Stable Two View Reconstruction Using the Six-Point Algorithm

Kazuki Nozawa, Akihiko Torii, and Masatoshi Okutomi

Tokyo Institute of Technology
2-12-1 O-okayama, Meguro-ku, Tokyo, 152-8550, Japan

Abstract. We propose a practical scheme for selecting a pair of images which can be a good initial seed for incremental SfM to accomplish a feasible reconstruction from input images with no external camera information such as EXIF. The key idea is the effective use of the 6-point algorithm by detecting infeasible pairs of images due to the degenerate configurations as well as the other conditions. We deeply analyze all the degenerate configurations of the 6-point algorithm and derive the algorithms for detecting image pairs fallen into those degenerate configurations. Further, we implement an efficient pipeline for selecting the initial pair, which can be easily plugged into the standard incremental SfM systems. Our experimental results on synthetic and real data show that our algorithms successfully detect and reject the pairs of images which are infeasible for 3D reconstruction. Further, we demonstrate 3D reconstruction by plugging our infeasible pair detection algorithm into the standard SfM pipeline.

1 Introduction

Incremental Structure-from-Motion (SfM) has achieved great successes for 3D reconstruction from photo collections [23] as well as sequential images [2], even for extremely large scale [1,7]. The resulted camera poses and scene structures (sparse 3D point clouds) are used for various applications, e.g. virtual navigation [17], camera localization [16,21], and dense reconstruction [8,3].

Since typical incremental SfM computes camera motions and scene structures by a seed-and-grow manner, it is critical to have an accurate and a stable initial seed reconstruction from a pair or a tuple of images. The initial seeds are determined by evaluating several conditions obtained from the results of pairwise image matching. For example, the commonly used conditions are quality of feature correspondences, i.e. number of matched features, and the geometric relationship among cameras and scene structures [23,10]. In this paper, we focus on the use of a pair of images as the initial seed of SfM in order to keep the simplicity and generality of the pipeline in contrast to [12].

For the pair of images selected as the initial seed, two-view reconstruction can be performed by using the 5-point algorithm [18] combining with RANSAC [6] (or its variants [4,19,20]). One of the advantages in this technique is that it has higher probability to hit a hypothesis not contaminated by outliers, i.e. robust

K.M. Lee et al. (Eds.): ACCV 2012, Part IV, LNCS 7727, pp. 122–135, 2013.

against mismatches in feature correspondences, than the 7- or the 8-point algorithm [11] since it requires fewer samples for computing each hypothesis. Another advantage is that the 5-point algorithm itself has only one degenerate configuration of cameras, which is pure rotation, thanks to the minimal computation of the essential matrix that encodes relative rotation and translation only. The natural drawback of the 5-point algorithm is to require camera intrinsic parameters by some other methods.

For most of the recent cameras, it is possible to assume zero skew, a known aspect ratio (set to 1), and a known optical center (center of an image) [25], in contrast, a focal length widely changes on every image by zooming. In order to obtain focal lengths, the popular SfM pipelines [22,28] use EXIF tags and camera manufacture specifications or, if such external information is unavailable, simply assume a certain preset such as a 60-degree field of view. With this approach, if the focal lengths are estimated with large errors, the quality of initial 3D reconstruction is very unpredictable: we cannot predict whether the errors might be compensated in bundle adjustment or the reconstruction could end up in a complete failure.

In this work, we choose the 6-point algorithm [24,15] which can compute the camera motion (fundamental matrix) and focal length from a six-tuple of correspondences. This is a natural extension of the 5-point algorithm [18] which solves minimal problems based on Gröbner basis [18] or polynomial eigenvalue problem [15]. Even though the 6-point algorithm can give the focal length estimate with only one additional correspondence w.r.t. the 5-point case, it has not been spotlighted since it requires careful treatments to degenerate configurations of camera pairs and scenes. All of the image pair becomes degenerated in some particular cases, i.e. a poster on single planar wall and nothing else is taken ("planar scene", described in Section 2.1), turntable sequences taken under the condition that the optical axis is intersected to the rotation axis of the turntable ("equidistant intersecting optical axes", Section 2.3), video sequences acquired by a vehicle-mounted camera running with no turn ("parallel axes", Section 2.4). In those cases, stable reconstruction cannot be achieved by starting an initial reconstruction using the focal length and the relative camera motion obtained by the 6-point algorithm. Torii et. al. [27] tackled this problem by detecting pairs of images acquired with degenerate configurations by adopting singular value ratio test (Section 3).

In this paper, we propose a practical scheme for selecting a pair of images which can be a good initial seed for incremental SfM to accomplish a feasible reconstruction from input images with no external camera information. The key idea is the effective use of the 6-point algorithm which gives camera motion (fundamental matrix) and focal length by efficiently detecting infeasible pairs of images due to the degenerate configurations (Section 2) as well as the other conditions (Section 3). The main contribution w.r.t. the most related work [27] is that we deeply analyze all the degenerate configurations of the 6-point algorithm, of which two are not considered in [27], and derive the algorithms for detecting image pairs fallen into those degenerate configurations. Further, we implement

an efficient pipeline for selecting the initial pair, which can be easily plugged into the standard incremental SfM systems.

