2016 23rd International Conference on Pattern Recognition (ICPR)
Cancún Center, Cancún, México, December 4-8, 2016

Geometric Verification Using Semi-2D Constraints for 3D Object Retrieval

Kohei Matsuzaki*, Yusuke Uchida*†, Shigeyuki Sakazawa* and Shin'ichi Sato‡
*KDDI R&D Laboratories, Inc., †University of Tokyo, ‡National Institute of Informatics
*Saitama, Japan, †Tokyo, Japan, ‡Tokyo, Japan
*{ko-matsuzaki, ys-uchida, sakazawa}@kddilabs.jp, ‡satoh@nii.ac.jp

***Abstract*—Geometric verification with epipolar geometry often results in a high score for an incorrect image pair due to ambiguity in its geometric constraints. The ambiguity is caused by a high degree of freedom in the epipolar geometry and a weak constraint from the fitting between a point and a line. In order to mitigate the ambiguity, we propose to filter geometrically inconsistent components, namely correspondences, a sample, a model, and inliers in a RANSAC-based geometric verification. For the filtering, we introduce novel semi-2D constraints whose geometric constraint is weaker than full-2D constraint, but stronger than pure-epipolar constraint. Additionally, an advantage of the proposed approach is that it requires only an image pair instead of neither additional information nor prior learning. Experiments on the public dataset containing 3D object images show that the proposed approach improves the true positive rate when the false positive rate is low, and greatly reduces computational time for the geometric verification of both a correct image pair and an incorrect image pair.**

I. Introduction

Large scale image retrieval and recognition [1] searches a similar image containing a rigid object from uncalibrated database images, given a query image containing that object. Typical image retrieval and recognition pipeline represents an image as a set of local features such as SIFT [2], encodes them with something like bag-of-words indexing [3], and ranks the similarity scores between the query image and the database images. Then it performs geometric verification (GV) based on RANdom SAmple Consensus (RANSAC) [4] on image pairs with top N similarity scores, and re-ranks them according to their GV scores.

In such pipelines, what is input to the GV is frequently an incorrect image pair. Therefore, to achieve image recognition, an additional GV score thresholding is performed to determine whether it is the same object in the database or not.

RANSAC [4] is a typical solution, which estimates the geometric model from correspondences containing outliers. RANSAC computes a specified type of model (e.g., affine, homography, and epipolar geometry (EG)) from a sample drawn randomly from correspondences. Then it classifies correspondences into inliers and outliers that satisfy the model constraints. Eventually, it chooses the model that obtains the greatest number of inliers. The GV score means the number of inliers in that instance.

In general, the suitable type of model is different depending on the shape of the target object. GV in many recherches and applications assumes an affine or a homography type of model [1], [3], [5]. However, neither affine nor homography is suitable for 3D objects since it is impossible to describe the correspondences of two non-planar objects with them. In contrast, the EG is suitable when the target is a 3D object since it is able to describe the correspondences of two non-planar objects.

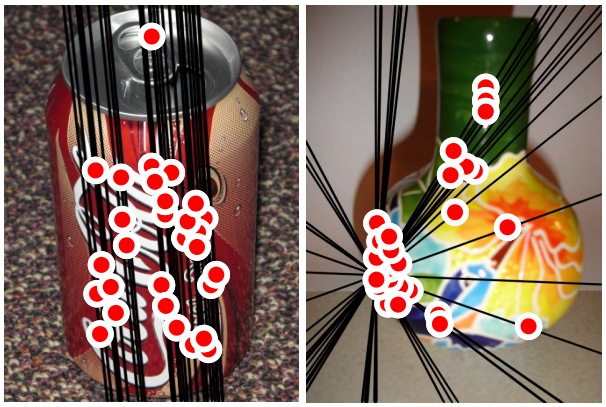

Fig. 1. Illustration of geometric verification with an epipolar geometry in an incorrect image pair. Red circles and black lines represent inliers and epipolar lines respectively.

