

Article



Pose Estimation of Automatic Battery-Replacement System Based on ORB and Improved Keypoints Matching Method

Jiabin Jiang, Fan Wu, Pengfei Zhang, Fanyi Wang and Yongying Yang *

State Key Laboratory of Modern Optical Instrumentation, Department of Optical Engineering, Zhejiang University, 38 Zheda Road, Hangzhou 310027, China; jiabinjiang@zju.edu.cn (J.J.); 3110102718@zju.edu.cn (F.W.); wfbz007@126.com (P.Z.); 11730038@zju.edu.cn (F.W.)

* Correspondence: chuyyy@zju.edu.cn; Tel.: +86-0571-87951514

Received: 9 December 2018; Accepted: 4 January 2019; Published: 10 January 2019



Abstract: This paper presents an improved Oriented Feature from Accelerated Segment Test (FAST) and Rotated BRIEF (ORB) keypoints matching method for pose estimation of automatic battery-replacement systems. The key issue of the system is how to precisely estimate the pose of the camera in respect to the battery. In our system, the pose-estimation hardware module is mounted onto the robot manipulator, composed of double high brightness LED light source, one monocular camera, and two laser rangefinders. The camera is utilized to take an image of the battery, the laser rangefinders on both sides of the camera are utilized to detect the real-time distance between the battery and the pose-estimation system. The estimation result is significantly influenced by the matching result of the keypoints detected by the ORB technique. The modified matching procedure, based on spatial consistency nearest hamming distance searching method, is used to determine the correct correspondences. Meanwhile, the iterative reprojection error minimization algorithm is utilized to discard incorrect correspondences. Verified by the experiments, the results reveal that this method is highly reliable and able to achieve the required positioning accuracy. The positioning error is lower than 1 mm.

Keywords: pose estimation; electric vehicle; keypoints matching; camera; laser rangefinders

1. Introduction

Techniques used to estimate the camera pose, with respect to an object, are widely employed in many fields, such as industrial control, object tracking, vehicle navigation, etc. However, a recurring issue is the recovery of the six unknown degrees of freedom of the pose. Various optical 3D pose estimation techniques have been developed in the past few years, such as laser scanning, binocular vision, structured light, and deep learning methods. Laser scanning employs the triangulation relationship in optics [1], however the conventional scanning measurement technique is not always fast enough. Binocular stereo vision technique observes one scene from different viewpoints [2] to recover the depth information of the scene. The basic principle of binocular stereo vision is similar to the mechanism of human eyes. The structured light method projects coded light onto the object [3] and uses the deformed fringes recorded from the surface of the object to calculate the depth and the pose information. Recently, there has been an increased research interest in the deep learning method, where oriented 3D bounding boxes of physical objects can be detected by using a convolutional neural network given 3D sensor data [4]. However, this statistical-learning method requires a large number of training samples [5–8]. Electric vehicle battery charging always time consuming, meanwhile the process of changing an automatic battery takes less than 5 min. The key issue surrounding the

automatic battery-replacement system is how to precisely estimate the pose of the camera in respect to the battery.

Pose-estimation algorithm, based on feature matching and Perspective-n-Point (PnP) method, is an attractive technique since new objects can easily be learned online. The single-shot method is stable and computationally efficient, showing great potential for time-critical applications. This method needs fewer cameras than the binocular stereo vision technique, which means that the system incurs lower-cost and is more compact. However, only the front side of the battery can be captured, which results in a planar pose-estimation problem, and the accuracy of the pose-estimation results will slightly decrease when considering the yaw angle. In this paper, laser rangefinders are utilized to detect the distance between the camera and the battery with higher precision, providing more precise yaw angle information. We focus on solving the problem for 3D pose estimation of textured objects with a monocular camera and laser rangefinders.