Related Works: Although most of the related works are already described, we summarize a few more strongly related works. Kanatani et.al. [13,14] presented a method for computing focal lengths from a fundamental matrix computed by the 7- or 8-point algorithm in a closed form [13] and further extended its stability including the detection of degeneracy conditions in their case [14]. The theory and algorithms presented in [14] are concrete but unfortunately, it is hard to assess practical performances in challenging dataset due to the experimental validations with limited examples.

Gherardi and Fusiello [9] proposed a practical autocalibration approach which repeats update of an initial guess of intrinsic parameters of an image pair by searching an inherently bounded parameter space and by scoring likelihood of the estimated intrinsic parameters using the other cameras. Due to its nature of estimating all intrinsic parameters, the stable estimation can be achieved with more than two cameras as they demonstrated in the experiments.

In this paper, we focus on the use of the 6-point algorithm for estimating focal length from a pair of images according to its efficiency when bundled with a RANSAC scheme and its simplicity to plug into incremental SfM pipelines.

2 Detection of Degenerated Image Pairs

In this chapter, we describe four types of degeneracy underlying the computation of a relative camera motion and a focal length from an image pair using the 6-point algorithm. One is due to the degenerate scene and the others are to the degenerate camera configurations.

In most of practical situations, it is hard to classify whether the images are degenerated by using the inlier ratio resulted from RANSAC. This is because RANSAC returns the best hypothesis arbitrary fitting to an inlier set even for degenerate configurations according to noisy measurements. Even worse, the hypothesis with degeneracy often gives high inlier ratio. Therefore, we develop the algorithms which are optimized for detecting degeneracy.

2.1 Planar Scene

The degenerated scene is "planar scene"; all feature points seen by two cameras lie on a plane in a 3D space, i.e. coplanar. The 6-point algorithm can neither obtain a valid fundamental matrix nor a focal length when the six-tuple corresponding is coplanar. Planar scene is also degeneracy for the 7- and the 8-point algorithms, so that the detection algorithm is well known [5]. We can detect this degeneracy by explicitly computing a homography from the corresponding feature points.

As the similar way proposed in DEGENSAC [5], the image pairs degenerated by planar scene is quickly detected by verifying whether a six-tuple of points used

for computing the fundamental matrix is related by a homography. In detail, we compute a homography using a four-tuple out of the six and verify whether the remaining two fit in the homography. Instead of testing this degeneracy for samples on each RANSAC loop as [5], we test 15 homographies obtained from the six-tuple resulted by RANSAC. The important idea in DEGENSAC [5] is to find the stable hypothesis even from the scene dominated by a plane using the plane-plus-parallax. In contrast, we simply reject such an image pair by assuming we have sufficiently large dataset and better seeds exist for the following SfM.

Additionally, we check whether the scene is dominated by a single homography by using the standard 4-point RANSAC for all of the input correspondences. If the scene is actually planar, the number of inliers resulted by RANSAC on a homography hypothesis increase. Therefore, we can detect the degeneracy by comparing the number of inliers which support the fundamental matrix and the homography. This is computationally costly but more robust to the noise than checking six-tuple of correspondences only.

2.2 Pure Rotation

The most simple degenerate camera configuration is "pure rotation"; two cameras are configured without translation. In this case, the 6-point algorithm fails to estimate both the fundamental matrix and the focal length. This is also detected by using the same algorithm based on homography as described in Section 2.1. Note that this is also degeneracy for the 5-point algorithm as well as the 7- and the 8-point algorithms.

2.3 Equidistant Intersecting Optical Axes

The third degeneracy is that the optical axes of the two cameras intersect and the two distances between the camera center and the intersection point are the same. This configuration can be interpreted as cameras lie on a sphere and their optical axes are oriented to the center of the sphere. This degeneracy often occurs in practice, for instance, if images taken by a fixed camera while a target object moved on a turntable and the optical axis intersected to the rotation axis of the turntable, the configuration falls into this degeneracy.

Further, it is impossible to obtain the correct focal length but the estimated fundamental matrix is still valid [24].[1] Therefore, we can detect this configuration by evaluating the projection of the optical axes. The detection is composed by two steps of evaluating the necessary conditions: (i) detection of coplanar optical axes; (ii) detection of isosceles triangle composed of the camera centers and the intersection point.

Step1: Detecting Coplanar Optical Axes. One of the necessary conditions of this degeneracy is that the optical axes must lie on the same plane in 3D space.

[1] This degenerate configuration gives 3D scene reconstruction by arbitrary but common focal length due to the remaining projective ambiguity.

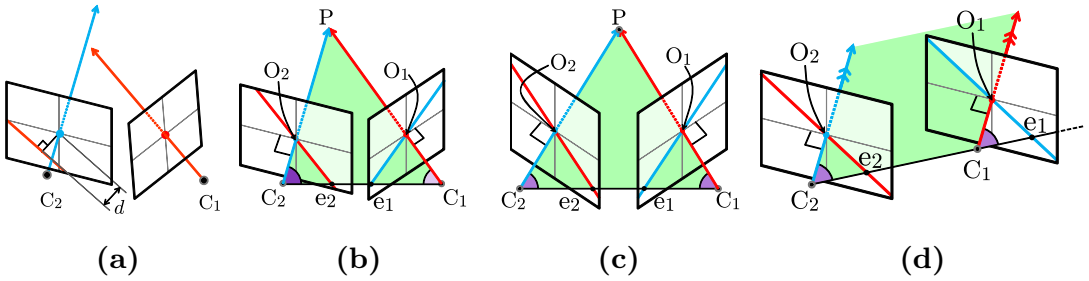


Fig. 1. (a), **(b)**: The optical axes are not coplanar or intersect at the difference distance from the cameras. The 6-point algorithm can estimate the focal length. **(c)**, **(d)**: The optical axes intersect at the same distance from the cameras or are parallel so that focal length cannot be recovered.

