Unsupervised Learning of Threshold for Geometric Verification in Visual-Based Loop-Closure

Gim Hee Lee, and Marc Pollefeys
Computer Vision and Geometry Lab, ETH Zürich, Switzerland
{glee@student, marc.pollefeys@inf}.ethz.ch

***Abstract*— A potential loop-closure image pair passes the geometric verification test if the number of inliers from the computation of the geometric constraint with RANSAC exceed a pre-defined threshold. The choice of the threshold is critical to the success of identifying the correct loop-closure image pairs. However, the value for this threshold often varies for different datasets and is chosen empirically. In this paper, we propose an unsupervised method that learns the threshold for geometric verification directly from the observed inlier counts of all the potential loop-closure image pairs. We model the distributions of the inlier counts from all the potential loop-closure image pairs with a two components Log-Normal mixture model - one component represents the state of non loop-closure and the other represents the state of loop-closure, and learn the parameters with the Expectation-Maximization algorithm. The intersection of the Log-Normal mixture distributions is the optimal threshold for geometric verification, i.e. the threshold that gives the minimum false positive and negative loop-closures. Our algorithm degenerates when there are too few or no loop-closures and we propose the χ^2 test to detect this degeneracy. We verify our proposed method with several large-scale datasets collected from both the multi-camera setup and stereo camera.**

I. Introduction

Loop-closure for Simultaneous Localization and Mapping (SLAM) refers to the problem of detecting whether the robot at its current location sees a previously visited location and the computation of the geometric constraint that relates these two locations. In the recent years, many works [1]–[5] have demonstrated the effectiveness of using a camera to do place recognition for loop-closure detection. These works made used of the vocabulary-tree [6] that consists of the training and query phases. In the training phase, the vocabulary-tree is trained offline with image features such as the SURF [7] extracted from a given set of training images. In the query phase, a database is built from the unique IDs assigned to the query images and the extracted image features according to the pre-trained vocabulary-tree in the form of an inverted file for efficient retrieval. The database is queried with the image features extracted from the current query image and the output is the image ID from the database with the highest similarity score. The current query image and the image from the database with the highest similarity score form a potential loop-closure image pair.

The selection of the potential loop-closure image pair with the vocabulary-tree is purely based on the appearance similarity between the images and does not take the geometric relation between the images into account. As a result, an additional step of geometric verification is taken to determine

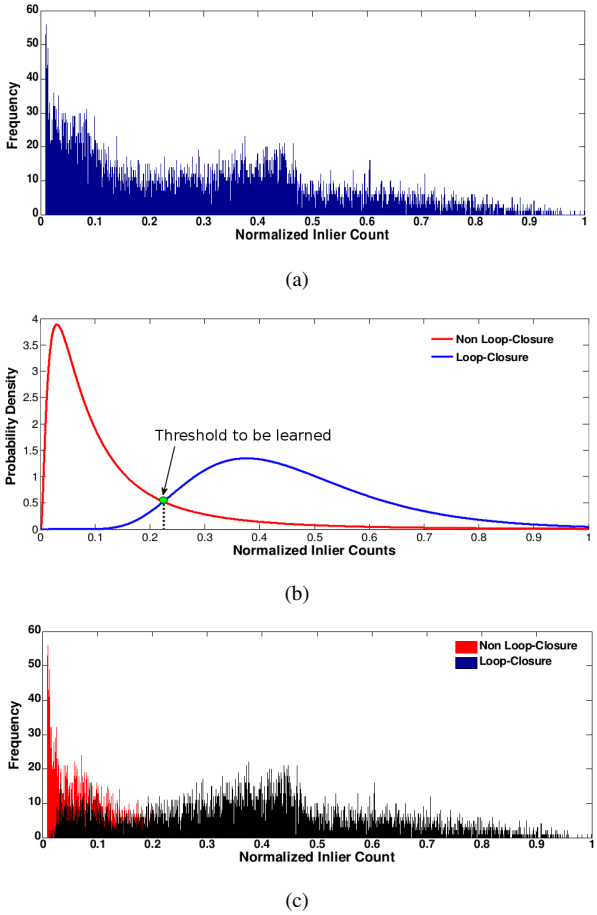


Fig. 1. (a) An example distribution of the all the inlier count from the geometric verifications. (b) Two components Log-Normal mixture model and geometric verification threshold learned from the inlier counts. (c) GPS/INS ground truth distribution of the loop-closure and non loop-closure inlier counts.

the correctness of each potential loop-closure image pair. Geometric verification refers to the process of checking the consistency of the feature correspondences between the image pairs. This is achieved by counting the number of inliers from the computation of the geometric constraint between each potential loop-closure image pair with RANSAC [8]. The 5-point [9] and 8-point [10] algorithms are commonly used to compute the geometric constraint between an image pair from a monocular camera. Similarly, the absolute orientation algorithm [11] is used for stereo cameras, and the 3-point [5] and pose estimation [12] algorithms for multi-camera setups. A potential loop-closure image pair with the

number of inliers from the geometric constraint computation exceeding a pre-defined threshold is selected as the loop-closure image pair. The choice for this threshold is critical to the success of identifying the correct loop-closures. A threshold which is too high results in missing out many correct loop-closure image pairs while a threshold which is too low results in many false positives. However, there is no fixed value for this threshold as it often varies with different datasets. The value for the threshold is also unknown and has been chosen empirically in the existing works for visual-based loop-closures [1]–[5].

