

Article

Automatic Pose Estimation of Uncalibrated Multi-View Images Based on a Planar Object with a Predefined Contour Model

Cailin Li ¹, Langming Zhou ^{2,*} and Wenhe Chen ¹

¹ School of Civil and Architectural Engineering, Shandong University of Technology, Zibo 255000, China; licailin@sdut.edu.cn (C.L.); chenwenhe123@163.com (W.C.)

² College of Electrical and Information Engineering, Hunan University, Changsha 410205, China

* Correspondence: langmingzhou@hnu.edu.cn; Tel.: +86-132-9865-0504

Academic Editor: Wolfgang Kainz

Received: 20 September 2016; Accepted: 9 December 2016; Published: 16 December 2016

Abstract: We have presented a framework to obtain camera pose (i.e., position and orientation in the 3D space) with real scale information of the uncalibrated multi-view images and the intrinsic camera parameters automatically. Our framework consists of two key steps. First, the initial value of the intrinsic camera and the pose parameters were extracted from homography estimation based on the contour model of some planar objects. Second, a refinement of the intrinsic camera and pose parameters was operated by the bundle adjustment procedure. Our framework can provide a complete flow of pose estimation of disorderly or orderly uncalibrated multi-view images, which can be used in vision tasks requiring scale information. Real multi-view images were utilized to demonstrate the robustness, flexibility and accuracy of the proposed framework. The proposed framework was also applied in 3D reconstruction.

Keywords: contour model; homography; pose estimation; bundle adjustment; machine vision; multi-view

1. Introduction

The pose estimation problem is one of the central problems in robotics, computer graphics, photogrammetry and computer vision. In robotics, pose estimation is commonly used in the hand-eye coordination systems [1]. In computer graphics, pose estimation plays an important role in tasks that combine computer-generated objects with photographic scenes, such as landmark tracking in determining head pose in augmented reality [2] or interactive manipulation of objects. In photogrammetry and computer vision, pose estimation is central to the procedure of 3D reconstruction [3] and object recognition [4].

The “Structure from Motion” (SfM) concept is the core method used in the automatic pose estimation of images [5]. In the late 1980s, effective SfM techniques were developed, which aimed to reconstruct the unknown 3D scene structure and the camera positions and orientations from a set of feature correspondences simultaneously. Longuet-Higgins introduced a still widely-used two-frame relative orientation technique in 1981 [6]; however, the development of a multi-frame structure in motion techniques, including factorization methods [7], and global optimization techniques [8–10] was delayed. In 2004, Nister matched small subsets of images to one another and then merged them for a complete 3D reconstruction in the form of sparse point clouds [11]. Vergauwen and Van Gool [12] developed a SfM tool for cultural heritage applications (hosted now by a web-based 3D reconstruction service). Recently, the SfM concept has made tremendous improvements, notwithstanding that the achievable 3D reconstructions were only useful for visualization, object-based navigation, annotation transfer or image browsing purposes. However, the procedure developed with the capability to

orient a large numbers of images. The two well-known packages are Bundler (or its graphical version Photosynth) [13] and Samantha [14]. Bundler is the implementation of the current state of the art in sequential SfM applications and extended toward a hierarchical SfM approach based on a set of key images [15]. Samantha appears to be faster because of the introduction of a local bundle adjustment procedure. The automatic methodologies to estimate image pose based on the SfM are also shown in [16–21].

In the SfM process, a direct result of the reconstruction of the cameras and scene structure can be obtained according to image point correspondences [21–23]. However, the method may have several limitations in measurable application: (1) SfM only estimates the relative pose of each camera without scale information; SfM can recover intrinsic parameters of non-degenerate configurations; however, this process can be unstable [21]; (2) the relative pose of the multi-view images can be obtained using SfM. When the images have geo-reference marks, the absolute pose can be obtained by a similarity transformation. However, if the accuracy of the relative pose is low, then the absolute pose with high accuracy cannot be acquired [22].

In this study, a multi-view image (captured from the same camera) pose estimation framework is presented, which can calculate the homography transformation by directly using the iterative optimization, and the intrinsic camera parameters (i.e., focus length, principle point radial distortion and tangential distortion) and the pose parameters (i.e., six parameters including the position (X_S, Y_S, Z_S) of the camera projection center in world coordinate system and the pose $(\varphi, \omega, \kappa)$ of every image) can be recovered simultaneously and robustly. This framework consists of the following two key steps: (1) the contour-based homography estimation is used for computing the homography relationship between models and images; the initial intrinsic camera and pose parameters can be recovered by the multiple homography matrices; moreover, the disorderly multi-view images can be rearranged by the camera pose derived from the initial pose parameters; (2) the bundle adjustment is used for the refinement of the initial intrinsic camera and pose parameters.

Therefore, the main contributions of this study include:

- (1) A robust contour model-based homography estimation (including recognition and tracking) of the planar object, which can transform disorderly multi-view images into orderly multi-view images in a general environment and provide good initial intrinsic camera and pose parameters.
- (2) A complete framework, which automatically provides both the intrinsic camera and pose parameters with a real scale for uncalibrated multi-view images. The framework can develop substantial measurable vision applications.

The remainder of this paper is organized as follows: Section 2 gives an overview of the approach. Section 3 presents the approach to obtain the initial parameters from the contour-based homography. Section 4 describes the parameter refinement using bundle adjustment. Section 5 presents the experimental results. Section 6 discusses the major advantages of the proposed framework. Section 7 presents the conclusion of the potentials and the limitations of the framework, as well as the objectives of future works.

2. Overview

In this section, we provide an overview of pose estimation, which includes camera localization and incremental bundle adjustment. Then, the approximate steps of our framework are shown.

Camera localization has received much interest in the last few years. Visual Simultaneous Localization and Mapping [24–26] and, in the computer vision, Structure from Motion with bundle adjustment optimization [27,28] are common ways of estimating the camera pose. These approaches reconstruct the environment and estimate the camera position simultaneously, but need to make a loop to correct the drift. Visual odometry is another way to retrieve the relative pose of the camera [29], but estimations drift irremediably. Royer had been shown that the use of 3D information on the environment ensures a better precision in pose estimation [30]. It makes the pose estimation of

the camera, embedded on a mobile platform, precise with no drifting, if the robot moves near these referenced [31] or even georeferenced [32] landmarks. For a few years, 3D models of cities or urban environments have been made available through various digitized town projects across the world. The French National Institute of Geography (IGN) digitalized streets and buildings of the 12th arrondissement of Paris in France. Model-based pose estimation is a problem tackled for several years working with various feature types: points [33,34], lines [35], both [36] or wireframe models [37–39]. These works dealt with geometrical features, but only a few other works take into account the photometric information explicitly in the pose estimation and tracking. Some of them mix geometric and photometric features [40,41]. Photometric features (image intensity) can directly be considered to estimate the homography and then the relative position between a current and a reference image [42]. A more recent approach proposes to estimate such a transformation using information theoretic approaches. Dame et al. proposed that mutual information shared by a planar textured model and images acquired by the camera is used to estimate an affine transformation or a homography [43,44].

Bundle adjustment optimization is a common way of estimating the image pose, and incremental bundle adjustment is a topic of recent research. Most existing work maintains real-time performance by optimizing a small subset of the most recent poses each time a new image is added. For example, Engels et al. [45] optimize a number of the most recent camera poses and landmarks using the standard bundle adjustment optimization, and Mouragnon et al. [46] perform a similar optimization, but including key frame selection and using a generic camera model. Zhang and Shan [47] perform an optimization over the most recent camera triplets, experimenting with both standard BA optimization and a reduced optimization where the landmarks are linearly eliminated. Other incremental bundle adjustment methods optimize more than just a fixed number of recent poses, instead adaptively identifying which camera poses to optimize [48–50]. However, in contrast with incremental smoothing [51], these methods result in approximate solutions to the overall least-squares problem and do not reuse all possible computations. In 2012, the Incremental Light Bundle Adjustment (ILBA) proposed by Vadim Indelman et al. [52] has been obtaining attention. They combine the two key ideas of structure-less SFM and incremental smoothing into a computationally-efficient bundle adjustment method, and additionally introduce the use of three-view constraints to remedy commonly encountered degenerate camera motions. In 2013, Vadim Indelman et al. [53] gave a probabilistic analysis of the incremental light bundle adjustment method and presented a computationally-efficient method for Vision-Aided Navigation (VAN) in autonomous robotic applications [54] with ILBA.

In this study, the framework calculated the homography transformation by using the bundle adjustment optimization.

As shown in Figure 1, the framework consists of the following four steps:

First, the structured line contour model was generated. The model of the object, i.e., A4 paper and a book, was measured in advance. The model can be asymmetric, as the Book 1 cover shown in Figure 2.

Second, the multi-view images of the scene (i.e., disorderly or orderly) were captured. In this study, a focus-fixed consumer-grade camera (or just a smartphone) was used.

Third, the intrinsic camera and pose parameters were calculated. The predefined structured line contour model was matched with the multiple line segments, which were extracted from the first frame of the set of pictures using the homography recognition procedure, to extract a correct homography. When at least three homographies were estimated, the intrinsic camera and pose parameters were computed linearly.

Finally, the parameters were refined by bundle adjustment. The sparse 3D points of the multi-view images were triangulated using the previously-mentioned parameters by the sparse correspondence procedure. Then, the parameters were set (i.e., the intrinsic camera and pose parameters), and the sparse 3D points were used in bundle adjustment to optimize the results.

The following sections describe the last two steps in detail.

