Keypoint Matching by Outlier Pruning with Consensus Constraint*

Yang Liu, Rong Feng and Hong Zhang

Abstract— **A simple and reliable keypoint matching method is proposed in this paper. Our research is motivated by the desire to improve the performance of multi-view geometry (MVG) based verification in visual loop closure detection under significant illumination change, where traditional methods may fail due to their inability to either find a sufficient number of correctly matched keypoints or identify correct underlying camera motion to verify the matches. Our method is inspired by research on the spatial statistics of optical flow. By observing that the displacement of matching keypoints between a pair of images is equivalent to the optical flow under the assumption of small camera motion (which is true in applications such as loop closure detection), we exploit the fact that the displacement of correctly matched keypoints between two images must follow a well-defined distribution. This paves the way to a keypoint matching method that uses this distribution to screen or prune potential matching keypoints, so as to remove the incorrect matches (outliers) and retain the true matches (inliers) without being overly and solely dependent on keypoint descriptors. The proposed method is validated on the outdoor image sequences and shows superior performance to the standard keypoint matching method based on distance ratio test.**

I. INTRODUCTION

Feature or keypoint matching between images is a basic step in many computer vision applications such as structure from motion and object recognition [1]. There are usually some invariance properties that need to be addressed in the development of a feature detector or descriptor, with respect to, for example, scale, rotation, translation and illumination. A good detector or descriptor is expected to achieve one or multiple invariance properties. In robotics however, not all the changes are equally important. For example, in visual loop closure detection by a ground robot, camera motion is small when loop closure happens. On the other hand, illumination change is a practical and challenging issue in visual robot navigation, especially in long-term mapping.

Our research in this paper addresses robust keypoint matching in visual loop closure verification in robot navigation. Verification of loop closures is usually achieved by multi-view geometry [2], with a fundamental step of finding sufficient true matching keypoints between two images. Even though the keypoint detectors and descriptors are designed to handle illumination change, keypoint matching based solely on descriptors can be still difficult when the lighting condition changes significantly. Figure 2 gives an example why illumination change can cause considerable difficulty in keypoint matching by the traditional distance ratio test. When there is no change in illumination in Figure 2(a), a majority of the true matches can be found and false matches rejected by choosing a proper threshold, as was reported in [1]. However, in the case of significant illumination change in Figure 2(c), it becomes difficult to select a threshold that finds sufficient true matches without including many false matches. A naive way to increase the number of true matches would be using a higher distance ratio, but this would introduce a large number of outliers, which causes difficulty in the RANSAC algorithm (discussed in detail later in identifying the underlying camera motion).

Exploiting small camera motion, our method resorts to geometry of matching keypoint locations as a way of finding likely matches, based on the well-established research on the spatial statistics of optical flow [3]. Since the displacement of matching keypoints between a pair of images is equivalent to optical flow under the assumption of small camera motion, these statistics can be employed as a prior to prune matches returned by the brute-force nearest neighbor method. Optical flow is an active research topic that has been extensively studied and well addressed [4]. Optical flow assumes differential motion of the image signal or, equivalently, that of the camera with respect to a (static) scene, with corresponding image points moving no more a few pixels. This assumption is usually well satisfied in the case of a loop closure when the camera position between the two separate visits is similar by definition. As a result, we are able to make use of the statistics of optical flow in visual loop closure detection. In fact, even when the two successive visits involve a camera motion with corresponding image points moving by tens of pixels, our proposed method can still work due to the planar nature of the camera motion, as will be shown later.

The paper proceeds as follows. Section II briefly discusses the related work of keypoint matching. In Section III we formulate our proposed method on the basis of the spatial statistics of optical flow and provide validation of these statistics in our application. Experimental results are shown in Section IV to compare our keypoint matching algorithm with the standard distance ratio test on outdoor image sequences involving significant lighting change, to demonstrate its superiority. We summarize the paper in Section V.