In this paper, we propose an unsupervised method that learns the threshold for geometric verification directly from the observed geometric verification inlier counts for all the potential loop-closure image pairs. We observed that the inlier counts from all the potential loop-closure image pairs, normalized with the maximum inlier count, can be approximated with a two components Log-Normal mixture distribution. An example of the distribution is shown in Figure 1(a). Each component represents the state of loop-closure or non loop-closure. The normalized inlier counts are the observed variables and we use the Expectation-Maximization (EM) algorithm to learn the hidden variables representing the parameters of the mixture model, and the latent variables representing the state of loop-closure or non loop-closure. The intersection of the Log-Normal mixture distributions is the optimal threshold for geometric verification, i.e. the threshold that gives the minimum false positive and negative loop-closures. Figure 1(b) shows an example of the learned model and Figure 1(c) shows the ground truth from GPS/INS. Our algorithm degenerates when there are too few or no loop-closures and we propose the χ^2 test to detect this degeneracy. We verify our proposed method with several large-scale datasets collected from both the multi-camera setup and stereo camera.

II. Unsupervised Learning of the Threshold

Let us denote the robot poses as $X = [x_1, x_2, ..., x_t]^T$ and the N potential loop-closure pose pairs as $\mathcal{X} = [..., \{x_i, x_j\}_n, ..., \{x_i, x_j\}_N]$. The potential loop-closure pose pairs $\mathcal{X}$ are the pose pairs with the most similar image pairs detected by the vocabulary-tree, i.e. $\{x_i, x_j\}_n$ is the n^{th} pose pair where the query image taken at x_i is most similar to the image at x_j from the vocabulary-tree database. Let $L = [l_1, l_2, ..., l_N]$ where $l_n \in \{0, 1\}$ denote the actual loop-closure state, i.e. $l_n = 1$ if the correspondence loop-closure robot pose pair $\{x_i, x_j\}_n$ is truly a loop-closure pair and vice versa. We further let $V = [v_1, v_2, ..., v_N]$ denote the inlier counts from the geometric verification of all the potential loop-closure image pairs. Each of the value v_n is normalized with the maximum value from V, i.e. $V = \frac{V}{max(V)}$. The normalization step makes it easier to model the distribution of the inlier counts since it is now within the range of 0 to 1.

A. Learning Mixture Model Parameters with EM

Assuming that each v_n is drawn independently, we can write the joint probability distribution of L and V as a two components Log-Normal mixture model given by

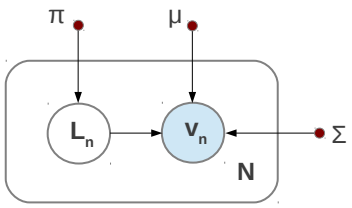


Fig. 2. Graphical model representing the threshold learning problem.

$$p(V, L|\mu, \Sigma, \pi) = \prod_{n=1}^{N} \prod_{k=1}^{2} (\pi_k ln\mathcal{N}(v_n|\mu_k, \Sigma_k))^{L_{n,k}} \quad (1)$$

where $\mu = [\mu_1, \mu_2]$, $\Sigma = [\Sigma_1, \Sigma_2]$ and $\pi = [\pi_1, \pi_2]$ are the means, standard deviations and mixing coefficients of the Log-Normal mixture model. Note that the mixing coefficients must sum to 1, i.e. $\pi_1 + \pi_2 = 1$. Figure 2 show the graphical model that represents the joint probability $p(V, L|\mu, \Sigma, \pi)$ where V is the observed variable, L is the latent variable and μ, Σ and π are the hidden variables. We shall denote the hidden variables jointly as $\phi = [\mu, \Sigma, \pi]$. The task is to find the hidden variables ϕ that maximizes the log likelihood $\ln p(V|\phi)$ given as

$$\underset{\phi}{\operatorname{argmax}} \ln p(V|\phi) = \underset{\phi}{\operatorname{argmax}} \ln\{\sum_{L} p(V, L|\phi)\} \quad (2)$$

which involves marginalizing out the latent variable L. One way to find the maximum likelihood in Equation 2 is by using the EM algorithm which iterates between the Expectation and Maximization steps. In the Expectation step, the posterior distribution $p(L|V, \phi')$ is computed based on the hidden variables ϕ' computed from the previous Maximization step. The posterior $p(L|V, \phi')$ is a $N \times 2$ matrix with each entry given by $\mathcal{T}(L_{n,k})$, where $n = 1, 2, ..., N$ and $k = 1, 2$.

$$\mathcal{T}(L_{n,k}) = p(L_{n,k}|v_n, \phi') = \frac{\pi_k ln\mathcal{N}(v_n|\mu_k, \Sigma_k)}{\sum_{j=1}^{2} \pi_j ln\mathcal{N}(v_n|\mu_j, \Sigma_j)} \quad (3)$$

The posterior probability $\mathcal{T}(L_{n,k})$ is used to evaluate the expectation of the log likelihood of the joint probability $p(V, L|\phi)$ given by

$$\begin{aligned} Q(\phi, \phi') &= \sum_{L} p(L|V, \phi') \ln p(V, L|\phi) \\ &= \sum_{n=1}^{N} \sum_{k=1}^{2} \mathcal{T}(L_{n,k}) \{\ln \pi_k + \ln ln\mathcal{N}(v_n|\mu_k, \Sigma_k)\} \end{aligned} \quad (4)$$