However, as shown in Fig. 1, it has been observed that the GV score using EG often returns a high number of inliers even when the image pair is entirely incorrect. This is due to both EG's many degrees of freedom (DoF) (i.e., seven DoF for the fundamental matrix and five DoF for the essential matrix), and ambiguous constraints on EG based on the fitting between a point and a line. This can happen when there are many correspondences that happen to fit a certain (mostly geometrically inconsistent) EG computed from an incorrect sample. Such accidental fits may happen frequently, since the epipolar constraint is satisfied when a point is located somewhere on a corresponding epipolar line.[1]

Furthermore, if we verify the query image on the large scale database, it is highly probable that an incorrect image pair with a large number of inliers will be found. For an application containing a recognition stage, it is ideal that no false positives

[1]In contrast, the number of inliers in an incorrect image pair always becomes small when we assume an affine or a homography model. Such accidental fits rarely happen with affine or homography constraints since they are based on the fitting between a point and a point.

978-1-5090-4846-5/16/$31.00 ©2016 IEEE

(i.e., recognition of incorrect images) occur. A naive approach is setting a high threshold. However, it reduces the true positive rate.

In order to address this problem, we aimed to reduce the number of inliers in an incorrect image pair. Many studies of two-view geometry aim to maximize the number of inliers in an correct image pair in order to derive a high quality EG [5], [6], [18]. However, it is necessary to increase the number of inliers in a correct image pair as well as reducing the number of inliers in an incorrect one during image retrieval and recognition. Nevertheless, the latter requirement has not been sufficiently investigated.

In this paper, we propose a novel approach that reduces accidental inliers in the RANSAC. This approach filters the four components of RANSAC, namely correspondences, a sample, a model, and inliers, all considered individually. This not only directly reduces the accidental inliers but also indirectly reduces them by filtering the other components in the early stages. Thus it significantly reduces inliers resulting from an incorrect image pair.

We show experimentally that the proposed approach significantly improves the accuracy of image retrieval and recognition. Furthermore, we show that our filtering of geometrically inconsistent components greatly reduces computational time for the GV of both a correct image pair and an incorrect image pair.

II. Related work

Here we summarize RANSAC's estimation of EG [8], [9], and also briefly describe recent extensions of RANSAC.

A. RANSAC

RANSAC estimates a model $\boldsymbol{T}$ from data Q based on iteration. It also classifies the data Q as a set of inliers I or set of outliers O simultaneously.

In each iteration, it randomly selects a sample, namely s sets of correspondences from Q, then it computes an EG $\boldsymbol{T}$ from the sample. When model $\boldsymbol{T}$ is assumed to be a fundamental matrix, a 7-point algorithm [10] ($s = 7$) or 8-point algorithm [11], [12] ($s = 8$) is typically employed. When model $\boldsymbol{T}$ is assumed to be an essential matrix, a 5-point algorithm [13] ($s = 5$) is typically employed.

For a correspondence $\boldsymbol{m}$, $\boldsymbol{m'}$ in a homogeneous coordinate system on data Q, it computes the square of the distance between a point and an epipolar line $d = \boldsymbol{m'}^{\top}\boldsymbol{T}\boldsymbol{m}$. When d is lower than a specified threshold value, the correspondence is classified as an inlier; otherwise an outlier. When a set of inliers with the maximum sizes so far is obtained, the number of inliers $|I|$ and the $\boldsymbol{T}$ of that time are stored.

In the standard approach, such processes are repeated until the number of iterations reaches the following:

$$M = log(1-p)/log[1-(1-k)^s] \tag{1}$$

where p is the confidence, and $k = |I|/|Q|$ is the inlier ratio. In order to guarantee a termination, it also terminates when the number of iterations reaches a specified maximum number.

B. RANSAC Extensions

DEGENSAC [6] produces a more robust fundamental matrix estimation by detecting homography-degenerate samples. QDEGSAC [7] yields more robust model estimations on (quasi-)degenerate data without requiring explicit knowledge about degeneracies. PROSAC [18] achieves a significantly fast fundamental matrix estimation by exploiting the linear ordering structure of the set of correspondences. However, while these approaches improve the estimation quality of a correct image pair, they do not reduce the number of inliers in an incorrect one.

SCRAMSAC [14] forms a reduced set of correspondences with more reliability by checking the spatial consistency of each correspondence. Then it produces a fundamental matrix estimation that is fast and more robust against degeneration by using RANSAC on the reduced set of correspondences.

