algorithms

MDPI

Article

Fundamental Matrix Computing Based on 3D Metrical Distance

Xinsheng Li * and Xuedong Yuan *

College of Computer Science, Sichuan University, Chengdu 610064, China
* Correspondence: lixinsheng@scu.edu.cn (X.L.); yxd@scu.edu.cn (X.Y.)

Abstract: To reconstruct point geometry from multiple images, computation of the fundamental matrix is always necessary. With a new optimization criterion, i.e., the re-projective 3D metric geometric distance rather than projective space under RANSAC (Random Sample And Consensus) framework, our method can reveal the quality of the fundamental matrix visually through 3D reconstruction. The geometric distance is the projection error of 3D points to the corresponding image pixel coordinates in metric space. The reasonable visual figures of the reconstructed scenes are shown but only some numerical result were compared, as is standard practice. This criterion can lead to a better 3D reconstruction result especially in 3D metric space. Our experiments validate our new error criterion and the quality of fundamental matrix under the new criterion.

Keywords: fundamental matrix; computer vision; RANSAC; metrical distance; structure from motion

Citation: Li, X.; Yuan, X. Fundamental Matrix Computing Based on 3D Metrical Distance. *Algorithms* **2021**, *14*, 89. https://doi.org/10.3390/a14030089

Academic Editor: Maurizio Patrignani

Received: 25 January 2021
Accepted: 2 March 2021
Published: 15 March 2021

1. Introduction

Nowadays, the demand for 3D graphical models in computer graphics, virtual reality and communication, etc., and their application are increasing rapidly. Much progress in multi-view stereo (MVS) algorithms has been made in recent years, both in terms of precision and performance. Fundamental matrix F computation is a basic problem in multi-view geometry for the reconstruction of 3D scenes. Once F is obtained, the projective reconstruction of a scene or object can be inferred from the fundamental matrix [1].

Usually, the eight-point method (8-pt) [2,3], seven-point method (7-pt), five-point method (5-pt) [4,5], and gold distance method (Gold) [6] are all based on the RANSAC (RANdom Sample And Consensus) framework. If picture quality and feature matching are precise enough, accurate results can be achieved; however, the estimation from the seven-point method is sensitive to noise. The five-point method makes use of all kinds of constraints but is not very robust. These methods often include a non-linear optimization as the last step, which is time-consuming. The sampling method is based on the distance from feature points to epipolar lines on the image to select the best F from the candidate F. The F with least average distance wins and is output as the final result. So the quality of F is determined by this distance number.

According to the epipolar theory in the multi-view geometry of computer vision, as demonstrated in Figure 1, all the epipolar lines, l_1, l_2, intersect at a single point, epipole, which is e or e' at each image plane. An epipolar line such as l_1 or l_2 is the intersection of an epipolar plane with the image plane. The epipolar plane, e.g., $O_1O_2X_i$, is a plane containing the baseline, e.g., O_1O_2. The epipole is the point of intersection of the line joining the camera centers (the baseline) with the image plane. Equivalently, the epipole is the image in one view of the camera centre of the other view. E.g. e is the projection of O_2 on the left view in Figure 1. It is also the vanishing point of the baseline (translation) direction.

This geometry relation between two images for X_i, $[u, u']$ is linked through F. Our method also utilizes the RANSAC framework to compute fundamental matrix F like most of the other ordinary algorithms.

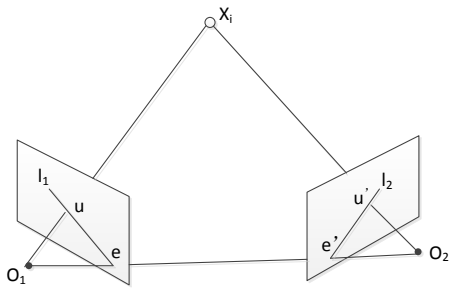


Figure 1. Two-view stereo to demonstrate the epipolar theory.

Although every result after RANSAC could have very small average distance from matching feature pixel to the epipolar lines, sometimes the reconstruction result is still unstable if the reconstructed scene is viewed by the other different standard. For example, epipoles of different final best samplings distribute anywhere but not concentrate in a small region. Even the epipoles could be located either within the image or out of the image. Sometimes, the epipole of each computation from F varies a long distance between them; however, both Fs have very small average distance to epipolar lines. When we utilized this F to reconstruct the two-view stereo, sometimes the reconstructed 3D scene was abnormal and occasionally kind of projective. This means the average distance to epipolar lines is not the only, nor very robust, criteria to decide the fundamental matrix. Actually, a fundamental matrix with very small average distance to epipolar lines sometimes led to very poor 3D reconstruction results, such as two-view reconstruction based on F. This problem is critical in affine and projective reconstruction in which there is no meaningful metric information about the object space.

A new error selection criterion is proposed in this paper to solve the problem that Fs with very small average distance to epipolar lines sometimes result in very bad 3D metrical reconstruction or even result in totally wrong results. The advantage of our criterion is that this method can visually present the 3D reconstructed scene to indicate the quality of F. The criterion is still the projective error to project X_i, which is in 3D metric space rather than projective space. The description of metric space or metric reconstruction implies that the structure is defined in the Euclidean coordinates where the reconstructed scene is highly similar with the real scene. Metric is used to compare with descriptions for projective and affine. One speaks of projective reconstruction, affine reconstruction, similarity reconstruction, and so on, to indicate the type of transformation involved. In this paper, the term metric reconstruction is normally used in preference to similarity reconstruction, being identical in meaning.

Especially, this modification to the error selection criterion can improve 3D reconstruction results if we assume the principle center of camera in the image coordinates lies at the center of the image. The F under our criterion is suitable for the 3D metric construction, whereas the other Fs are always used for projective reconstruction. When the re-projection error is extracted from metric reconstruction based on F is feed back to judge the quality of the F and then recompute F, the quality of F is improved accordingly. This is the novelty of this paper.