In the Maximization step, we compute the current values of the hidden variable ϕ by maximizing the expectation from Equation 4.

$$\underset{\phi}{\operatorname{argmax}}\ Q(\phi, \phi') \quad (5)$$

The Maximization step given by Equation 5 can be expressed in closed-form by setting the partial differentiation of $Q(\phi, \phi')$ with respect to the individual variables in ϕ to 0 and enforcing the constraint of $\pi_1 + \pi_2 = 1$. This gives

$$\mu_k = \frac{\sum_{n=1}^{N} \mathcal{T}(L_{n,k}) \ln(v_n)}{\sum_{n=1}^{N} \mathcal{T}(L_{n,k})} \tag{6a}$$

$$\Sigma_k = \frac{\sum_{n=1}^{N} \mathcal{T}(L_{n,k}) (\ln(v_n) - \mu_k)^2}{\sum_{n=1}^{N} \mathcal{T}(L_{n,k})} \tag{6b}$$

$$\pi_k = \frac{\sum_{n=1}^{N} \mathcal{T}(L_{n,k})}{N} \tag{6c}$$

The EM steps are iterated until convergence, i.e. there is minimal or no more changes to the hidden variables μ_k, Σ_k and π_k.

B. Solving for the Threshold

The parameters of the two components Log-Normal mixture distribution which we have solved with the EM algorithm models the distributions of the inlier counts from the non loop-closure and loop-closure image pairs. The intersection point v_T of the two Log-Normal distributions defines the point where $v_n < v_T$ is the region with a higher probability of non loop-closure, i.e. $l_n = 0$. Similarly, $v_n > v_T$ defines the region with a higher probability of loop-closure, i.e. $l_n = 1$. We solve for the threshold v_T by finding the root of the difference in the two Log-Normal distributions

$$f(v) = \pi_1 ln\mathcal{N}(v|\mu_1, \Sigma_1) - \pi_2 ln\mathcal{N}(v|\mu_2, \Sigma_2) = 0 \tag{7}$$

Since $f(v)$ is a non-linear function, we solve for the root with the Brent's method [13] which by default uses the less robust but faster Secant method to iteratively search for the root and fall back to the slower but more robust Bisection method if needed. Finally, we compute the threshold for geometric verification as $T_{GV} = max(V)v_T$ after denormalization.

C. Degenerated Case

Our algorithm for unsupervised learning of the threshold for geometric verification degenerates when there is too few or no loop-closures. In this case, there is only one dominant Log-Normal distribution from the non-loop closure inlier counts. The EM algorithm would not be able to detect the missing Log-Normal distribution that represents the loop-closure in the degenerated case. We propose to detect the degeneracy by performing the χ^2 test on the optimization results after SLAM optimization. Intuitively, a degenerated case includes wrong loop-closure constraints which causes wrong convergence in the optimization. As a result, the total Mahalanobis distance D^2 after optimization is high and will fail the χ^2 test.

A popular approach to perform the optimization with the loop-closure constraints is the pose-graph SLAM which represents the SLAM problem as a graph. The pose-graph SLAM optimization minimizes the errors between the predicted poses $X = [x_1, ..., x_i, ...x_j, ...x_t]^T$ represented by the vertices and the $\mathcal{K}$ set of measured constraints $Z = [..., z_{i,j}, ...]^T$ represented by the edges given by

$$\underset{X}{\operatorname{argmin}} \sum_{\mathcal{K} \in \{i,j\}} ||z_{i,j} - h(x_i, x_j)||^2_{Q_{i,j}} \tag{8}$$

where $h(.)$ is the function which computes the relative transformation between two given poses x_i and x_j, and $Q_{i,j}$ is the error covariance of the measurement $z_{i,j}$. The total Mahalanobis distance D^2 of the optimized poses $X^* = [x_1^*, ..., x_i^*, ...x_j^*, ...x_t^*]^T$ which reflects the "goodness-of-fit" is given by

$$D^2 = \sum_{\mathcal{K} \in \{i,j\}} ||z_{i,j} - h(x_i^*, x_j^*)||^2_{Q_{i,j}} \tag{9}$$

We compare the Mahalanobis distance D^2 with the critical value $\chi^2_{\mathcal{K},\alpha}$ of the χ^2 distribution where $\mathcal{K}$ is the degree of freedom which is equal to the number of measured constraints, and α is the significance level, i.e. there is a probability of α that D^2 would be greater than the critical value $\chi^2_{\mathcal{K},\alpha}$ by chance. α is usually set at 0.05 or 0.025. A degenerated case is detected when the total Mahalanobis distance is higher than the critical value, i.e. $D^2 > \chi^2_{\mathcal{K},\alpha}$.

D. Summary

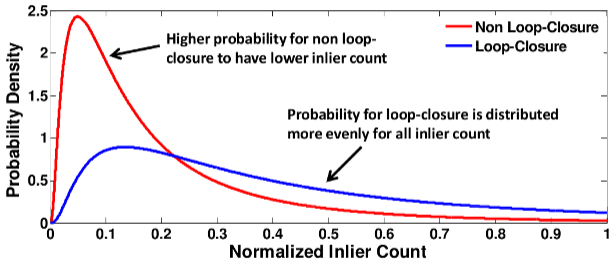


Fig. 3. Initial two components Log-Normal mixture distribution.

