Geometric Verification using semi-2D Constraints for 3D Object Recognition

Kohei Matsuzakil*, 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 high score for an incorrect image pair due to an ambiguity of its geometric constraints. The ambiguity is caused by high degree of freedom of epipolar geometry and 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 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, 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 of the geometric verification for both a correct image pair and an incorrect image pair.**

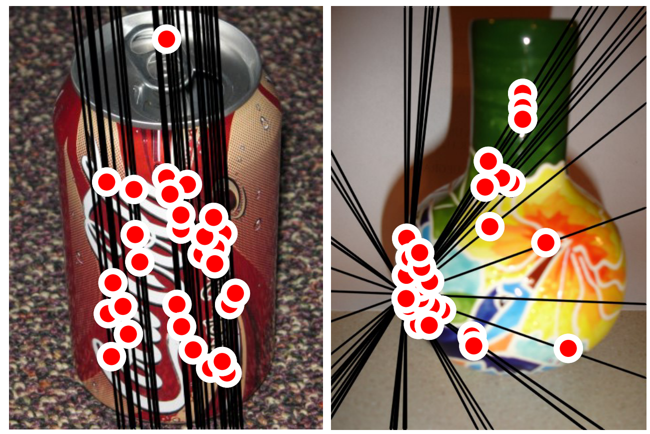

Fig. 1. Illustration of a result of geometric verification with an epipolar geometry on an incorrect image pair. Red circles and black lines represent inliers and epipolar lines respectively.

I. Introduction

Identifying correct correspondences between two or more images is an important task in many computer vision problems such as large scale image retrieval [1], object recognition [2], visual SLAM [3], and structure-from-motion [4]. The 2D-2D correspondence problem is solved using a global geometric model. For example, correspondences on two planes are described by a homography, and two-view of a rigid 3D object are described by an epipolar geometry (EG).

A typical solution to estimate the geometric model from correspondences containing outliers is RANSAC (RANdom SAmple Consensus) [5]. RANSAC computes a geometric model from a sample drawn randomly from correspondences. Then it classifies correspondences into inliers and outliers according to fitting to the model. Eventually, it chooses a geometric model which obtains the greatest number of inliers.

Many of researches about two-view geometry aim to maximize the number of inliers for an correct image pair in order to derive a high quality EG [6]–[8]. In contrast, we aim to reduce the number of inliers for an incorrect image pair since we are interested in discrimination of correct / incorrect image pairs containing a rigid 3D object by thresholding their number of inliers. It is necessary to increase the number of inliers for a correct image pair as well as reduce the number of inliers for the incorrect one in such discrimination. However, it has not been sufficiently researched so far for the latter.

Our motivation is to improve a geometric verification (GV) in a pipeline of large scale image retrieval / recognition. Given a query image containing a rigid object, large scale image retrieval [1] method retrieves a similar image containing the object from uncalibrated database images. Many of image retrieval methods typically represent an image as a set of local features such as SIFT [9], encode them with such as bag-of-words indexing [2], and rank their similarity scores between the query image and the database images. Then, they perform RANSAC-based GV to image pairs with top-N similarity scores, and re-rank them according to their GV scores. In the case of image recognition, an additional GV score thresholding is performed to determine whether the same object is in the database or not.

In such a pipeline, input to the GV is often an incorrect image pair. When we assume an affine matrix or a homography matrix as the geometric model, the number of inliers on an incorrect image pair always becomes small. However, as shown in Figure 1, it has been observed that GV with an EG often returns a high number of inliers although the image pair is entirely incorrect. This is due to both the high degree of freedom (DoF) of EG (i.e., seven DoF of fundamental matrix or five DoF of essential matrix) and ambiguous constraint of EG based on the fitting between a point and a line. Namely, this is because there is a case where many correspondences happen to fit to a certain (mostly geometrically inconsistent) EG computed from an incorrect sample.

Such accidental fits may happen frequently since epipolar constraint is satisfied when a point is locates somewhere on a corresponding epipolar line. On the other hand, it rarely happen such a case with affine constraint or homography constraint since they are based on the fitting between a point and a point. Furthermore, if we verify the query image on the large scale database, it is highly probable that an incorrect image pair with high number of accidental inliers is found. For an application containing the recognition stage, it is ideal that false positive (i.e., recognition of incorrect image) does not exist. However, setting a large threshold simply results in decrease of true positive rate.