Johns *et al.* [15] arrive at a more accurate fundamental matrix estimation than SCRAMSAC in terms of both image retrieval and place recognition by learning generative place models from a significant number of training images per place. However, this approach requires many training images with various changes of view point per target, and it is not easy to collect the training images in actual large scale image retrieval and recognition applications.

The closest approach to our goal is SCRAMSAC [14]. However, it is limited to improving the correspondences, therefore the problem of accidental inliers in an incorrect image pair still remains unsolved. In contrast, we propose to explicitly filter the accidental inliers without any additional information and prior learning.

III. Proposed Approach

The proposed approach introduces novel constraints for RANSAC-based GV using the components of local features (i.e., orientation, scale, and coordinates). Fig. 2 shows an overview of the proposed approach. In Fig. 2, blocks in single lines and blocks in double lines represent the standard RANSAC process and our original process respectively. Note that the four processes of the proposed approach (A, B, C, and D in Fig. 2) are applicable independently. The notations for the unique parameters required across our approach are summarized in Table I.

The proposed approach first filters an initial set of correspondences Q based on weak geometric consistency (WGC) [16] (A in Fig. 2). Throughout the subsequent process, it operates on the resulting set of correspondences Q'. In the RANSAC iteration, it detects a geometrically inconsistent sample, then terminates the iteration at an early stage if necessary (B in Fig. 2). It detects an EG that is not spatially consistent with the sample, then terminates the iteration at an early stage if necessary (C in Fig. 2). After the model fitting to Q', it checks the spatial consistency of the resulting set of inliers I against the sample and EG (D in Fig. 2). It filters the I through this check, then determines the final set of inliers I'.

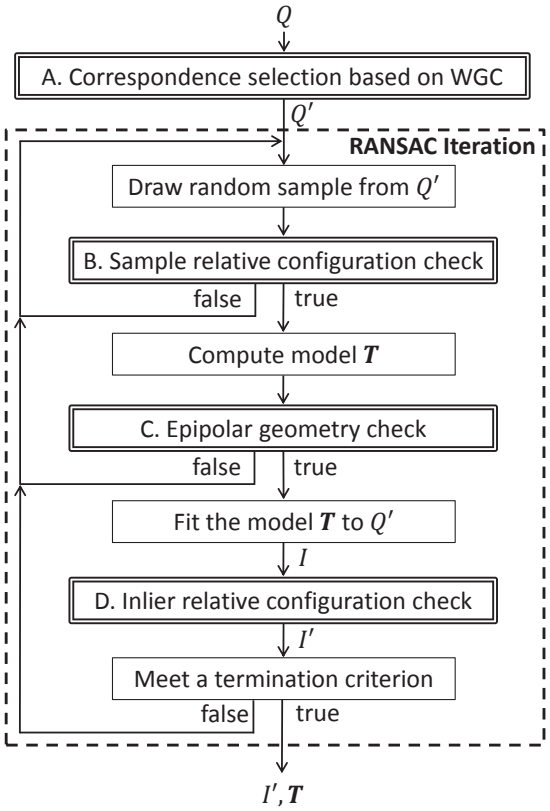

Fig. 2. Overview of the proposed approach

Filters B, C, and D in Fig. 2 are based on semi-2D constraints that can always be satisfied if the target is a 2D object. They impose weak 2D constraints to correspondences on the 3D object surface. That is to say, these constraints are weaker than a full-2D constraint, but stronger than a pure-epipolar constraint. Although these constraints involve the risk of rejecting correct assumptions (i.e., a sample, an EG, and inliers) ascribable to parallax, most of the cases can find an adequate solution from the partial regions that satisfy these constraints of the 3D object surface using the RANSAC iterations. In contrast, assumptions on the incorrect image pair are almost certainly rejected. In the following, we describe the four elements of the proposed approach.

A. Correspondence Selection based on WGC

WGC [16] was originally proposed to improve a scoring algorithm in the pipeline of a large-scale image retrieval. Specifically, WGC votes the matching descriptors to the bins of orientation difference and scale ratio. Then it improves the accuracy of image retrieval by filtering the bins, except for the bin with the maximum voting score. It is based on the assumption that correct correspondences are consistent with regard to orientation difference and scale ratio.

