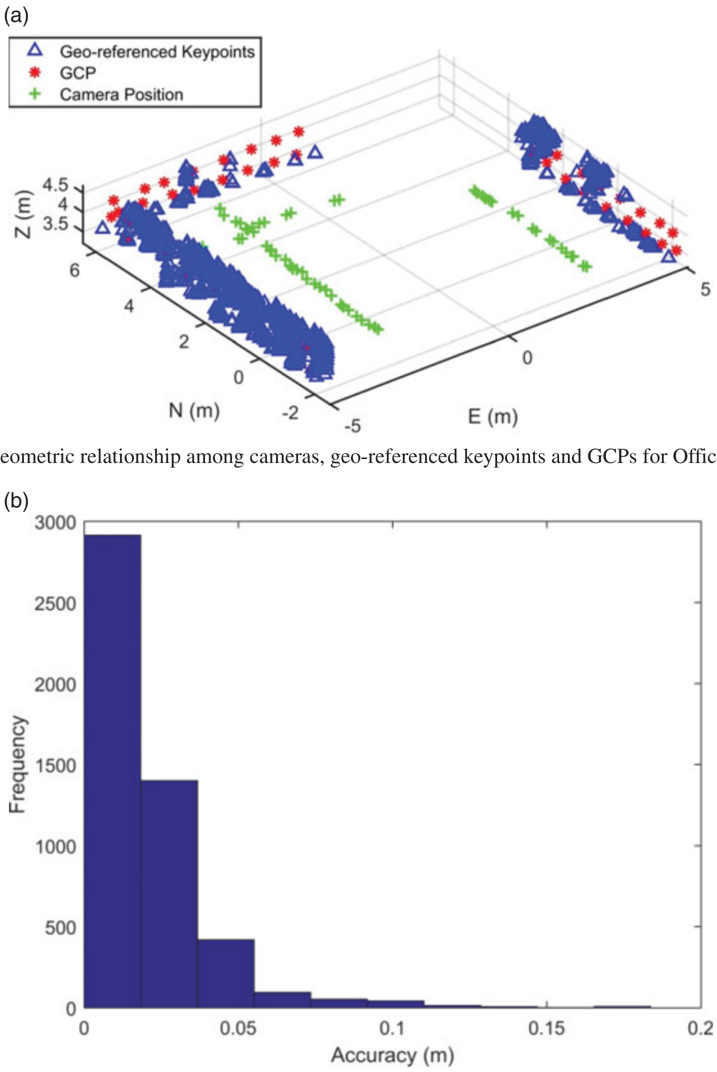
NoFs of LSIF-based approaches are larger than those of LICF-based approaches. In terms of NoM, the LSIF-based approach shows similar performances to LICF-based approaches. However, LSIF/PCD does not contain any mismatches. The two groups' matching accuracy is similar, which are all around one pixel. LICF/PCD and LSIF/PCD have the highest matching accuracy. LICF/SC, LICF/CO and LSIF/CO contained images whose numbers of matched keypoints are less than three, which are caused by the limited ability of description and matching.

NoFs for SIFT-based approaches are similar to those of LICF-based approaches. However, NoMs of SIFT-based approaches are the largest among the three groups, which is caused by the deteriorated distinctiveness of the descriptor. Also, their continuity performances are the worst among the three groups. Using SIFT/SIFT, the NoFs for most image pairs are fewer than four. Therefore it has the minimum continuity. SIFT/PCD has the largest NoF among the four algorithms but its NoM is larger than those of LICF/PCD and LSIF/PCD.

The NoM for LSIF/PCD is the smallest among all the algorithms, showing the superiority of the proposed description and matching approach. Both NoF and ATE for LSIF/PCD are the second best among all the algorithms. The processing time for LSIF/PCD is moderate and acceptable and it still can be optimised. The comparison demonstrates that LSIF has the most balanced performances over the five studied aspects.