We consider a pair of cameras; camera 1 and camera 2. On the image plane of camera 2, we draw the epipolar line corresponding to the image center of camera 1. The epipolar line is represented as:

$$\begin{pmatrix} x & y & 1 \end{pmatrix} \boldsymbol{F} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \boldsymbol{F}_{13}x + \boldsymbol{F}_{23}y + \boldsymbol{F}_{33} = 0 \tag{1}$$

This line passes through the image center only when the optical axes intersect or parallel (Fig. 1 (b), (c), (d)). It is possible to detect if the image pair has intersectional or parallel optical axes by measuring the distance d between the image center and the epipolar line corresponding to the image center of the other,

$$d = \frac{|\boldsymbol{F}_{33}|}{\sqrt{\boldsymbol{F}_{13}^2 + \boldsymbol{F}_{23}^2}} \tag{2}$$

The epipolar line on the other image is also verified. This detection is similar to the detection described in [14].

Step2: Detecting Isosceles Triangle. There is another necessary condition for the degeneracy of equidistant intersecting optical axes (Fig. 1 (c)). If the optical axes intersect equidistantly, the triangle composed of the camera centers and the intersection point is isosceles. This degeneracy could be detected by evaluating whether the triangle is isosceles or not. However, of course, the angles between the optical axes and the epipole cannot be computed in Euclidean space because the correct focal lengths are not estimated from the 6-point algorithm due to the degeneracy.

This degeneracy can be still detected by using the following geometrical relationships assuming an unknown but a common focal length. Let us consider the two triangles $\triangle(\mathrm{O}_1, \mathrm{C}_1, \mathrm{e}_1)$ and $\triangle(\mathrm{O}_2, \mathrm{C}_2, \mathrm{e}_2)$, where C_1 and C_2 are the camera centers, O_1 and O_2 are the image centers, e_1 and e_2 are the epipoles, respectively. Note that the angles $\angle(\mathrm{C}_1, \mathrm{O}_1, \mathrm{e}_1) = \angle(\mathrm{C}_2, \mathrm{O}_2, \mathrm{e}_2) = \pi/2$ because they are the intersection of optical axes and the image planes. Further, the lengths

O_1C_1 and O_2C_2 are equal since we assumed a common focal length. Consequently, if the lengths O_1e_1 and O_2e_2 are equal, the two triangles $\triangle(O_1, C_1, e_1)$ and $\triangle(O_2, C_2, e_2)$ are congruent, and the angles $\angle(O_1, C_1, e_1)$ and $\angle(O_2, C_2, e_2)$ are the same. Thus, the triangle $\triangle(P, C_1, C_2)$ is isosceles. Since the distances O_1e_1 and O_2e_2 can be computed on the image planes, it is possible to use these distances for detecting equidistant intersecting optical axes.

2.4 Parallel Optical Axes

The last case is when two optical axes are parallel. Unfortunately, this configuration also occurs frequently since the camera motion under pure translation plus rotation around the optical axis is included in this degeneracy.

As in the case of the intersectional and equidistant optical axes, focal length cannot be recovered; on the other hand, the estimated fundamental matrix is still valid. When the optical axes are parallel, they are coplanar and the corresponding angles consisted by the two optical axes and the epipole are equivalent (Fig. 1 (d)). Note that the parallel optical axes can be considered as intersecting at the point at infinity. Therefore, we can detect them using the same algorithm as detecting the intersectional and equidistant optical axes.

3 Detection of Invalid Essential Matrices

For the practical use of the 6-point algorithm with RANSAC, the degenerate configurations of cameras and scenes are not the only reason for contaminating the estimation of a relative camera motion and a focal length. The estimation fails if the measurements of features are too noisy or if some pairs of images with different focal lengths are included in the image set. Torii et al. [27] found that the quality of the estimation of the algorithm is correlated with the ratio of the two non-zero singular values of the essential matrix and thus the ratio can be a criterion of evaluating the validity of estimation.

Singular Value Test (SVT). The two non-zero singular values of an essential matrix is ideally equivalent. The 6-point algorithm uses this property as one of the constraint for minimal solution, so that the ratio of two non-zero singular values of essential matrix decomposed from the fundamental matrix $\boldsymbol{F}_{6pt}$ using the focal length f_{6pt} is always one. Here, if the estimation of a fundamental matrix and a focal length is successful and the inlier set $\mathrm{inl}_{6\mathrm{pt}}$ supporting them is geometrically correct, the fundamental matrix $\boldsymbol{F}_{LS}$ re-estimated from $\mathrm{inl}_{6\mathrm{pt}}$ using least squares should be valid. Then, the essential matrix $\boldsymbol{E}_{LS}$ obtained by factorizing the fundamental matrix $\boldsymbol{F}_{LS}$ using f_{6pt} should have the ratio of two non-zero singular values s_1 and s_2 to be one. We use the singular value ratio (SVR) $\tau = s_2/s_1$, where $s_1 \geq s_2$, of $\boldsymbol{E}_{LS}$ for classifying if the estimation is valid.