We apply this assumption to our method named correspondence selection based on WGC (CSW). That is to say, we vote the Q to the two-dimensional bins consisting of orientation difference and scale ratio. Then we filter the correspondences except for those belonging to the bin with the maximum voting score. For softer voting, we establish the bins with 50% overlapping resolution with regard to orientation difference O_{res} and scale ratio S_{res}. Thus when we set the resolution of orientation difference O_{res} to 30 degrees, 24 bins that overlap every 15 degrees are established.

By using CSW, it is expected that the voting score of each bin becomes random, and the size of Q' is consequently significantly reduced in an incorrect image pair. Reducing the size of Q' reduces the probability of accidental inliers. On the other hand, potential inliers concentrate in a specific bin for a correct image pair, maintaining the final number of inliers. The improved inlier ratio reduces the expected number of iterations required to find the correct solution.

TABLE I
NOTATIONS

Notation	Definition
O_{res}	The resolution of orientation difference
S_{res}	The resolution of scale ratio
TH_{tri}	The threshold value of the number of inside-outs

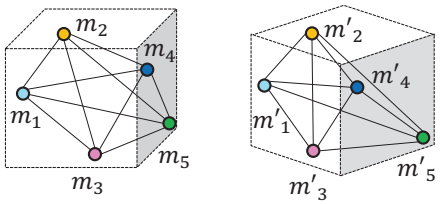

Fig. 3. An example of sample configuration in the case of $s = 5$. Colored circles represent the 5 sets of correspondences m_i, m'_i $(i = 1, 2, ..., 5)$. The ${}_5\mathrm{C}_3 = 10$ corresponding triangles are formed from them. The dotted lines of the cubes represent the target 3D object.

B. Sample Relative Configuration Check

The filter named the sample relative configuration check (SRCC) is based on the assumption that the relative configuration of a correct sample (i.e., s sets of correspondences) is consistent within a correct image pair. Geometrically inconsistent samples are rejected by this filter, so the potential inliers from the sample are rejected.

As shown in Fig. 3, ${}_s\mathrm{C}_3$ corresponding triangles can be formed from a sample. For each corresponding triangle, SRCC detects the inside-out. Specifically, it examines the direction of rotation (i.e., clockwise or anti-clockwise) from the three vertices of the triangle. It determines that the corresponding triangles do not have the inside-out problem if their direction matches. For example, corresponding triangles $\triangle m_2 m_4 m_5$ and $\triangle m'_2 m'_4 m'_5$ have the inside-out problem in Fig. 3.

For a 3D object, corresponding triangles from a correct sample can generate an inside-out ascribable to parallax. However, it is expected that the number of inside-outs will be significantly smaller than that from an incorrect sample. Therefore, SRCC counts the number of inside-outs among the ${}_s\mathrm{C}_3$ corresponding triangles, then returns true if the count is smaller than the threshold value TH_{tri}; and otherwise it returns false.