II. RELATED WORK

While appearance based robot localization and mapping has been extensively studied in recent years, the issue of loop closure verification is not widely discussed in the literature due to the existing standard techniques for solving the problem. Finding the putative matching keypoints first

*This work was supported by the Natural Sciences and Engineering Research Council (NSERC) through the NSERC Canadian Field Robotics Network (NCFRN).

Yang Liu, Rong Feng and Hong Zhang are with the Department of Computing Science, University of Alberta, Edmonton, AB, Canada {liu17, rfeng2, hzhang@cs.ualberta.ca}

and then performing RANSAC [5] for inlier identification form the two steps in the standard solution, followed by applying a simple threshold on the number or the percentage of inliers. Putative matches are typically generated among nearest neighbors using the keypoint descriptors in combination with a further constraint on the nearest neighbors. One example constraint is mutual consistency, which accepts a true matching pair only when two keypoints are the best correspondences of each other. A more widely used technique is the distance ratio test [1] where a nearest neighbor is accepted as true match when the ratio of distances to the nearest and second nearest neighbor is lower than a pre-defined threshold (0.6, for example). The idea is to emphasize the distinctness of the keypoints being matched. Apparently, both methods depend on the descriptors only and may not work well when the descriptor values change greatly such as when a scene is subject to varying lighting conditions.

Attempts from several different perspectives can be made to improve the keypoint matching performance under illumination change. Developing keypoint descriptors that are invariant to illumination change is one such attempt. This can be achieved by choosing a proper color space in which to compute the descriptors. A comprehensive study in [6] evaluated the performance of color descriptors including SIFT in the application of scene recognition with illumination change. It was shown that color SIFT generally performs better than the original SIFT descriptor, which is computed on a grayscale image.

Alternatively, accurate keypoint matching can be achieved by learning a distance metric in the feature space, where the correlations among the features are characterized in terms of the statistics of co-occurrence. This approach is based on the assumption that Euclidean distance may not be optimal in capturing the similarity between two descriptors. An example of this approach is presented in [7] with its application to visual localization under illumination change in [8].

Yet another approach to address illumination invariance is through feature selection. The idea is to retain keypoints that are distinct, representative and easy to match, and eliminate the less reliable ones that impact matching negatively. Scale dependent feature selection [9] is one example with the observation that keypoints extracted at coarse scales usually provide better matching performance than those at fine scales, although the numbers are fewer. In general, keypoint matching repeatability can be learned and used to improve the result [10].

III. KEYPOINT MATCHING BY OUTLIER PRUNING WITH CONSENSUS CONSTRAINT

The study on the spatial statistics of optical flow [3] serves as the basis of our keypoint matching method. Since correctly matched keypoints are simply a subset of the pixels whose movements between two images define the optical flow, the statistics of which naturally apply to inlier matches, and can function as a means to prune putative matches. Specifically, under the assumption of small camera motion, the horizontal and vertical components of the optical vectors are found to follow a well-defined distribution (Laplacian in this case). This distribution is used as a prior in our algorithm to remove outliers in the initial nearest neighbor matches. Since this prior is defined in the displacement space of the pixel coordinates, it is insensitive to illumination and can largely improve the robustness of a matching algorithm in case of significant illumination change.

To confirm the findings in [3], we have conducted experiments on multiple datasets with and without illumination change. The results are shown in Figure 3. Clearly and not surprisingly, the statistics of displacement of the matching features in both *x* and *y* closely follow a Laplacian distribution with a small diversity value, typically around 1.0 (detailed analysis will be discussed in Section IV). On the other hand, we have observed that the statistics of false matches follow a distribution with extremely long tails (Figure 4). Note that the study in [3] involves statistics of optical flow in multiple image pairs, whereas we deal with keypoint matching between one image pair. Fortunately, as will be shown in Section IV-B, a similar conclusion holds, with the difference only in the mean of the optical flow displacement distribution but not the Laplacian model itself. Hence the characteristic distribution of the displacements for true matches provides us a simple prior in a keypoint matching method as summarized in Algorithm 1.