The pose estimation aims to calculate the 3D position and orientation of objects via a camera view-point. With oriented Feature from Accelerated Segment Test(FAST) [9] detecting the keypoints, and rotated BRIEF [10] extracting the descriptors, the ORB method is a very fast binary descriptor which is rotation invariant and resistant to noise [11]. With the matched 2D coordinates of the keypoints in the image, and the corresponding 3D coordinates obtained in real world, the pose of the battery will be calculated on the basis of the theory of the pinhole camera model. The Fast Library for Approximate Nearest Neighbors (FLANN) [12] technique is one of the most popular matching techniques for ORB key-points and descriptors because of its efficiency and simplicity. However, this method is based on a global searching technique, the number of incorrect correspondences grows significantly if there are a large number of keypoints with similar descriptors.

This paper presents an improved matching technique based on a spatial consistency searching method to determine the correct correspondences. Finding the corresponding keypoints is still based on the hamming distance, but is improved by combining cluster searching and an iterative reprojection algorithm. Before keypoints matching process, the center-based spatial consistency determining step is executed to narrow the search area for every keypoint. Since the pinhole camera model relates all the correspondences, if too many incorrect correspondences are involved, the accuracy of the pose-estimation results will decrease and the reprojection error will increase. The iterative algorithm is utilized to discard most of the incorrect correspondences.

The end to end solution of battery replacement for electric vehicle is presented in this paper. The whole system consists of the pose-estimation module, the battery-extract module, and the Kuka robot manipulator. All the modules are controlled by the computer server. Experiments indicate that the method proposed is highly appropriate for the estimation task.

2. System Layout

The layout of pose-estimation system is shown in Figure 1, the pose-estimation module is mounted onto the robot manipulator, composed of one calibrated camera, a pair of high brightness LED light source, and two laser rangefinders. The LED light provides a bright and clear view of the battery for the Charge Coupled Device (CCD)camera to capture the image, and the data is then sent to the computer server for pose estimation. The keypoint P_n in the image is extracted from the battery with printing pattern, which is the key step for pose estimation. Laser rangefinders provide the real-time distance from the battery. The 3D position and orientation information of the battery, relative to the camera, is obtained and transformed to the pose relationship between the battery and the robot manipulator. The pose information is transmitted to the robot cabinet, which controls the robot manipulator to complete the process of battery replacement. The demand of the positioning precision would have to be within ± 2 mm in the x and y direction, and ± 3 mm in z direction. The positioning error of our system is lower than 1 mm, while the results are inside about 2 mm without spatial consistency, and the yaw errors is 1° without laser rangefinders.



Figure 1. System layout of pose-estimation system; the pose-estimation module is fixed on the robot manipulator; The pose-estimation module, composed of Charge Coupled Device (CCD) camera, two laser rangefinders and two LED light sources; The battery of electric car with printing pattern.

3. Improved ORB-Based Matching Method for Pose Estimation

3.1. The Pinhole Camera Model and Camera Calibration

The imaging system is the most important part of the whole pose-estimation system for obtaining high quality images. The schematic diagram of the imaging model is shown in Figure 2. A large format CCD camera with wide-angle lens provides high resolution images with more than 2000 pixels in three dimensions. The keypoint P_n in the 3D world coordinate is projected to the image plane. The lenses used in the system sustain a small amount of distortion since they are not perfectly manufactured. The distortion will be corrected in the camera calibration process.



Figure 2. Schematic diagram of the imaging model of the pose-estimation system.

The pinhole camera model [13,14] relates the correspondences between n reference points of the scene in a 3D world coordinate system and their projections on the image plane, as is shown in Figure 3. The model assumes that all the rays of light generated by these pairs of corresponding points would converge to the projection center point, away from the image plane at a distance of *f* which is the focal length. The found scene keypoint $P_c(x_c, y_c, z_c)$ in a camera coordinate system, and the corresponding point *p*(*u*,*v*) on image plane are related by:

$$\begin{bmatrix} u - u_0 \\ v - v_0 \end{bmatrix} = \frac{f}{z_c} \begin{bmatrix} x_c \\ y_c \end{bmatrix}$$
(1)

The 3D transformation between the real-world coordinate $P_c(x_w, y_w, z_w)$ and camera coordinate $P_c(x_c, y_c, z_c)$ can be defined by a 3 × 3 rotation matrix *R* and the translation vector *T*:

$$\begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & T \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}$$
(2)