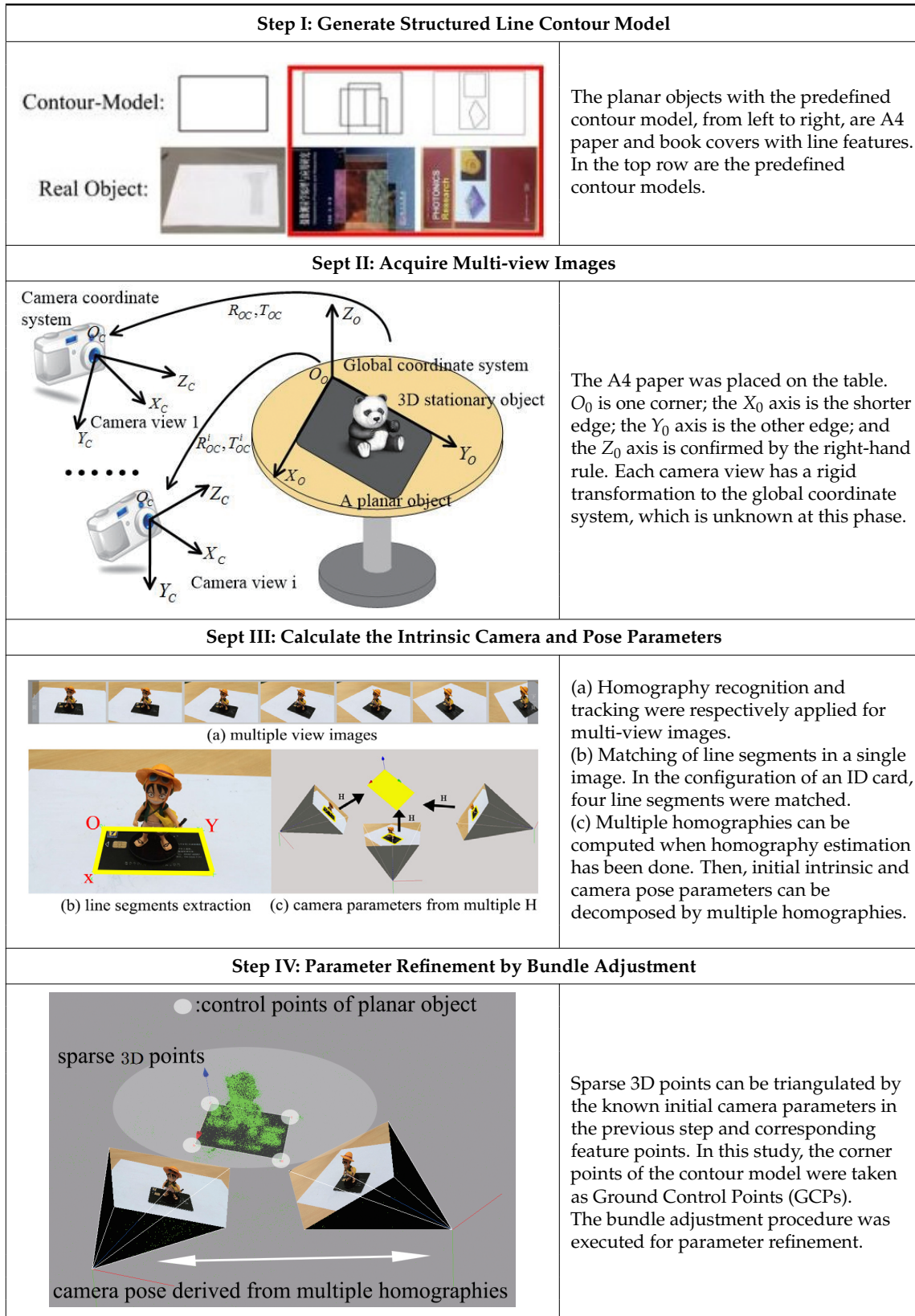


Figure 1. Overview of our framework.

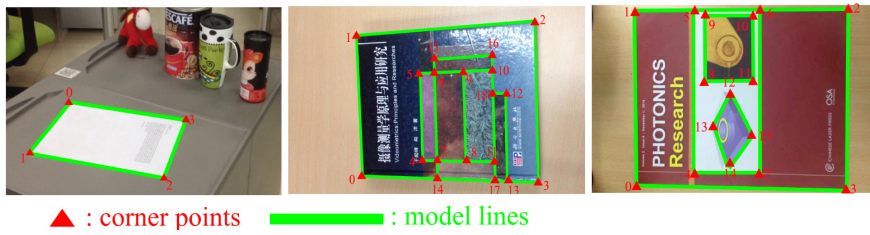


Figure 2. The lines and corners of the planar objects (from left to right: A4 paper, Book 1 cover and Book 2 cover). The model lines are shown in green, and the corner points are shown in red.

3. Initial Parameters Obtained from Contour-Based Homography

In this section, the problem of homography estimation was initially elucidated. Then, homography estimation that uses contour models for multi-view images was recognized. Subsequently, the homographies were optimized by minimizing the errors between the sample points and their corresponding image points yielded by utilizing the 1D search along the normal direction. At last, initial camera parameters, including the intrinsic camera and pose parameters, were retrieved from the multiple homographies based on the multiple view geometry.

3.1. Problem Statement

The problem considered in this section was the estimation of the homography transformation between the model and image features. The relationship between a 2D planar model P and its corresponding image point p is given as:

$$p = \Psi(P, H) = \begin{bmatrix} \frac{(HP)_x}{(HP)_z} & \frac{(HP)_y}{(HP)_z} \end{bmatrix}^T \tag{1}$$


$$P = (P_x \ P_y \ 1)^T, p = (p_x \ p_y \ 1)^T$$

Furthermore, the model line segment L is defined by its two endpoints (P_1, P_2) . Therefore, the homography between a model line segment L and its projection in the image plane I is given by the projection of its two endpoints:

$$I = \Psi(L, H) = (\Psi(P_1, H), \Psi(P_2, H)) \tag{2}$$

When noise appears in the measurement data, \tilde{P} denotes the noisy observation of 2D point P , and \tilde{I} denotes the noisy observation of 2D line segment I . The contour models utilized in this study include the geometrical and textural edges of the planar objects (see Step I of Figure 1), where they were modeled as lines and intersected corners. The lines and corners are illustrated in Figure 2, and the corner points were taken as the planar GCPs in the bundle adjustment procedure. The details (i.e., the coordinates of the corners and the index of the line segments) of the contour model of the cover of Book 1 are defined in Table 1.

Table 1. The contents of the planar models.

Real Object	Features	
Book 1 Cover 	Points index and coordinates (mm): 0: (0,0,0) 1: (174.5,0,0,0.0) 2: (174.5,245.1,0,0) ...	Line index: 0: 0 1 1: 1 2 2: 2 3 ...

3.2. The Disorderly Images and the First Frame: Recognition of the Contour-Based Homography

For the disorderly images or the first frame of orderly images, the homographies were recognized based on the contour model of the planar object, which includes two steps, namely hypothesizing and verifying. In the first step, a quadrangle-like structure formed by two image corners was selected and utilized to generate a number of approximate homography hypotheses. In the second step, the homography hypotheses were quickly ranked by matching the model line set with the image line set in the local region around the base line according to the distance function between the line segments derived from a series of noisy edge points in the probabilistic approach.

Although the image lines were fragmented because of the occlusion or the faulty line detection, the homography was determined with a closed-form solution from the four correspondences of the model lines and the image lines. A small set of homographies with approximately high certainty must be generated, including at least one homography close to the accurate transformation. Similar to [55], an approximate homography was obtained by exploiting the corner-like structures. However, the proposed method does not rely on the assumption that particular image lines, called base lines, are unfragmented [55]. In this study, corner-like structures were formed by pairs of line segments that terminate at a common point with a certain distance and intersection angle between each other. In contrast to the proposed method in [55], which only utilized one correspondence of the corner-like structure to generate the affine hypotheses, the present method generated the homography hypotheses from two correspondences of the corner-like structure (denoted as the quadrangle-like structure). In the two corner-like structures that share the same line segment (see Corner 1 and Corner 3 of Figure 3), the base lines are chosen if the lines match the corner-like structure of the other two line segments in the neighborhood, as shown in Figure 3.

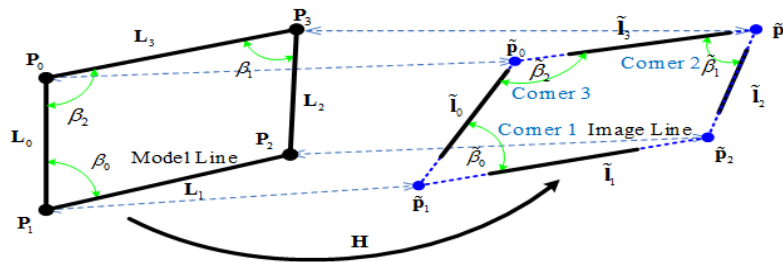


Figure 3. Homography transformation between the model and image lines.

The endpoints of the base line cannot be utilized directly because the base line may be fragmented. The endpoints of the base line can be intersected with the other two lines, because they are invariant in the perspective transformation [55]. As shown in Figure 3, the base image line is \tilde{I}_0 , and the points \tilde{P}_0, \tilde{P}_1 are the intersections with the other two lines \tilde{I}_3, \tilde{I}_1 . Given the single correspondence of the quadrangle-like structure in the model plane and image plane, the two point pairs $\{P_0, \tilde{P}_0\}$ and $\{P_1, \tilde{P}_1\}$ can be obtained. The homography transformation has eight DOF; however, the method has only four constraints (i.e., two for each point pair). Therefore, a ratio constraint that the ratio of the length of the other two lines \tilde{I}_1, \tilde{I}_3 with that of the base line remains unchanged was imposed in homography transformation. Therefore, the second endpoint \tilde{P}_2 of the line \tilde{I}_1 can be obtained according to the following equation:

$$\frac{\|P_2 - P_1\|}{\|P_2 - P_0\|} = \frac{\|\tilde{P}_2 - \tilde{P}_1\|}{\|\tilde{P}_1 - \tilde{P}_0\|} \quad (3)$$

where P_0 and P_1 are the endpoints of the base model line L_0 and P_2 is the second endpoint of the line L_1 . \tilde{P}_2 is given by the following equation:

$$\tilde{P}_2 = \tilde{P}_1 + \frac{|P_2 - P_1|}{|P_1 - P_0|} \begin{bmatrix} \cos\tilde{\beta}_O & -\sin\tilde{\beta}_O & 0 \\ \sin\tilde{\beta}_O & \cos\tilde{\beta}_O & 0 \\ 0 & 0 & 1 \end{bmatrix} (\tilde{P}_1 - \tilde{P}_0) \quad (4)$$

The second endpoint \tilde{P}_3 of the image line \tilde{I}_2 can be obtained in the same manner. Then, the four correspondences of the model and image points can be obtained according to the following equation:

$$\tilde{P}_i = \frac{1}{\lambda_i} H P_i, i = 0, 1, 2, 3 \quad (5)$$

If two corner-like structures form a quadrangle (see Corner 1 and Corner 2 in Figure 3), then four point pairs $\{P_0, \tilde{P}_0\}$, $\{P_1, \tilde{P}_1\}$, $\{P_2, \tilde{P}_2\}$, $\{P_3, \tilde{P}_3\}$ can be obtained without any assumption. For each correspondence $\{P_i, \tilde{P}_i\}$, two linear equations were in the eight unknowns in H . H can be solved linearly by four corresponding points, which the planar object of A4 paper can fulfill.