The time complexity analysis for LSIF/PCD is as follows. Assume the number of pixels in the left and right image is N_p pixels. In LSIF detection, the time complexity of line segment detection is $O(N_p)$ (Von Gioi et al., 2008). The computation time for filtering and intersection is proportional to the number of detected line segments. Assume there are m_L and m_R detected LSIFs respectively ($m_L > m_R$). In the description stage, the computation time for SIFT and Shape Context is proportional to m_L and m_R respectively. In the coarse matching stage, the complexity of Hungarian optimisation is $O({m_L}^3)$. In the fine matching stage, assume there are n_f tentative correspondences to be further matched. The computation time for the modified CPD algorithm is $k_{EM}O(n_f^2)$ if the EM process converges after k_{EM} iterations. In the validation stage, assume the global test is passed after k_G iterations. The time complexity is $k_GO(n_f)$. Therefore for LSIF/PCD, the processing time mainly depends on the number of detected keypoints since it directly affects the computation time of the Hungarian optimisation.

https://doi.org/10.1017/S0373463317000224 Published online by Cambridge University Press



Geometric relationship among cameras, geo-referenced keypoints and GCPs for Office

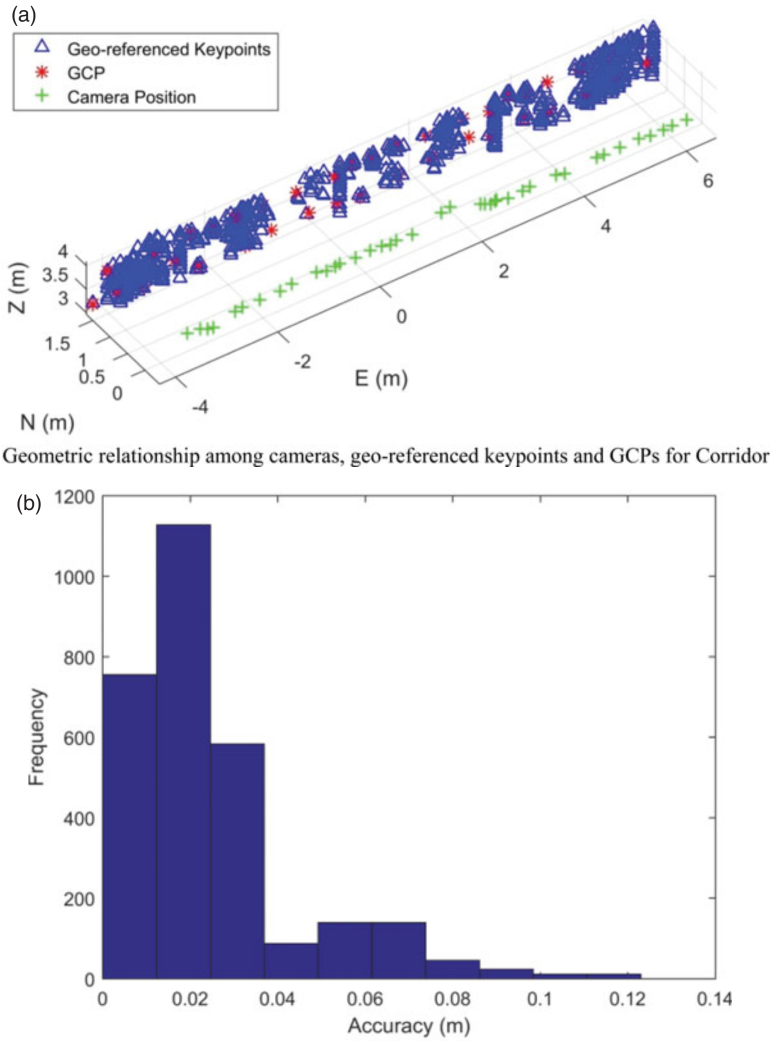
Histogram of LSIFs' accuracy for Office

Figure 10. Geometric relationship and corresponding accuracy histogram for Office

6.3. *Navigation Performance Comparison.* In the framework of navigation using reality-based 3D maps (Li et al., 2011), this section aims to compare the navigation performance of LSIF/PCD with three other competitive algorithms (LICF/PCD, SIFT/SIFT and SIFT/PCD).