The final projection can be expressed as Equation (3),

$$Z_{c} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = A \begin{bmatrix} R & T \end{bmatrix} \begin{bmatrix} x_{w} \\ y_{w} \\ z_{w} \\ 1 \end{bmatrix}$$
(3)

where *A* is the intrinsic matrix, (u_0, v_0) is the intersection pixel point of the optical axis and image plane, *s* is the skew angle, which is very close to 0 in most cases, d_u and d_v represent the scale factors in each direction since the pixels in the CCD are not absolutely square. The camera intrinsic parameters listed above, link the keypoint 3D coordinates to pixel coordinates on the captured image.

$$A = \begin{bmatrix} d_u f & s & u_0 \\ 0 & d_v f & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$
(4)

The intrinsic matrix is assumed to be known before estimating the pose of objects. Therefore, the camera needs to be calibrated with Matlab calibration toolbox [15] by exploiting a chessboard.



Figure 3. System layout of the pinhole camera model.

3.2. ORB Detector and Improved Matching Technique

Before pose estimation, the keypoints coordinates on both the image p(u,v) and in the real world $P_c(x_w, y_w, z_w)$ are needed to be known prior to the estimation. A feature matching method performs well on obtaining keypoints. There are many feature matching techniques that have proved to be remarkably successful for machine vision applications. The Scale-Invariant Feature Transform (SIFT) keypoint detector and descriptor is widely used in visual features applications, and has proved to be quite successful [16]. However, this technique requires high computation cost. The Oriented FAST and Rotated BRIEF (ORB) show good performances and low computation costs in keypoint detection [11,17]. ORB utilizes the Feature from Accelerated Segment Test (FAST) to extract keypoints.

The point *F* in Figure 4b is a candidate keypoint, surrounded by a circle of sixteen points. The FAST determines whether the point F can be a keypoint by the following equation:

$$n = \sum_{i=1}^{16} \begin{cases} \frac{max(|I_F - I_i| - I_{threshold}, 0)}{|I_F - I_i| - I_{threshold}|} |I_F - I_i| \neq I_{threshold} \\ 0 & otherwise \end{cases}$$
(5)

where I_F denotes the grayscale intensity of center point F, I_i denotes the grayscale of one of the sixteen points, and $I_{threshold}$ denotes the threshold. The point F can be marked as a keypoint when n is larger than nine and they're all consecutive points. Harris corner detection algorithm [18] is applied to remove bad keypoints since FAST has large responses along edges.



Figure 4. (a) The highlighted patch contains the pixels used for the keypoints detection; (b) The point *F* is a candidate keypoint [19], surrounded by a circle of sixteen points; (c) A number of pair-wise intensity comparisons describe point *F* [20], \overrightarrow{FC} is the vector from keypoint *F* to the centroid *C* of the patch.

BRIEF computes binary descriptors to describe all the keypoints, which is not hindered by lighting, blur, or perspective distortion. However, it has no orientation component and is very sensitive to in-plane rotation. Therefore, we used ORB to calculate the intensity centroid [21] of the keypoint to estimate a patch's orientation [11].

$$m_{pq} = \sum_{u,v} x^p y^q I(u,v) \tag{6}$$

Apparently, the point *C* is the centroid:

$$C = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}}\right) \tag{7}$$

Now the orientation of the patch is determined by the following Equation:

$$\theta = \operatorname{atan2}(m_{01}, m_{10}) \tag{8}$$

Before computing the BRIEF binary descriptor [10], ORB needs to rotate the patch to the same orientation with θ , making it rotation–invariant in BRIEF. To determine whether the object in an image is perfectly matched, ORB needs to compute 256 bits binary descriptors to describe every keypoint of the reference object. Pre-smoothing on the patch with Gaussian kernel helps to achieve better matching performance.