The second step is to rank all of the homography hypotheses from the first step. A geometric measure of the model line set (projected by an approximate homography) and the image line set were computed. Only a subset of the lines around the base line (if the two corners form a quadrangle, then the diagonal line from one corner to the other is chosen as the base line) was compared with the image line set to compute the similarity measure quickly. Let $M = \{L_1, L_j, \dots, L_M\}$ be the model lines and $N = \{\tilde{I}_1, \tilde{I}_2, \dots, \tilde{I}_N\}$ be the image lines. $N(I) = \{\tilde{I}_{\sigma_1(\tilde{I})}, \tilde{I}_{\sigma_2(\tilde{I})}, \dots, \tilde{I}_{\sigma_{N_{nbr}(\tilde{I})}}\}$ is the subset of image lines surrounding the base image line \tilde{I} , where $N_{nbr} > M$. Then, the geometric similarity between the model line set M and the image line set $N(I)$ around the base image line \tilde{I} is given by:

$$S(M, S, H) = \sum_{L_j \in M} \{S_{max}, \min_{I_i \in N(\tilde{I})} d(\tilde{I}_i, \Psi(L_j, H))\} \quad (6)$$

where $d(\bullet)$ denotes the distance between two lines in the image, which was defined in the literature [56]. The smaller the value of $S(M, S, H)$, the more similar the model line set is to the image lines under the current homography H . The parameter S_{max} was employed to ensure that the correct homographies are not penalized too severely when a model line is fully occluded in the image. In Equation (6), the homography hypotheses generated from the quadrangle-like structures can be ranked. Then, the ranked hypotheses were refined by the proposed homography optimization method (which is presented in Section 3.3), and the hypothesis with the smallest alignment error was chosen as the optimal homography.

3.3. Orderly Images: Contour-Based Homography Optimization (Tracking)

In Section 3.2, the homographies are of the first frame of the orderly images (i.e., video images). Then, for the rest of the images, optimization (i.e., tracking strategy) was proposed to obtain the optimal (the subsequent) homographies. As shown in Figure 4b, the 2D model edge of the ID card was projected to the image plane using the prior homography of the planar object. Instead of dealing with the line segment itself, the projected line segment (black solid line in Figure 4a) with a series of points (brown points in Figure 4a) was sampled. Then, the visibility test was performed for each of the sample points, because some of these sample points may be out of the field of view of the camera. In each of the visible sample points, the 1D search along the normal direction of the projected model line was employed to find the edge point with the strongest gradient or the closest location at its correspondence. Finally, the sum of the errors between the sample points and their corresponding image points was minimized to solve for the homography between frames.

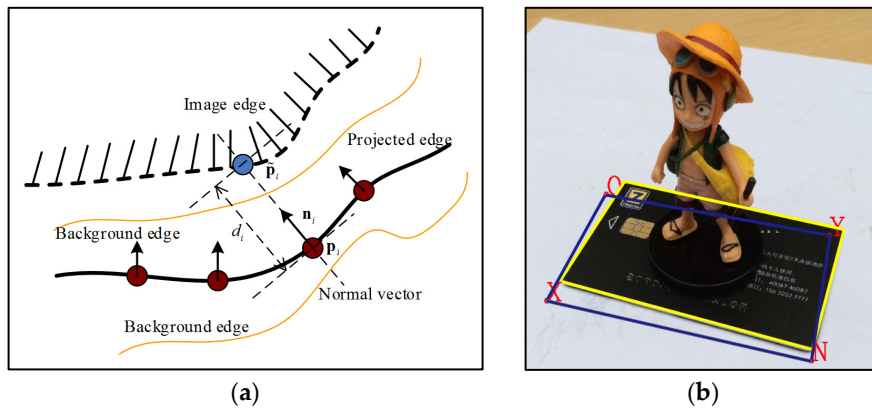


Figure 4. 1D search from the model line to the image line. (a) Sketch map of the image edge and the projected edge. The black solid line is the sampled line segments, and brown points are the sampled points. (b) Real image of the image edge and the projected edge. The yellow line is the image edge projected by a correct homography or the prior homography. The blue line is the projected edge by the current estimated homography.

As shown in Figure 4a, P_i and \tilde{p}_i are the set of projected sample points and their corresponding image points with the presence of the observation noise along the normal direction, respectively. Then, a function is defined to measure the normal distance between P_i and \tilde{p}_i :

$$d_i = n_i^T (P_i, \tilde{p}_i) \quad (7)$$

where n_i is the unit normal vector of the projected sample point P_i . Assuming a Gaussian distribution for d_i , the problem of homography optimization can be presented as:

$$\hat{H} = \arg \min_H \left(- \sum_{i=1}^S \log(p(\Psi(P_i, H) | \tilde{p}_i)) \right) = \arg \min_H \sum_{i=1}^S (d_i)^2 \quad (8)$$

where S is the number of model points. The optimization problem can be solved by the nonlinear technique. The iterative least square technique was used in this study.

3.4. Initial Intrinsic Camera and Pose Parameter Retrieval

With a series of homography matrices, i.e., H -matrices (more than three orientations), the intrinsic camera and pose parameters can be estimated using a closed-form solution [57].

The degenerate case, such as the regular calibration method using a chessboard, also exists [57]. The multiple view images were taken as a pure planar motion [58] when the camera was fixed and the object was on the turntable, as shown in Figure 5. In that situation, the method failed to retrieve camera parameters, although multiple homographies were estimated. The hand-held moving cameras that break pure planar motion can guarantee that such motion of the images will be avoided.

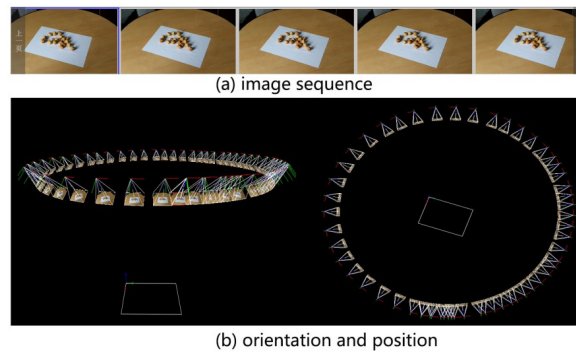


Figure 5. Degenerate situation for camera parameter calculation. (a) Image sequence of a flat scene with a fixed camera and the scene moving around a single-axis turntable; (b) the orientation of the camera motion of (a); actually, the pose of the images has a large offset from the real value, which means the pose estimation failed.

4. Parameter Refinement by Bundle Adjustment

The intrinsic camera and pose parameters of multiple view images are presented in Section 3. However, these parameters may not be reasonably accurate because of the image noises and the soft constraint of the contour-based model [59]. Thus, in this section, the parameters were refined and considered in the scale information for obtaining a measurable framework. Specifically, the parameters were refined using the bundle adjustment model [60] with planar control points of the contour-based model, which were also used in the homography estimation stage in Section 3 [61]. The planar control points play the role of GCPs to obtain the real scale of the pose parameters. With regard to the optimization problem on the 3D structure and viewing parameters, the bundle adjustment obtained the optimal reconstruction based on the assumption that the noise from the observed image features was considered white Gaussian noise [62].

Generally, the following conditions were necessary for the bundle adjustment procedure: (1) the input images must have sufficient overlapping degrees and (2) the initial parameter value must be known. These conditions can all be guaranteed in this framework stated in the previous sections. First, the images were arranged by homography recognition specified in Section 3.2. Second, the overlapping degree was easily guaranteed by the shooting mode (i.e., video images or small movement of the camera). Finally, the initial intrinsic camera and pose parameters were estimated in Section 3.4. Therefore, the second condition was also satisfied. A routine sparse reconstruction similar to SfM was initially implemented based on these conditions, such that the images can obtain a relative orientation structure in the subsequent section. Then, a bundle adjustment model was built to refine all of the parameters by using the sparse reconstruction and intrinsic camera and pose parameters, as shown in Section 3.4.

4.1. Sparse Reconstruction

In this section, a part of the commonly-used flow of SfM was taken to obtain the correspondences and tracks of images. The first step was to find the point features using the SIFT [63] or the SURF [64] key point detector. The second step was to match those key points to each pair of multiple view images using the FLANN [65] method. Then, the corresponding fundamental matrix was robustly estimated using the RANSAC iteration [66]. During each RANSAC iteration, the candidate fundamental matrix was computed using the eight-point algorithm, followed by the nonlinear refinement [58]. Subsequently, the matches that were outliers to the recovered fundamental matrix were removed. When the number of inliers was less than the present threshold (i.e., 20 is used in our framework), all of the matches from the consideration were removed [22]. Finally, all matches were geometrically organized into tracks after finding a set of consistent matches between each image pair using the union-find-based tracking method [67].

In contrast to the procedure of SfM, which aimed to recover a set of camera parameters and a 3D location for each track by the incremental strategy [22], the proposed method already had the initial camera parameters solved by the homographies estimation in Section 3.4. Therefore, the sparse 3D points can be directly obtained by the triangulating pairs, a large number of matches and base lines. Those feature points were defined as Type I features. Other type points (defined as Type II features) were the corner points of the contour model (as shown in Figure 2). Type II feature points were taken as GCPs in the subsequent bundle adjustment procedure.

4.2. Bundle Adjustment

A dataset was already obtained that conforms to three conditions of the bundle adjustment procedure mentioned previously. A bundle adjustment model with planar control points can be built and solved further, which will be described in this section to refine the parameters. As shown in Figure 6, a ray exists connecting camera C and space point P among the bundle, and this ray intersects with the corresponding image plane at p .