2. Previous Research

Lots of works have been done on the fundamental matrix computation in the past 30 years [2,7–12]. It's a well and deeply developed area in computer vision; however, there still are some papers on the fundamental matrix published every year. Some traditional methods yield good results, which means very low projective errors to the epipolar lines. This average distance to the epipolar lines is a universal or almost unique criterion to judge the quality of the fundamental matrix.

The five-point method [4,5] is the most popular method that finds the possible solutions for relative camera pose between two calibrated views based on five given corresponding points. The algorithm consists of computing the coefficients of a tenth degree

polynomial in closed form and subsequently finding its roots. Kukelova et al. [13] proposed polynomial eigenvalue solutions to the five-point and six-point relative pose problems, which can be solved using the standard efficient numerical algorithms. It is somewhat more stable than solutions proposed by Nister [4,5] and Stewenius [14]. Another five-point method [15] was proposed to estimate the fundamental matrix between two non-calibrated cameras from five correspondences of rotation invariant features. Three of the points have to be co-planar and two of them in general position. The solver, combined with Graph-Cut RANSAC, was superior to the seven and eight-point algorithms both in terms of accuracy and the needed sample number on the evaluated 561 publicly available real image pairs.

Reference [11] proposed a novel technique for estimating the fundamental matrix based on Least Absolute Deviation (LAD), which is also known as the L_1 norm method. Banno et al. [16] proposed a parameterization for the nonlinear optimization which includes the rank 2 constraint. Double quaternion and a scalar are adopted as the parameter set. Menglong et al. A two-point fundamental matrix method [17] is even proposed. It makes use of the epipolar theory to estimate the fundamental matrix based on three corresponding epipolar lines instead of seven or eight corresponding points. Sengupta et al. [12] introduce a novel rank constraint, $F = A + AT$ and $rank(A) = 3$, on collections of fundamental matrices in multi-view settings. Then iterative re-weighted least squares and alternate direction method of multiplier (ADMM) are used to minimize a $L1$-cost function. Reference [18] developed the algorithm MAPSAC to obtain the robust MAP estimate of an arbitrary manifold to compute F. But by its open source code, the SfM result is not stable which means that the reconstructed scene varies from one to another totally different one.

As [19] points out, the use of symmetric epipolar distance should be avoided. The error criterion, i.e., the Kanatani distance (REK) [20], is the most effective error criterion found in their experiment for use as a distance function during the outlier removal phase of the F estimation. Some methods [21,22] searched for corresponding epipolar lines using points on the convex hull of the silhouette of a single moving object. These methods fail when the scene includes multiple moving objects. Reference [23] extends previous work to scenes having multiple moving objects by using the "Motion Barcodes", a temporal signature of lines. Reference [24] developed a new method for calculating the fundamental matrix combined with the feature lines. In [24], the camera orientation elements and relative orientation elements are used as the parameters of the fundamental matrix, and the equivalent relationships are deduced based on the horizontal and vertical feature lines.

Recently, after overwhelming studies of deep learning by neural networks, fundamental matrix estimation without correspondences is proposed by novel neural network architectures in an end-to-end manner [25]. New modules and layers are designed to preserve mathematical properties of the fundamental matrix as a homogeneous rank-2 matrix with seven degrees of freedom.

Most of the methods above focus on more constraints or new optimization methods. Unlike these methods, our algorithm uses a new criterion visually for metric reconstruction under the RANSAC framework and achieves a result similar to [4,5] i.e., our F is better to connect with 3D Euclidean coordinate. In other words, our F is better to reconstruct the real world of Euclidean coordinate.

3. Traditional Algorithm

3.1. F Computation in Multi-View Geometry

The most basic F computation is based on the linear equations

$$Af = 0 \tag{1}$$

where

$$f = \begin{bmatrix} F_{11} & F_{12} & F_{13} & F_{21} & F_{22} & F_{23} & F_{31} & F_{32} & F_{33} \end{bmatrix}^T \tag{2}$$

$$A = \begin{bmatrix} x_1^{'}x_1 & x_1^{'}y_1 & x_1^{'} & y_1^{'}x_1 & y_1^{'}y_1 & y_1^{'} & x_1 & y_1 & 1 \\ x_2^{'}x_2 & x_2^{'}y_2 & x_2^{'} & y_2^{'}x_2 & y_2^{'}y_2 & y_2^{'} & x_2 & y_2 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n^{'}x_n & x_n^{'}y_n & x_n^{'} & y_n^{'}x_n & y_n^{'}y_n & y_n^{'} & x_n & y_n & 1 \end{bmatrix} \quad (3)$$

where $u = [x\ y\ 1]^T$ and $u' = [x'\ y'\ 1]^T$ are the coordinates of the matching feature on two-view images. It is usually called the eight-point method because at least eight matching point pairs are needed to determine this equation. Of course more than eight matching point pairs can compose an over-determined system of equations to minimize $|Af|$ that could produce more robust result.

In RANSAC framework, increasing iteration times n_I is to find a better F with less error. But as the n_I increases, the errors of projective distance or metric distance become smaller and smaller at beginning of the iterations, then keep almost at same level after beginning iterations.

If constraint $\det(F) = 0$ is enforced to Equation (1), only seven matching pairs are necessary to solve this equation [26]. So this is usually called seven-point method. With more constraints being added in Equation (1), less matching pairs become indispensable. In the extreme occasion as mentioned in Section 2, only two matching pairs are needed [17].