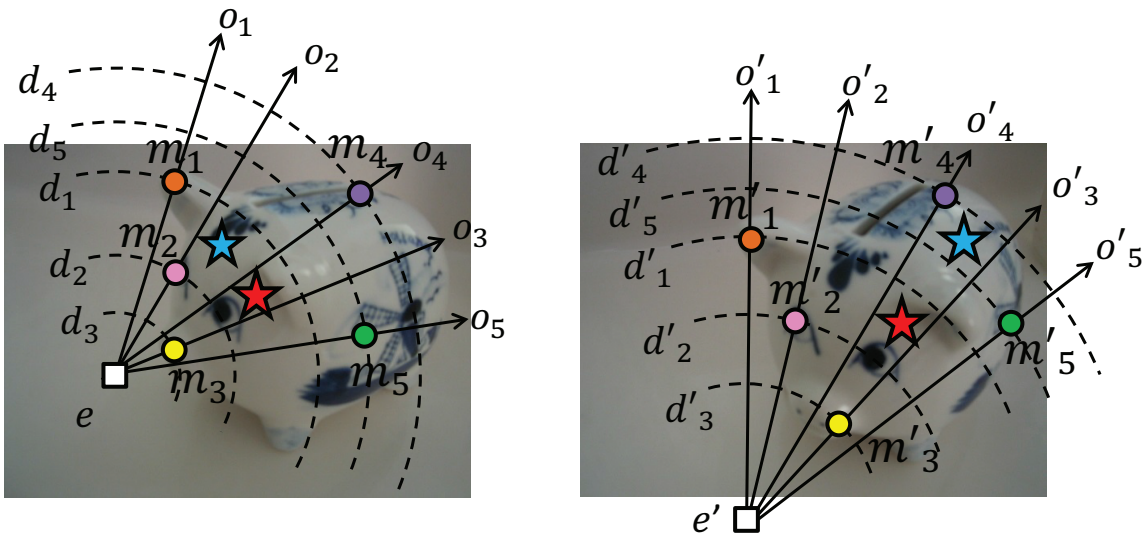

Fig. 4. Illustration of the relative configuration of components. White quadrangles, colored circles, black straight lines, and dotted lines respectively represent epipoles projected onto the image, sample correspondences, epipolar lines, and equidistant curves from the epipole respectively. Red stars and blue stars represent a correct inlier correspondence and an incorrect inlier correspondence respectively.

This filter is likely to return the wrong result if the sample points are dense in an extremely narrow range or if they lie on almost the same straight line, as in the character string image. To avoid such cases, SRCC rejects insufficiently scattered samples in x-y space. Specifically, it returns false if the area of the convex hull of the sample is extremely small on each image.

C. Epipolar Geometry Check

The epipolar geometry check filter （EGC） is based on the assumption that the relative configuration of epipoles from a correct EG and the sample are consistent within a correct image pair. Geometrically inconsistent EGs are rejected by this filter, so the potential inliers from the EG are rejected.

As shown in Fig. 4, s sets of epipolar lines are computed from the EG and the sample in each image. It is known that epipole e in the image can be calculated from the fundamental matrix [10]. Let m_i denote sample points $(i = 1, 2, ..., s)$. Let o_i and d_i denote orientations and distances from e to m_i respectively.

In Fig. 4, the corresponding orientations o_i and o'_i are arranged in the same order from end to end (i.e., o_1, o_2, o_4, o_3, o_5) in the image pair. The corresponding distances d_i and d'_i are also the same (i.e., d_3, d_2, d_1, d_5, d_4). In this way, EGC generates permutations of the ID with respect to orientations and distances in the image pair. If the permutations with regard to both orientation and distance match, it returns true.

In Fig. 4, both epipoles are located on the same side of the object, but they can be located on the opposite sides. In such a case, EGC returns true if the permutations with respect to both orientation and distance match in reverse order.

If the epipole is located inside the convex hull of the sample correspondences, it means that the camera is not moving only back and forth, and then epipolar lines become radial. In that case, EGC returns true if the circular permutations match with respect to orientation, since it does not find both ends.

D. Inlier Relative Configuration Check

The inlier relative configuration check filter (IRCC) is based on the assumption that the relative configuration of correct inliers, the epipoles, and the sample is consistent within a correct image pair. Geometrically inconsistent inliers are rejected by this filter.

As shown in Fig. 4, an image can be divided into disjoint regions by using both epipolar lines and curves equidistant from the epipole. We propose to accept an inlier only if both inlier points are located in corresponding regions in each image.

Specifically, IRCC first computes the orientation and distance from the epipole to the inlier points on each image. Then it finds the IDs on both sides of a point on an inlier with respect to both orientation and distance on each image. If these IDs match in both image of a pair, the inlier is accepted; otherwise the inlier is re-classified as an outlier.

For example, an inlier represented by red stars in Fig. 4 is accepted since they are located in corresponding regions of each image. However, an inlier represented by blue stars is re-classified as an outlier, since they are not located in corresponding regions in each image.

IV. Experiments

We evaluated the proposed approach (Prop) experimentally, then compared it with SCRAMSAC [14]. In order to draw a correct sample with fewer iterations, we employed PROSAC [18] strategy in the Prop.