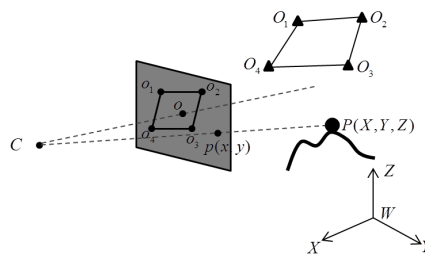


Figure 6. Bundle adjustment model with planar control points.

In Figure 6, C was the projective center of the camera, o was the corresponding principal point with coordinate of (c_x, c_y) and while the equivalent focal length of the camera was (f_x, f_y) . The lens distortion parameters were $(k_0, k_1, k_2, k_3, k_4)$. The pose parameters of the camera were a translation vector $T(T_X, T_Y, T_Z)$ and the rotation matrix R with the following two equivalent expressions: (1) rotation angle (A_X, A_Y, A_Z) and (2) nine elements $r_0 \sim r_8$ of the rotation matrix.

P was the space feature point corresponding to image point p . The distortion of p was signed as (δ_x, δ_y) . Furthermore, $O_1 \sim O_4$ were Type II feature points with corresponding image points $o_1 \sim o_4$. As shown in Figure 1, the relationship between the space feature point and corresponding image point was described by collinearity equations [68]:

$$\begin{cases} x = \delta_x + c_x + f_x \frac{r_0 X + r_1 Y + r_2 Z + T_X}{r_6 X + r_7 Y + r_8 Z + T_Z} \\ y = \delta_y + c_y + f_y \frac{r_3 X + r_4 Y + r_5 Z + T_Y}{r_6 X + r_7 Y + r_8 Z + T_Z} \end{cases} \quad (9)$$

Equation (9) was called the observation equations in the bundle adjustment model [60], and the observations refer to coordinate (x, y) of the image feature points. The unknown parameters of the bundle adjustment procedure contain space feature points (including Types I and II) and camera parameters. The constraint of the bundle adjustment procedure is the collinearity condition described in Equation (9), and this constraint is often used in computer vision.

The bundle adjustment procedure was illustrated using a video image, as shown in Figure 7. First, the initial camera parameters were obtained by the method described in Section 3. Then, feature matching and tracking were executed, as shown in Figure 7a,b, and the sparse 3D points were obtained by triangulation of the feature tracks. Subsequently, the initial camera parameters, sparse 3D points and GCPs (planar control points) were subjected to the bundle adjustment procedure. A converged result of the camera parameters and sparse 3D points were obtained, as shown in the Figure 7c. The sparse 3D points exhibited evident improvement in visual sensation. A further quantitative comparison will be shown in this experiment.

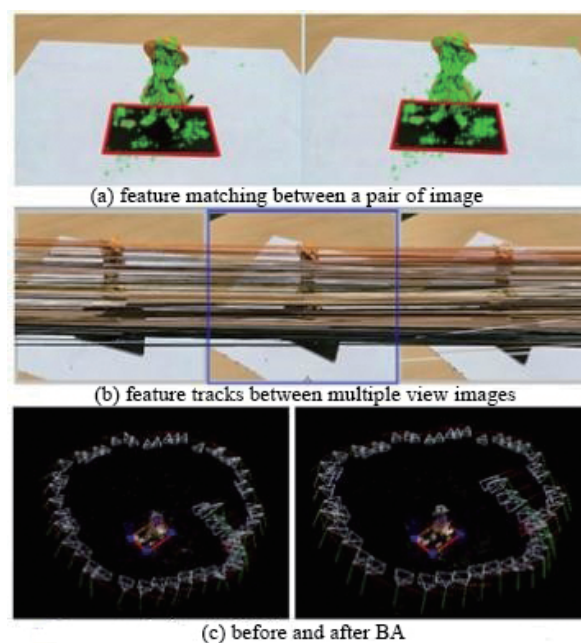


Figure 7. Bundle adjustment procedure with planar control points. (a) A feature matching between a pair of images. The green cross flags are matched key points of SIFT. The four line segments with red color are the edges of an ID card; (b) Feature tracks between multiple view images; (c) A bundle adjustment model is generated and executed. The left is the sparse reconstruction result using initial camera parameters. The right is the sparse result refined by bundle the adjustment procedure.

5. Experimental Results

5.1. Homography Recognition and Tracking

The homography from the planar objects to the image plane in an environment under a wide variety of viewpoints was recognized and tracked to validate the proposed method. All images were captured by a mobile phone. The line segments utilized for homography recognition were detected in the images by the LSD method [69].

A. Performance of homography recognition for disorderly images: In this experiment, the asymmetric objects shown in Step I of Figure 1 were placed in the environment, and the disorderly multiple view images were acquired by a moving camera at different positions and orientations. As shown in Figure 8a, the lines detected in the images were drawn in red, whereas the thick green lines correspond to the projections of the model lines mapped by the recognized homography. Figure 8a shows a large number of image line segments that were distributed on the background, such as the other books and the desk, whereas only a small number of line segments was detected in the object region. In such a case, the proposed method can provide an accurate homography recognition. Moreover, the angle, with respect to the orientation of the mode plane, was comparatively large, and the proposed method can still recognize the planar objects. Although deviation exists in homography recognition, it is sufficient for homography optimization and can lead to an accurate homography estimation. When all homographies were recognized, the images can be rearranged by the pose of the camera to obtain a spatial order set, that is a polygon shape without intersection, as shown in Figure 8b. The operation can build pairwise images with as much overlapping areas as possible to ensure successful feature matching and tracking.

B. Performance of homography tracking for orderly images: The objects were placed in the environment undergoing large rotation and translation, which were captured in the form of video images, to validate the performance of the proposed homography tracking method. Some sampled results from video images were exhibited. As showed in Figure 9, the proposed method can provide a

good match with the contour of the objects in the images. The pose parameters of the video images were computed by the proposed framework, and the dense matching procedure [21] was executed based on the camera parameters.

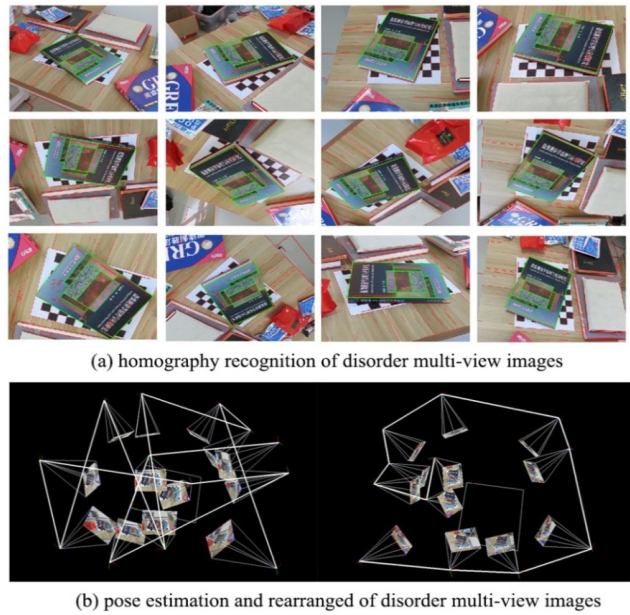


Figure 8. Homography recognition for disorderly multi-view images. (a) Homography recognition of the disorder multi-view image captured by a smartphone; the correct recognition homographies are shown in a green lines format, while other error lines are shown in red color; (b) the left is the results of pose estimation for the disorder multi-view images; polylines are connected according to the capture sequence. The right is the rearranged sequence by the pose of the cameras.

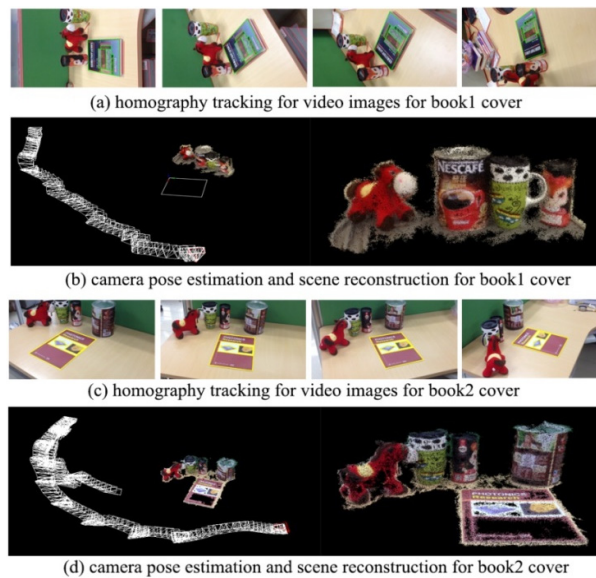


Figure 9. Homography tracking for Book 1 and Book 2 cover contour models in an environment. (a) The projections of the contour models of the Book 1 cover are drawn in blue by the recovered homography. The four images are the 100th, 200th, 300th and 400th frame in the video, respectively; (b) The left is the camera pose result of the video images, and the right is the 3D point clouds of the scene containing the Book 1 cover; (c) The projections of the contour models of the Book 2 cover are drawn in yellow by the recovered homography. The four images are the 100th, 200th, 300th and 400th frame in the video, respectively; (d) The left is the camera pose result of the video images, and the right is the 3D point clouds of the scene contain the Book 2 cover.

5.2. Accuracy Evaluation

In this section, the pose estimation technique and the corner-based method were initially compared using the four images sampled from the video captured with a smartphone, as shown in Figure 10. Then, a 3D reconstruction result was used to verify the parameter accuracy.



Figure 10. Four images of the model plane for camera calibration.

The image captured by the smartphone was 960×540 . In the experiment, the chessboard plane contained 10×13 interior corners and 23 lines. The results are shown in Table 2. The results of the proposed method exhibited a slight difference from the results of the corner-based method. When only four edges of the plane pattern were utilized, the proposed method provided consistent results with the corner-based method, and the offset of the camera parameters, which is five pixels, was small with respect to the corner-based method. The last column of Table 2 shows the reprojection RMS of the three methods. When all of the 23 lines were utilized, the proposed method provided almost the same reprojection error with the corner-based method. The four-line-based method returns a slightly larger reprojection error, because only the minimum of the model lines was utilized.

Table 2. Calibration results with the real data of four images.