In order to address this problem, we propose a novel approach that reduces accidental inliers in the RANSAC. This approach performs filtering to four components of RANSAC, namely correspondences, a sample, a model, and inliers individually. It does not only directly reduce the accidental inliers but also indirectly reduce them by filtering the other components in early stages. Thereby, it significantly reduces inliers resulting from the incorrect image pair.

The purpose of the proposed approach is to improve the accuracy of image recognition based on GV. Therefore, we aim to reduce GV score for the incorrect image pair while maintaining GV score for the correct one as possible. We show experimentally that the proposed approach improve the accuracy of image recognition conclusively. Furthermore, we show that our filtering of a geometrically inconsistent components achieves the speed-up of the RANSAC.

II. Related work

We summarize the RANSAC to estimate the EG [10], [11], and also briefly describe recent extensions of the RANSAC.

A. RANSAC

RANSAC estimates a model T on data Q based on iteration. It also classifies the data Q as 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 T from the sample. In the case that the model T is assumed as a fundamental matrix, the 7-point algorithm [12] ($s = 7$) or 8-point algorithm [13], [14] ($s = 8$) is typically employed. In the case that the model T is assumed as an essential matrix, the 5-point algorithm [15] ($s = 5$) is typically employed. For a correspondence m, m' in homogeneous coordinate system on data Q, it computes a square of the distance between a point and an epipolar line $d = {}^t m' T m$. In the case that d is lower than a specified threshold value, the correspondence is classified as an inlier; otherwise an outlier. When a set of inliers with maximum size so far is obtained, the number of inliers $|I|$ and T of that time are stored. In the standard approach, such processes are repeated until the number of iterations reaches following:

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

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

B. RANSAC Extensions

DEGENSAC [7] achieves more robust fundamental matrix estimation against degeneration by detecting a homography-degenerate sample. QDEGSAC [8] achieves more robust model estimation on (quasi-)degenerate data without explicit knowledge about degeneracies. However, these approaches improve the estimation quality on an correct image pair but do not reduce the number of inliers on incorrect one.

SCRAMSAC [16] forms a reduced set of correspondences with more reliability by checking a spatial consistency of each correspondence in a circular region around the correspondence. Then, it achieves a fundamental matrix estimation that is fast and more robust against degeneration by operating RANSAC on the reduced set of correspondences. However, the risk of explosion of the accidental inliers is still remaining since this approach does not include the explicit countermeasure for them.

Johns *et al.* [17] achieve 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, light environment, and so forth per target. Therefore, it is not easy to collect the training images in practical use of the large scale image retrieval / recognition applications.

The closest approach to our motivation is SCRAMSAC [16]. However, it is limited to improve the correspondences, thus the problem of the accidental inliers on an incorrect image pair is still remaining unsolved. In contrast, we propose to explicitly filter geometrically inconsistent components and 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 feature (i.e., orientation, scale, and coordinates). Figure 2 shows the overview of the proposed approach. In Figure 2, blocks of the single line and blocks of the double line 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 Figure 2) are applicable independently.

The proposed approach firstly filters an initial set of correspondences Q based on the weak geometric consistency (WGC) [18] (A in Figure 2). This filtering improves the inlier ratio of the set of correspondences on a correct image pair, thus increases the probability of finding the correct solution. Furthermore, it leads speed-up for the RANSAC by reducing the size of the resulting set of correspondences Q'. In all of the subsequent process, the proposed approach operates on the Q'. In the inside of the RANSAC iteration, the proposed approach detects a geometrically inconsistent sample, then terminates the iteration early if necessary (B in Figure 2). The proposed approach detects an EG that has not spatial consistency with the sample, then terminates the iteration early if necessary (C in Figure 2). After the model fitting to Q', the