Algorithm 1 shows the pseudo-code of our algorithm for unsupervised learning of the threshold for geometric verification. The input is the inlier counts $V = [v_1, v_2, ..., v_N]$ from all the potential loop-closure image pairs generated by the vocabulary-tree. The output is the threshold T_{GV} for geometric verification. Line 2 normalizes V with its maximum value so that the values lie in the range of 0 to 1. Line 4 initializes the parameters ϕ from the two components Log-Normal mixture distribution. The mixing coefficients π are initialized to 0.5 which give the two components of the Log-Normal distribution equal weights. Figure 3 shows the initial Log-Normal mixture distribution with means $\mu = [-2, 1]$ and standard deviations $\Sigma = [1, 1]$. The non loop-closure Log-Normal distribution has a high peak at 0.05 and a long tail with values close to 0. In comparison, the loop-closure Log-Normal distribution is distributed more evenly with lower values near the peak and higher values at the tail region. We chose this initial distribution that is close to the final solutions, where there is always a higher occurrence of low inlier counts which are non loop-closure and higher occurrence of high inlier counts which are loop-closure.

Lines 7 to 31 are the EM loop which terminates when the change in ϕ is small. Lines 10 to 20 are the Expectation

Algorithm 1 Unsupervised Learning of Threshold for Geometric Verification.

Require: Inlier counts $V = [v_1, v_2, ..., v_N]$ from all the potential loop-closure image pairs.

Ensure: Threshold for geometric verification T_{GV}.

1: // Normalize inlier count list
2: $V = \frac{V}{max(V)}$;
3: // Parameters initialization
4: $\mu = [-2.0, 1.0]$; $\Sigma = [1.0, 1.0]$; $\pi = [0.5, 0.5]$;
5: $\phi = [\mu, \Sigma, \pi]$; $\phi' = 0$;
6: // EM algorithm
7: **while** $|\phi' - \phi| > \delta$ **do**
8: $\phi' = \phi$;
9: // Expectation step, Equation 3
10: **for** $n = 1$ to N **do**
11: // Compute denominator of Equation 3
12: $\mathcal{Z}_n = 0$;
13: **for** $j = 1$ to 2 **do**
14: $\mathcal{Z}_n = \mathcal{Z}_n + \frac{\pi'_j}{v_n \Sigma'_j \sqrt{2\pi}} \exp(-\frac{(\ln(v_n) - \mu'_j)^2}{2\Sigma'^2_j})$;
15: **end for**
16: **for** $k = 1$ to 2 **do**
17: // Compute posterior probability $\mathcal{T}(L_{n,k})$
18: $\mathcal{T}(n,k) = \frac{\pi'_k}{\mathcal{Z}_n v_n \Sigma'_k \sqrt{2\pi}} \exp(-\frac{(\ln(v_n) - \mu'_k)^2}{2\Sigma'^2_k})$;
19: **end for**
20: **end for**
21: // Maximization step, Equation 6
22: **for** $k = 1$ to 2 **do**
23: $\mu_k = 0$; $\Sigma_k = 0$; $\pi_k = 0$;
24: **for** $n = 1$ to N **do**
25: $\mu_k = \mu_k + \mathcal{T}(n,k) \ln(v_n)$;
26: $\Sigma_k = \Sigma_k + \mathcal{T}(n,k)(\ln(v_n) - \mu_k)^2$;
27: $\pi_k = \pi_k + \mathcal{T}(n,k)$;
28: **end for**
29: $\mu_k = \frac{\mu_k}{\pi_k}$; $\Sigma_k = \frac{\Sigma_k}{\pi_k}$; $\pi_k = \frac{\pi_k}{N}$;
30: **end for**
31: **end while**
32: Solve for the root v_T in Equation 7.
33: $T_{GV} = max(V) v_T$; // Denormalization
34: **return** T_{GV};

step where the posterior distribution $\mathcal{T}(L_{n,k})$ is computed, and Lines 22 to 30 are the Maximization step where the parameter ϕ is updated. Line 32 solves for the intersection of the two Log-Normal distributions and Line 33 does denormalization to compute the threshold T_{GV}.

III. Results

We verify our proposed algorithm with several large-scale real-world datasets - two datasets from a multi-camera setup and one dataset from a stereo camera. We chose the multi-camera and stereo setups because these two setups allow us to compute the loop-closure constraint with metric scale [5]. We show the χ^2 test results for all the three datasets. In addition, we show an example of the degenerated case and its detection with the χ^2 test.

A. Multi-Camera System

Fig. 4. Our car equipped with a multi-camera system with minimal overlapping field-of-views and GPS/INS system for ground truth.