As shown in Figure 4c, BRIEF descriptor [20] is composed of 256 pair-wise intensity comparisons on patch *p*:

$$f(p) = \sum_{1 \le i \le 256} 2^{i-1} \tau(p; x_i, y_i)$$
(9)

where τ is defined by:

$$\tau(F; x, y) = \begin{cases} 1 & if \ p(x) < p(y) \\ 0 & otherwise \end{cases}$$
(10)

where p(x,y) is grayscale intensity at point x = (u,v) in the smoothed patch p. These pair-wise points $n_d(x,y)$ are sampled from an isotropic Gaussian distribution, making the distribution of Hamming

distances between descriptors separated, resulting in better performance of the matching task based on Hamming distances.

Robust matching of keypoints is the key step for this computer vision problem, aiming to extract keypoints from the captured image to their corresponding points in a reference image [22]. In the traditional matching method, all the keypoints of the reference model are considered as the candidate points for a keypoint in the scene image. Unfortunately, too many keypoints with a similar descriptor will increase the number of incorrect correspondences. This paper presents an improved matching algorithm based on a spatial consistency searching method to determine the correct correspondences. Moreover, the iterative reprojection error minimization algorithm is proposed to discard incorrect correspondences.

The batteries are printed with the same pattern and the pose of batteries does not vary much since the electric automobile pulls up in the guide channel. Therefore, the distribution of the keypoints on the battery is similar. Before keypoints matching, the coordinates of the keypoints are rotated by Varimax rotation for spatial consistency, given by:

$$X_{rotation} = X * V \tag{11}$$

where $X = \begin{bmatrix} \overrightarrow{u}^T, \overrightarrow{v}^T \end{bmatrix}$ is the keypoint coordinate, and *V* is the varimax rotational matrix, which is given by,

$$[V,D] = eig(X_m^T X_m) \tag{12}$$

where the algorithm *eig* produces the matrices of eigenvectors *V* and eigenvalues *D* of matrix $X_m^T X_m$. The eigenvalues of matrix *D* should be in reverse order, and the matrix X_m is zero-centered with the following equation:

2

1500

$$\mathbf{X}_{m} = \left[\overrightarrow{u}^{T} - mean(\overrightarrow{u}), \overrightarrow{v}^{T} - mean(\overrightarrow{v}) \right]$$
(13)

As shown in Figure 5, the center point C_s of scene image and center point of the reference model C_m can be easily obtained. For every keypoint in the scene image, it is important to determine the corresponding searching cluster in the reference image to narrow the search area. The searching area is restricted to a circle area with radius of *R* according to the spatial consistency $P_1Cs = Cc_1Cm$, where Cc_1 is the corresponding cluster center point for keypoint P_1 .

keypoint in scen center point in scene 1450 center point in model 1400 1350 Y axis 1300 Cluster2 Cluster3 1250 1200 Cluster1 1150 900 1000 1100 1400 1500 1600 1700 1200 1300 X axis

Figure 5. Keypoints matching based on spatial-consistency; corresponding searching cluster of every keypoint in scene image is restricted to a circular areal according to the spatial consistency $\overrightarrow{P1Cs} = \overrightarrow{Cc1Cm}$.

3.3. Improved RANSAC Algorithm

The main purpose of the pose estimation is to calculate the rotation matrix *R* and the translation vector *T*, given a set of 3D world points and 2D projections. This is so called PnP problem, which assumes the camera is calibrated. To solve this problem, there are many iterative solutions [23,24] and non-iterative solutions [13,25,26]. In this paper, an improved random sample consensus (RANSAC) algorithm based on spatial consistency is proposed to solve this problem. The RANSAC [27] algorithm categorizes all the corresponding keypoints into inliers and outliers. The inliers can be explained by the pinhole camera model and the outliers do not fit to this model. Spatial relations are used here to remove some incorrect correspondence so as to increase the inlier percentage, which can be considered as a preprocessing step for RANSAC algorithm.

Since the matched keypoints are obtained with spatial consistency, the process of the proposed method is composed of three steps: (1) Determining the most reliable corresponding keypoints, (2) Building the spatial evaluation model by obtaining the distance between normal keypoints and most reliable keypoints, (3) Removing the outliers to increase the inlier percentage according to spatial consistency. The evaluation model is built by using the most reliable correspondence, also called the base-points.