Using LSIF/PCD, the geometric relationship between the camera, geo-referenced keypoints (e.g. Pseudo GCPs - PGCPs) and GCPs of the three data sets are shown in Figure 10, where the crosses show the mapping cameras' moving path, the asterisks are the manually configured targets which act as GCPs, and the triangles are geo-referenced LSIFs. The approximate reference is set as the planar determined by GCPs. The coarse mapping accuracy indicator is defined as the distance from the geo-referenced LSIFs to the

https://doi.org/10.1017/S0373463317000224 Published online by Cambridge University Press



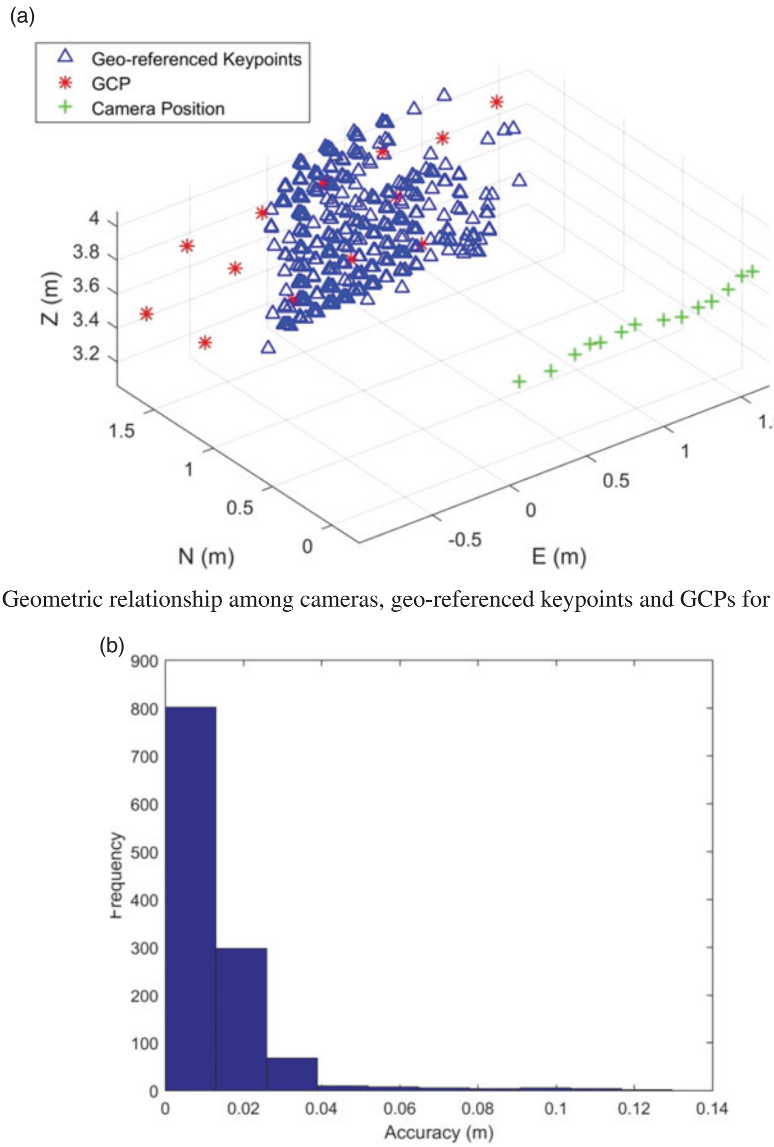
Geometric relationship among cameras, geo-referenced keypoints and GCPs for Corridor

Histogram of LSIFs' accuracy for Corridor

Figure 11. Geometric relationship and corresponding accuracy histogram for Corridor.

defined reference since the ground-truth position of each geo-referenced LSIF is difficult to obtain.

With the designed detection, description, matching and validation procedure, a total of 4968 LSIFs are geo-referenced from 59 images of the Office dataset. The LSIF's distribution and density are shown in Figure 10(a). It is noted that LSIFs cover most areas of one wall for the Office data set. There are fewer geo-referenced LSIFs for the other two walls. The reason for this is physically there are fewer line segments from the actual environment, reducing the number of matched LSIFs. It is observed that the position errors of most geo-referenced LSIF lie between −0·05 m to 0·05 m as illustrated in Figure 10(b), indicating that the mapping accuracy is promising. Some of geo-referenced LSIFs' accuracies are



Geometric relationship among cameras, geo-referenced keypoints and GCPs for Kitchen

Histogram of LSIFs' accuracy for Kitchen

Figure 12. Geometric relationship and corresponding accuracy histogram for Kitchen

larger than 0·1 m. This can be explained by the fact that the physical positions of these geo-referenced LSIFs are not strictly on the wall.