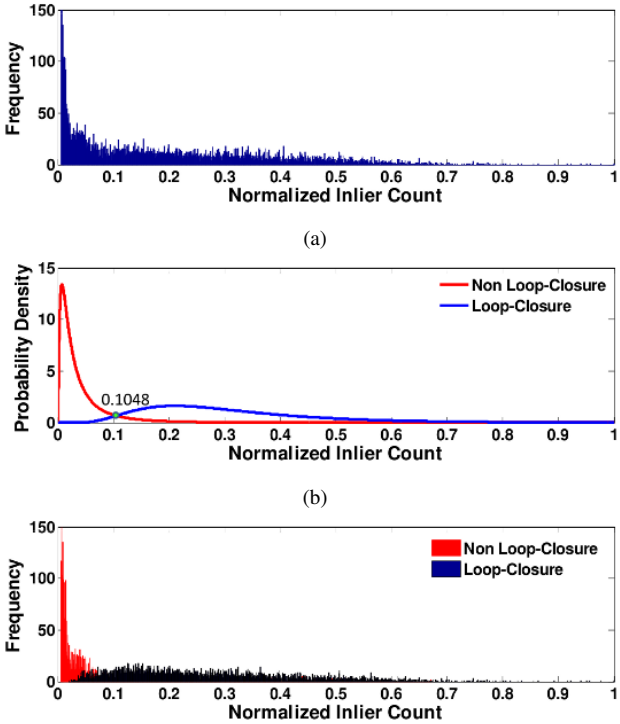
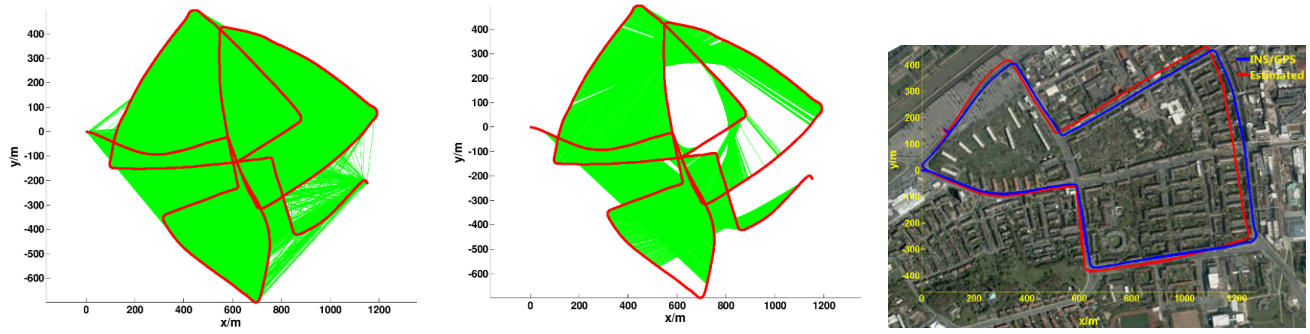


Fig. 5. Wolfsburg dataset with a multi-camera system. (a) Distribution of the all the inlier counts from the geometric verifications. (b) Two components Log-Normal mixture model and geometric verification threshold learned from the inlier counts. The threshold is 220 after denormalization. (c) Ground truth distribution from INS/GPS.

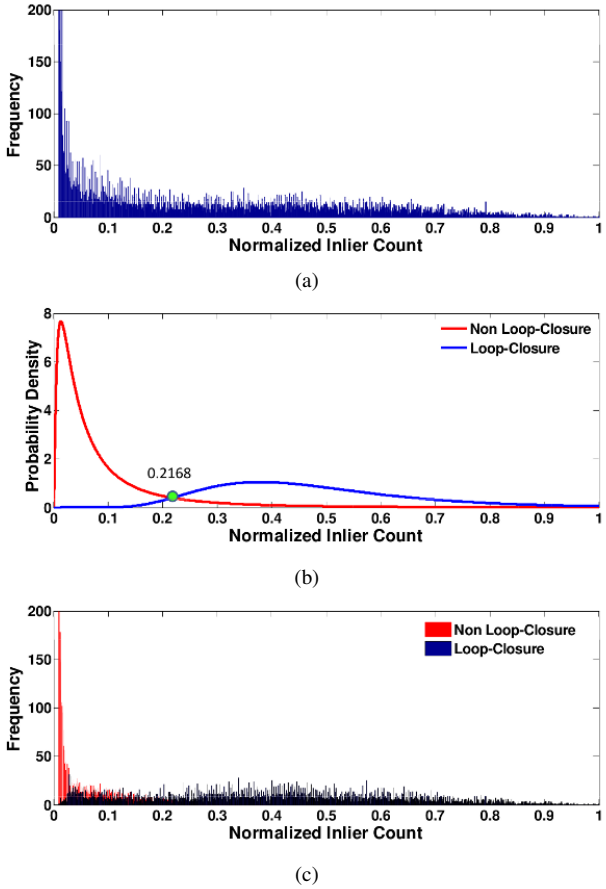
Figure 4 shows the car platform we used to collect the datasets. The car is equipped with four fish-eye cameras looking front, left, rear and right with minimal overlapping field-of-views. The car is also equipped with wheel odometry and INS/GPS system for ground truth. The first dataset - the Wolfsburg dataset was collected by driving the car along public roads in Wolfsburg Germany where it is largely urban scenes. The Wolfsburg dataset consists of a total of 13250×4 images from a trajectory that spans across approximately $9\ km$ forming two large closed loops. We form the pose-graph with the wheel odometry readings. The loop-closure candidates are found from the vocabulary-tree as mentioned in Section I and the loop-closure constraints are computed with the 3-point algorithm described in [5]. Similar to [5], we maintain only one vocabulary-tree for all the cameras



(a) (b) (c)

Fig. 6. Results from the Wolfsburg dataset. (a) Pose-graph from wheel odometry trajectory (red) with all the detected loop-closures (green). (b) Pose-graph from wheel odometry (red) with the loop-closures (green) after applying the geometric verification threshold learned from our algorithm. (c) Final result after pose-graph optimization (red) compared with the INS/GPS ground truth (blue).

from our multi-camera system. We do so by assigning unique image IDs given by imageID = frameID $\times$ n + cameraID, where n is the total number of cameras in the multi-camera system.