In Section IV-B, we evaluated each element of the Prop, namely CSW, SRCC, EGC, and IRCC (see Section III). We employed PROSAC with an EG as a baseline method. Let "Base(EG)" denote this method. Let

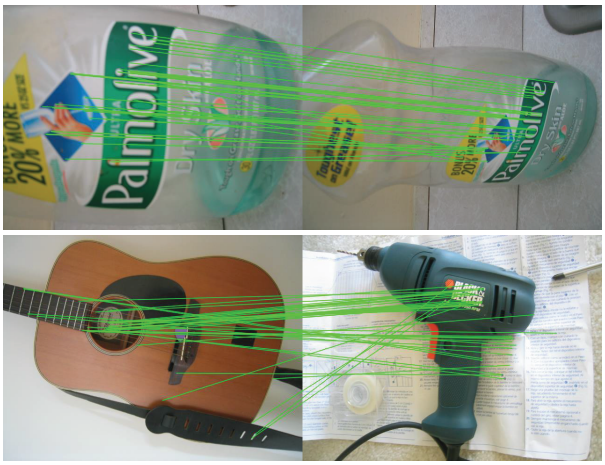

Fig. 5. An example of true positive and false positive when the threshold value is 40. Green lines represent resulting inliers from geometric verification with an epipolar geometry. Top: A correct image pair with 49 inliers (true positive). Bottom: An incorrect image pair with 41 inliers (false positive).

"Base(EG)+CSW", "Base(EG)+SRCC", "Base(EG)+EGC", and "Base(EG)+IRCC" denote the method added to each element in the Base(EG) respectively. Prop is an accumulation of the Base(EG) and all elements.

In Section IV-C, we compared Prop with SCRAMSAC. With reference to the full-2D constraint, we also measured the PROSAC with a homography. Let "Base(H)" denote this method. That is to say, we also compared both Prop and SCRAMSAC with Base(H) for comparison of epipolar constraint and full-2D constraint.

A. Dataset and Experimental Setup

In our experiments, we used the University of Kentucky Benchmark[2] (UKB), which is a standard object retrieval and recognition benchmark. UKB consists of 10,200 images of 2,550 objects taken from four different viewpoints. We generated ${}_4\mathrm{C}_2 = 6$ correct image pairs per object from the four images of the same object. For the incorrect image pairs, we generated $2,550 \times 6$ sets of random image pairs from UKB, so there were no correct image pairs. Thus we conducted GV on $2,550 \times 6 \times 2$ sets of image pairs.

When the GV score exceeds a certain threshold value it is considered a "positive". Fig. 5 shows an example of true positive and false positive when the threshold value is 40. We define the "recognition rate" as the maximum true positive rate under the constraint that the false positive rate is zero, and "TH_{rec}" as the threshold value for that time. In other words, TH_{rec} becomes identical to maximum number of inliers among the all incorrect image pairs.

For local features, we adopted ORB [17], which is efficient and suitable for mobile devices. On average, 900 features are extracted from 8 scales. In order to give the initial set of correspondences to each image pair, we performed nearest-neighbor matching with the cross-check method.

[2]http://vis.uky.edu/~stewe/ukbench/

TABLE II
COMPARISON OF GEOMETRIC VERIFICATION PERFORMANCES

	Recognition rate	TH_{rec}	Computation time [ms]	
			correct pair	incorrect pair
Base(EG)	0.669	76	170.5	294.6
Base(EG)+CSW	0.672	65	37.3	137.8
Base(EG)+SRCC	0.750	43	13.6	13.1
Base(EG)+EGC	0.734	27	245.9	400.2
Base(EG)+IRCC	0.748	24	404.8	644.2
Prop	**0.819**	**16**	**11.6**	**9.8**
SCRAMSAC [14]	0.704	50	15.5	20.2
Base(H)	0.751	19	183.5	343.1

We used the fundamental matrix as the model in RANSAC. In order to compute it, we employed the 7-point algorithm [10]. The threshold value of the distance between a point and an epipolar line was set to three. When we used the homography matrix as the model in RANSAC, the threshold value of the re-projection error was set to three. The maximum number of iterations of RANSAC was set to 10,000.