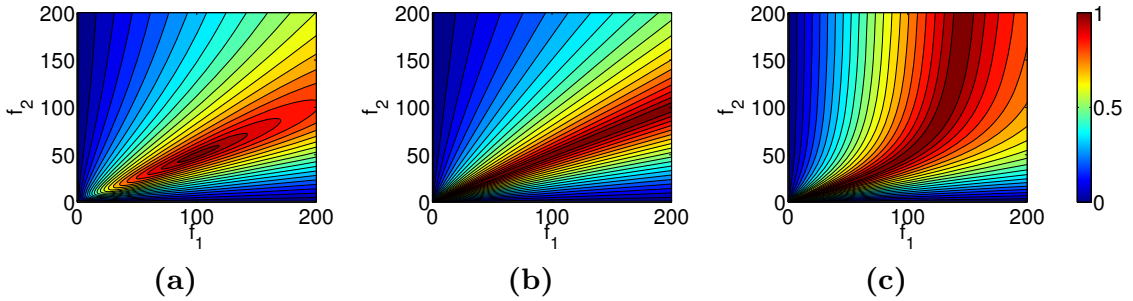


Fig. 2. The SVR of essential matrix decomposed from valid fundamental matrix using various focal lengths. The ground truth is $f_{t1} = 100, f_{t2} = 50$. (a) The camera configuration is non-degenerate. SVR is high only when it is computed with the correct focal lengths. (b) The optical axes intersect at the same distance from the cameras or are parallel. SVR can be high when $f_1 : f_2$ is equivalent to $f_{t1} : f_{t2}$. The algorithm described in Section 2.3 can detect this case. (c) The optical axes intersect at the different distances from the cameras. SVR can be high even though the focal length estimation via the 6-point algorithm fails.

Limitation of the SVT. For the scene and the camera configuration with non-degeneracy, SVR is close to one only if the fundamental matrix is factorized using correct focal lengths (Fig. 2 **(a)**). Therefore, SVT can detect the invalid estimation of the 6-point algorithm due to the measurement noise or input of image pairs with different focal lengths.

Meanwhile, SVT is confused by two conditions: (i) the optical axes of the cameras are parallel or intersect equidistantly; (ii) the optical axes intersect at the different distance from the cameras.

In the case of (i), note that this is degenerate condition described in Section 2.3 and Section 2.4, SVR is high when $f_1 : f_2$ is correctly estimated as well as f_1 and f_2 respectively are (Fig. 2 **(b)**). When the genuine focal lengths are equivalent each other, SVT cannot detect the failure of focal length estimation because the solution of the 6-point algorithm always satisfies $f_1 : f_2 = f_{6pt} : f_{6pt} (= 1 : 1)$ and the SVR is one. However, this case can be detected by checking the projections of the optical axes as Section 2.3 because fundamental matrix is estimated correctly when the images have a common focal length. Besides, if the focal lengths are different, the solution is always to be $f_1 : f_2 \neq f_{6pt} : f_{6pt}$, so the SVR decrease.

The problematic case is (ii): SVR is high with certain combination of the focal lengths (Fig. 2 **(c)**). This combination is associated by the focal lengths and the distances to the intersection point. In this situation, we cannot evaluate the validity of estimation using SVR. Moreover, the 6-pooint algorithm can estimate neither the focal length nor the fundamental matrix from the image pair with various focal lengths. We cannot detect this case by the detectors we described above without the fundamental matrix and the reliability of SVT, so that we need another algorithm to detect it.

In order to detect the case (ii), we re-estimate fundamental matrix via linear 8-point algorithm with RANSACing all the tentative matches. All the image pairs which deteriorate the 8-point algorithm are detected by the detector with

homography, so that the estimation will be successful. We can detect the case by checking the co-planarity of the optical axes in Section 2.3 using the correct fundamental matrix computed by the 8-point algorithm.

4 The Pipelines of Initial Image Pair Selection

We consider two types of pipelines to select an initial seed for stable initial reconstruction. The first pipeline assumes the input images have a fixed focal length. The second one assumes more general situation, i.e. the input images with various focal lengths. We combine all the detectors described above while taking into account the efficiency.

4.1 A Fixed Focal Length

1. Pick a pair of input images.
2. Estimate the focal length f_6 and the fundamental matrix $\boldsymbol{F}_6$ using the 6-point algorithm with RANSAC and obtain the set of inliers inl_6 which support $\boldsymbol{F}_6$.
3. If the number of inliers inl_6 is less than a threshold (30 in the experiments) the pair is rejected from initial pair candidate then go to step1. This is the detection of image pairs which have no or small common field of view.
4. Compute a homography $\boldsymbol{H}_6$ from the four-tuple out of the 6 points used for computing $\boldsymbol{F}_6$ and f_6.
5. If the six-tuple of points is coplanar, then go to step1.
6. Re-estimate the fundamental matrix $\boldsymbol{F}_{LS}$ from inl_6 using the least squares.
7. Draw the epipolar line corresponding to the center of the other image.
8. If the distance is smaller than the threshold (5% of image width in the experiments), they have coplanar optical axes, move to step9, otherwise move to step1.
9. Calculate the distances between the image center and the epipole.
10. If the distances are almost the same, reject the pair and go to step1.
11. Compute SVR from the essential matrix $\boldsymbol{E}_{LS}$ obtained from $\boldsymbol{F}_{LS}$ and f_6.
12. If the SVR is smaller than threshold (0.98 in the experiments) reject the pair and move to step1.
13. Decompose $\boldsymbol{E}_{LS}$ into a rotation matrix and a translation vector using the result of SVD.
14. Triangulate all the correspondents which support the estimation.
15. Compute the dominant apical angle [26] of the reconstruction points.
16. If the DAA is small (smaller than 0.1 deg), reject the pair and go to step1. This is because a point whose apical angle is extremely small tends to magnify the noise.
17. Run four-point-sampling RANSAC and obtain the likely homography and the inlier set inl_H which support them.
18. If the $\text{inl}_\text{H} > \text{inl}_6$, reject the pair and move to step1
19. If this is the last pair, quit the procedure; if not, move back to step1