Normalization [6,27] of matching feature coordinates is an essential step and recommended to reduce the final projection error. The suggested normalization is a translation and scaling of each image so that the centroid of the matching feature points is at $[0, 0]$, the origin point of the coordinates, and the root mean square (RMS) distance of the points to the origin point is $\sqrt{2}$. Usually, the last step is to optimize F for to minimize some sort of average error by nonlinear optimization method. The error is always the squared distance between a point and its epipolar line, which will be computed for both points of one match and averaged over all N matches.

3.2. Essential Matrix Computation and Decomposition

The relationship between F and essential matrix E is as listed by Equation (4).

$$E = K^{'T} F K \quad (4)$$

If K, $K^{'}$ are intrinsic parameter matrices of two cameras. E implies the relative position of a two-camera pair if R and t mean the disposition of two cameras. Thus E can be represented as Equation (5).

$$E = [t]_{\times} R \quad (5)$$

For a given essential matrix E which can be decomposed by, Equation (6)

$$E = U\ diag(1, 1, 0)\ VT \quad (6)$$

and the first camera matrix $P_E = [\ I \mid 0\ ]$, there are four possible choices for the second camera matrix $P_E^{'}$, namely

$$P_E^{'} = \begin{matrix} [UWV^T | u_3] & [UWV^T | -u_3] \\ [UW^TV^T | u_3] & [UW^TV^T | -u_3] \end{matrix} \quad (7)$$

where u_3 is the third column of the U, but only one of the four choices is correct. As we know, reconstructed X by the correct $P_E^{'}$ should locate in front of two cameras. Choosing one or more X_i to check if they are in front of two cameras is the way to select the correct $P_E^{'}$ out. This is the key point that our method can show the 3D metric reconstructed scene because our method can estimate E to obtain $P_E^{'}$ through the given F.

3.3. Optimization Based on Projective Geometric Distance

After an initial F is achieved by the methods as that in Section 3.1, an optimization always follows as the last step. One goal of optimization is to minimize the re-projection error—the (summed squared) distances in projective space. Usually the cost function is like Equation (8), which minimizes the distance of a point from its corresponding projected point, computed in each of the images.

$$\begin{aligned} \epsilon^{proj} &= \frac{1}{N}\sum_{i=1}^{N} \|u_i - \tilde{u}_i\| + \|u_i^{'} - \tilde{u}_i'\| \\ &= \frac{1}{N}\sum_{i=1}^{N} \|u_i - PX_i\| + \|u_i^{'} - P'X_i\| \end{aligned} \tag{8}$$

$$P = [\ I \mid 0\ ], P' = [\ [e']_{\times}Fe'\ ] \tag{9}$$

With projection matrices of cameras denoted by P, P', the feature coordinates on image should be $\tilde{u} = PX$, $\tilde{u}' = P'X$; however, there must be some error for $\tilde{u}$, $\tilde{u}'$. $|u_i - PX_i|$ and $|u_i^{'} - P'X_i|$ represent this re-projection error, which means the projective distances between the projections of X to the measured image points $[u, u']$. Obviously, P, P' are the projective matrices under projective coordinate but not the metric reconstruction under Euclidean coordinate. e' is the epipole on the second image with respect to F.

The initial X can be computed by P, P' and $[u, u']$ by methods such as linear or nonlinear triangulation. Then, the Levenberg–Marquardt method is applied to minimize cost of the Equation (8), which has $3N + 12$ variables: $3N$ for N 3D points X_i, and 12 for the camera matrix $P' = [M|t]$, with $F = [t]_{\times}M$. After optimization, a better result is produced most of the time. It is worth noting that sometimes it has only better numerical index on the average projective distance rather than reasonable 3D metric reconstruction.

4. Computing *F* in Metric Space Based on Geometric Distance

As we know, F includes the information in the projective coordination but not the Euclidean coordination. If metric reconstruction is necessary, it can only be estimated based on F via camera's intrinsic matrix K. Auto- (or self-) calibration, which can result in K automatically, is the computation of metric properties of the cameras and/or the scene from a set of uncalibrated images. Our method does not need any intrinsic parameter measured previously to reconstruct under the metric frame of Euclidean coordinate. In other words, we use a simplified auto-calibration method to estimate K.

In this paper, some constraints [28] are enforced on the intrinsic parameters of camera to estimate K. These constraints are: (1) Typically the aspect ratio is around 1. (2) The skew ratio can be assumed to be 0. (3) The principal point of the camera is projected to the center of image, $[w/2,\ h/2]$ where w, h is the width and height of image in unit pixel, respectively. After moving the origin of image coordinate to $[w/2,\ h/2]$, the intrinsic parameter matrix is the K defined in Equation (10), where f is the focal length of camera. This is a basic assumption for our method because this condition can be satisfied easily now and easier in the future with improvements in the manufacturing. Our method tries to fully make use of this assumption. Theoretically, it can achieve a perfect result if the real images satisfy this constraint of K ideally.

$$K = \begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{10}$$

4.1. 3D Metric Distance Criterion of F

As mentioned above, almost all the computation of F uses the distance of feature pixel to epipolar line as criterion to discard outliers of features matching. However, our algorithm uses the 3D metric distance criterion and is detailed as follows.

Compared with the traditional methods of F, the 3D metric distance criterion is more robust and has a similar performance on precision. It is the average distance between the projection of X on two image plains, $P_E X$ and $P'_E X$, and feature pixels $[u,\ u']$ in Equation (11), denoted by ϵ_i^{metric}. If ϵ_i^{metric} of one point X_i is less than threshold presented, it is an inlier. Then Equation (11) is designed to compute the average distance ϵ^{metric}, our 3D metric distance criterion of F, with all N matching inliers.

$$\epsilon^{metric} = \frac{1}{N}\sum_{i=1}^{N} \|u_i - P_E X_i\| + \|u_i^{'} - P'_E X_i\| \tag{11}$$

This 3D metric distance criterion can be applied into any RANSAC framework of F computation to judge whether one feature matching is an outlier or not. This criterion is more robust and has better performance especially for the images satisfied the assumption that the optical center lies at the center of image, $[w/2, h/2]$. The validation of this criterion by experiment is discussed in Section 5.