Methods/Pixels	f_x	f_y	c_x	c_y	Reprojection RMS
Corner-based method	1147.23	1146.68	475.39	258.04	0.41
Line-based method	1150.76	1151.49	474.61	262.60	0.42
Four line-based method	1141.14	1139.99	480.25	253.07	0.67

The number of lines from four to 23 was varied to investigate the stability of the proposed method further. The results are shown in Figure 11. c_x, c_y recovered by the proposed method were approximately the same as the values estimated by the corner-based method, with only a small deviation. The reprojection errors of the proposed method decreased significantly from four to 17. When the number was greater than 17, the reprojection error was close to that of the corner-based method.

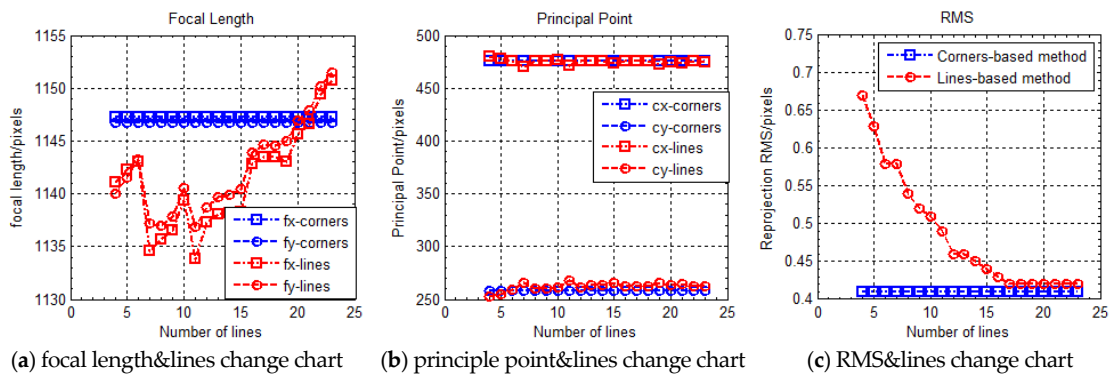


Figure 11. Results versus the number of model lines.

The robust homography recognition and tracking can provide good initial parameters for multi-view images; therefore, the bundle adjustment model can converge effectively. Comparison

equipment was set to validate the accuracy of the proposed method. A chessboard was placed in the scene. Multi-view images (total count = 24) were captured around the scene (as shown in Figure 12).

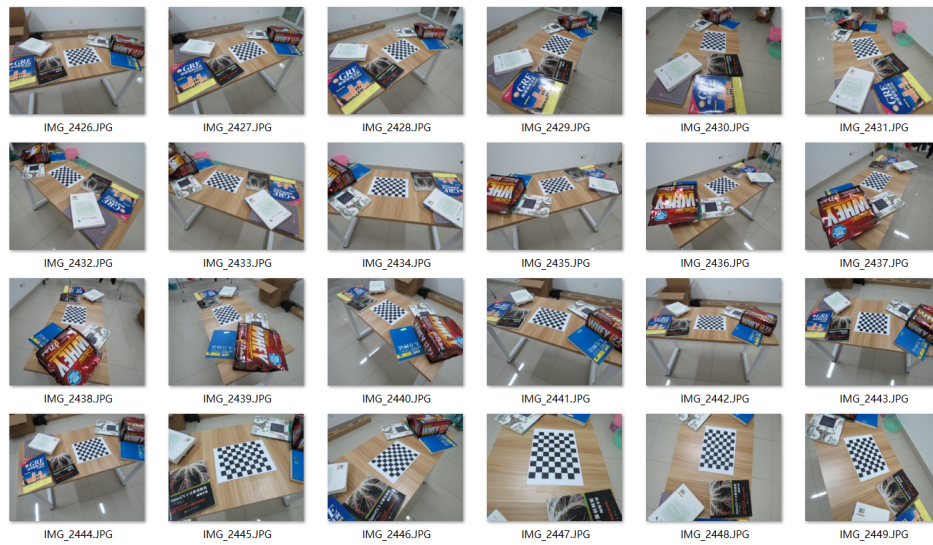


Figure 12. Multi-view images of the scene. The images are captured by a smartphone, and a $7 \times 9 \times 25$ mm chessboard was put in the scene.

Then, the camera parameters (i.e., intrinsic camera and pose parameters) were calculated by the classical chessboard calibration method [57] and the proposed method. The coupling of the intrinsic camera and pose parameters has less significance in comparing the values of the parameters. Therefore, a comparison in 3D space was set because the accuracy of the pose is the most important element in dense matching. As shown in Figure 13, three methods were designed to calculate the camera parameters of the same multi-view image set: (1) homo_exp: the camera parameters were calculated from the homography estimation, which is called the initial parameters of this framework; (2) homo&ba_exp: the initial camera parameters of homo_exp were refined by the bundle adjustment procedure; and (3) chessbd_exp: the camera parameters were calculated by the chessboard method. The three sets of parameters were used in the same dense matching procedure [21]. The point cloud results were compared to evaluate the accuracy of the camera pose.

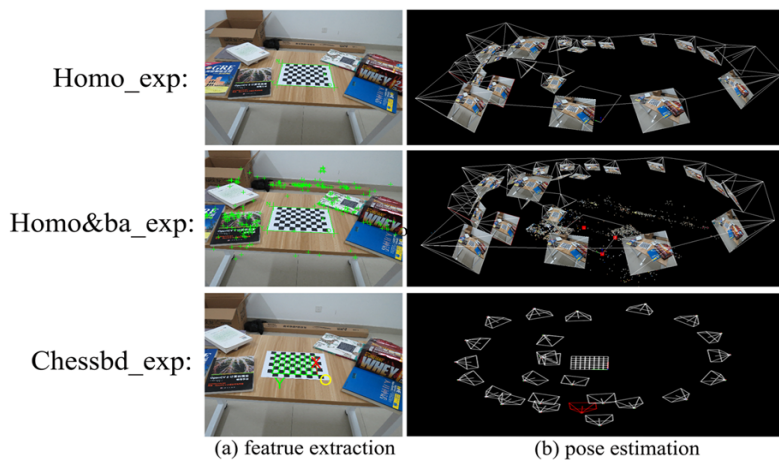


Figure 13. Pose estimation of three methods. (a) The left column is the features used in the estimation procedure, which, from top to bottom, are the four outer edge lines of A4 paper, the four outer edge lines of A4 paper with bundle adjustment and the corners of the chessboard, respectively; (b) the right column is the pose results of the three methods.

As shown in Figure 13, the outer four edge lines of the chessboard are taken as a contour model for our framework. The initial parameters are decomposed from multiple homographies by the homography tracking strategy described in Section 3.3.

As shown in Figure 14, the depth image of chessbd_exp was taken as the ground truth. Notably, the depth image of the homo_exp method was the worst because of the discontinuous depth values, which were in conflict with the real scene, and the depth image of homo&ba_exp was close to the ground truth.

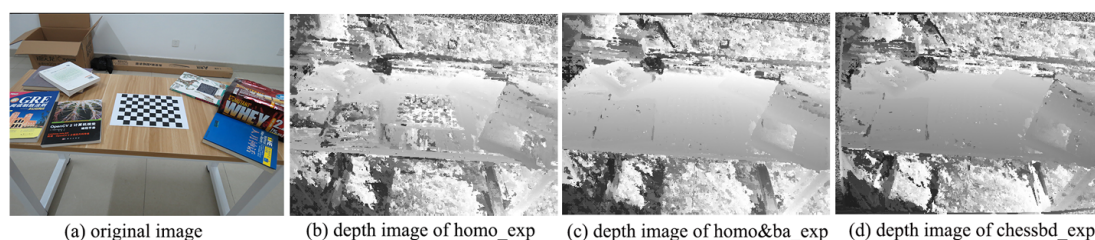


Figure 14. Comparison of depth images. We pick one depth image corresponding to the original image of (a), and (b–d) are the depth images of the methods homo_exp, homo&ba_exp and chessbd_exp.

As shown in Figure 15, the vertices of the three point clouds are 570, 790, 799, 157 and 820, 155, from top to bottom, of (a) and (b). When the camera parameters are accurate, the point clouds are denser.

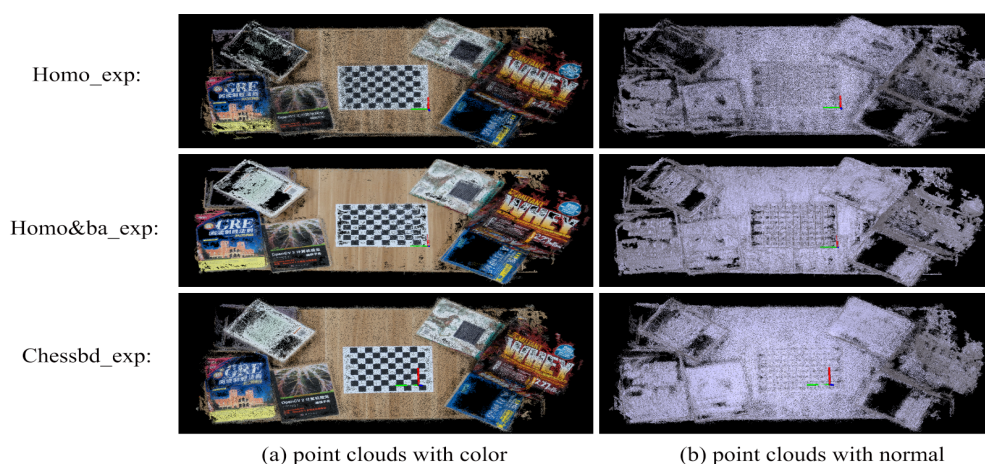


Figure 15. Point clouds of dense reconstruction using the camera parameters of the three methods. (a,b) The total point clouds with color and normal.

The particularly interesting parts were marked by the green ellipse, as shown in Figure 16. In the first column, the contour of the text “2.72 kg” marked by the green ellipse was sharper and clearer in the homo&ba_exp and chessbd_exp methods than in the homo_exp method. The same phenomenon occurred in the second and third columns of the figure. In the last column, the side view of the table in the scene was shown. A clear misalignment was observed in the homo_exp and chessbd_exp methods, and a good alignment was observed in the homo&ba_exp method. Such misalignment is shown on the left corner of the book cover in the second column. The fault of the point clouds can be attributed to the camera parameters provided by the three methods because the dense reconstruction and input images were all the same in the three methods.