Algorithm 1: Keypoint Matching with Consensus Constraint

1. For the current image I_c and a query image I_q, perform keypoint matching by nearest neighbor search.
2. Calculate Δu and Δv for each nearest neighbor of a keypoint. $\Delta u = u_q - u_c$, $\Delta v = v_q - v_c$, where (u_c, v_c) and (u_q, v_q) are the image coordinates of the matching keypoints in I_c and I_q.
3. Run a mode-seeking algorithm such as *mean shift* to find the mode of $(\Delta u, \Delta v)$ and estimate the model parameters (assuming a Laplacian distribution).
4. Based on the estimated parameters of the *pdf*, choose the decision boundary that can include a certain amount (e.g., 90%) of $(\Delta u, \Delta v)$.
5. Generate a set of candidate matching pairs PM that correspond to the selected $(\Delta u, \Delta v)$ in the previous step.
6. Performing RANSAC (if applicable) on PM to identify the inliers.

In our implementation, we first used mean shift with a Gaussian kernel to find the mode. To prevent mean shift from converging to a local maximum, we quantized all $(\Delta u, \Delta v)$ by building a 2D histogram with fixed bin width (5 pixels for example) in each dimension, and the top bin was used as the initialization. The diversity parameter in the Laplacian model was calculated using the elements that are weighted above 0.7 to the mode. In Step 4 of Algorithm 1, we can assume that the estimated model is a good approximation to the true *pdf* of $(\Delta u, \Delta v)$ corresponding to inliers and the boundary values $b_{x_1}, b_{x_2}(b_{x_1} < b_{x_2})$ and $b_{y_1}, b_{y_2}(b_{y_1} < b_{y_2})$ were selected symmetrically with respect to the mode m_x and m_y in x and y direction respectively, i.e., $b_{x_2} - m_x = m_x - b_{x_1} = \sigma_x$, $\int_{b_{x_1}}^{b_{x_2}} L_x dx = 0.9$ and $b_{y_2} - m_y = m_y - b_{y_1} = \sigma_y$, $\int_{b_{y_1}}^{b_{y_2}} L_y dy = 0.9$, where L_x and L_y are the estimated *pdf* of the Laplacian models in x and y direction respectively. We then used the elements that fall into the elliptical area with the axles σ_x and σ_y, and the center (m_x, m_y) to form PM, assuming that there is no correlation between the two directions (covariance is 0). In other words, all the $(\Delta u, \Delta v)$ that satisfy $\frac{(\Delta u - m_x)^2}{\sigma_x^2} + \frac{(\Delta v - m_y)^2}{\sigma_y^2} = 1$ were considered and their corresponding matches were selected.

Apparently, our algorithm does not need to solve a chicken-and-egg problem as RANSAC does. We refer to this constraint of statistics as *consensus constraint*, in contrast to distance ratio or mutual consistency constraint. In the verification of loop closures, I_c and I_q match if there is a sufficient number of inliers t_{in} in PM determined by RANSAC.

IV. EXPERIMENTAL RESULTS

To evaluate our proposed keypoint matching method, we conducted comprehensive experiments on a dataset in changing illumination environments. We collected the dataset with five image sequences at five different times of a day in a campus environment. The images were collected with a Husky A200 mobile robot equipped with a Xtion Pro camera. In each sequence the robot was driven along the same trajectory of approximately 700 meters at the speed of $1m/s$. Therefore, the image numbers of each sequence can be used in building the ground truth of positive loop closures. Several weather and illumination conditions are covered in this dataset, including cloudy, sunny, rainy and night. Figure 1 shows a set of matching images in the five sequences representing the same location.

