Article

Multi-View-Based Pose Estimation and Its Applications on Intelligent Manufacturing

Haiwei Yang , Peilin Jiang and Fei Wang *

School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China; yanghw.2005@stu.xjtu.edu.cn (H.Y.); pljiang@mail.xjtu.edu.cn (P.J.)

* Correspondence: wfx@mail.xjtu.edu.cn

Received: 27 June 2020; Accepted: 27 August 2020; Published: 7 September 2020

Abstract: Pose estimation is a typical problem in the field of image processing, the purpose of which is to compare or fuse images acquired under different conditions. In recent years, many studies have focused on pose estimation algorithms, but so far there are still many challenges, such as efficiency, complexity and accuracy for various targets and conditions, in the field of algorithm research and practical applications. In this paper, a multi-view-based pose estimation method is proposed. This method can solve the pose estimation problem effectively for large-scale targets and achieve good performance accuracy and stability. Compared with existing methods, this method uses different views (positions and angles), each of which only observes some features of large-size parts, to estimate the six-degree-of-freedom pose of the entire large-size parts. Experimental results demonstrate that the accurate six-degree-of-freedom pose for different targets can be obtained by the proposed method which plays an important role in many actual production lines. What is more, a new visual guidance system, applied into intelligent manufacturing, is presented based on this method. The new visual guidance system has been widely used in automobile manufacturing with high accuracy and efficiency but low cost.

Keywords: pose estimation; multiple views; intelligent manufacturing

1. Introduction

Pose estimation has been widely used in aerospace [1,2], unmanned driving [3], augmented reality [4,5], intelligent robots [6,7], thermal analysis [8–10], and automobile manufacturing [11,12]. It is a vital research direction in the field of computer vision. Now, six-degree-of-freedom pose estimation (6D pose estimation) is the dominant trend. It refers to the technique of acquiring RGB image or depth image of a target and using the image to estimate three position parameters and three orientation parameters of the object in a specific coordinate system.

6D pose estimation has played a prominent part in some typical fields. For example, in the augmented reality, it is used to perceive the real world, that is, to estimate the position and posture of objects in the real world, so that the information of the virtual world can be reasonably superimposed on the real world. In the aerospace field, it is also widely used. Especially at the space rendezvous and docking, in order to complete the docking work successfully, the relative attitude between the docking spacecraft and the target spacecraft needs to be adjusted, using the pose estimation technology. Besides, 6D pose estimation technology also can help the removal of space debris. It is applied when the satellite manipulator estimates the pose of space debris and captures the debris. In the unmanned driving, 6D pose estimation technology is used to estimate the relative pose between unmanned vehicles and people, which provides a guarantee for the safety of unmanned driving. In the field of intelligent robots, visual simultaneous locating and mapping (vSLAM) uses 6D pose estimation

technology to achieve intelligent robot path planning. In the automobile manufacturing, 6D pose estimation technology is used to implement tasks such as intelligent grasping, painting and welding.

Although 6D pose estimation technology has been widely used in many fields, the applications of 6D pose estimation technology still face many difficulties, such as, the occlusion problem, the cluster problem, the flexibility for different objects, the dependence on the data set, and the low measurement precision. These difficulties put forward higher requirements on the accuracy and robustness of 6D pose estimation algorithm. With the computing power constantly increasing and achievements in deep learning emerging, the 6D pose estimation technology has gained many exciting developments. However, most of them aim at special or small targets, limited by the camera's field of view. Few studies are focused on large-scale or different-size targets, although there are many urgent needs for accurate and efficient 6D pose estimation of large-scale targets. For example, in the intelligent manufacturing of automobile, there are a lot of large-size parts to be handled. Therefore, it is necessary to determine their positions and orientations, which are relative to coresponding robots, in order to achieve intelligent grasping, transporting, welding, spraying and so on. To sum up, the traditional 6D pose estimation technology needs further exploration and development, in order to meet the demand of intelligent manufacturing.