(a)

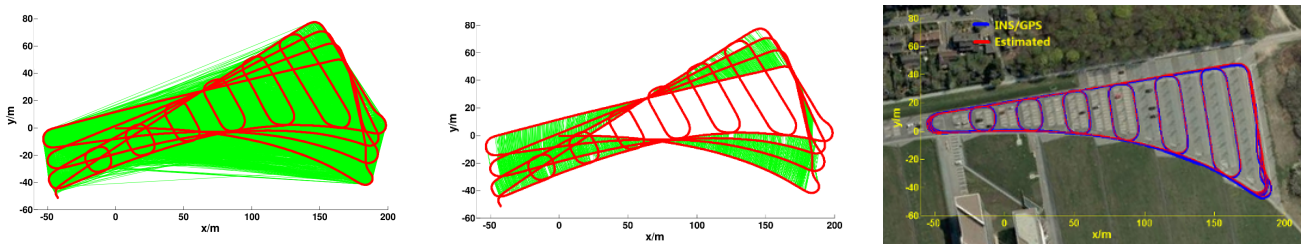
(b)

(c)

Fig. 7. Carpark dataset with a multi-camera system. (a) Distribution of the all the inlier counts from the geometric verifications. (b) Two components Log-Normal mixture model and geometric verification threshold learned from the inlier counts. (c) Ground truth distribution from INS/GPS.

Figure 5(a) shows the distribution of the normalized inlier counts obtained from the computations of the geometric constraints with the 3-point algorithm for the Wolfsburg dataset. Figure 5(b) shows the two components Log-Normal mixture distribution learned from the observations in Figure 5(a) with our proposed algorithm. The intersection of the two Log-Normal distributions is found to be 0.1048, which is 220 after denormalization. Figure 5(c) shows the ground truth distribution from the INS/GPS. We can see that the intersection of the two distributions is very close to the estimate we have obtained from our algorithm. Figure 6(a) shows the pose-graph from wheel odometry (red) with all the detected loop-closures (green) from the vocabulary-tree where there are many wrong loop-closures. Figure 6(b) shows the pose-graph with the loop-closures after applying the geometric verification threshold learned from our algorithm where majority of the wrong loop-closures are removed. Figure 6(c) shows the pose-graph (red) after optimization overlaid on the satellite image. We apply the robust pose-graph optimization proposed in [14] to minimize the effects of the a small number of non-loop closures with higher inlier count than the threshold. We compare the estimated pose-graph with the INS/GPS ground truth (blue) where we can see that our estimated pose-graph is sufficiently close to the ground truth.

We apply our algorithm on another dataset - the Carpark dataset collected from the same car with the multi-camera setup. The dataset was collected by driving the car around a huge car park besides an office building. A total of 12000×4 images are collected over a trajectory that spans across approximately 3.5 km forming 3 large and 6 nested loops. Figure 7(a) shows the distribution of the normalized inlier counts obtained from the computations of the geometric constraints with the 3-point algorithm for the Carpark dataset. Figure 7(b) shows the two components Log-Normal mixture distribution learned from the observations from Figure 7(b) with our algorithm. The intersection of the two Log-Normal distributions is found to be 0.2168, which is 277 after denormalization. Figure 7(c) shows the ground truth distribution from the INS/GPS. We can see that the intersection of the two distributions is very close to the estimate we have obtained from our algorithm. It is important to note that that the threshold for geometric verification for the datasets differs slightly even though they are collected from the same platform. This explains our approach to learn the threshold from each dataset in a batch process for the best results. However, it is also possible to take the most



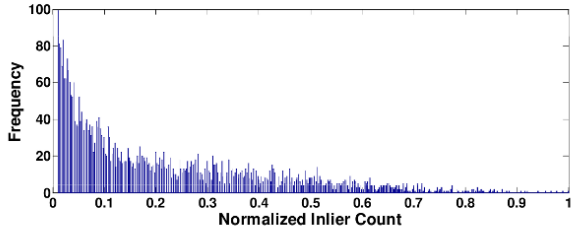
(a) (b) (c)

Fig. 8. Results from the Carpark dataset. (a) Pose-graph from wheel odometry trajectory (red) with all the detected loop-closures (green). (b) Pose-graph from wheel odometry (red) with the loop-closures (green) after applying the geometric verification threshold learned from our algorithm. (c) Final result after pose-graph optimization(red) compared with the INS/GPS ground truth (blue).

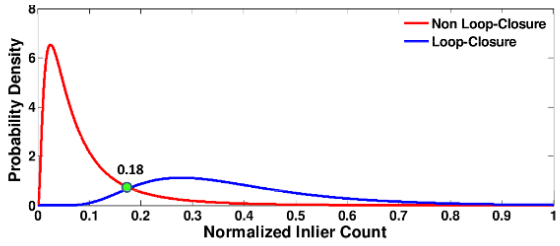
conservative threshold, i.e. highest threshold learned from multiple datasets collected from the same platform and use it for online geometric verifications.