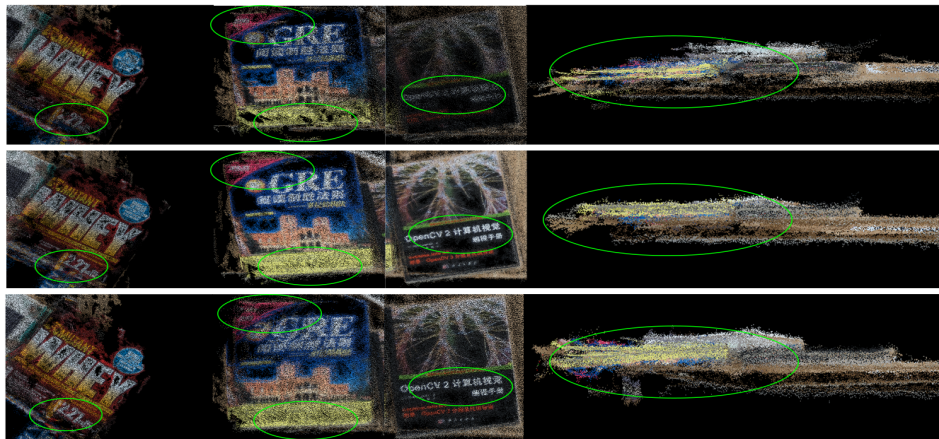
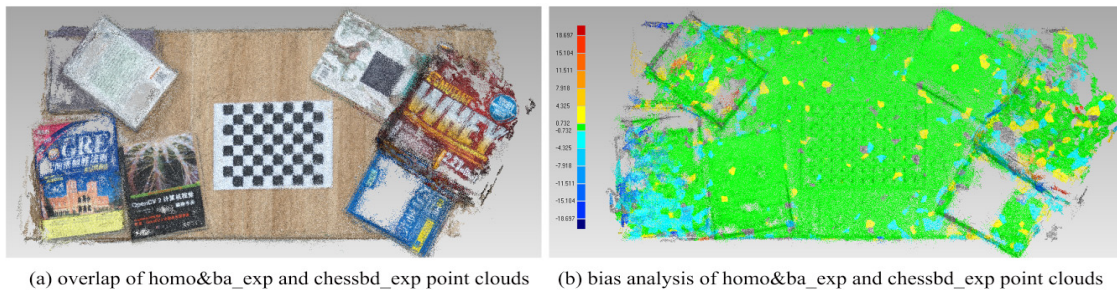


Figure 16. The detailed parts of the three sets in zoomed-in view. Four parts are selected to show the diversity of the three methods. The detail regions are marked by the green ellipse.

The point clouds of the *homo&ba_exp* and *chessbd_exp* methods were aligned to the same coordinate system. Only a small translation between the two methods was observed because of the different original points. The bias analysis between the two point clouds was computed in a color error map, as shown in Figure 17, which was measured as the distance in mm. Approximately 90% of points were in the 0.72 mm to 0.72 mm areas (green color), and the average distance and standard deviation of the two aligned point clouds were 0.90 mm and 1.3 mm, respectively. The completeness value that measures how well the parameters provided by the *homo&ba_exp* method covers the ground truth (the traditional chessboard method). Compared with the SfM method, the camera parameters and the follow-up 3D reconstruction all had scale information because the real size of the contour model was used. Therefore, the proposed framework can be used in vision tasks, which need real-scale information.



(a) overlap of *homo&ba_exp* and *chessbd_exp* point clouds (b) bias analysis of *homo&ba_exp* and *chessbd_exp* point clouds

Figure 17. Bias analysis when the two point clouds are aligned.

5.3. 3D Reconstruction Application

Camera pose is vital to 3D reconstruction, particularly the measurable framework, which can provide the real size of the 3D model. As shown in Figure 18, the aim is to recover the 3D shape of the statue outdoors. Similar to the A4 paper, the platform edge of the statue (size = 4 m × 4 m) was used as the planar object in this framework, and a video around the statue was captured using a smartphone (iPhone 5S; video size = 1920 × 1080). The intrinsic camera and pose parameters of the multi-view images (approximately 100 frames) decomposed from the video were obtained from this framework. A mixed silhouette-based and photo-consistent 3D reconstruction approach [70] was executed with a set of silhouette images segmented by the level set method [71]. The point cloud was constructed into mesh format using the Poisson method [72], and then, the mesh is textured with the color images [73]. Considerable indoor 3D reconstruction results are shown in Figures 19

and 20 (the images also captured from 1080p video). Table 3 is designed to show the experiments in different scenarios.

It can be concluded from the experiments that the pose estimation method was applied in 3D reconstruction successfully. It was not only suitable for small objects indoors (as shown in Figures 19 and 20), but also for large objects outdoors (as shown in Figure 18). Additionally, the contour models with a standard size and planar structure (as shown in Table 3) made the experiment easy to operate. In addition, the photography method was different in the three experiments (from the third column of Table 3), and it did not affect the results. In the end, the computational time of sparse reconstruction, which was measured on a 3.00-GHz Intel Xeon E3-1220 v5 architecture, is shown in the fifth column in Table 3.

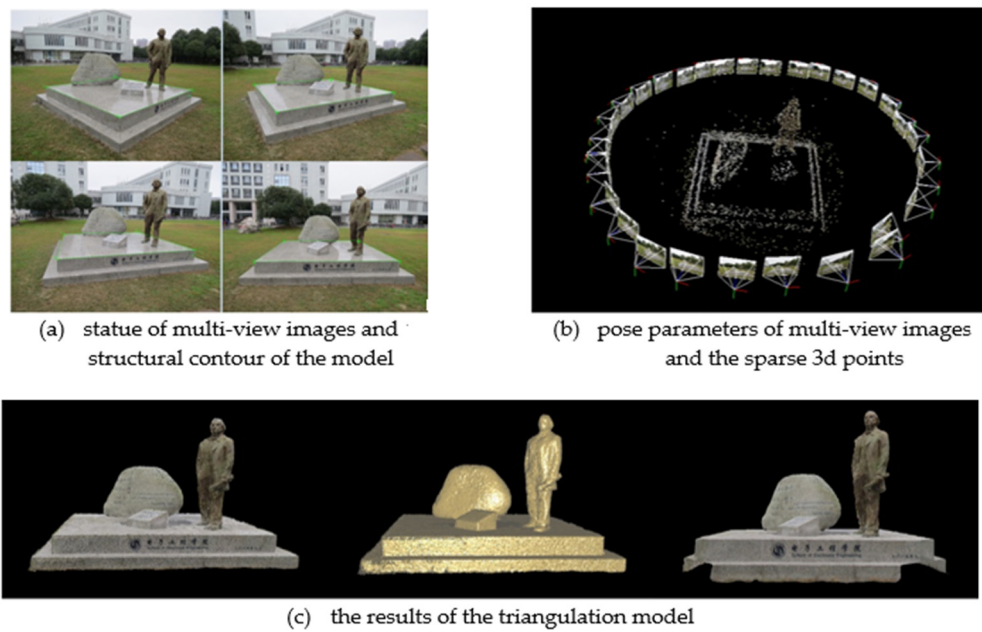


Figure 18. 3D shape of the statue from uncalibrated multi-view images. (a) Several frames sampled from a 1080p video captured by a smartphone; (b) recoverable results of the pose parameters of multi-view images and the sparse 3D points; (c) the left is the results of the color point clouds, and the normal point clouds are shown in the middle; the right is the model of the triangulation result.

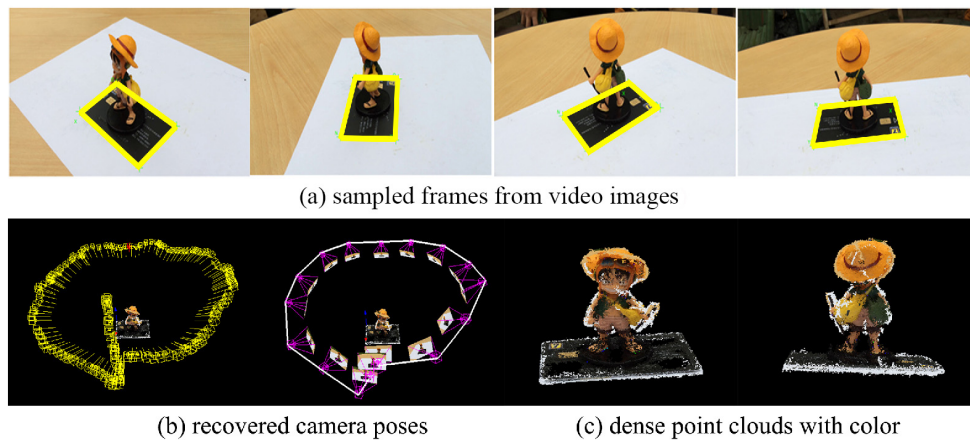


Figure 19. 3D reconstruction results of One Piece.

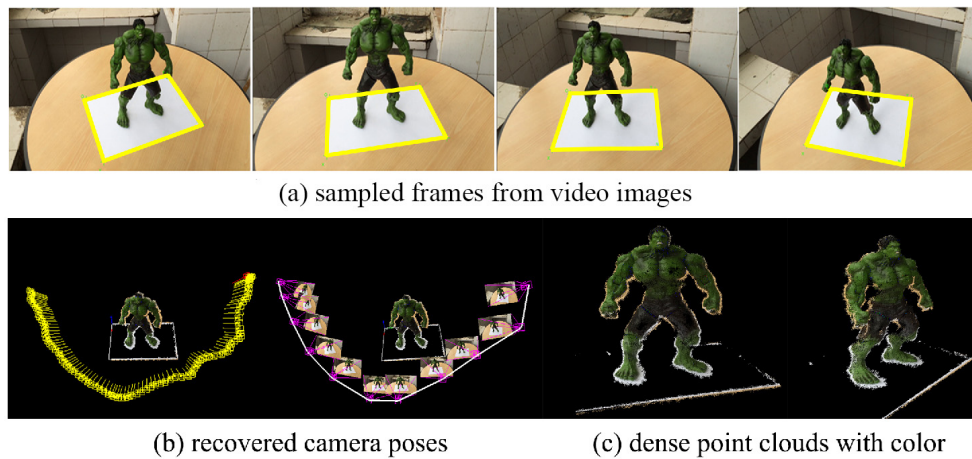


Figure 20. 3D reconstruction results of The Hulk.