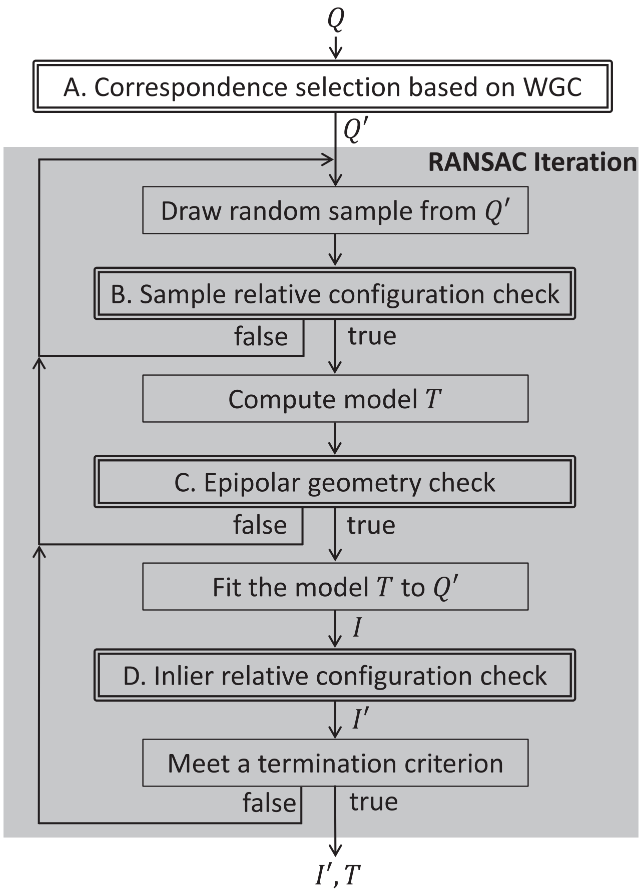

Fig. 2. Overview of the proposed approach

proposed approach check the spatial consistency of resulting set of inliers I against the sample and EG (D in Figure 2). The proposed approach filters the I through this check, then determines the final set of inliers I'.

The filters of B, C, and D in Figure 2 are based on semi-2D constraints that always can be satisfied if the target is 2D object. That is to say, these constraints result in a geometric constraint that is weaker than a full-2D constraint, but stronger than an pure-epipolar constraint. Although these constraints involve the risk of rejecting correct components (i.e., a sample, an EG, and inliers) ascribable to the parallax, most of the cases may find the sufficient solution through the RANSAC iterations. In the following, we describe the four elements of the proposed approach.

A. Correspondence Selection based on WGC

WGC [18] is originally proposed to improve a scoring algorithm in the pipeline of large scale image retrieval. Specifically, it votes the matching descriptors to the bins of orientation difference and the bins of scale ratio. Then it improves the accuracy with regard to image retrieval by filtering the bins excepting a bin with a maximum voting score. WGC is based on the assumption that correct correspondences have a consistency with regard to orientation difference and scale ratio.

We apply this assumption to the correspondence selection. 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 excepting correspondences belonging to a bin with a maximum voting score. For softer voting, we establish the bins with overlapping half of resolution with regard to orientation difference and scale ratio. When we set the resolution of orientation difference to 30 degrees, the 24 bins with overlapping every 15 degrees are established. We name this method the correspondence selection based on WGC (CSW).

By using CSW, it is expected that voting score of each bin becomes random, and the size of Q' is significantly reduced consequently on an incorrect image pair. Reducing the size of Q' result in reducing the possibility of the accidental inliers. On the other hands, potential inliers concentrate on a specific bin on a correct image pair, maintaining the final number of inliers. As a result of improved inlier ratio, expected number of iteration until finding the correct solution gets fewer.

B. Sample Ralative Configuration Check

The filter named the sample relative configuration check (SRCC) is based on the assumption that the relative configuration of correct sample (i.e., s sets of correspondences) has a consistency between a correct image pair. Geometrically inconsistent sample is rejected by using this filter, thus the potential inliers resulting from the sample are rejected.

As with shown in Figure 3, ${}_s\mathrm{C}_2$ corresponding triangles can be formed from the 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 vertexes of the triangle. It determines that the corresponding triangles do not have the inside-out if their direction is match.

For the 3D object, corresponding triangles from correct sample can generate the inside-out ascribable to parallax. However, it is expected that the frequency is significantly less than the one from the incorrect sample. Therefore, SRCC counts the number of inside-out from among the ${}_s\mathrm{C}_2$ corresponding triangles, then it returns true if the count is lower than the threshold value TH_{tri}; otherwise it returns false.