4.2. Optimization Based on Metric Geometric Distance

Our optimization based on metric geometric distance is in the same framework as the gold standard method [6] but in a metrical space for X and $P_E, P_E^{'}$, rather than the projective space described as Equation (8) in Section 3.3. In other words, we use a different method to compute projective matrix labeled as $P_E,\ P_E^{'}$ which represents the metric scene in Euclidean coordinates.

The input of our optimization is an initial F, image size $[w, h]$, and matching feature points $[u, u']$. The initial F of the one images pair is computed by the eight-point method as described in Section 3.1 in the RANSAC framework based on the traditional distance from $[u, u']$ to epipolar line. The output is the final result of F. Goal of the optimization is to minimize the Equation (11). The steps of this algorithm are listed as follows:

1. Estimate $E = {K'}^T F K$ with K in Equation (10).
2. Decompose E to $P_E,\ P_E^{'}$ by the method in Section 3.2.
3. Normalize the matching points $[u, u']$ to $[\hat{u},\ \hat{u}']$ with respect to K and $[w,\ h]$. if $u = [u_1,\ u_2,\ 1]^T$, then $\hat{u} = K^{-1} \times [u_1 - w/2, u_2 - h/2, 1]^T$.
4. With these correspondence $[\hat{u},\ \hat{u}']$ and $P_E,\ P_E^{'}$, estimate X by the triangulation method [6].
5. Minimize energy function Equation (11) over X, P_E and $P_E^{'}$. The cost is minimized using the Levenberg–Marquardt algorithm over $3N + 24$ variables: $3N$ for N 3D points X_i, and 24 for the camera matrix $P_E, P_E^{'}$.
6. Retrieve F from the K, K', P_E, P'_E. At first, compute the epipole e' of the second image by $e' = P_E^{'} \times null(P_E)$ where $null(P_E)$ is zero space of P_E. Then return the final $F = {K'}^{-1}[e']_\times P_E^{'} P_E^{-1} K^{-1}$.

As mentioned above, the K is estimated through the constraints, but not a precisely measured one. Of course if the precisely measured K is applied to our algorithm, better result will be achieved. For most of the scenarios, the precise K could not be known in advance. Therefore, this estimated and feasible K is adopted in our optimization. In fact it works very well and always achieves a satisfying result.

Compared with normalization in Section 3.1, our method uses a different normalization method for coordinates of features to compute final F. In Section 3.1, its origin point is a centroid that lies at the barycenters of u or u', and its average distance from u or u' to the origin point is $\sqrt{2}$ which is fixed. However, in our algorithm, origin point is at the center of the images, and the average distance from origin point is less than 1, which is not fixed. Our normalization is more efficient than the normalization in Section 3.1. Our normalization method is a linear operation that does not need to calculate the barycenter. Furthermore, our method can visualize metric 3D reconstruction of X as described in

Section 5.3 but the other normalization methods cannot produce reasonable visualization of reconstruction directly.

4.3. The Framework of Computing F Visually

The input data of our algorithm are a pair of images of a scene that need to be reconstructed, which are taken with a hand-held camera with a fixed focal length. The framework of the experiment is as follows:

Step 1: Extract SIFT features on both images. Compute the matchings of features $[u, u']$.
Step 2: Use a traditional method such as the eight-point method to compute F as the initial input of next step.
Step 3: Optimize F using algorithm in Section 4.2.
Step 4: If the metrical projection error is smaller than Step 2, output the result as the final F. If not, go to Step 2 to repeat.

In step (3), we can deduce P_E, $P_E^{'}$ from E. Then, the figure of reconstructed 3D points X can be shown in a Cartesian coordinate system even in each iteration.

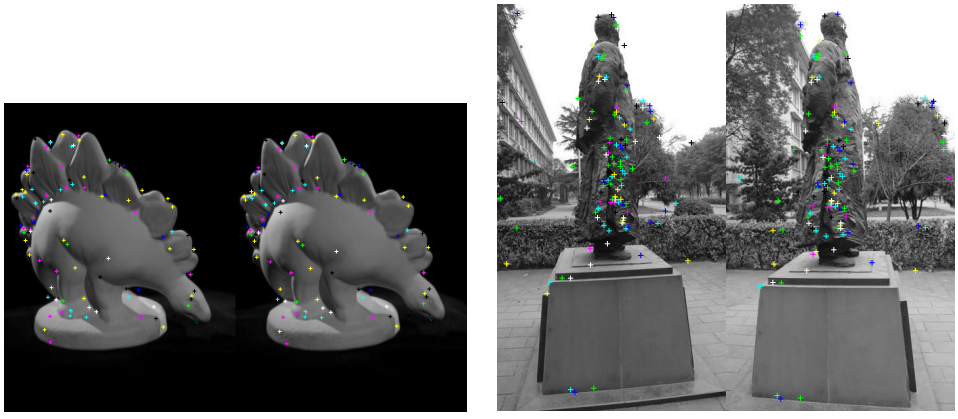
5. Experiment

5.1. Experimental Settings

Three different types of datasets, *Dino*, *ZhangLan*, and *Simu* were used in our experiment. *Dino* is a public-accessed dataset with an image size of 480×640. *ZhangLan* has an image size of 338×600 and was captured by us on our own campus. *Simu* is a human-made simulation of camera projection with an image size of 702×496.