It is not very hard to find the base-points, the smaller hamming distance of the descriptor means more reliable matches. However, it is impossible every chosen base-point is 100% reliable, therefore we should choose more than one base-point. Sixteen matched keypoints with the minimum hamming distance are chosen as the candidate-points, as shown in Figure 6a. The candidate-points are sorted by the distances between two candidate-points in scene image and reference image. Eight candidate-points with closer value to the median distance are chosen as the base-points, as is shown in Figure 6b, and the battery of an electric automobile is shown in Figure 6c.



Figure 6. Most reliable matched keypoints (**a**) Sixteen candidate-points (**b**) Eight base-points (**c**) Battery of electric automobile.

After obtaining the base-points, the hypothesis is generated to categorize the correspondences into inliers and outliers. Suppose that P_s and P_r are a pair of corresponding keypoints in scene image and reference image, B_s^i and B_r^i donate i^{th} corresponding base-points in scene image and reference image. The distance in scene image between every undetermined keypoints to a base-point can be determined by $\sqrt[2]{(x(P_s) - x(B_s^i))^2 + (y(P_s) - y(B_s^i))^2}$, and the distance in reference image

is $\sqrt[n]{(x(P_r) - x(B_r^i))^2 + (y(P_r) - y(B_r^i))^2}$. With n base-points, the sum of distance ratios *Sr* for an undetermined keypoint is given by:

$$Sr = \sum_{i=0}^{n} \frac{\sqrt[2]{(x(P_s) - x(B_s^i))^2 + (y(P_s) - y(B_s^i))^2}}{\sqrt[2]{(x(P_r) - x(B_r^i))^2 + (y(P_r) - y(B_r^i))^2}}$$
(14)

the value Sr is used as the criterion to determine the outliers. The sum of distance ratio for every inlier is close to a certain value, therefore the process of finding outliers is similar with that of finding base-points. The undetermined keypoints are sorted by the sum of distance ratios S_r . Top 80% corresponding keypoints with closer value to the median value of S_r are chosen as the inliers, with the remaining 20% keypoints as outliers. The overall flow chart of improved RANSAC method is illustrated in Figure 7. The corresponding 3D coordinates of the chosen non-collinear points is given by the reference model, which is introduced in next subsection.



Figure 7. Flow chart of improved RANSAC method (**a**) Preprocess step of RANSAC to reject some outliers and improve the inliers percentage. (**b**) Ordinary RANSAC algorithm.

3.4. Model Registration and Pose Estimation

Sufficient 2D–3D correspondences are important for the RANSAC and EPnP algorithm to precisely estimate the pose of objects. The 2D pixel coordinates of the keypoints can be easily obtained with an ORB detector, however the corresponding 3D world coordinates cannot be detected from the image. Therefore, the model registration is a critical step that must be conducted before pose estimation. Every battery is printed with the same pattern at the same spot, which ensures the feasibility of our pose-estimation method. One of them was chosen as the reference battery for model registration.

As shown in Figure 8, at the beginning of the registration, it is necessary to load the geometry information of the battery and manually obtain the 2D coordinates of the battery's vertices on the image; the vertices can also be called the control points whose corresponding 3D coordinates can be extracted from the geometry information of the battery. The registration process is offline and there

are not many vertices. Furthermore, we can also obtain the 2D coordinates of vertices with corner detection method. The 2D–3D correspondences of control points are used to estimate the pose of the object. After computing all the ORB keypoints' coordinates and descriptors, it is easy to extract the corresponding 3D world coordinates.



Figure 8. The flow chart of model registration and pose estimation. Model registration is the first step before pose estimation.

The descriptors and 2D–3D corresponding coordinates of the keypoints are obtained when the reference image is registered. For the pose-estimation step, the ORB features and descriptors in the input scene image are computed first, then the pose of the object can be estimated using the improved RANSAC method.