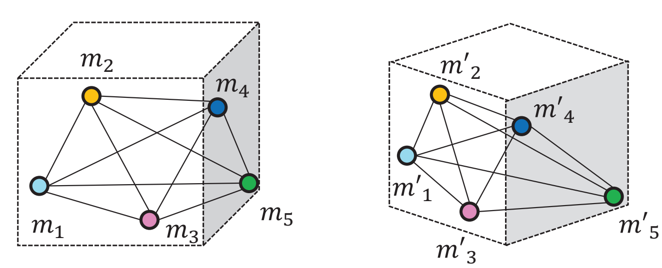

Fig. 3. An example of sample configuration in the case of $s = 5$. Colored circles represent the 5 sets of correspondences m_i, m'_i $(i = 1, 2, ..., 5)$. The ${}_5\mathrm{C}_2 = 10$ corresponding triangles are formed from them. The dotted line of the cubes represent the target 3D object

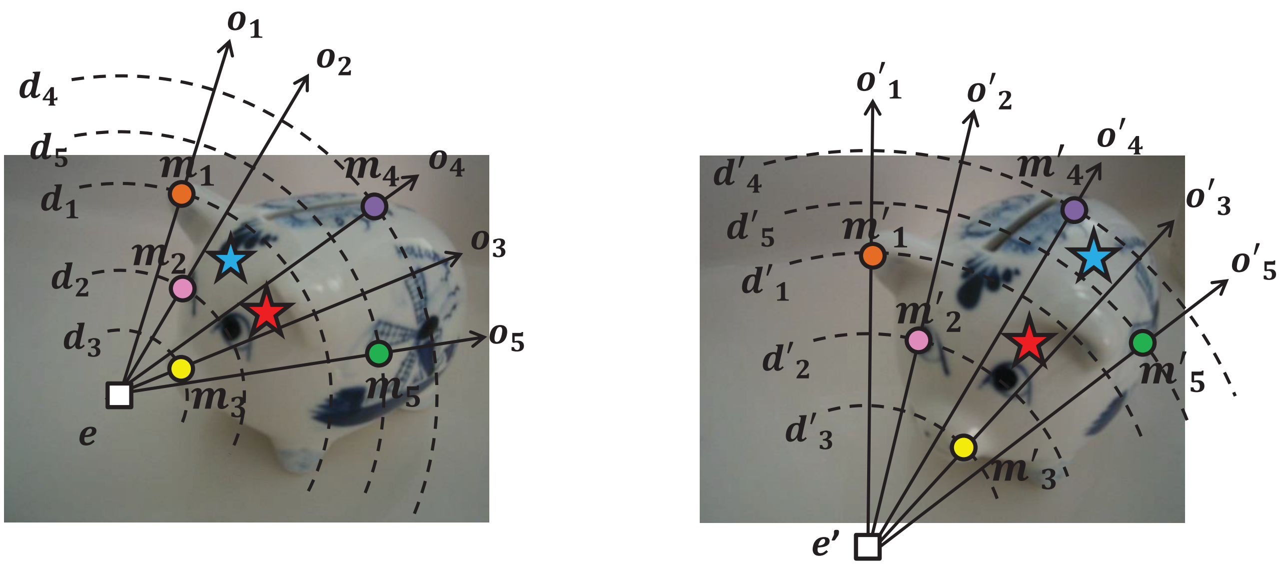

Fig. 4. Illustration of the relative configuration of components. White quadrangles, colored circles, black straight lines, and dotted lines represent epipoles projected onto the image, sample correspondences, epipolar lines, and equidistant curves from the epipole respectively. Red stars and blue stars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

This filter has a high possibility to return a wrong result if the sample points are dense in extremely narrow range or the sample points are almost on a same straight line such as on the character string image. As an exception process to avoid such cases, SRCC rejects the too less scattered sample in the 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 filter named the epipolar geometry check（EGC） is based on the assumption that the relative configuration of epipoles from correct EG and the sample has a consistency between a correct image pair. Geometrically inconsistent EG is rejected by using this filter, thus the potential inliers resulting from the EG are rejected.