Similarly, Figure 11(a) shows the coordinates of 2972 geo-referenced LSIFs from 40 images of the Corridor dataset, and Figure 11(b) shows the geo-reference accuracy of LSIFs, demonstrating that the majority of LSIFs' position error lies in [−0·05, 0·05] m as well. Figure 12(a) illustrates the 1186 geo-referenced LSIFs from 14 images of the Kitchen. Since physically most of the geo-referenced LSIFs lie on the wall, the coarse accuracy indicator shown in Figure 12(b) is higher than the other two data sets.

https://doi.org/10.1017/S0373463317000224 Published online by Cambridge University Press

Table 2. Navigation solution accuracy of Office using LSIF.

Position and orientation	Mean	σ	Max	Min
E (m)	0·010	0·016	0·097	0·000
N (m)	0·028	0·024	0·093	0·002
Z (m)	0·020	0·023	0·093	0·000
Omega (degree)	0·524	0·592	2·405	0·004
Phi (degree)	0·083	0·073	0·270	0·002
Kappa (degree)	0·755	0·639	2·400	0·048

Table 3. Navigation solution accuracy of Corridor using LSIF.

Position and orientation	Mean	σ	Max	Min
E (m)	0·018	0·019	0·061	0·001
N (m)	0·006	0·006	0·022	0·000
Z (m)	0·011	0·013	0·051	0·002
Omega (degree)	0·356	0·407	1·632	0·058
Phi (degree)	0·065	0·047	0·195	2·027
Kappa (degree)	0·591	0·614	2·027	0·001

Table 4. Navigation solution accuracy of Kitchen using LSIF.

Position and orientation	Mean	σ	Max	Min
E (m)	0·015	0·011	0·034	0·000
N (m)	0·004	0·003	0·009	0·000
Z (m)	0·018	0·014	0·044	0·000
Omega (degree)	0·548	0·415	1·278	0·008
Phi (degree)	0·052	0·034	0·103	0·013
Kappa (degree)	0·426	0·309	1·020	0·002

It can be concluded that LSIFs have the feasibility to be correctly matched and geo-referenced. The generated reality-based 3D map acts as the resource to provide geo-spatial information for navigation. The navigation solution is obtained in an epoch-by-epoch manner.

35, 26 and 13 querying images from the three datasets are matched with the reference images of generated maps to obtain geo-referenced keypoints' world coordinates respectively. All the querying images can obtain their navigation solutions. As each querying image contains at least four GCPs with known image coordinates and world coordinates, the ground truth of the navigation solution can be obtained uniquely by space resection, which can be used for accuracy evaluation. Four statistics for navigation error including average value, standard deviation, maximum value and minimum value are calculated for the position and orientation error.

These statistics for Office, Corridor and Kitchen are illustrated in Tables 2, 3 and 4 respectively. Due to the larger number of matched keypoints with favourable distribution, most position errors are less than 5 cm, and the largest error is less than 10 cm. All the errors in orientation generally were less than 3°.

LSIF's navigation performance is compared with LICF/PCD, SIFT/SIFT and SIFT/PCD in five aspects, namely continuity, number of PGCPs (NoPGCP), availability, position error and orientation error. The definition of continuity is the same as that in Section 6.2. NoPGCP reflects the geometric strength of navigation. Based on the typical accuracy in

Table 5. Continuity and average number of PGCPs for the four algorithms.

	SIFT/SIFT		SIFT/PCD		LICF/PCD		LSIF/PCD	
Datasets	Continuity	NoPGCP	Continuity	NoPGCP	Continuity	NoPGCP	Continuity	NoPGCP
Office	17·14%	6	85·71%	12	100%	12	100%	82
Corridor	6·67%	6	53·33%	14	60%	20	100%	51
Kitchen	0%	0	84·61%	9	100%	17	100%	53