Figure 8(a) shows the pose-graph from wheel odometry (red) with all the detected loop-closures (green) from the vocabulary-tree where there are many wrong loop-closures. Figure 8(b) shows the pose-graph after applying the geometric verification threshold learned from our algorithm where majority of the wrong loop-closures are removed. Figure 8(c) shows the final pose-graph (red) after robust pose-graph optimization. We compare the estimated pose-graph with the INS/GPS ground truth (blue) where we can see that our estimated pose-graph follows the ground truth very closely.

B. Stereo Camera



(a)



(b)

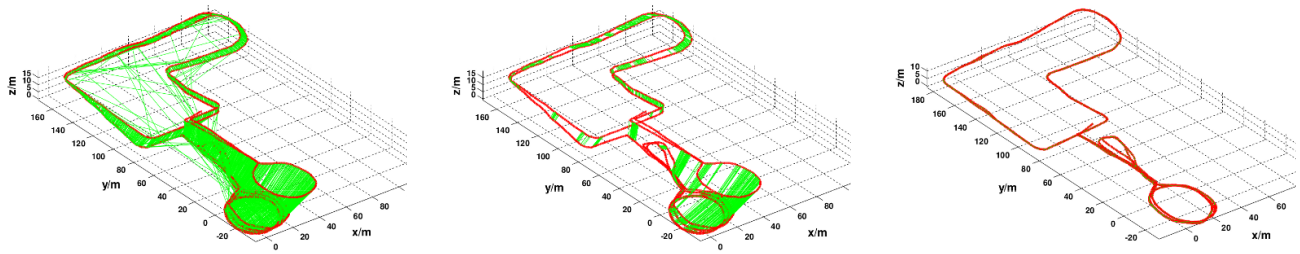
Fig. 9. New College dataset with a stereo camera. (a) Distribution of the all the inlier counts from the geometric verifications. (b) Two components Log-Normal mixture model and geometric verification threshold learned from the inlier counts.

We also test our algorithm on the New College dataset [15] which was collected with a stereo camera mounted on a ground robot. A total of 48241 stereo frames from a trajectory that spans across approximately 1.25 km are used. The vocabulary-tree for visual loop-closure is formed from the left stereo images and the loop-closure constraints are computed from the absolute orientation algorithm [11]. Figure 9(a) shows the distribution of the distribution of the normalized inlier counts obtained from the computations of the geometric constraints with the absolute orientation for the New College dataset. Note that this distribution is similar to Figure 5(a) and 7(a) even though a different camera setup is used. Figure 9(b) shows the two components Log-Normal mixture distributions learned with our algorithm. The intersection of the two Log-Normal distributions is found to be 0.18, which is 76 after denormalization. Figure 10(a) shows the pose-graph from stereo visual odometry [16] (red) with all the detected loop-closures (green) from the vocabulary-tree where there are many wrong loop-closures. Figure 10(b) shows the pose-graph with the loop-closures after applying the geometric verification threshold learned from our algorithm where majority of the wrong loop-closures are removed. Figure 10(c) shows the pose-graph after robust robust pose-graph optimization. Although there is no INS/GPS ground truth for the New College dataset, we can see that the pose-graph after optimization appears reasonable with all the deviations in the z-axis removed.

C. Degenerated Case

We show an example of the degenerated case with a dataset collected from our car platform with the multi-camera setup. The dataset consists of 1600×4 images from a trajectory that spans across approximately 300 m forming one loop with the starting and ending points at approximately the same location. The small number of loop-closures means that there is not enough information for our algorithm to learn the threshold correctly. Figure 11(a) shows the pose-graph from wheel odometry (red) and all the loop-closures from the vocabulary-tree. Figure 11(b) shows the pose-graph with the loop-closures after applying the threshold learned from our algorithm where it is clearly visible that some wrong loop-closures remain. Figure 11(c) shows the pose-graph (red) after robust pose-graph optimization compared with the ground truth (blue). The ground truth is obtained by performing robust pose-graph optimization on the manually chosen correct loop-closures. It can be observed that the estimated pose-graph is slightly distorted by the wrong loop-closures even after applying the robust optimization because the ratio of correct to wrong loop-closures is too low.

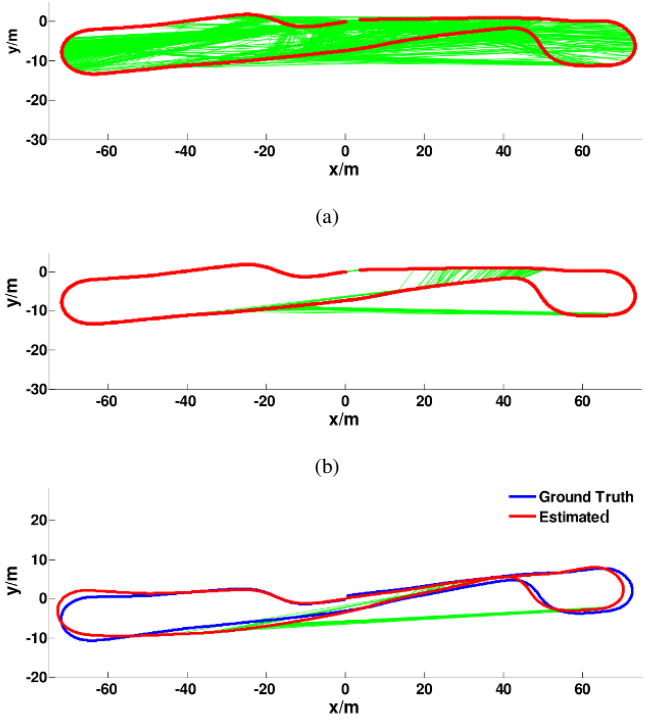
As mentioned in Section II-C, we do the χ^2 test to detect the degenerated case. Table I shows the χ^2 test results for all the datasets at a significance level $\alpha = 0.05$. We can see from the χ^2 test results that the first three datasets -