5.2. Features Matching

Figure 2 shows the features extracted by SIFT on dataset *Dino* and *ZhangLan*. Many different stereo matching methods are fully and deeply researched [29]. Thus, our algorithm adopts SIFT [30] for the feature matching $[u, u']$ of two images. It is the first step to compute F. Here, Euclidean distance is the criterion to measure the features' difference. The ratio of the nearest distance over the second nearest distance was set to be less than threshold $T_{matching} = 0.55$. Otherwise it will not be labeled as a feature.



(a) Dino (b) ZhangLan

Figure 2. Feature matching for F matrix computation. (**a**) *Dino*. (**b**) *ZhangLan*. The matching of *Simu* are shown in Figure 3, and were produced from a human-made scene. Different feature colors distinguish the different feature matching.

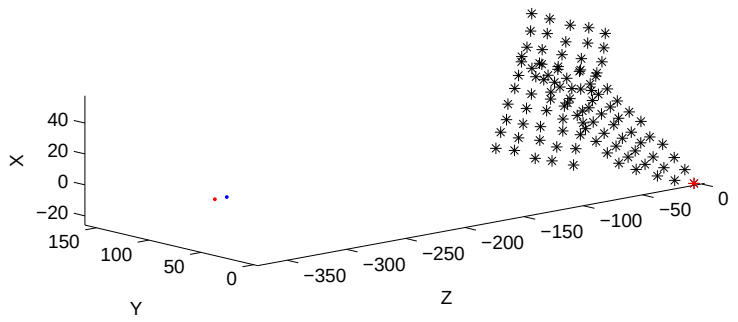
Actually, the other feature matching algorithms such as Harris, BRISK, SURF, ORB, etc., can still work. We choose SIFT only because it is typical and good enough to compare with other F computation methods, not because it is the best or most accurate matching method. SIFT has some mismatching features most of the time. Because of the RANSAC

framework used in this paper, the mismatching can be culled out by the sampling and iteration mechanism as shown in Figure 2. This is one of the differences to most of F computation methods that always carefully remove the wrong feature matching manually.

For the matching of simulated scene *Simu* displayed in Figure 3, all the points in the image are perfectly matched because the scene was manually made. So the simulation matching of scene *Simu* in Figure 3 is not listed in Figure 2. The 3D points X and camera positions are displayed in Figure 3a and the projection image of X at the first and second viewpoint are displayed in Figure 3b and Figure 3c, respectively.

For feature matching from SIFT in Figure 2, there must be some outliers of matching with long distances to epipolar lines or obvious re-projection error of X. If outliers are validated by the criteria mentioned in this paper, the incorrect matching should be removed from the feature matching sequence and not be counted into error calculation. This is an important step to improve precision and robustness of our method.

Figure 4 shows part of the typical epipoles and their corresponding epipolar lines chosen randomly and computed with our 3D metrical criterion for the final F under the RANSAC framework like the other traditional methods. As we can see, the epipoles are close to the epipolar lines. The distances are less than one pixel distance on average as shown in Table 1.



(a)

(b) (c)

Figure 3. The simulated scene. (**a**) 100 points on two planes are add some random disposition. The noise make the lines and two planes have obvious but slight vibration. The red '.' and blue '.' are the positions of cameras. In order to make the figure clear, we use '$*$' to indicate the points. (**b**,**c**) are the projection of 3D points on two images w.r.t the cameras.

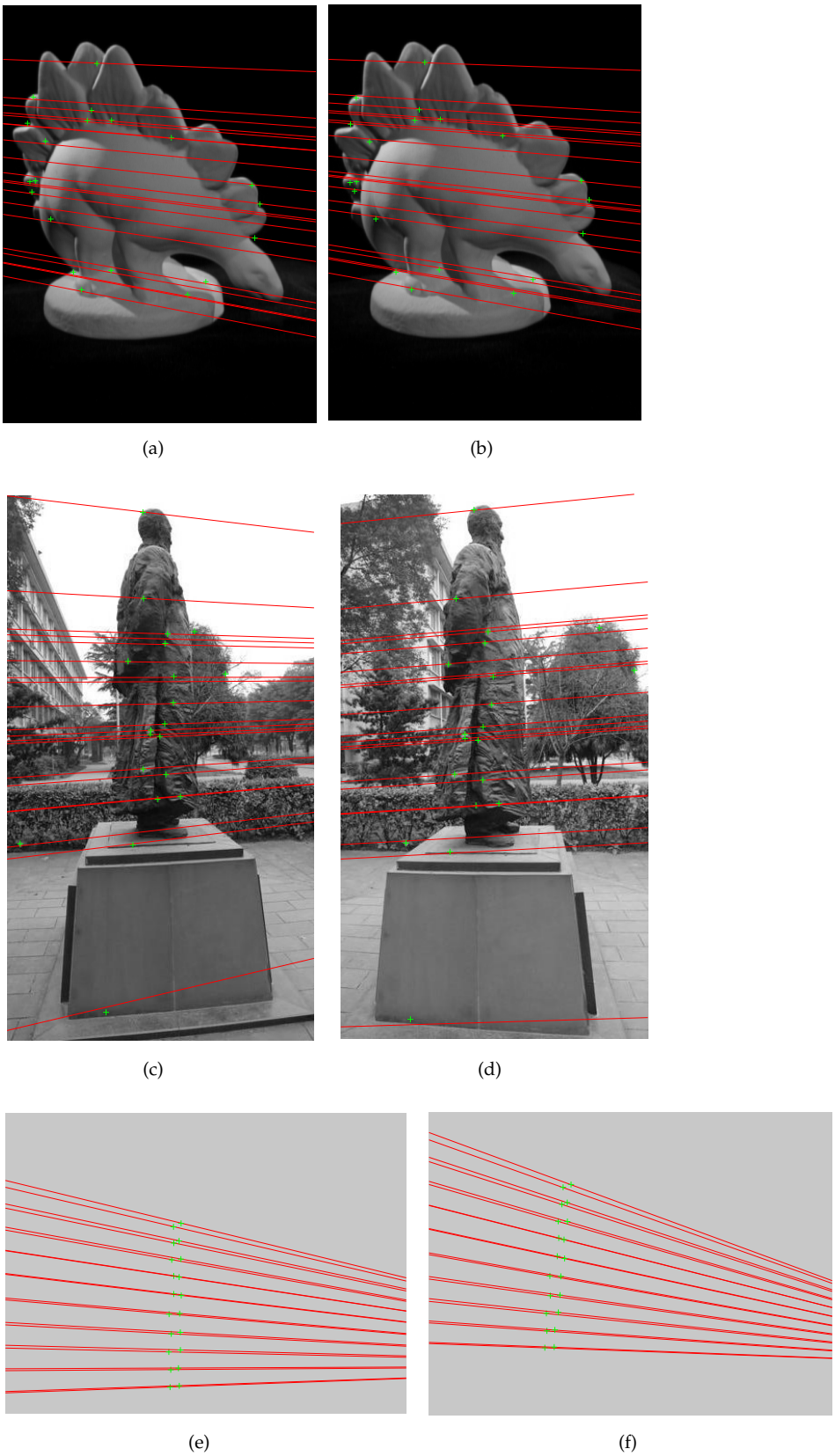
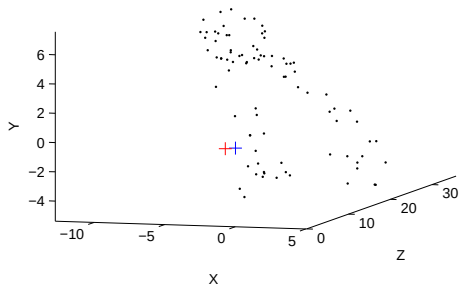