position and orientation, it is reasonable to consider the epochs whose position and orientation errors are less than 0·1 m and 5° respectively are available for indoor navigation. Therefore, the availability is defined as the ratio between the number of epochs whose navigation solutions meet the above criteria and the number of epochs that can obtain navigation solutions. Continuity indicates whether the epoch can obtain a navigation solution or not, while on the basis of continuity, availability puts more concerns on accuracy. As it is meaningless to involve incorrect navigation solutions in the discussion about accuracy, the accuracy is calculated and compared using only available epochs. Continuity of the four algorithms with regard to the three datasets is illustrated in Table 5. Through the comparison, classical SIFT/SIFT has the lowest continuity and NoPGCP due to the weak texture. SIFT/PCD and LICF/PCD have slightly higher continuity and NoPGCP. Using LSIF/PCD we can obtain a navigation solution for all the epochs, and its NoPGCP is the highest among the four algorithms.

As illustrated in Table 6, SIFT/SIFT completely loses availability. Therefore its position and orientation error are not applicable. The availability's variation for SIFT/PCD and LICF/PCD depends on the number of correctly matched keypoints. The availability of LSIF/PCD for all the datasets is 100%. The accuracy of available epochs for SIFT/PCD, LICF/PCD and LSIF/PCD are similar.

Continuity is closely related to the number of correctly matched keypoints. The accuracy of navigation solutions are influenced by a number of factors, such as the geometry between geo-referenced keypoints and camera, the accuracy of geo-referenced keypoints' world coordinates in the mapping stages, and the matching accuracy of keypoint matching algorithms. It could be concluded that by using LSIFs as the environmental feature, most navigation solutions could achieve centimetre-level accuracy in positioning, and the accuracy of orientation could be controlled to a few degrees.

6.4. *Navigation Performance Evaluation Based on the Trajectory.* This section aims to evaluate the performance of LSIF/PCD in a comparatively larger indoor environment. A total of 152 mapping images and 277 querying images were collected from a lobby with weak texture in the Civil Engineering Building on the University of New South Wales (UNSW) campus.

The 2D trajectory of navigation solutions and the simplified surrounding environment is illustrated in Figure 13. The 2D trajectory is consistent with the reality of the motion. All the epochs can obtain their navigation solutions, which means the continuity for LSIF is 100%. On average, 91 LSIFs can be matched on each querying image. If the same definition of availability as that in Section 6.3 is followed, its availability is 97.47% as the navigation solutions of eight epochs are not available. However, their position error is slightly larger in only a certain direction (e.g. E direction) and navigation solutions are still reasonable. One possible reason is the weak geometric distribution of PGCPs in the mapping stage.

https://doi.org/10.1017/S0373463317000224 Published online by Cambridge University Press

Table 6. Availability, position error and orientation error for the four algorithms.

Dataset		Office	Corridor	Kitchen
SIFT/SIFT	Position (m)	N/A	N/A	N/A
	Orientation (degree)	N/A	N/A	N/A
	Availability	0%	0%	0%
SIFT/PCD	Position (m)	0·041	0·065	0·031
	Orientation (degree)	1·025	2·741	0·932
	Availability	26·67%	12·50%	81·82%
LICF/PCD	Position (m)	0·023	0·013	0·024
	Orientation (degree)	0·437	0·406	0·731
	Availability	20%	66·67%	30·77%
LSIF/PCD	Position (m)	0·019	0·011	0·012
	Orientation (degree)	0·454	0·337	0·344
	Availability	100%	100%	100%

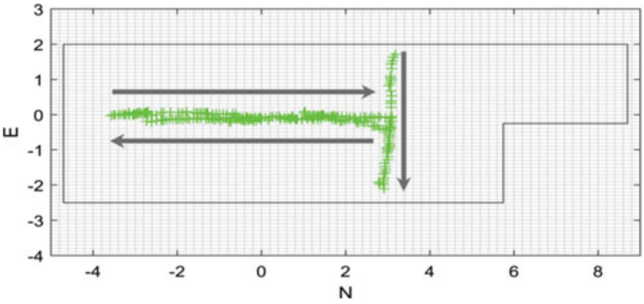


Figure 13. Aerial view of the camera's trajectory

Table 7. Navigation accuracy of position and orientation for the large indoor environment.