When all the image pairs have been gone through this procedure, we can start the initial reconstruction from one of the initial pair candidates. If there is no pair available for stable reconstruction with the 6-point algorithm, our pipeline successfully selects "none" as an initial seed and avoids performing meaningless reconstruction.

4.2 Various Focal Lengths

When we include images taken with different focal lengths, we have to add a few more steps to the algorithm in Section 4.1 since it is necessary to detect and to reject the image pairs with different focal lengths.

As we described in Section 3, the image pair with coplanar optical axes must be simply detected and rejected, thus the step8 on the pipeline is replaced with

8. If the distance is smaller than the threshold, reject the pair and go to step1.

so the step9 and 10 in the pipeline in Section 4.1 are skipped.

Finally, we add the following steps,

19. Re-estimate the fundamental matrix $\boldsymbol{F}_{L8}$ via linear 8-point algorithm with RANSACing all the tentative matches.
20. Re-execute the step7 and 8 using $\boldsymbol{F}_{L8}$
21. Re-execute the step11 and 12 using $\boldsymbol{F}_{L8}$
22. If this is the last pair, quit the procedure; if not, move back to step1

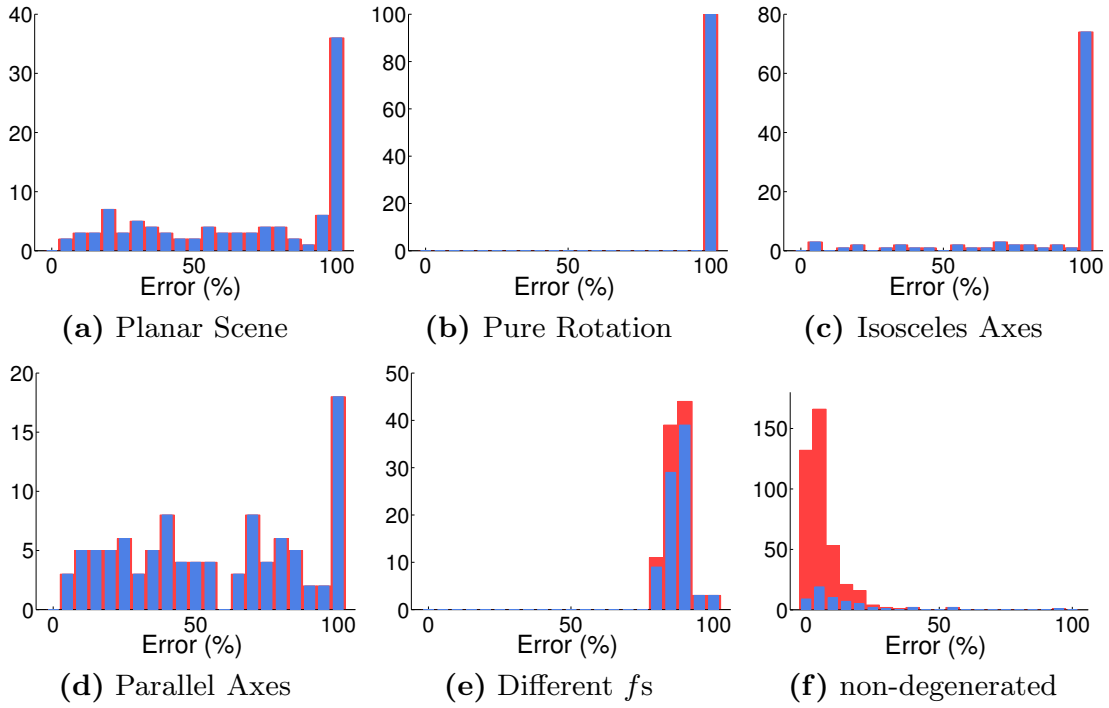
Since the additional process using the 8-point algorithm with RANSAC is costly, the pipelines are explicitly separated.

5 Experiments

5.1 Synthetic Data

We demonstrate the proposed degeneracy detection on synthetic data. First, we consider the six different cases, as shown in Fig. 3. There are four degenerated sets: (a) planar scene, (b) pure rotation, (c) isosceles axes, and (d) parallel optical axes. The other case is (e) different fs and (f) non-degenerated which is designed absolutely not to degenerate.