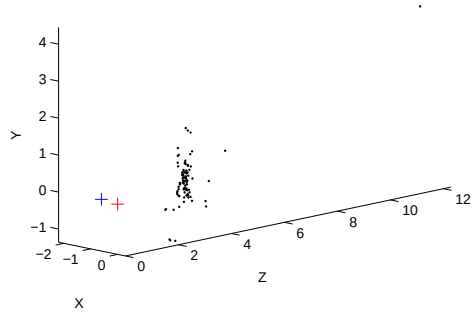
Figure 4. The epipolar lines and features of *Dino* (**a**,**b**), *ZhangLan* (**c**,**d**), and *Simu* (**e**,**f**). Here, only 20 epipolar lines were drawn in case all the lines concentrated in one image are too dense for reading.

5.3. Reconstructed 3D Points

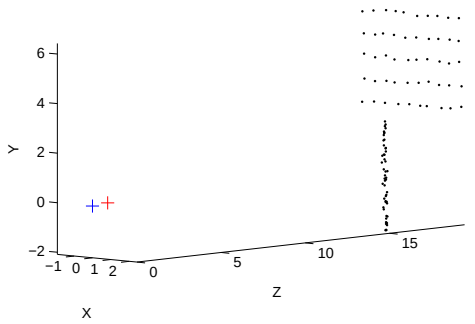
The visual results of Figure 5 show the 3D reconstruction with respect to our 3D metric projection error criterion based on E from F. However, eight-point [6], seven-point [6], and Gold [6] methods can only compute the projective F and cannot reconstruct the metrical 3D result directly.



(a)



(b)



(c)

Figure 5. 3D metric reconstruction result with ours metric gold criterion for (**a**) *Dino*, (**b**) *Zhanglan* and (**c**) *Simu*. The red and blue '+' are the positions of two cameras. In (**c**), we rotate the scene to a special view point to see that the 50 points lie at a same plane and thus nearly only a line can be seen at the lower part of the figure, which validate that our method is reasonable and robust. The precision analysis is described in the last paragraph of Section 5.4.

In our experiment, the eight-point method always has good reconstruction results for each criterion. Every point of X is recovered precisely. The projection errors are 1.50, 0.5, and 0.45 for the three different scenarios, respectively.

On the other hand, for the seven-point method, no matter how carefully the parameters are adjusted, the reconstruction result is not good enough because seven-point method is very sensitive to noise. The reconstructed 3D points X of *Simu* almost lie on one plane which in fact are two planes. The reason is that restriction of $\det(F) = 0$ is too arbitrary and only has mathematical meaning. For the reconstruction here, $\det(F) = 0$ has a kind of side effect when computing E. In our experiment, $\det(F) = 0$ was a useful and necessary but weak constraint. It is useful because it adds one equation to reduce one point pair

input and the epipole has to be calculated under this constraint. It is weak because that it affects the result significantly when this constraint is applied. Most of the time without this constraint, the better projective error or average epipolar lines distance are obtained, although this does not mean they will result in better 3D construction results. Actually, $\det(F) = 0$ limits the space of the solution of F. When this constraint is applied to the algorithms above, there is an advantage to compute the epipoles on corresponding images.

The appended material are the result from the five-point method that computes E as output directly. So the epipolar criterion can not be applied to five-point method. Compared with the other two methods over their distance to the epipolar line, Gold and Ours almost have results similar to the seven-point method. The five-point method's projection error of X, which is listed on the fourth row of Table 1, is obviously bigger than the other four. The five-point method is not robust or reliable enough in our experiments.

For the scene *Simu*, if only 0.5 pixel random projection error or less is added on $[u, u']$, it could have better reconstruction than the other scenarios because the optical center of camera lies at image center $[w/2,\ h/2]$ precisely.