Position and orientation	Mean	σ	Max	Min
E (m)	0·005	0·010	0·086	0·000
N (m)	0·009	0·013	0·078	0·000
Z (m)	0·007	0·011	0·071	0·000
Omega (degree)	0·176	0·291	2·185	0·001
Phi (degree)	0·027	0·036	0·370	0·000
Kappa (degree)	0·237	0·365	2·168	0·001

Similarly, the accuracy of all the available navigation solutions can be evaluated using the GCPs, which is illustrated in Table 7. The accuracy is similar with that in Section 6.3, where position and orientation errors are limited to a few centimetres and degrees, respectively.

7. CONCLUDING REMARKS.

This paper has proposed a new environmental line feature named as LSIF and its corresponding detection, description, matching and validation algorithms to overcome the continuity and correctness challenges in weakly textured areas of indoor environments. The proposed LSIF outperforms other keypoints in terms of the

https://doi.org/10.1017/S0373463317000224 Published online by Cambridge University Press

number of matched keypoints and matching correctness. Under the framework of navigation using reality-based 3D maps, high accuracy, continuity and availability have been achieved, showing its feasibility and potential as a new feature for vision-based navigation.

LSIF originates from the detection of line segments so that the detection of LSIF mainly depends on the line segment detection algorithm, whose invariance to scaling and illumination is still not fully explored. Further investigation is underway to improve its repeatability.

REFERENCES

Belongie, S., Malik, J. and Puzicha, J. (2002). Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **24**, 509–522.

Hartley, R. and Zisserman, A. (2003). *Multiple view geometry in computer vision*. Cambridge University Press.

Humenberger, M., Engelke, T. and Kubinger, W. (2010). A census-based stereo vision algorithm using modified Semi-Global Matching and plane fitting to improve matching quality. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops* (CVPRW). 77–84.

Jin, Z., Wang, X., Moran, B., Pan, Q. and Zhao, C. (2016). Multi-region scene matching based localisation for autonomous vision navigation of UAVs. *Journal of Navigation*, **69**, 1215–1233.

Kim, H. and Lee, S. (2012). Simultaneous line matching and epipolar geometry estimation based on the intersection context of coplanar line pairs. *Pattern Recognition Letters*, **33**, 1349–1363.

Knight, N.L., Wang, J. and Rizos, C. (2010). Generalised measures of reliability for multiple outliers. *Journal of Geodesy*, **84**, 625–635.

Kuhn, H.W. (1955). The Hungarian method for the assignment problem. *Naval research logistics quarterly*, **2**, 83–97.

Lewis, J. (1995). Fast normalized cross-correlation. *Vision interface*, 120–123.

Li, X., Wang, J., Knight, N. and Ding, W. (2011). Vision-based positioning with a single camera and 3D maps: accuracy and reliability analysis. *Journal of Global Positioning Systems*, **10**, 19–29.

Li, Z., Wang, J., Alqurashi, M., Chen, K. and Zheng, S. (2016). Geometric analysis of reality-based indoor 3D mapping. *Journal of Global Positioning Systems*, **14**, 1.

Liu, C., Zhou, F., Sun, Y., Di, K. and Liu, Z. (2012). Stereo-image matching using a speeded up robust feature algorithm in an integrated vision navigation system. *Journal of Navigation*, **65**, 671–692.

Lowe, D.G. (1999). Object recognition from local scale-invariant features. *The proceedings of the 7th IEEE international conference on Computer vision*, 1999. 1150–1157.

Myronenko, A. and Song, X. (2010). Point set registration: Coherent point drift. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **32**, 2262–2275.

Thompson, E. (1966). Space resection: Failure cases. *The Photogrammetric Record*, **5**, 201–207.

Von Gioi, R.G., Jakubowicz, J., Morel, J.-M. and Randall, G. (2008). LSD: A fast line segment detector with a false detection control. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 722–732.

Xian, Z., Hu, X. and Lian, J. (2015). Fusing stereo camera and low-cost inertial measurement unit for autonomous navigation in a tightly-coupled approach. *Journal of Navigation*, **68**, 434–452.

Zhou, H., Zou, D., Pei, L., Ying, R., Liu, P. and Yu, W. (2015). StructSLAM: Visual SLAM With Building Structure Lines. *IEEE Transactions on Vehicular Technology,* **64**, 1364–1375.

https://doi.org/10.1017/S0373463317000224 Published online by Cambridge University Press