The image size is 2288×1520 and Gaussian noise ($\sigma = 3$) is added on the images. We reconstructed the scene with the 6-point algorithm and analyzed the "error". Here, an "error" is defined for each image pair as the average distance between the reconstructed point and the ground truth. The distance is normalized by the distance from the camera. For (a)-(e), we applied a detector which is related to the degeneracy or infeasible condition. For (f), we applied all the detectors to the non-degenerated dataset. Figure. 3 shows that the reconstruction from all the dataset but the non-degenerated one is unreliable. However, all the degenerated pairs and most of the different focal lengths pairs are detected. There are few false positives in well-reconstructed dataset, the non-degenerated one.



(a) Planar Scene **(b)** Pure Rotation **(c)** Isosceles Axes

(d) Parallel Axes **(e)** Different fs **(f)** non-degenerated

Fig. 3. The histograms of the average error of the 3D reconstruction using the focal length estimated with the 6-point algorithm. The error is defined as the distance from the reconstructed point to the original position divided by the distance from the camera as normalization. **(a)** – **(e)** Each graph represents the 100 image pairs satisfies different kinds of infeasible condition. (*Red*) all the image pairs. (*Blue*) image pairs detected as infeasible for reconstruction. **(f)** 400 non-degenerate pairs. (*Red*) all the image pairs. (*Blue*) the image pairs failed at least one examination.

5.2 Real Data

Next, we evaluate the performance of the proposed pipeline on the real photographs. We make two real image datasets. "Campus" dataset is composed of 200 pictures of a campus building with a fixed focal length, which gives 19900 image pairs. "Trevi" dataset is composed of 46 images available on Flickr, which gives 1035 pairs. We reject all the image pairs which deteriorating 3D reconstruction as we described in Section 4.

The histograms in Fig. 4 (a), (b) show the results of focal length estimation. We define the error as the difference between the estimated focal lengths and the ones obtained from the EXIF tag (by regarding as the ground truth) normalized by the ground truth. The errors of estimation from the image pairs without common field of view are meaningless so that they, few matching pairs, are rejected from the evaluation. Note that the gray histogram is made up with the image pairs which are NOT detected. There are a large number of image pairs that cannot provide the accurate focal lengths with the 6-point algorithm. Specifically, the focal length estimation from Trevi dataset is very difficult because most of them are compound different focal lengths pair. We can see that most of the image pairs which provide large error are detected. Although there

are some false positives remained, the initial seed selection is not contaminated combining with the scoring process after the removal of most of the infeasible image pairs.

We score the image pairs by

$$s = \frac{s_1 + s_2 + s_3 + s_4}{4}, \qquad \text{where} \begin{cases} s_1 = 1 - \dfrac{|\boldsymbol{m}^{\mathrm{H}}|}{|\boldsymbol{m}|} \\ s_2 = \dfrac{|\boldsymbol{m}|}{500} \\ s_3 = 1 - \dfrac{1 - SVR}{1 - 0.98} \\ s_4 = \dfrac{CH(C_i) + CH(C_j)}{A_i + A_j} \end{cases} \tag{3}$$

First three terms are described on [27]. $|\boldsymbol{m}|$ represents the number of the inlier supporting the fundamental matrix. $|\boldsymbol{m}^{\mathrm{H}}|$ denotes the number of the inlier supporting the homography obtained via RANSAC. SVR of the s_3 denotes the singular value ratio of the obtained essential matrix. s_4 refer a part of the score described in [10]. $CH(\cdot)$ is the area of the convex hull of a set of points and $C_{\cdot}$ represents the point that corresponding is confirmed. $A_{\cdot}$ is the area of the image. Then we make Bundler [22] start initial reconstruction from the pair with best score using the focal length estimated with the 6-point algorithm. We qualitatively compare the result with the reconstruction by Bundler with and without EXIF tag. The reconstructed point clouds are shown in Fig. 5 and the numerical data are on Table 1. Without EXIF, bundler obtains the focal lengths of the initial pair supposing the angle of view but sometimes, like these cases, it does not work well. The reconstruction of campus without EXIF (Fig. 5 (c)) is damaged. On the other hand, the result of Trevi without EXIF (Fig. 5 (d)) still keeps a shape of facade but it is skewed and the number of reconstructed points are much less than the other two (Figs. 5 (f), (h)). Using the 6-point algorithm with our proposed criteria, 3D reconstruction succeeds without any external information and the output is promising.

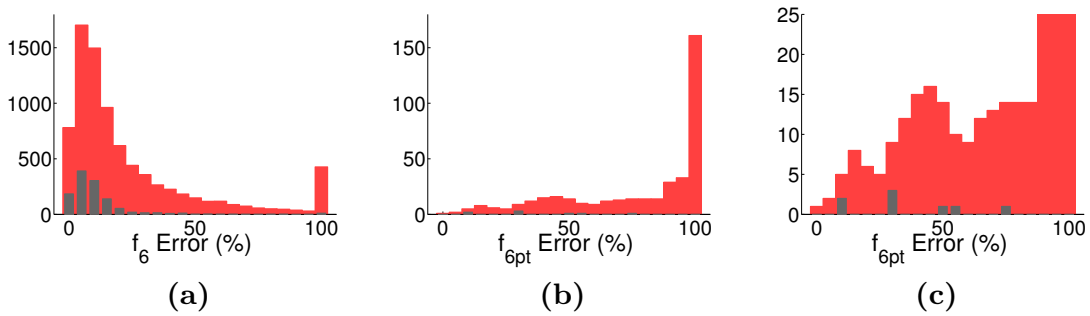


Fig. 4. The histogram of the error of focal length estimation which are composed of the results from (*Red*) image pairs whose enough numbers of matches are counted. (*Gray*) image pairs that pass all the examinations. (a) Campus; fixed focal length. (b) Trevi; images from Flickr. (c) is a close-up of (b) at lower frequency.