As with shown in Figure 4, s sets of epipolar lines are computed from an EG and s sets of sample correspondences on each image. It is known that the epipole e on the image can be calculated from the fundamental matrix [12]. 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 Figure 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 sample correspondence ID with respect to orientation and distance in the image pair. If the permutations with regard of both orientation and distance are match, EGC returns true.

In Figure 4, both epipoles are located in the same side of the object, but they can be located in the opposite sides of the object with each other. Taking account into such a case, EGC returns true if the permutations with regard of both orientation and distance are match in reverse order.

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

D. Inlier Relative Configuration Check

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

As with shown in Figure 4, an image can be disjointed to unequal regions by using both epipolar lines and equidistant curves from the epipole. We propose to accept the inlier only if both inlier points are located in the corresponding region of each image.

Specifically, IRCC firstly computes the orientation and distance from the epipole to the inlier points on each image. Then, it finds the sample correspondence IDs of both sides of a point of inlier with regard of both orientation and distance on each image. If these sample correspondence IDs are match between the image pair, the inlier is accepted; otherwise the inlier is re-classified to outlier.

For example, an inlier represented by red stars in Figure 4 is accepted since they are located in the corresponding divided region of each image. On the contrary, an inlier represented by blue stars is re-classified to outlier since they are located in the incorresponding divided region of each image.

IV. Experiments

We experimentally evaluate the proposed approach (Prop), then compare it with SCRAMSAC [16].

A. Dataset and Experimental Setup

In our experiments, we use the University of Kentucky Benchmark[1] (UKB) which is a standard object retrieval / recognition benchmark. UKB consists of 10,200 images taking 2,550 objects from four different viewpoints. We generate ${}_4\mathrm{C}_2 = 6$ correct image pairs per object from the four images taking the same object. For the incorrect image pairs, we generate $2,550 \times 6$ sets of random image pairs from UKB as there is no correct image pair. Therefore, we conduct the GV on $2,550 \times 6 \times 2$ sets of image pair.

We regard that the GV score exceeds a certain threshold value as a "positive". We define the "recognition rate" as true positive rate when the false positive rate = 0, and "TH_{rec}" as the threshold value of that time.

For local features, we adopt the ORB [19] 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 perform the nearest neighbor matching with cross-check method. These correspondences are firstly matched to the nearest neighbor to the other side from one side, then they are filtered excepting the nearest neighbor to one side from other side.

We assume the fundamental matrix as the model of RANSAC. In order to compute it, we employ the 7-point algorithm [12]. The threshold value with respect to the distance between a point and an epipolar line is set to three. In the case that we assume the homography matrix as the model of RANSAC, the threshold value with respect to the re-projection error is set to three. The maximum number of iterations of RANSAC is set to 10,000. In order to draw a correct sample with fewer iterations, we employ the PROSAC [20] strategy with respect to the sample drawing in the Prop. For parameters specific to SCRAMSAC, we employ the same values of experiments in the literature [16] (i.e., $s_{min} = 0.5, s_{max} = 2, \theta = 0.55, r = 7$ in [16]).

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

For our CSW (A in Figure 2), the resolution with regard of orientation difference and scale ratio is set to 60 degrees and four-fold respectively. They are fairly coarse resolutions to avoid filtering correct correspondences. For our SRCC (B in Figure 2), the threshold value TH_{tri} with regard of the number of inside-out is set to three. These parameters provided excellent results in our preliminary experiments.

B. Impact of Each Element of the Proposed Approach

We evaluate each element of the Prop, CSW (A in Figure 2), SRCC (B in Figure 2), EGC (C in Figure 2), and IRCC (D in Figure 2). We employ the PROSAC [20] with an EG as a baseline method. Let "Base(EG)" denote this method. Let "Base(EG)+CSW", "Base(EG)+SRCC", "Base(EG)+EGC", and "Base(EG)+IRCC" denote the method added to each element to the Base(EG) respectively. We summarize the recognition rate, TH_{rec}, and computation time on each of the correct / incorrect image pair obtained from these methods in Table I.