3.5. The Laser Rangefinders

Only the front side of the battery can be captured, which results in a planar pose-estimation problem, and, as a result of this, the accuracy of the pose-estimation results will slightly decrease in the yaw angle. In this paper the laser rangefinders on both side of the camera are utilized to detect the distance between the camera and the battery, providing the yaw angle information with higher precision. The yaw angle α can be computed with the two detected distances d_1 and d_2 , as shown in:

$$\alpha = \arctan(\frac{d_2 - d_1}{L}) \tag{15}$$

where *L* indicates the distance between the two laser rangefinders. The measurement resolution of the 1D laser rangefinder is 1 mm.

4. Results

A series of experiments were carried out to verify the validity of the proposed pose-estimation system. As shown in Figure 9, the automatic battery-replacement system was composed of the pose-estimation module, battery-extract module, and the robot manipulator. The electric automobile pulls over in the restricted parking spot, in front of the KUKA robot manipulator, then the image of the battery is captured by the camera. Meanwhile the laser rangefinder provides the real-time distance away from the car. The suction cups of the battery-extract module will take out the battery when the robot arm receives the pose information of the battery.



Figure 9. (a) The pose-estimation module and battery-extract module are mounted onto the KUKA robot manipulator, the suction cup is used to pull the battery out of the electric automobile; (b) The scene of automatic battery-replacement process.

The image of the battery was recorded by the CCD camera (BM-500GE, GigE Vision, JAI), with the resolution being 2456 \times 2058 and the unit cell size of the CCD sensor being 3.45 um \times 3.45 um. The camera was calibrated with a black and white chessboard by 96 corner points in eight rows and twelve columns. On the chessboard, 30 images were captured with different positions and orientations.

The calibrated intrinsic parameters matrix of the camera is:

$$A = \begin{bmatrix} d_u f & s & u_0 \\ 0 & d_v f & v_0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2401.1 & 0 & 1251.8 \\ 0 & 2410.6 & 1041.1 \\ 0 & 0 & 1 \end{bmatrix}$$
(16)

The lens distortion coefficient is $[K_1, K_2, P_1, P_2] = [-0.009549, 0.12061, 0.00015, -0.00017]$, the higher order of distortion was too small to be considered.

The feature matching result between reference image and scene image is shown in Figure 10. Compared to the ordinary method, the method based on spatial consistency rejected more than 40% of the outliers most of time.



Figure 10. Feature matching result between reference image and image in scene; (**a**) Keypoints matching result based on ordinary method; (**b**) Keypoints matching result based on spatial consistency.

The KUKA robot manipulator with repeat positioning accuracy 0.2 mm was utilized to verify the validity of the system. The unique Euclidean 3D transformation between the camera coordinate system and the robot manipulator coordinate system was defined as a rotation matrix R' and a 3D translation vector T'. The chessboard was used to calibrate the transformation model. To verify the accuracy of pose-estimation algorithm, the robot manipulator moved along the *X*-axis, *Y*-axis, and *Z*-axis. Each time the robot manipulator moved forward for a certain distance, the camera captured the image of the battery. For every moving step along the axis, the pose-estimation module obtained the pose of the object relative to the robot manipulator. Then the relative computed distance between the consecutive two movements along each axis was easily obtained.

Since the movement of the robot manipulator was easy to control, it was reasonable to compare it to the calculated distance along each axis and use the difference to indicate the accuracy of the pose estimation. As shown in Figure 11a, no matter which axis the robot manipulator moved along, the positioning error was lower than 1 mm. The estimation errors of the rotation matrix *R* are shown in Figure 11b, the rotation matrix *R* can be defined by the three angles: θ (roll), ϕ (pitch), ψ (yaw), which are the rotation angles around the *Z*, *X*, and *Y* axis, respectively. The errors of yaw angle were larger than that of roll angle and pitch angle, however that is acceptable for the estimation task in most cases.



Figure 11. Experimental results of the estimation error in 3D coordinates; (a) Errors in translation vector *T*; (b) Errors in rotation matrix *R*, which can be defined by the three angles: θ (roll), ϕ (pitch), ψ (yaw).