For parameters specific to SCRAMSAC, we employed the same values found in experiments in the literature (i.e., $s_{min} = 0.5, s_{max} = 2, \theta = 0.55, r = 7$).

For our CSW, the resolution of orientation differences O_{res} and scale ratios S_{res} were respectively set to 60 degrees and four-fold. These are fairly coarse resolution, to avoid filtering correct correspondences. For our SRCC, the threshold value TH_{tri} of the number of inside-outs was set to three. These parameters provided excellent results in our preliminary experiments.

All experiments were performed on a 3.6 GHz Intel Core i7 with 4 GB of RAM. Computation time included only the process shown in Fig. 2, and not the time for local feature extraction and initial correspondences generation.

We summarize the recognition rate, TH_{rec}, and computation time for each of the correct and incorrect image pairs obtained from all methods in Table II.

B. Impact of Each Element of the Proposed Approach

As shown in Table II, all elements result in a higher recognition rate and smaller TH_{rec} than those of the Base(EG). Base(EG)+CSW reduces the probability of accidental inliers by reducing the size of the Q. Base(EG)+SRCC and Base(EG)+EGC indirectly reduce accidental inliers by rejecting geometrically inconsistent samples and EGs respectively. Base(EG)+IRCC directly reduces accidental inliers. The EGC and IRCC especially effective to reduce TH_{rec}. Furthermore, the accumulation of Base(EG) and all elements (i.e., Prop) achieves the highest recognition rate and smallest TH_{rec}.

Regarding the computation time, it is shown that Base(EG)+CSW and Base(EG)+SRCC achieve greater speed than Base(EG), while Base(EG)+EGC and Base(EG)+IRCC

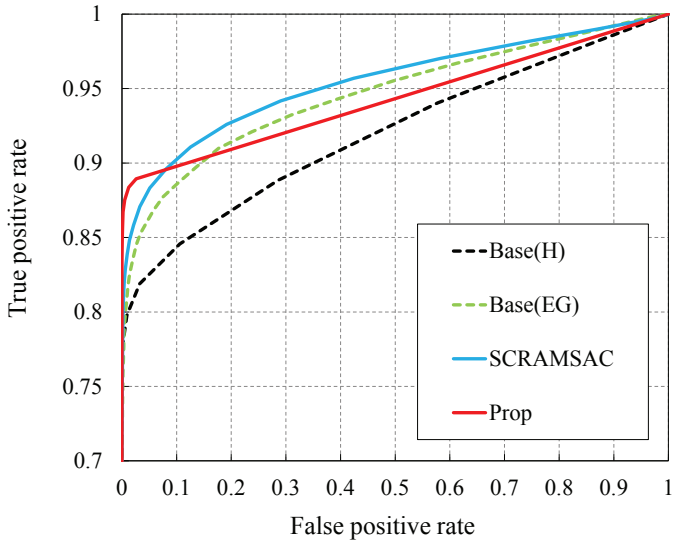

Fig. 6. ROC curves

are computationally expensive. However, the latter two have only a negligible impact on the computation time of the Prop, since most iterations are terminated at an early stage by combining with the other elements. The accumulation of CSW and SRCC greatly reduces computational time of Prop.

In general, the processing of a correct pair is faster than that of an incorrect pair, since the former terminates at an early stage when it obtains a high inlier ratio k in Equation 1. In contrast, note that only methods including SRCC achieve comparable computation time on an incorrect pair. This is because the frequency with which the iteration is terminated at an early stage by SRCC is higher for the incorrect pair.

C. Comparing the Proposed Method with SCRAMSAC

As an indicator of GV performance, we plotted the ROC (Receiver Operating Characteristic) curve. Its vertical axis shows the true positive rate and the horizontal shows the false positive rate. For reference, we also plotted the results of Base(EG). Fig. 6 shows all results.

As shown in Fig. 6, Prop has the highest true positive rate when the false positive rate is low. This is because Prop can reduce the accidental inliers in an incorrect image pair. In contrast, SCRAMSAC has the highest true positive rate when the false positive rate is high. This characteristic of Prop provides more reliable results in practical use with large-scale image retrieval and recognition applications.