TABLE I
IMPACT OF EACH ELEMENT OF THE PROPOSED APPROACH

	Recognition rate	TH_{rec}	Computation time [ms]	
			correct pair	incorrect pair
Base(EG)	0.695	68.6	170.5	294.6
Base(EG)+CSW	0.705	57.2	37.3	137.8
Base(EG)+SRCC	0.766	39.1	13.6	13.1
Base(EG)+EGC	0.745	23.2	245.9	400.2
Base(EG)+IRCC	0.797	18.2	404.8	644.2

We can see that all elements result in 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 sample and EG respectively. Base(EG)+IRCC directly reduces accidental inliers.

In regard to the computation time, it is shown that Base(EG)+CSW and Base(EG)+SRCC achieve the speed-up compared with the Base(EG), while Base(EG)+EGC and Base(EG)+IRCC are computationally expensive. However, thet later two have negligible impact on computation time of the Prop since the most iterations are terminated in an early stage by combining with the other elements. This fact is shown in the next subsection.

In general, the process on the correct pair is faster than the one on the incorrect pair since the former terminates in an early stage when it obtains a high inlier rate k in Equation 1. In contrast, note that only Base(EG)+SRCC achieves comparable results on the incorrect pair. This is because the frequency with which the iteration is early terminated by SRCC is higher on the incorrect pair.

C. Compare the Proposed Method with SCRAMSAC

We compare the proposed approach (Prop) with SCRAMSAC [16]. Given the same initial set of correspondences extracted from the same image pair, we measure the GV performances of them. As a reference of the full-2D constraint, we also measure the PROSAC [20] with a homography constraint. Let "Base(H)" denote this method.

As an indicator of GV performance, we plot the ROC (Receiver Operating Characteristic) curve. Its vertical line shows the true positive rate and the horizontal one shows false positive rate. For reference, we also plot the results of Base(EG) in the previous subsection. All results are shown in Figure 5. Plot shows the resulting average value. We summarize the recognition rate, TH_{rec}, and computation time on each of the correct / incorrect image pair obtained from these methods in Table II.

[1] http://vis.uky.edu/ stewe/ukbench

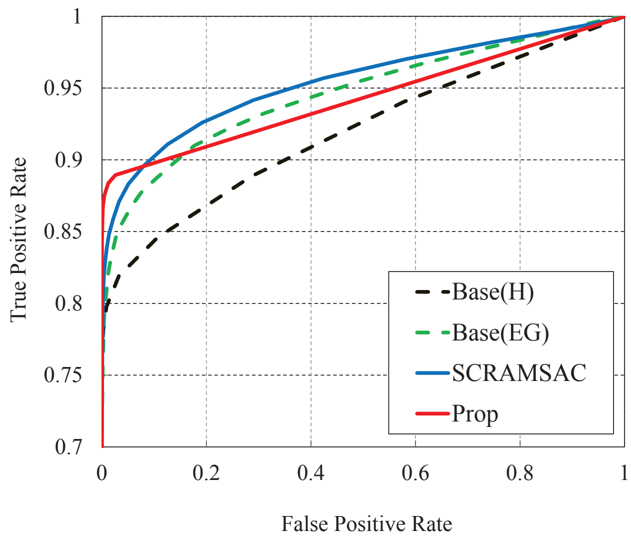

Fig. 5. ROC curves

As can be seen from Figure 5, Prop has the highest true positive rate when false positive rate is low. This is because that Prop achieves to explicitly reduce the accidental inliers on an incorrect image pair. On the other hand, SCRAMSAC has the highest true positive rate when false positive rate is high. Such characteristic of Prop provides the more reliable results with respect to the practical use of the large scale image retrieval / recognition applications.

As can be seen from Table II, SCRASAC has lower true positive rate in the case of false positive rate = 0 (i.e., recognition rate) than Base(H). In contrast, Prop always has higher true positive rate than Base(H). This is because that Prop imposes a geometric constraint that is weaker than full-2D constraint, but stronger than pure-epipolar constraint for the database containing 3D objects. That is to say, the full-2D constraint often does not fit to the 3D object on the correct image pair, and the pure-epipolar constraint often explode the accidental inliers on the incorrect one. In contrast, Prop is able to find the EG to fit the 3D object on the correct image pair while reducing the accidental inliers on the incorrect one.