5.4. Numerical Result Comparison

Table 1 compares metric criterion with traditional epipolar distance in 8-pt [6], 7-pt [6], 5-pt [13] and Gold [6] method. The second to sixth row of Table 1 are result based on the average distance from feature pixels to epipolar lines w.r.t the 5 methods to compute F. The columns are three different scenes of *Dino*, *ZhangLan* and *Simu*. The total initial matching features number are 100, 123 and 100. The numbers under the column 'Error' are the average distance from feature pixels to epipolar lines while the numbers under the column 'Inliers #' are numbers of inliers. Here unit of the distance is pixel in image. Our method and the method Gold have the best re-projection error than the other four.

Table 1. Error comparison based on the average distance from feature pixels to epipolar lines.

Methods	***Dino***		***ZhangLan***		***Simu***	
	Error	**Inliers #**	**Error**	**Inliers #**	**Error**	**Inliers #**
8-pt	1.50	86	**0.50**	100	0.45	**100**
7-pt	4.50	**90**	1.26	82	1.36	**100**
5-pt	5.24	88	4.35	101	5.65	**100**
Gold	1.38	89	1.69	**123**	**0.41**	**100**
Ours	**1.05**	88	1.22	98	1.05	**100**

Table 2 is the average re-projection error from feature pixels $u,\ u'$ to $P_E X,\ P'_E X$ and the 3D metric criterion $\|u - P_E X\|$ and $\|u' - P'_E X\|$ as described in Section 4.1, for the same datasets *Dino*, *ZhangLan*, and *Simu*. The last column is our method, which outperforms the other four methods. In order to compare with each other, the same F in Table 1 were used to decompose out E with K to reconstruct the scene, i.e., their inliers are recorded to compute the average re-projection error. Our method has comparatively good results compared with the other four.

Table 2. Error comparison based on the average re-projection error from feature pixels $u,\ u'$ to $P_E X,\ P'_E X$. It shows that our method outperform other methods.

Scenes	**8-pt**	**7-pt**	**5-pt**	**Gold**	**Ours**
Dino	7.02	13.77	5.24	7.42	**4.41**
ZhangLan	6.15	15.29	4.35	6.91	**4.07**
Simu	1.54	5.36	5.65	2.31	**1.05**

Projective validation is a common way to judge the quality of F. In this paper, we propose metric validation to perform the comparison. Metric validation is based on the assumption that the intrinsic matrix K is subject to the constraints in Section 4. Then, the

reconstruction is achieved by the deduced essential matrix E. The criteria to judge the inlier becomes the average re-projection error of X on images to $[\widehat{u},\ \widehat{u}']$. If images are captured with the intrinsic matrix K under these constraints, the achieved result F will have the better final reconstruction in the metric coordinate. Actually, this is the case for most of the cameras in making metric validation work. The biggest benefit of our metric validation is that it can produce reasonable and visual metric reconstruction results rather than a sort of projective reconstruction with small numerical projection errors that could be used to compare. If a more precise result is required, the intrinsic matrix K should be measured in advance. Here, we focus on an automatic approach so that the pre-measured K is not adopted. As we can see in Table 1, our method has better results in metric space than the five-point method.

In our experiments, thresholds of inliers have two types. (1) For metric validation, it is the metric re-projection errors of X on images to $\widehat{u}$ and $\widehat{u}'$. (2) For RANSAC inliers, it is the projective errors from $\widehat{u}$ or $\widehat{u}'$ to the corresponding epipolar lines. Usually, the smaller threshold means less inliers and better re-projection errors especially for real image pairs. For example, for the result of *Zhanglan* with the five-point method, the re-projection errors decreased from 5.5 to 4.7, on the other hand the number of inliers dropped from 94 to 86.

In Section 4.3, the bigger metrical projection error could be produced in Step 3, which optimize F using algorithm in Section 4.2, because the initialization of global optimization is crucial. Once the initialization of Step 3 is recalculated, the result with smaller result will be obtained in three iterations most of the time.

In order to measure the precision of reconstruction in metric space, we add some noise into matching features' position and then reconstruct the scene *Simu* to compare the distance between the points with the ground-truth distance simulated out. Because there are 100 points in *Simu*, $(99+1)*99/2 = 4950$ distances can be used to compare the error with the ground-truth. We calculate the mean error of 4950 distances and repeat this process 100 times with different noise at the same level to obtain a average error distance. Considering the scene area is a unbounded factor to affect the error of distances, we apply the ratio of the above average error distance to maximum distance of these 100 points. The ratio result is within $1 \pm 0.4\%$ most of the time when the noise onto features is 0.5 pixel with Gaussian distributions. Our method provides a reasonable and high precision for 3D reconstruction.

6. Conclusions

We propose a novel RANSAC criterion for the fundamental matrix computation. The results of experiments have shown that our method under metric space re-projection distance has good results for different test images compared with other methods such as eight-point, seven-point, five-point, and Gold standard methods. Our method can present the result visually to judge the quality of F. In future work, we would like to explore theoretical proof to provide stronger evidence to support our 3D metrical criterion of quality of fundamental matrix.