As shown in Table II, SCRAMSAC has a lower true positive rate under the constraint that the false positive rate is zero (i.e., recognition rate) than does Base(H). In contrast, Prop always has higher true positive rate than Base(H). This is because Prop imposes a geometric constraint that is weaker than a full-2D constraint, but stronger than a pure-epipolar constraint for a database containing 3D objects. That is to say, the full-2D constraint often does not fit the 3D object of a correct image pair, and the pure-epipolar constraint often explodes the accidental inliers in an incorrect one. In contrast, Prop can find the EG to fit the 3D object in the correct image pair while reducing the accidental inliers in the incorrect one.

Regarding the computation time, as shown in Table II, Prop is faster than SCRAMSAC on both correct and incorrect image pairs. The main reason is shown by the early rejection of samples by our SRCC. For an incorrect image pair, Prop rejects almost all samples, requiring no subsequent processing. On the other hand, Prop often performs subsequent processing that is computationally expensive on the correct image pair, but it often terminates early by obtaining a high inlier ratio k in Equation 1. As a result, Prop is fast even with a correct image pair.

Therefore, Prop is superior as a GV method in the context of large scale image retrieval and recognition, both in terms of accuracy and efficiency.

V. Conclusion

In this paper, we proposed a novel geometric verification using the semi-2D constraints for 3D objects. The proposed approach reduces the accidental inliers on an incorrect image pair. Experimental results show that the proposed approach is superior to recent geometric verification approaches in the context of image retrieval and recognition.

References

[1] J. Philbin *et al.*, "Object retrieval with large vocabularies and fast spatial matching," in *CVPR*, 2007, pp. 1–8.
[2] D. G. Lowe, "Distinctive image feature from scale-invariant keypoints," *IJCV*, vol. 60, no. 2, pp. 91–110, 2004.
[3] J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos," in *ICCV*, 2003, pp. 1470–1477.
[4] M. Fischler and R. 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, 1981.
[5] O. Chum, J. Matas, and J. Kittler, "Locally optimized RANSAC," in *DAGM*, 2003, pp. 236–243.
[6] O. Chum, T. Werner, and J. Matas, "Two-view geometry estimation unaffected by a dominant plane," in *CVPR*, 2005, pp. 772–779.
[7] J. M. Frahm and M. Pollefeys, "RANSAC for (quasi-) degenerate data (QDEGSAC)," in *CVPR*, 2006, pp. 453–460.
[8] P. H. S. Torr and D. W. Murray, "Outlier detection and motion segmentation," in *SPIE*, 1993, pp. 432–443.
[9] P. H. S. Torr and D. W. Murray, "The development and comparison of robust methods for estimating the fundamental matrix," *IJCV*, vol. 24, no. 3, pp. 271–300, 1997.
[10] R. Hartley and A. Zisserman, "Multiple view geometry in computer vision," in *Cambridge University Press*, 2000.
[11] Z. Zhang, "Determining the epipolar geometry and its uncertainty: A review," *IJCV*, vol. 27, no. 2, pp. 161–195, 1998.
[12] R. Hartley, "In defense of the 8-point algorithm," in *ICCV*, 1995, pp. 1064–1070.
[13] D. Nistér, "An efficient solution to the five-point relative pose problem," *IEEE PAMI*, vol. 26, no. 6, pp. 756–770., 2004.
[14] T. Sattler, B. Leibe, and L. Kobbelt, "SCRAMSAC: Improving RANSAC's efficiency with a spatial consistency filter," in *ICCV*, 2009, pp. 2090–2097.
[15] E. Johns and G. Z. Yang, "RANSAC with 2D geometric cliques for image retrieval and place recognition," in *CVPR Workshop*, 2015, pp. 4321–4329.
[16] H. Jégou, D. Matthijs, and S. Cordelia, "Hamming embedding and weak geometric consistency for large scale image search," in *ECCV*, 2008, pp. 304–317.
[17] E. Rubee *et al.*, "ORB: an efficient alternative to sift or surf," in *ICCV*, 2011, pp. 2564–2571.
[18] O. Chum and J. Matas, "Matching with PROSAC - progressive sample consensus," in *CVPR*, 2005, pp. 220–226.