Even though the errors of yaw angle were acceptable for most cases in Figure 11, it was not reliable enough with errors of $\pm 1^{\circ}$. Therefore, we used laser rangefinders to detect the yaw angle. The experimental result of yaw angle errors is shown in Figure 12. It can be seen that the laser rangefinders improved the measurement precision, with a performance of $\pm 0.5^{\circ}$.



Figure 12. Experimental results of yaw angle errors with laser rangefinders.

The experimental results of pose estimation for three different electric automobiles with different positions and poses are presented in Figure 13. The four control points of the battery were reprojected to the battery based on the pose information acquired by the pose-estimation system. To further verify the validation of the system, the KUKA robot manipulator was used to grasp the battery based on the relative pose information between the robot and the battery. The results show that the robot can complete the battery replacement with the calculated pose.



Figure 13. Experimental results of pose estimation for three different electric automobiles with different poses. (a) Battery in the minicar; (b) Battery in the compact car; (c) Battery in the truck.

5. Conclusions

This paper develops a pose-estimation method based on an improved keypoints matching method and laser rangefinders. The end-to-end solution of automatic battery-replacement system was also built to automatically replace the battery for the electric automobile. The theoretical framework that calculated the relative pose was proposed by introducing the pinhole camera model, ORB descriptor, improved feature matching methods, and EPnP algorithm. A large format CCD camera was used to capture the image the battery and the ORB detected the keypoints of the printing pattern in the battery to obtain the 3D-to-2D points correspondences. Following this, the pose of the battery was estimated on the basis of the theory of pinhole camera model. The laser rangefinders were used to detect the yaw angle with higher precision. The KUKA robot manipulator, equipped with the battery-extract module, automatically replaced the battery. Comparison experiments between the calculated pose and the movement of robot manipulator showed that the proposed pose-estimation method can satisfy the requirement of positioning accuracy. It can be concluded that the system can be used to solve the battery-replacement problem since the system showed good stability and reliability, which means promising prospects for industrial automation. The system can also be applied to solve some similar pose-estimation problems such as automatic assembly based on computer vision. Our method needs further improvement, especially if the detected object under research has a reduced number of keypoints. Moreover, research is required into how our system can be more compact.

Author Contributions: All the authors contributed to the research work. The major experiments and analyses were undertaken by J.J. Y.Y. guided this study. F.W. and P.Z. designed the experiments and analyzed the experiments. F.W. participated in the experiments. All the authors have read and approved the final manuscript.

Funding: This work was supported by National Natural Science Foundation of China (grant numbers 61627825) and the State Key Laboratory of Modern Optical Instrumentation of Zhejiang University.