Table 1. The number of cameras whose pose is estimated, reconstructed points, and the comparison of (estimated) focal lengths of initial pair and that on the EXIF as ground truth

	Campus		Initial f	Trevi		Initial fs
	Cameras	Points	(f on EXIF)	Cameras	Points	(fs on EXIF)
Bundler (without EXIF)	200	31,217	532 (1026)	22	3,077	532 (1591) 532 (1818)
Bundler (with EXIF)	200	66,281	1026 (1026)	24	6,149	1591 (1591) 1818 (1818)
Proposed + Bundler (without EXIF)	200	66,577	1039 (1026)	24	6,227	2013 (1652) 2013 (1624)

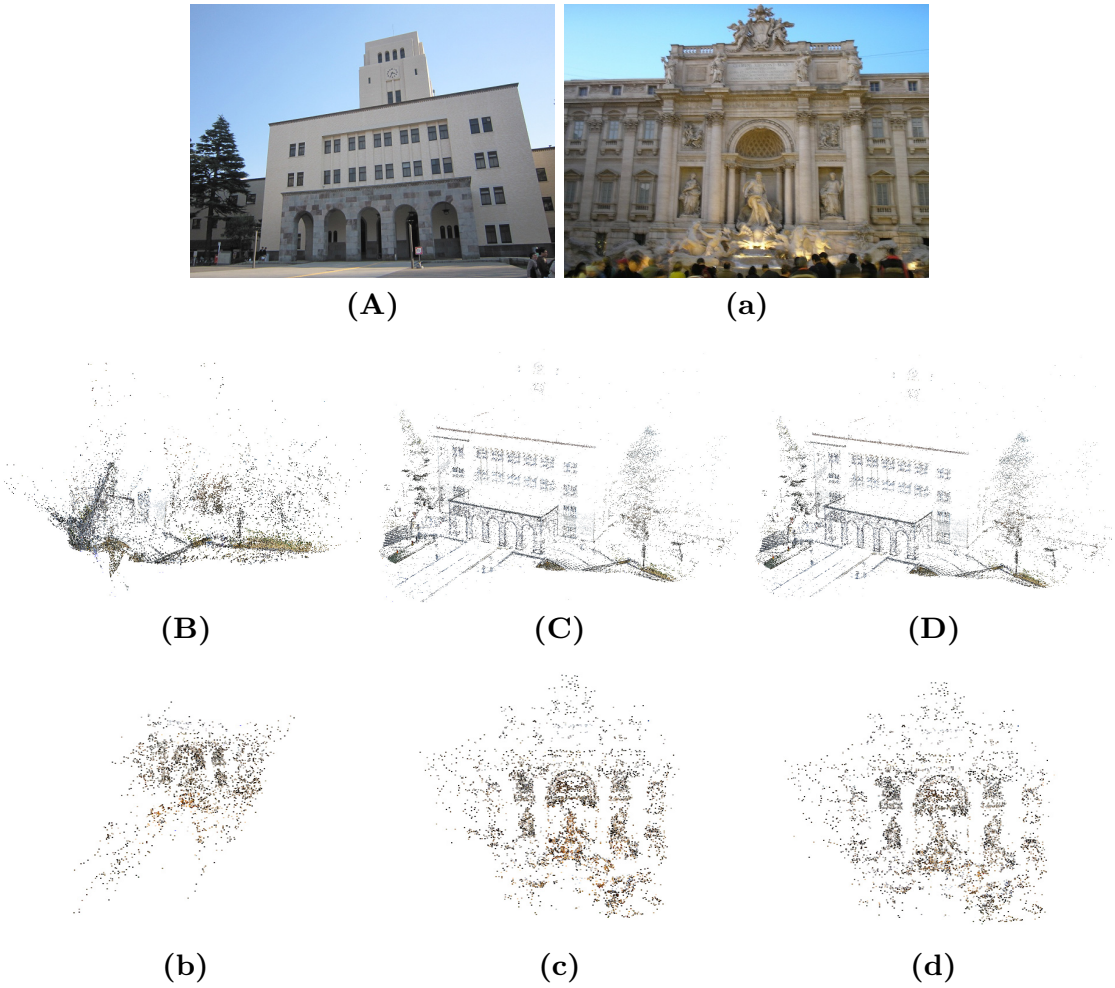


Fig. 5. **(A)**: The picture of the campus building. **(a)**: The picture of Trevi Fountain downloaded from Flickr. **(B)**,**(b)**: Reconstruction by Bundler without EXIF. **(C)**,**(c)**: Reconstruction by Bundler using EXIF. **(D)**,**(d)**: Reconstruction by Bundler without EXIF starting from the image pair and using the focal length which are obtained by our proposal.

6 Conclusions

We list up all the conditions deteriorating the 6-point algorithm and proposed the criteria for detecting the entire image pairs infeasible for it; especially, in the focal length estimation, the intersection of optical axes is very critical so we showed how to detect the intersection. We showed the performance of our tests and that we can start initial reconstruction stably without any ancillary information or invalid assumptions on angle of view.

Acknowledgement. This work was partly supported by Grant-in-Aid for Scientific Research (21240015) from the Japan Society for the Promotion of Science.