Table 3. Experiments in different scenarios.

Object	Planar Model	Photography Way	Number of Images	Runtime
statue	platform edge	a circle around the object	100	30 min
One Piece	ID card	more than a circle around the object	140	36 min
The Hulk	A4 paper	a half circle around the object	100	27 min

6. Discussion

In this study, a complete framework that aims to estimate the pose of multi-view images has been presented. Under the proposed framework, the intrinsic camera and pose parameters of disorderly or orderly multi-view images can be recovered. Compared with other methods, the proposed method has the following advantages:

- (1) The homography estimation method can be considered the revised version of the model-based 3D tracking [74,75], which was developed to estimate the six DOF pose of the camera. Rather than initially estimating the affine transformation parameters and then the remaining non-affine parameters [76], this method used an iterative optimization process to refine the recognized homography directly.
- (2) Compared with a similar work [77], in which mapping was modeled as affine transformation and line correspondences were utilized in the refining process, the proposed method recognized the eight DOF of homography and optimized the initial transformation iteratively by dealing with the object contour as a series of sample points in a manner that the curved edge can be integrated. In this approach, the initial homography was recognized in the framework of hypothesizing and verifying the unmatched set of lines. Moreover, the optimized homography was obtained by minimizing the errors between the sample points and their corresponding image points obtained by utilizing the 1D search along the normal direction.
- (3) The robust approximate homography estimation is a vital stage, which can provide good initial parameters for the bundle adjustment procedure and can transform disorderly multi-view images to orderly multi-view images. The proposed method focused on obtaining the intrinsic camera and pose parameters with scale information and improving the precision of those parameters by the bundle adjustment procedure.

In addition, for general users that will perform vision tasks, the prepared planar calibration may not always be available. However, common items in daily life are of a standard size and planar structure. An easier and practical pose estimation method was proposed by exploiting the edge information.

7. Conclusions

In this study, a framework was designed that can provide the intrinsic camera and pose parameters of uncalibrated multi-view images via the bundle adjustment procedure and homography estimation of the contour model of the planar objects. The framework aims to help the general user to perform vision tasks without a prepared planar calibration pattern (i.e., chessboard). However, common items in daily life are of a standard size and planar structure (see Step I of Figure 1). By exploiting edge information, an easy, practical and automatic pose estimation method for uncalibrated multi-view images was proposed. In practice, the method can be used to measure the actual size of the object. For example, the 3D shape of the human feet can be used to measure several key data in designing shoes. In addition, the differences in the size between two objects can be measured according to the same prepared planar model. The method can also play an important role in the measurement of industrial parts and digital preservation of antiquities.

The approximate homography was obtained in the framework of hypothesizing and verifying. The quadrangle-like structure was used to ensure automatic and stable recognition of the homography in common environments. The robust and approximate homography can provide good initial intrinsic camera and orientation parameters. Moreover, disorderly images can be rearranged into an orderly set, which is helpful in multi-view image processing. Subsequently, a refinement procedure of the intrinsic camera and pose parameters was operated by the bundle adjustment procedure. The experimental results revealed that the proposed method can conveniently generate a multi-view framework with scale information as compared with the traditional absolute orientation method and the SfM procedure. The proposed method has substantial features, e.g., circle and quadric. In practice, all of these features can be added to the homography estimation stage. However, only the line features of planar objects were used in this framework. The configuration must fulfill the demands that is the number of lines must be at least four and must have intersection points. Fulfilling the demands of configuration was the main challenge that we met. The objective of future study is to use all of the features, which can result in a more convenient and accurate pose estimation.

Acknowledgments: This work is supported by the National Natural Science Foundation of China (Grant No. 41601496), the China Postdoctoral Science Foundation (Grant No. 2015M572706), the Project of Shandong Province Higher Educational Science and Technology Program of China (Grant No. J15LN32) and the Young Teacher Development Support Foundation of Shandong University of Technology in China (Grant No. 114016).

Author Contributions: Cailin Li and Langming Zhou proposed the core idea of the pose estimation method of multi-view images automatically based on a planar object with a predefined contour model and designed the technical framework of the paper. Cailin Li and Langming Zhou performed the experiments, the data analysis, results interpretation and manuscript writing. Wenhe Chen assisted with the experimental result analysis and refining manuscript writing.

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

Abbreviations

The following abbreviations are used in this manuscript:

DOF	Degree of Freedom
SIFT	Scale-Invariant Feature Transform
FLANN	Fast Library for Approximate Nearest Neighbors
RANSAC	Random Sample Consensus
ILBA	Incremental Light Bundle Adjustment
VAN	Vision-Aided Navigation
IGN	The French National Institute of Geography

References

1. Mammarella, M.; Campa, G.; Napolitano, M.R.; Fravolini, M.L. Comparison of point matching algorithms for the UAV aerial refueling problem. *Mach. Vis. Appl.* **2010**, *21*, 241–251. [[CrossRef](#)]
2. Vese, L.A.; Osher, S.J. Modeling textures with total variation minimization and oscillating patterns in image Processing. *J. Sci. Comput.* **2003**, *19*, 553–572. [[CrossRef](#)]

3. Tommaselli, A.M.G.; Berveglieri, A. Automatic orientation of multi-scale terrestrial images for 3D reconstruction. *Remote Sens.* **2014**, *6*, 3020–3040. [[CrossRef](#)]
4. Yun, X.; Bachmann, E.R. Design, Implementation, and experimental results of a quaternion-based kalman filter for human body motion tracking. *IEEE Trans. Robot.* **2006**, *22*, 1216–1227. [[CrossRef](#)]
5. Pollefeys, M.; Gool, L.V. From images to 3D models. *Commun. ACM* **2002**, *45*, 50–55. [[CrossRef](#)]
6. Longuet-Higgins, H.C. A computer algorithm for reconstructing a scene from two projections. *Nature* **1981**, *293*, 133–135. [[CrossRef](#)]
7. Tomasi, C.; Kanade, T. Shape and motion from image streams under orthography: A factorization method. *Int. J. Comput. Vis.* **1992**, *9*, 137–154. [[CrossRef](#)]
8. Spetsakis, M.; Aloimonos, J.Y. A multiframe approach to visual motion perception. *Int. J. Comput. Vis.* **1991**, *6*, 245–255. [[CrossRef](#)]
9. Szeliski, R.; Kang, S.B. Recovering 3D shape and motion from image streams using nonlinear least squares. *J. Vis. Commun. Image Represent.* **1994**, *5*, 10–28. [[CrossRef](#)]
10. Oliensis, J. A multi-frame structure-from-motion algorithm under perspective projection. *Int. J. Comput. Vis.* **1999**, *34*, 163–192. [[CrossRef](#)]
11. Nister, D. Automatic passive recovery of 3D from images and video. In Proceedings of the 2nd IEEE International Symposium on “3D Data Processing, Visualization and Transmission”, Thessaloniki, Greece, 6–9 September 2004.
12. Vergauwen, M.; Gool, L.V. Web-based 3D reconstruction service. *Mach. Vis. Appl.* **2006**, *17*, 411–426. [[CrossRef](#)]
13. Snavely, N.; Seitz, S.M.; Szeliski, R. Modelling the world from internet photo collections. *Int. J. Comput. Vis.* **2008**, *80*, 189–210. [[CrossRef](#)]
14. Farenzena, M.; Fusiello, A.; Gherardi, R. Structure and motion pipeline on a hierarchical cluster tree. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Kyoto, Japan, 27 September–4 October 2009.
15. Snavely, N.; Seitz, S.M.; Szeliski, R. Skeletal graphs for efficient structure from motion. In Proceedings of the IEEE Computer Society Conference on Computer Vision & Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008.
16. Barazzetti, L.; Scaioni, M.; Remondino, F. Orientation and 3D modeling from markerless terrestrial images: Combining accuracy with automation. *Photogramm. Rec.* **2010**, *25*, 356–381. [[CrossRef](#)]
17. Pierrot-Deseilligny, M.; Cléry, I. APERO, an open source bundle adjustment software for automatic calibration and orientation of a set of images. In Proceedings of International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 38(5/W16), Trento, Italy, 2–4 March 2011.
18. Del Pizzo, S.; Troisi, S. Automatic orientation of image sequences in cultural heritage. In Proceedings of International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 38(5/W16), Trento, Italy, 2–4 March 2011.
19. Agarwal, S.; Snavely, N.; Simon, I.; Seitz, S.M.; Szeliski, R. Building rome in a day. In Proceedings of the International Conference on Computer Vision, Kyoto, Japan, 27 September–4 October 2009.
20. Strecha, C.; Pylvainainen, T.; Fua, P. Dynamic and scalable large scale image reconstruction. In Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010.
21. Fuhrmann, S.; Langguth, F.; Goesle, M. MVE—A multi-view reconstruction environment. In Proceedings of the Eurographics Workshop on Graphics and Cultural Heritage, Darmstadt, Germany, 6–8 October 2014.
22. Snavely, N.; Seitz, S.M.; Szeliski, R. Photo tourism: Exploring photo collections in 3D. *ACM Trans. Graph.* **2006**, *25*, 835–846. [[CrossRef](#)]
23. Pollefeys, M.; Gool, L.V.; Vergauwen, M.; Cornelis, K.; Verbiest, F.; Tops, J. 3D recording for archaeological fieldwork. *IEEE Comput. Graph. Appl.* **2003**, *23*, 20–27. [[CrossRef](#)]
24. Karlsson, N.; Bernardo, D.B.; Ostrowski, J.; Goncalves, L.; Pirjanian, P.; Munich, M.E. The vslam algorithm for robust localization and mapping. In Proceedings of the IEEE International Conference on Robotics and Automation, Barcelona, Spain, 18–22 April 2005.
25. Lemaire, T.; Lacroix, S. Monocular-vision based SLAM using line segments. In Proceedings of the IEEE International Conference on Robotics and Automation, Roma, Italy, 10–14 April 2007.