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

References

- 1. Ji, Z.; Leu, M.C. Design of optical triangulation devices. Opt. Laser Technol. 1989, 21, 339–341. [CrossRef]
- 2. Hartley, R.; Zisserman, A. *Multiple View Geometry in Computer Vision*; Cambridge University Press: Cambridge, UK, 2000; pp. 1865–1872.
- Jalkio, J.A.; Kim, R.C.; Case, S.K. Three dimensional inspection using multistripe structured light. *Opt. Eng.* 1985, 24, 966–974. [CrossRef]
- 4. Qi, C.R. Frustum PointNets for 3D Object Detection from RGB-D Data. In Proceedings of the Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
- 5. Ferrari, V.; Jurie, F.; Schmid, C. *From Images to Shape Models for Object Detection*; Kluwer Academic Publishers: Norwell, MA, USA, 2010; pp. 284–303.
- Dalal, N.; Triggs, B. Histograms of Oriented Gradients for Human Detection. Paper presented at the IEEE Computer Society Conference on Computer Vision & Pattern Recognition, San Diego, CA, USA, 20–25 June 2005; pp. 886–893.
- 7. Bosch, A.; Zisserman, A.; Munoz, X. Image Classification using Random Forests and Ferns. Paper presented at the IEEE International Conference on Computer Vision, Rio de Janeiro, Brazil, 14–21 October 2007.
- Cao, Z.; Simon, T.; Wei, S.E.; Sheikh, Y. Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. In Proceedings of the Conference on Computer Vision and Pattern Recognition, Las Vegas Valley, NV, USA, 26 June–1 July 2016.
- 9. Dollár, P.; Appel, R.; Belongie, S.; Perona, P. Fast Feature Pyramids for Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 1532–1545. [CrossRef] [PubMed]
- 10. Calonder, M.; Lepetit, V.; Ozuysal, M.; Trzcinski, T.; Strecha, C.; Fua, P. BRIEF: Computing a Local Binary Descriptor very Fast. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *34*, 1281–1298. [CrossRef] [PubMed]
- 11. Rublee, E.; Rabaud, V.; Konolige, K.; Bradski, G. ORB: An efficient alternative to SIFT or SURF. In Proceedings of the IEEE International Conference on Computer Vision, Barcelona, Spain, 3–6 November 2011; pp. 2564–2571.
- 12. Muja, M.; Lowe, D.G. Scalable Nearest Neighbor Algorithms for High Dimensional Data. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 2227–2240. [CrossRef] [PubMed]
- 13. Lepetit, V.; Moreno-Noguer, F.; Fua, P. EPnP: An Accurate O(n) Solution to the PnP Problem. *Int. J. Comput. Vis.* **2009**, *81*, 155–166. [CrossRef]
- 14. Zhang, Z. A Flexible New Technique for Camera Calibration. *Tpami* 2000, 22, 1330–1334. [CrossRef]
- 15. Bouguet, J.Y. Camera calibration toolbox for MATLAB. 2013. Available online: www.vision.caltech.edu/ bouguetj (accessed on 12 December 2013).
- 16. Lowe, D.G. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]
- 17. Sharif, H.; Hölzel, M. A comparison of prefilters in ORB-based object detection. *Pattern Recognit. Lett.* **2017**, 93, 154–161. [CrossRef]
- 18. Harris, C. A combined corner and edge detector. In Proceedings of the Fourth Alvey Vision Conference, Manchester, UK, 31 August–2 September 1988; pp. 147–151.
- 19. Rosten, E.; Drummond, T. Machine learning for high-speed corner detection. In Proceedings of the European Conference on Computer Vision, Graz, Austria, 7–13 May 2006; pp. 430–443.
- 20. Calonder, M.; Lepetit, V.; Strecha, C.; Fua, P. BRIEF: Binary robust independent elementary features. In Proceedings of the European Conference on Computer Vision, Heraklion, Crete, Greece, 5–11 September 2010; pp. 778–792.
- 21. Rosin, B.P.L. Measuring corner properties. Comput. Vis. Image Underst. 1999, 73, 291–307. [CrossRef]
- 22. Mount, D.M.; Netanyahu, N.S.; Moigne, J.L. Efficient algorithms for robust feature matching. *Pattern Recognit.* **1997**, *32*, 17–38. [CrossRef]
- 23. Horaud, R.; Dornaika, F.; Lamiroy, B.; Christy, S. Object Pose: The Link between Weak Perspective, Paraperspective, and Full Perspective. *Int. J. Comput. Vis.* **1997**, *22*, 173–189. [CrossRef]
- 24. Lu, C.P.; Hager, G.D.; Mjolsness, E. *Fast and Globally Convergent Pose Estimation from Video Images*; IEEE Computer Society: Washington, DC, USA, 2000.
- Ansar, A.; Daniilidis, K. Linear pose estimation from points or lines. *IEEE Trans. Pattern Anal. Mach. Intell.* 2003, 25, 578–589. [CrossRef]

- 26. Morenonoguer, F.; Lepetit, V.; Fua, P. Accurate Non-Iterative O(*n*) Solution to the P*n*P Problem. In Proceedings of the 2007 IEEE 11th International Conference on Computer Vision, Rio de Janeiro, Brazil, 14–21 October 2007.
- 27. Fischler, M.A.; Bolles, R.C. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography; ACM: New York, NY, USA, 1981; pp. 726–740.



© 2019 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/).