A. Distance Ratio in Illumination Change

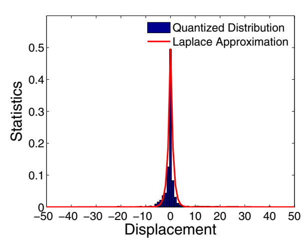
Our first experiment was to determine the extent to which the matching performance of the traditional distance ratio test is affected by illumination change. We examine three cases of change: no illumination change between cloudy1 and cloudy2, minor change between rainy and sunny, and significant change between sunny and night. For each case, we randomly selected 100 matching image pairs and studied the statistics of inliers and outliers returned by RANSAC with respect to distance ratio *dr*. Results are shown in Figure 2. As can be seen, with increasing illumination change, it becomes difficult or impossible to choose a *dr* that prunes most of the outliers while retaining most of the inliers. In other words, a higher number of inliers by increasing *dr* comes only at the expense of a higher number of outliers, at a level that causes difficulty to RANSAC as will be shown later.

B. Spatial Statistics of Matching Keypoints

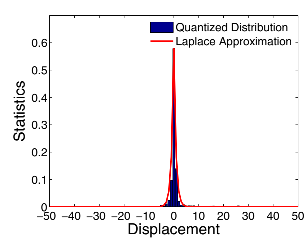
In next experiment, we wanted to verify that the spatial statistics of optical flow hold true for matching keypoints. Specifically, we are able to show that displacement of the matching keypoints in both x and y coordinates follow a Laplacian distribution with a small diversity value around 1.0. As mentioned, this distribution serves as a prior for our method to prune nearest neighbor matches, with the important property that it is independent of illumination change.

We used the extreme case of matching between dark night and sunny day (see Figure 1 (d) and (e)) as they are the most challenging situations. From all the matching image pairs, we randomly selected 100 of them in this study. In each selected pair, the keypoint displacement distributions of true matches along x and y image coordinates were considered respectively. All the displacements of matching keypoints were collected and the displacement histograms were constructed. The histograms were then normalized to have a zero mean, and averaged over the 100 image pairs in both x and y, to produce the final histograms of displacement and represent the spatial statistics.

Figure 3 (a) and (b) show the normalized histograms of keypoint displacement in x and y coordinates respectively. Also shown in the figures are the fitted Laplacian distributions (in this example, the diversity parameter $b_x = 1.0$ and $b_y = 0.86$). This result is consistent with the findings in [3] and confirms the critical assumption of the proposed method. In addition, we repeated the procedure multiple times and obtained similar results as in Figure 3.



(a) x coordinate



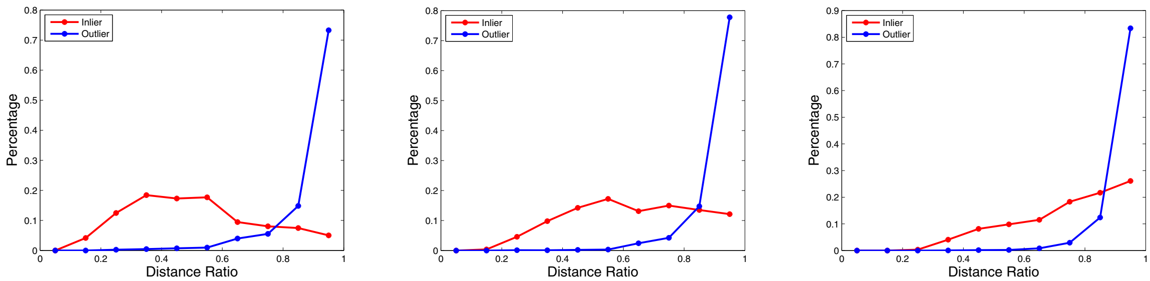
(b) y coordinate

Fig. 3. Keypoint displacement distribution for inliers. Both statistics along x and y can be fitted to a Laplacian distribution.