References

1. Agarwal, S., Snavely, N., Simon, I., Seitz, S.M., Szeliski, R.: Building rome in a day. In: Proc. IEEE 12th Int. Computer Vision Conf., pp. 72–79 (2009)
2. Akbarzadeh, A., Frahm, J.M., Mordohai, P., Clipp, B., Engels, C., Gallup, D., Merrell, P., Phelps, M., Sinha, S., Talton, B., Wang, L., Yang, Q., Stewénius, H., Yang, R., Welch, G., Towles, H., Nistér, D., Pollefeys, M.: Towards urban 3D reconstruction from video. In: Proc. 3DPVT (2006)
3. Chauve, A.L., Labatut, P., Pons, J.P.: Robust piecewise-planar 3d reconstruction and completion from large-scale unstructured point data. Proc. CVPR, 1261–1268 (2010)
4. Chum, O., Matas, J.: Matching with prosac: Progressive sample consensus. In: Proc. CVPR, pp. I:220–I:226 (2005)
5. Chum, O., Werner, T., Matas, J.: Two-view geometry estimation unaffected by a dominant plane. In: Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition, CVPR 2005, vol. 1, pp. 772–779 (2005)
6. Fischler, M.A., Bolles, R.C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Comm. ACM 24, 381–395 (1981)
7. Frahm, J.-M., Fite-Georgel, P., Gallup, D., Johnson, T., Raguram, R., Wu, C., Jen, Y.-H., Dunn, E., Clipp, B., Lazebnik, S., Pollefeys, M.: Building Rome on a Cloudless Day. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part IV. LNCS, vol. 6314, pp. 368–381. Springer, Heidelberg (2010)
8. Furukawa, Y., Ponce, J.: Accurate, dense, and robust multiview stereopsis. PAMI 32, 1362–1376 (2010)
9. Gherardi, R., Fusiello, A.: Practical Autocalibration. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part I. LNCS, vol. 6311, pp. 790–801. Springer, Heidelberg (2010)
10. Gherardi, R., Toldo, R., Farenzena, M., Fusiello, A.: Samantha: Towards automatic image-based model acquisition. In: Proc. Conf. Visual Media Production, CVMP, pp. 161–170 (2010)
11. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press (2000) ISBN: 0521623049
12. Havlena, M., Torii, A., Pajdla, T.: Efficient structure from motion by graph optimization. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part II. LNCS, vol. 6312, pp. 100–113. Springer, Heidelberg (2010)

13. Kanatani, K., Matsunaga, C.: Closed-form expression for focal lengths from the fundamental matrix. In: Proc. 4th Asian Conf. Computer Vision, pp. 128–133 (2000)
14. Kanatani, K., Nakatsuji, A., Sugaya, Y.: Stabilizing the focal length computation for 3-d reconstruction from two uncalibrated views. Int. J. Comput. Vision 66, 109–122 (2006)
15. Kukelova, Z., Bujnak, M., Pajdla, T.: Polynomial eigenvalue solutions to the 5-pt and 6-pt relative pose problems. In: Proc. BMVC 2008 (2008)
16. Li, Y., Snavely, N., Huttenlocher, D.P.: Location Recognition Using Prioritized Feature Matching. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part II. LNCS, vol. 6312, pp. 791–804. Springer, Heidelberg (2010)
17. Microsoft: Photosynth (2008), http://livelabs.com/photosynth
18. Nistér, D.: An efficient solution to the five-point relative pose problem. IEEE Transactions on Pattern Analysis and Machine Intelligence
19. of RANSAC, Y.: 26, 756–770 (2004) (2008), http://cmp.felk.cvut.cz/ransac-cvpr2006/
20. Raguram, R., Frahm, J.-M., Pollefeys, M.: A Comparative Analysis of RANSAC Techniques Leading to Adaptive Real-Time Random Sample Consensus. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part II. LNCS, vol. 5303, pp. 500–513. Springer, Heidelberg (2008)
21. Sattler, T., Leibe, B., Kobbelt, L.: Fast image-based localization using direct 2d-to-3d matching. In: ICCV, pp. 667–674 (2011)
22. Snavely, N.: Bundler: Structure from motion (sfm) for unordered image collections (2008), http://phototour.cs.washington.edu/bundler/
23. Snavely, N., Seitz, S., Szeliski, R.: Modeling the world from internet photo collections. International Journal of Computer Vision 80, 189–210 (2008)
24. Stewenius, H., Nister, D., Kahl, F., Schaffalitzky, F.: A minimal solution for relative pose with unknown focal length. In: Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition, CVPR 2005, vol. 2, pp. 789–794 (2005)
25. Szeliski, R.: Computer vision: algorithms and applications. Springer, New York (2010)
26. Torii, A., Havlena, M., Pajdla, T., Leibe, B.: Measuring camera translation by the dominant apical angle. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition, CVPR 2008, 1–7 (2008)
27. Torii, A., Kukelova, Z., Bujnak, M., Pajdla, T.: The Six Point Algorithm Revisited. In: Koch, R., Huang, F. (eds.) ACCV Workshops 2010, Part II. LNCS, vol. 6469, pp. 184–193. Springer, Heidelberg (2011)
28. Wu, C.: Visualsfm: A visual structure from motion system (2011), http://www.cs.washington.edu/homes/ccwu/vsfm/