26. Silveira, G.; Malis, E.; Rives, P. An efficient direct approach to visual SLAM. *IEEE Trans. Robot.* **2008**, *24*, 969–979. [[CrossRef](#)]
27. Triggs, B.; McLauchlan, P.; Hartley, R.; Fitzgibbon, A. *Bundle Adjustment: A Modern Synthesis*; Springer: Berlin, Germany, 2000.
28. Lhuillier, M. Automatic scene structure and camera motion using a catadioptric system. *Comput. Vis. Image Underst.* **2008**, *109*, 186–203. [[CrossRef](#)]
29. Comport, A.; Malis, E.; Rives, P. Real-time quadrifocal visual odometry. *Int. J. Robot. Res. Spec. Issue Robot Vis.* **2010**, *29*, 245–266. [[CrossRef](#)]
30. Royer, E.; Lhuillier, M.; Dhome, M.; Lavest, J.M. Monocular vision for mobile robot localization and autonomous navigation. *Int. J. Comput. Vis.* **2007**, *74*, 237–260. [[CrossRef](#)]
31. David, P. Vision-based localization in urban environments. In *Proceeding of the Army Science Conference*, Orlando, FL, USA, 29 November–2 December 2010.
32. Frontoni, E.; Ascani, A.; Mancini, A.; Zingaretti, P. Robot localization in urban environments using omnidirectional vision sensors and partial heterogeneous a priori knowledge. In *Proceeding of the IEEE/ASME International Conference on Mechatronics and Embedded Systems and Applications (MESA)*, Qingdao, China, 15–17 July 2010.
33. Haralick, R.; Lee, C.; Ottenberg, K.; Nolle, M. Analysis and solutions of the three point perspective pose estimation problem. In *Proceeding of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Hamburg, Germany, 3–6 June 1991.
34. Lepetit, V.; Fua, P. Monocular model-based 3D tracking of rigid objects: A survey. *Found. Trends Comput. Graph. Vis.* **2005**, *1*, 1–89. [[CrossRef](#)]
35. Jiang, B. Calibration-free line-based tracking for video augmentation. In *Proceeding of the International Conference on Computer Graphics & Virtual Reality (CGVR)*, Las Vegas, NV, USA, 26–29 June 2006.
36. Rosten, E.; Drummond, T. Fusing points and lines for high performance tracking. In *Proceeding of the IEEE International Conference on Computer Vision*, Beijing, China, 17–20 October 2005.
37. Comport, A.; Marchand, E.; Pressigout, M.; Chaumette, F. Real-time markerless tracking for augmented reality: The virtual visual servoing framework. *IEEE Trans. Vis. Comput. Graph.* **2006**, *12*, 615–628. [[CrossRef](#)] [[PubMed](#)]
38. Drummond, T.; Cipolla, R. Real-time visual tracking of complex structures. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 932–946. [[CrossRef](#)]
39. Lowe, D.G. Fitting parameterized three-dimensional models to images. *IEEE Trans. Pattern Anal. Mach. Intell.* **1991**, *13*, 441–450. [[CrossRef](#)]
40. Georgel, P.; Benhimane, S.; Navab, N. A unified approach combining photometric and geometric information for pose estimation. In *Proceeding of the British Machine Vision Conference (BMVC)*, Leeds, UK, 1–4 September 2008.
41. Pressigout, M.; Marchand, E. Real-time hybrid tracking using edge and texture information. *Int. J. Robot. Res.* **2007**, *26*, 689–713. [[CrossRef](#)]
42. Baker, S.; Matthews, I. Lucas-kanade 20 years on: A unifying framework. *Int. J. Comput. Vis.* **2004**, *56*, 221–255. [[CrossRef](#)]
43. Dame, A.; Marchand, E. Accurate real-time tracking using mutual information. In *Proceeding of the IEEE International Symposium on Mixed and Augmented Reality*, Seoul, Korea, 13–16 October 2010.
44. Dowson, N.; Bowden, R. Mutual information for lucas-kanade tracking (milk): An inverse compositional formulation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *30*, 180–185. [[CrossRef](#)] [[PubMed](#)]
45. Engels, C.; Stewénius, H.; Nistér, D. Bundle adjustment rules. In *Proceeding of the Photogrammetric Computer Vision*, Bonn, Germany, 20–22 September 2006.
46. Mouragnon, C.; Lhuillier, M.; Dhome, D.; Dekeyser, F.; Sayd, P. Generic and real time structure from motion using local bundle adjustment. *Image Vis. Comput.* **2009**, *27*, 1178–1193. [[CrossRef](#)]
47. Zhang, Z.; Shan, Y. Incremental motion estimation through local bundle adjustment. In *Proceedings of the International Conference on Image Processing*, Barcelona, Spain, 14–17 September 2003.
48. Steedly, D.; Essa, I. Propagation of innovative information in non-linear least squares structure from motion. In *Proceeding of the IEEE International Conference on Computer Vision (ICCV)*, Vancouver, BC, Canada, 7–14 July 2001.

49. Sibley, G.; Mei, C.; Reid, I.; Newman, P. Adaptive relative bundle adjustment. In Proceedings of the Robotics Science & Systems, Seattle, WA, USA, 28 June–1 July 2009.
50. Mei, C.; Sibley, G.; Cummins, M.; Newman, P.; Reid, I. RSLAM: A system for largescale mapping in constant-time using stereo. *Int. J. Comput. Vis.* **2011**, *94*, 198–214. [[CrossRef](#)]
51. Kaess, M.; Johannsson, H.; Roberts, R.; Ila, V.; Leonard, J.; Dellaert, F. ISAM2: Incremental smoothing and mapping using the bayes tree. *Int. J. Robot. Res.* **2012**, *31*, 217–236. [[CrossRef](#)]
52. Indelman, V.; Roberts, R.; Beall, C.; Dellaert, F. Incremental light bundle adjustment. In Proceeding of British Machine Vision Conference (BMVC), Guildford, UK, 3–7 September 2012.
53. Indelman, V.; Roberts, R.; Dellaert, F. Probabilistic analysis of incremental light bundle adjustment. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Tokyo, Japan, 3–7 November 2013.
54. Indelman, V.; Melim, A.; Dellaert, F. Incremental light bundle adjustment for robotics navigation. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Tokyo, Japan, 3–7 November 2013.
55. Wu, F.Z. *Mathematical Methods in Computer Vision*; Science Press: Beijing, China, 2008.
56. Zhang, Y.Q.; Zhou, L.M.; Shang, Y.; Zhang, X.H.; Yu, Q.F. Contour model based homography estimation of texture-less planar objects in uncalibrated images. *Pattern Recognit.* **2015**. [[CrossRef](#)]
57. Zhang, Z. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1330–1334. [[CrossRef](#)]
58. Hartley, R.; Zisserman, A. *Multiple View Geometry in Computer Vision*; Cambridge University Press: Cambridge, UK, 2000.
59. Zhu, B. Determining intrinsic and pose parameters of camera based on concentric circles. In Proceedings of the Third International Conference on Digital Manufacturing & Automation, Changsha, Hunan, China, 18–20 December 2010.
60. Triggs, B.; Mclauchlan, P.; Hartley, R. Bundle adjustment—A modern synthesis. *Vis. Algorithms Theory Pract.* **2000**, *1883*, 153–177.
61. Bakken, R.H.; Eilertsen, B.G.; Matus, G.U.; Nilsen, J.H. *Semi-Automatic Camera Calibration Using Coplanar Control Points*; Norsk Informatikkonferanse: Trondheim, Norway, 2009.
62. Yuuki, I.; Yasuyuki, S.; Kenichi, K. *Bundle Adjustment for 3-D Reconstruction: Implementation and Evaluation*; Memoirs of the Faculty of Engineering, Okayama University: Okayama, Japan, 2011.
63. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [[CrossRef](#)]
64. Bay, H.; Tuytelaars, T.; Gool, L.V. SURF: Speeded up robust features. *Comput. Vis. Image Underst.* **2006**, *110*, 404–417.
65. Muja, M.; Lowe, D.G. Fast approximate nearest neighbors with automatic algorithm configuration. In Proceedings of the International Conference on Computer Vision Theory & Applications, Lisboa, Portugal, 5–8 February 2009.
66. Fischler, M.A.; Bolles, R.C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Comput. Vis.* **1987**, *24*, 726–740. [[CrossRef](#)]
67. Moulon, P.; Monasse, P. Unordered feature tracking made fast and easy. In Proceedings of the 9th European Conference on Visual Media Production, London, UK, 5–6 December 2012.
68. Qifeng, Y. *Videometrics: Principles and Researches*; Science Press: Beijing, China, 2009.
69. Grompone, V.G.R.; Jakubowicz, J.; Morel, J.M.; Randall, G. LSD: A fast line segment detector with a false detection control. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *32*, 722–732.
70. Mulayim, A.Y.; Yilmaz, U.; Atalay, V. Silhouette-based 3-D model reconstruction from multiple images. *IEEE Trans. Man Cybern. Soc.* **2003**, *33*, 582–591. [[CrossRef](#)] [[PubMed](#)]
71. Li, C.M.; Xu, C.Y.; Gui, C.F.; Fox, M.D. Distance regularized level set evolution and its application to image segmentation. *IEEE Trans. Image Process.* **2010**, *19*, 3243–3254. [[PubMed](#)]
72. Kazhdan, M.; Bolitho, M.; Hoppe, H. Poisson surface reconstruction. In Proceedings of the Symposium on Geometry Processing (SGP), Cagliari, Sardinia, Italy, 26–28 June 2006.
73. Waechter, M.; Moehle, N.; Goesele, M. *Let There Be Color! Large-Scale Texturing of 3D Reconstructions*; Springer: Zurich, Switzerland, 2014.

74. Choi, C.; Christensen, H.I. Real-time 3D Model-based tracking using edge and keypoint features for robotic manipulation. In Proceedings of the IEEE International Conference on Robotics & Automation, Anchorage, AK, USA, 3–7 May 2010.
75. Petit, A.; Marchand, E.; Kanani, K. A robust model-based tracker combining geometrical and color edge information. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Tokyo, Japan, 3–7 November 2013.
76. Drummond, T.; Cipolla, R. Visual tracking and control using Lie algebras. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Fort Collins, CO, USA, 23–25 June 1999.
77. David, P.; Dementhon, D. Object recognition in high clutter images using line features. In Proceedings of the Tenth IEEE International Conference on Computer Vision, Beijing, China, 17–21 October 2005.



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).