On the other hand, for our method to work well, outliers must not have similar displacement statistics to those of the inliers in order for the statistics to be useful in a pruning technique. By considering the displacement distribution of false matches to be similar to that of the distance between two random points in a rectangle, a closed-form solution to this distribution in fact exists [12], which is unimodal with a large variance, and far from the peaked Laplacian distribution with a small diversity value. As a result, we can expect that the keypoint displacement distribution of outliers covers a wide range with long tails. The final distribution by nearest neighbor matches is a mixture of the two: the Laplacian followed by the inliers, and the one defined in [12] followed by the outliers. Figure 4 indicates one example of statistics

(a) Cloudy1 (b) Cloudy2 (c) Rainy (d) Sunny (e) Night

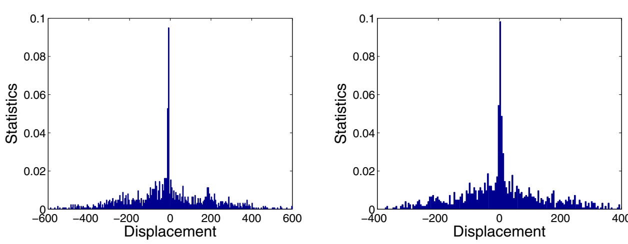
Fig. 1. Sample images at the same location in five sequence, cloudy (two cases), rainy, sunny after the rain and dark night at 10 o'clock.



(a) No change (b) Minor change (c) Significant change

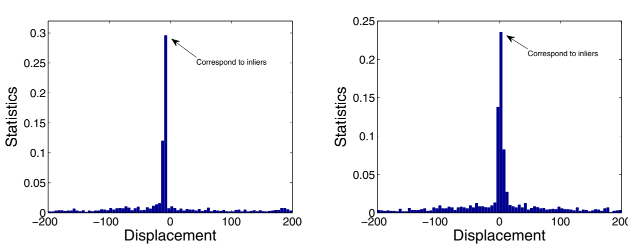
Fig. 2. Inlier/Outlier percentage w.r.t dr in different illumination conditions. As illumination changes more obviously, it becomes more difficult to find a sufficient number of inliers with a low threshold. Using a higher threshold will introduce more outliers.

of outliers along x and y image coordinates, and the mixture statistics for both inliers and outliers are shown in Figure 5, where the peak bin(s) that correspond to a majority of inliers (and possibly a few outliers as well) can be identified. In fact, we speculate that the center bins in Figure 4 contain both inliers and outliers since obtaining perfect ground truth matches is difficult.



(a) x coordinate (b) y coordinate

Fig. 4. Keypoint displacement statistics for outliers in one image pair. The statistics of outliers can be fitted to a distribution with long tails.



(a) x coordinate (b) y coordinate

Fig. 5. Mixture statistics for both inliers and outliers. The peak bin(s) correspond to a majority of inliers, and possibly with a few outliers.

Another important fact as we can observe is that, as shown in Figure 4 and Figure 5, the Laplacian model in general holds true for every single matching image pair, with different parameters in mean and diversity, implying that our method can and should be directly applied online without obtaining any pre-trained parameters. The key is to find the correct model and parameter estimation for each case.

C. Performance Comparison with Significant Illumination Change

We then compared the proposed method with traditional keypoint matching algorithms, including nearest neighbor matching and distance ratio test. Nearest neighbor matching usually provides the highest number of true matches, although the result contains a large number of outliers. In applications such as loop closure detection, RANSAC can be used to exclude most outliers. However, in the case of non-matching images, RANSAC would still find a geometric transformation between two images and is therefore unable to reject the non-matching images. In other words, with these false matches, RANSAC can also find sufficient support to construct the camera motion. The consequence is that the trade-off between precision and recall of loop closure verification can be extremely difficult to make. In such a case, a pruning technique of removing the false matches is highly desired.