(a) (b) (c)

Fig. 10. Results from the New College Dataset. (a) Pose-graph from visual odometry trajectory (red) with all the detected loop-closures (green). (b) Pose-graph from visual odometry (red) with the loop-closures (green) after applying the geometric verification threshold learned from our algorithm. (c) Final result after pose-graph optimization.

Wolfsburg, Carpark and New College pass the χ^2 test with $D^2 < \chi^2_{\alpha=0.05}$ while the degenerated dataset fails the χ^2 test with $D^2 > \chi^2_{\alpha=0.05}$.



(a)

(b)

(c)

Fig. 11. Results from the Degenerated dataset. (a) Pose-graph from wheel odometry trajectory (red) with all the detected loop-closures (green). (b) Pose-graph from wheel odometry (red) with the loop-closures (green) after applying the geometric verification threshold learned from our algorithm. (c) Wrong convergence after pose-graph optimization(red) compared with the ground truth (blue).

IV. Conclusion

The threshold for geometric verification is crucial in identifying the correct loop-closures. However, the threshold varies for different datasets and has been chosen empirically in the existing works for visual-based loop-closure. We described a method for unsupervised learning of the threshold for geometric verification in this paper. Our method is based on the EM algorithm which learns the threshold from the inlier counts generated from the RANSAC computation of the geometric constraints between all potential loop-closure image pairs. We verified our method with multiple large-scale datasets from both the multi-camera and stereo setups.

TABLE I

χ^2 Test Results for the Datasets

Dataset	DOF	$\chi^2_{\alpha=0.05}$	D^2
Wolfsburg	12840	1.3105e+04	4.1249e+03
Carpark	17506	1.7815e+04	1.0325e+04
New College	9713	9.9434e+03	644.7097
Degenerated	1589	1.6828e+03	4.8715e+03

V. Acknowledgement

This work is supported in part by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant #269916 (v-charge) and 4DVideo ERC Starting Grant Nr. 210806.

References

[1] M. Cummins and P. Newman, "Appearance-only slam at large scale with fab-map 2.0," *IJRR*, vol. 30, no. 9, pp. 1100–1123, August 2011.
[2] D. Galvez-Lopez and J. D. Tardos, "Real-time loop detection with bags of binary words," in *IROS*, September 2011, pp. 51 –58.
[3] N. Sünderhauf and P. Protzel, "BRIEF-Gist - Closing the loop by simple means," in *IROS*, 2011, pp. 1234–1241.
[4] D. Galvez-Lopez and J. D. Tardos, "Bags of binary words for fast place recognition in image sequences," *IEEE Transactions on Robotics*, vol. 28, no. 5, pp. 1188–1197, October 2012.
[5] G. H. Lee, F. Fraundorfer, and M. Pollefeys, "Structureless pose-graph loop-closure with a multi-camera system on a self-driving car," in *IROS*, 2013.
[6] D. Nistér and H. Stewénius, "Scalable recognition with a vocabulary tree," in *CVPR*, 2006, pp. 2161–2168.
[7] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *CVIU*, vol. 110, no. 3, pp. 346–359, June 2008.
[8] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, June 1981.
[9] D. Nister, "An efficient solution to the five-point relative pose problem," in *PAMI*, vol. 26, no. 6, 2004, pp. 756–770.
[10] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, ISBN: 978-0-521-54051-3, 2004.
[11] S. N. Berthold K. P. Horn, Hugh M. Hilden, "Closed-form solution of absolute orientation using unit quaternions," *Journal of the Optical Society of America A*, 1987.
[12] G. H. Lee, B. Li, M. Pollefeys, and F. Fraundorfer, "Minimal solutions for pose estimation of a multi-camera system," in *ISRR*, 2013.
[13] R. P. Brent, *Algorithms for Minimization without Derivatives*. Englewood Cliffs, NJ: Prentice-Hall, ISBN: 0-13-022335-2, 1973.
[14] G. H. Lee, F. Faundorfer, and M. Pollefeys, "Robust pose-graph loop-closures with expectation-maximization," in *IROS*, 2013.
[15] M. Smith, I. Baldwin, W. Churchill, R. Paul, and P. Newman, "The new college vision and laser data set," *IJRR*, vol. 28, no. 5, pp. 595–599, May 2009.
[16] D. Nistér, O. Naroditsky, and J. R. Bergen, "Visual odometry for ground vehicle applications," *JFR*, vol. 23, no. 1, pp. 3–20, 2006.