Author Contributions: Conceptualization, X.L. and X.Y.; methodology, X.L.; software, X.Y.; validation, X.L. and X.Y.; formal analysis, X.L.; investigation, X.L.; resources, X.L.; data curation, X.L.; writing—original draft preparation, X.L.; writing—review and editing, X.L. and X.Y.; visualization, X.L.; funding acquisition, X.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Sichuan Science and Technology Program grant number 2019YFG0117.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Feng, C.L.; Hung, Y.S. A Robust Method for Estimating the Fundamental Matrix. In Proceedings of the International Conference on Digital Image Computing, Sydney, Australia, 10–12 December 2003; pp. 633–642.
2. Christopher Longuet-Higgins, H. A Computer Alorithm for Reconstructing a Scene from Two Projections. *Nature* **1981**, *293*, 133–135. [CrossRef]
3. Hartley, R.I. In defense of the eight-point algorithm. *IEEE Trans. Pattern Anal. Mach. Intell.* **1997**, *19*, 580–593. [CrossRef]
4. Nister, D. An efficient solution to the five-point relative pose problem. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 756–777. [CrossRef] [PubMed]
5. Nister David, Hartley, R.I.; Henrik, S. Using Galois Theory to Prove Structure from Motion Algorithms are Optimal. In Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Minneapolis, MN, USA, 17–22 June 2007.
6. Hartley, R.; Zisserman, A. *Multiple View Geometry in Computer Vision*; Cambridge University Press: Cambridge, UK, 2003.
7. Torr, P.H.S.; Zisserman, A.; Maybank, S.J. Robust Detection of Degenerate Configurations while Estimating the Fundamental Matrix. *Comput. Vis. Image Underst.* **1998**, *71*, 312–333. [CrossRef]
8. Torr, P.; Murray, D. The Development and Comparison of Robust Methods for Estimating the Fundamental Matrix. *Int. J. Comput. Vis.* **1997**, *24*, 271–300. [CrossRef]
9. Li, W.; Li, B. Map estimation of epipolar geometry by em algorithm and local diffusion. In Proceedings of the International Conference on Image Processing, ICIP, Atlanta, GA, USA, 8–11 October 2006; Volume 5. [CrossRef]
10. Zhou, H.; R. Green, P.; Wallace, A. Estimation of epipolar geometry by linear mixed-effect modelling. *Neurocomputing* **2009**, *72*, 3881–3890. [CrossRef]
11. Yang, M.; Liu, Y.; You, Z. Estimating the fundamental matrix based on least absolute deviation. *Neurocomputing* **2011**, *74*, 3638–3645. [CrossRef]
12. Sengupta, S.; Amir, T.; Galun, M.; Goldstein, T.; Jacobs, D.W.; Singer, A.; Basri, R. A new rank constraint on multi-view fundamental matrices, and its application to camera location recovery. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4798–4806.
13. Kukelova, Z.; Bujnak, M.; Pajdla, T. Polynomial Eigenvalue Solutions to the 5-pt and 6-pt Relative Pose Problems. In Proceedings of the British Machine Vision Conference, Leeds, UK, 10–13 September 2008.
14. Stewenius, Henrik, C.E.; Nister, D. Recent developments on direct relative orientation. *ISPRS J. Photogramm. Remote Sens.* **2006**, *60*, 284–294. [CrossRef]
15. Barath, D. Five-point fundamental matrix estimation for uncalibrated cameras. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 235–243.
16. Banno, A.; Ikeuchi, K. Estimation of F-Matrix and image rectification by double quaternion. *Inf. Sci.* **2012**, *183*, 140–150. [CrossRef]
17. Benartzi, G.; Halperin, T.; Werman, M.; Peleg, S. Two Points Fundamental Matrix. *arXiv* **2016**, arXiv:1604.04848.
18. Torr, P. Bayesian Model Estimation and Selection for Epipolar Geometry and Generic Manifold Fitting. *Int. J. Comput. Vis.* **2002**, *50*, 35–61. [CrossRef]
19. Tolba, M.E.F.A.S.H.M.F. Fundamental matrix estimation: A study of error criteria. *Pattern Recognit. Lett.* **2011**, 383–391. [CrossRef]
20. Kanatani, K.; Sugaya, Y.; Niitsuma, H. Triangulation from Two Views Revisited: Hartley-Sturm vs. Optimal Correction. In Proceedings of the British Machine Vision Conference, Leeds, UK, September 2008; BMVA Press: Durham, UK, 2008; pp. 18.1–18.10. [CrossRef]
21. Sinha, S.N.; Pollefeys, M. Camera network calibration and synchronization from silhouettes in archived video. *Int. J. Comput. Vis.* **2010**, *87*, 266–283. [CrossRef]
22. Ben-Artzi, G.; Kasten, Y.; Peleg, S.; Werman, M. Camera calibration from dynamic silhouettes using motion barcodes. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 4095–4103.
23. Kasten, Y.; Ben-Artzi, G.; Peleg, S.; Werman, M. Fundamental matrices from moving objects using line motion barcodes. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 220–228.
24. Zhou, F.; Zhong, C.; Zheng, Q. Method for fundamental matrix estimation combined with feature lines. *Neurocomputing* **2015**, *160*, 300–307. [CrossRef]
25. Poursaeed, O.; Yang, G.; Prakash, A.; Fang, Q.; Jiang, H.; Hariharan, B.; Belongie, S. Deep Fundamental Matrix Estimation without Correspondences. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
26. Luong, Q.; Faugeras, O.D. The Fundamental Matrix: Theory, Algorithms, and Stability Analysis. *Int. J. Comput. Vis.* **1996**, *17*, 43–75. [CrossRef]
27. Agrawal, A.; Ramalingam, S. Single image calibration of multi-axial imaging systems. In Proceedings of the 2013 IEEE Conference onComputer Vision and Pattern Recognition (CVPR), Portland, OR, USA, 23–28 June 2013; pp. 1399–1406.
28. Pollefeys, M.; Koch, R.; Van Gool, L. Self-calibration and metric reconstruction inspite of varying and unknown intrinsic camera parameters. *Int. J. Comput. Vis.* **1999**, *32*, 7–25. [CrossRef]

29. Scharstein, D.; Szeliski, R.; Zabih, R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. Comput. Vis.* **2001**, *47*, 131–140.
30. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]