In regard to the computation time, as can be seen from Table II, Prop is faster than SCRAMSAC on both correct / incorrect image pair. The main reason for this is the early rejection of sample with our SRCC (B in Figure 2). On the incorrect image pair, Prop rejects almost all samples, then it does not perform the subsequent process. Therefore, Prop on incorrect image pair is faster than the one on correct image pair. On the other hands, Prop often perform the subsequent process which is computationally expensive on the correct image pair, but it often terminates early by obtaining a result with high inlier rate k in Equation 1. As a result, Prop is fast even on the correct image pair.

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

TABLE II
COMPARE THE PROPOSED METHOD WITH SCRAMSAC

	Recognition rate	TH_{rec}	Computation time [ms]	
			correct pair	incorrect pair
Base(H)	0.768	17.3	183.5	343.1
SCRAMSAC [16]	0.735	47.0	15.5	20.2
Prop	**0.840**	**13.2**	**11.6**	**9.8**

V. CONCLUSION

In this paper, we propose a novel geometric verification using the semi-2D constraints intended for 3D object. The proposed approach reduces the accidental inliers on the incorrect image pair without additional information and prior learning. Experimental results show that the proposed approach is superior to recent geometric verification approaches in terms of image retrieval / recognition.

REFERENCES

[1] J. Philbin *et al.*, "Object retrieval with large vocabularies and fast spatial matching," in *CVPR*, 2007, pp. 1–8.
[2] J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos," in *ICCV*, 2003, pp. 1470–1477.
[3] K. Konolige and M. Agrawal, "Frameslam: From bundle adjustment to real-time visual mapping," *IEEE Transactions on Robotics*, vol. 24, no. 5, pp. 1066–1077, 2008.
[4] M. Pollefeys *et al.*, "Detailed real-time urban 3d reconstruction from video," *IJCV*, vol. 78, no. 2–3, pp. 143–167, 2008.
[5] 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.
[6] O. Chum, J. Matas, and J. Kittler, "Locally optimized RANSAC," in *DAGM*, 2003, pp. 236–243.
[7] O. Chum, T. Werner, and J. Matas, "Two-view geometry estimation unaffected by a dominant plane," in *CVPR*, 2005, pp. 772–779.
[8] J. M. Frahm and M. Pollefeys, "RANSAC for (quasi-) degenerate data (QDEGSAC)," in *CVPR*, 2006, pp. 453–460.
[9] D. G. Lowe, "Distinctive image feature from scale-invariant keypoints," *IJCV*, vol. 60, no. 2, pp. 91–110, 2004.
[10] P. H. S. Torr and D. W. Murray, "Outlier detection and motion segmentation," in *SPIE*, 1993, pp. 432–443.
[11] 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.
[12] R. Hartley and A. Zisserman, "Multiple view geometry in computer vision," in *Cambridge University Press*, 2000.
[13] Z. Zhang, "Determining the epipolar geometry and its uncertainty: A review," *IJCV*, vol. 27, no. 2, pp. 161–195, 1998.
[14] R. Hartley, "In defense of the 8-point algorithm," in *ICCV*, 1995, pp. 1064–1070.
[15] D. Nistér, "An efficient solution to the five-point relative pose problem," *IEEE PAMI*, vol. 26, no. 6, pp. 756–770., 2004.
[16] T. Sattler, B. Leibe, and L. Kobbelt, "SCRAMSAC: Improving RANSAC's efficiency with a spatial consistency filter," in *ICCV*, 2009, pp. 2090–2097.
[17] E. Johns and G. Z. Yang, "RANSAC with 2D geometric cliques for image retrieval and place recognition," in *CVPR Workshop*, 2015, pp. 4321–4329.
[18] H. Jégou, D. Matthijs, and S. Cordelia, "Hamming embedding and weak geometric consistency for large scale image search," in *ECCV*, 2008, pp. 304–317.
[19] E. Rubee *et al.*, "ORB: an efficient alternative to sift or surf," in *ICCV*, 2011, pp. 2564–2571.
[20] O. Chum and J. Matas, "Matching with PROSAC - progressive sample consensus," in *CVPR*, 2005, pp. 220–226.