We randomly selected 500 true matching and non-matching image pairs respectively in this round of experiment. The non-matching pairs were selected so that they represent different locations but still share some similar structures such as tree and building, and hence are possible to be considered matching locations by a loop closure detection algorithm. Using nearest neighbor matching, distance ratio test ($dr = 0.6$) and our own method to generate matches

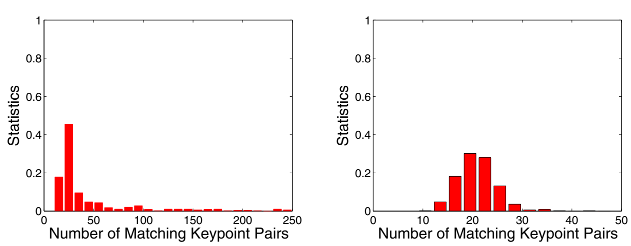
TABLE I

AVERAGE NUMBERS OF MACHING KEYPOINTS IN THREE METHODS. NEAREST NEIGHBOR (NN), DISTANCE RATIO (DR) AND CONSENSUS CONSTRAINT (CC)

		NN	DR	CC
Campus	Positive	47	17	66
	Negative	20	2	4

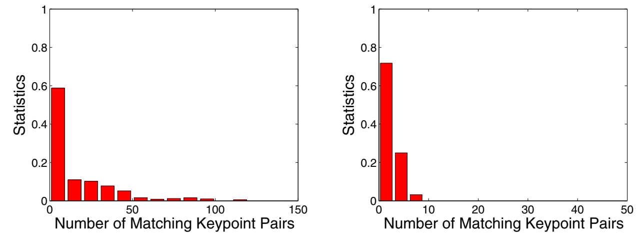
as the input to RANSAC, we obtained the statistics of the number of matches returned by RANSAC, for both postive and negative 500 cases (For the pruning techniques, if the input does not contain sufficient putative matches, all the matches before RANSAC will be considered). Figure 6 show that with significant illumination change, nearest neighbor matching does not work well for false cases since a significant number of matches can be returned for many non-matching image pairs and therefore, it is almost impossible to choose a threshold that can well separate the positive and negative cases in loop closure detection. Our proposed method significantly outperforms distance ratio test. For example, we can observe from Figure 8(a) that a majority of the positive image pairs have enough matches, which is not the case in distance ratio method shown in Figure 7(a). On the other hand, the proposed method can perform as well as distance ratio in identifying the non-matching cases by returning only a few matches as shown in Figure 8(b).

Precision-recall curves are shown in Figure 9, where at 100% precision, consensus constraint surpasses distance ratio by more than 30% in the recall value. We also observed that when using nearest neighbor matching, the precision-recall curve is much worse because outliers can be accepted by RANSAC, although this seldom happens in the other two methods. As mentioned, this is possibly due to deficient model (fundamental matrix) created by RANSAC with a large portion of outliers. It further emphasizes the importance of outlier pruning. Table I describes the average number of matching keypoints for the three methods. We can see that with significant illumination change, the proposed method finds four times as many matches as the distance radio method, and also even more true matches than nearest neighbor matching, which further confirms a side effect of outliers in RANSAC.



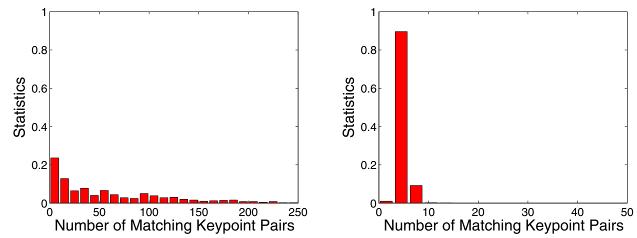
(a) Positive cases (b) Negative cases

Fig. 6. Number of matches returned by RANSAC – nearest neighbor. The postive and negative cases are not distinguishable.



(a) Positive cases (b) Negative cases

Fig. 7. Number of matches returned by RANSAC – distance ratio with $dr = 0.6$. 60% postive cases have only a few matches due to the significant illumination change.



(a) Positive cases (b) Negative cases

Fig. 8. Number of matches returned by RANSAC – consensus constraint. Most positive cases have tens of or more matches.

Regarding the time of the proposed method, the only time-consuming part is the iterative mode-seeking algorithm. However, as we already provided a good initialization by histogram binning, mean shift usually converges in a few steps to the global maximum. Using MATLAB implementation on a normal laptop, the proposed method completes the outlier pruning in about $0.01s$ on 3000 initial matches provided by nearest neighbor matching. This satisfies the real-time requirement in most robotics applications. We will make comprehensive comparison between our method and the state-of-the-art in terms of efficiency in our future work.

Figure 10 gives an example of a positive image pair. In spite of the “night and day” difference in illumination, our method successfully finds corresponding keypoint pairs in the tree, while distance ratio cannot return any true match.

V. CONCLUSION AND DISCUSSION

We have presented a simple and effective method for keypoint matching. The method can be used in computer vision and robotics applications such as structure from motion and loop closure detection, where camera motion between two frames is assumed to be relatively small. The key to the method lies in finding these correspondences in spite of variation in the keypoint descriptors, which in many cases cannot be handled by the traditional matching techniques based on distance ratio test, and simply increasing the distance ratio to include more inliers would not fix the problem because it will also introduce too many outliers to prevent RANSAC from rejecting the non-matching images. The proposed technique is based on the fact that spatial statistics of optical flow in general follows a Laplacian distribution in terms of image coordinates and most importantly, these

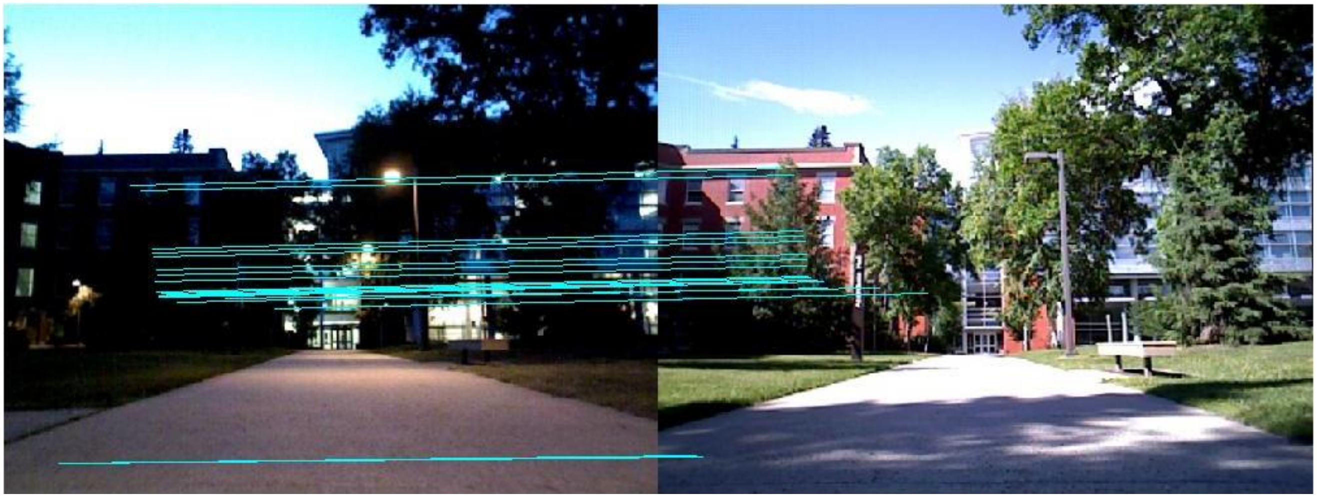


Fig. 10. An example of keypoint matching with consensus constraint in the case of a loop closure with dramatically different lighting conditions. Some true matches are found on the trees and the building.

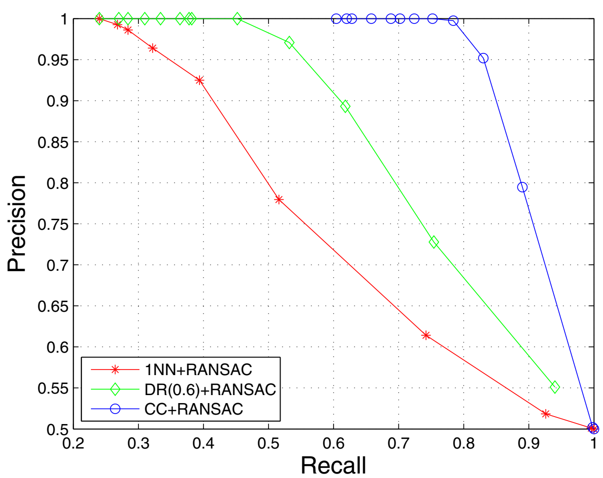


Fig. 9. Precision-recall curves of the three methods. Under illumination change, the proposed method significantly outperforms distance ratio test by more than 30% at 100% precision. Both pruning techniques are superior to nearest neighbor matching without pruning.

statistics are invariant with respect to illumination change. Therefore, outliers can be easily identified and discarded by using the spatial statistics or the consensus constraint. Through experiments, we have shown that in the case of significant illumination change, our method significantly improves the keypoint matching performance, leading to an obvious increase of the recall by above 30% in verifying true loop closures at 100% precision in a custom dataset involving significant illumination change.

In the future, we will compare our method with the most recent techniques that have been developed in dealing with similar problems. One of the representatives is vector field consensus [13], which identifies the inliers by vector field interpolation using *EM* algorithm. Both the matching performance and efficiency will be included in our comparison.

REFERENCES

[1] D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints”, *International Journal of Computer Vision*, 60(2), pp. 91-110, 2004.

[2] A. Angeli, D. Filliat, S. Doncieux, *et al.*, “Fast and Incremental Method for Loop-Closure Detection Using Bags of Visual Words”, *IEEE Transactions on Robotics*, 24(5), pp. 1027-1037, 2008.

[3] S. Roth M. J. Black, “On the Spatial Statistics of Optical Flow”, *IEEE International Conference on Computer Vision*, Beijing, China, pp. 42-49, 2005.

[4] J. L. Barron, D. J. Fleet and S. S. Beauchemin, “Performance of Optical Flow Techniques”, *International Journal of Computer Vision*, 12(1), pp. 43-77, 1994.

[5] M. A. Fischler and R. C. Bolles, “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography”, *Communication of the ACM*, 24(6), pp. 381-395, 1981.

[6] K. E. A. van de Sande, T. Gevers and C. G. M. Snoek, “Evaluating Color Descriptors for Object and Scene Recognition”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9), pp. 1582-1596, 2010.

[7] A. Mikulík, M. Perdoch, O. Chum, *et al*, “Learning a Fine Vocabulary”, *European Conference on Computer Vision*, Crete, Greece, 2010.

[8] A. Ranganathan, S. Matsumoto and D. Ilstrup, “Towards illumination invariance for visual localization”, *IEEE International Conference on Robotics and Automation*, Karlsruhe, Germany, pp. 3791-3798, 2013.

[9] H. Zhang, “BoRF: Loop-Closure Detection with Scale Invariant Visual Features”, *IEEE International Conference on Robotics and Automation*, Shanghai, China, pp. 3125-3130, 2011.

[10] W. Hartmann, M. Havlena and K. Schindler, “Predicting Matchability”, *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, USA, 2014.

[11] M. Cummins and P. Newman, “FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance”, *The International Journal of Robotics Research*, 27(6), pp. 647-665, June 2008.

[12] J. Philip, “The probability distribution of the distance between two random points in a box”, *TRITA MAT*, 7(10), December 2007.

[13] J. Ma, J. Zhao, J. Tian, A. L. Yuille and Zhuowen Tu, “Robust Point Matching via Vector Field Consensus”, *IEEE Transactions on Image Processing*, 23(4), pp. 1706-1721, 2014.