With the development of technology, industrial robots have been widely used in industrial production. In traditional way, the industrial robots are taught to finish so many grasping tasks, in which each part is fixed tightly by complex fixture. The disadvantages are obvious. Firstly, this process requires manual involvement because it requires parts to be accurately fixed before being grasped. Secondly, the process of placing parts usually takes a long time, which affects the efficiency of production. Thirdly, fixtures will be worn out during the production cycle, which will cause deviations in the fixed position of the part and then lead to the failure of the grasping task. Fourthly, it is very inefficient and complicated to re-customize fixtures when new parts are also needed to be accurately fixed. Therefore, industrial manufacturing has a very urgent need for intelligent technology which can measure the pose of each part and guide the robot to automatically handle them. In industrial manufacturing, the requirement of grasping accuracy is usually at the millimeter level [13]. There are some real-time pose measurement devices that can be used for guiding robot, such as the indoor global positioning system (iGPS) [14,15], and the workspace measuring and positioning system (wMPS) [16,17]. iGPS and wMPS have the characteristic of high precision. However, these equipments are large and expensive, not suitable for general industrial applications. Vision system for robot guidance has been used in automotive industry since the end of 20th century. In recent years, vision measurement technology has been developed rapidly. The robotic visual system, which uses vision measurement technology for robot guiding, has been applied extensively in the aerospace, aircraft and automotive manufacturing industries [18–22].

There are several different vision measurement solutions such as monocular vision [23], binocular vision [24], structured light method [25] and multiple-sensor method [26]. In general, the three dimensional information about the part cannot be obtained directly with monocular vision [27]. The structured light method needs the assistance of laser, but it is limited by laser triangular measurement principle. So it can't be used to estimate the pose of large-size parts [28]. In Reference [29], a monocular-based 6-DOF pose estimation technology is proposed for robotic intelligent grasping systems. It can estimate the pose of large-size parts through the camera movement. However, this movement is restricted to translation only, which limits their applications. What's more, this method cannot obtain an accurate initial value of the part pose, which makes it possible to fall into a local minimum in subsequent nonlinear optimization.

In this paper, we focus on the applications of multi-view-based pose estimation. The ultimate goal of this paper is to seek a 6D pose estimation method with high accuracy, high efficiency but low cost, so that it can be widely used in the field of intelligent manufacturing. The multi-view-based pose estimation method, proposed by us, can handle the 6D pose estimation of large-scale targets. Its applications on intelligent manufacturing and are explored deeply. Compared with existing

mainstream methods or applications in industrial manufacturing, our multi-view-based 6D pose estimation method can be directly used in many applications, due to its efficiency for different industrial parts.

The contributions of this work are three aspects: First, we setup a visual guidance system for robotic intelligent grasping, which can estimate the 6-degree-of-freedom (6-DOF) pose of the part and guide the robot to grasp it accurately. Second, we introduce a fast hand-eye calibration method, which can quickly calibrate the relative transformation between the camera and the robot end-effector. Third, we propose a multi-view-based 6-degree-of-freedom pose estimation method for large size parts. Compared with existing methods, the present method can estimate the 6-DOF pose of large-size parts through capturing large-size parts from several different positions and orientations. Specially, we use binocular reconstruction technology to calculate the initial value of the pose before nonlinear optimization, to ensure the validity and accuracy of the final pose. Experimental results demonstrate the proposed method can obtain accurate 6-DOF pose. Moreover the visual guidance system can accomplish intelligent grasping tasks.

2. Methods

2.1. Visual Guidance System

In automotive manufacturing, different parts or workpieces are assembled through robots. Due to the lack or abrasion of fixtures which are used for positioning parts, the position and orientation of these part are not fixed before the robot grasps the part. Therefore, in order to guide the robot, it is necessary to introduce a visual guidance system to measure the 6-degree-of-freedom pose between the parts and the robot, which can ensure the grasping accuracy and the production efficiency. Specially, we use binocular vision system to enhance stability.

As shown in Figure 1, before the robot grasps the part, it moves the camera to measure several feature points on it, and then determines the pose of the part. Parts vary in size and differ in shape, as small-sized parts may be less than $0.5\text{ m} \times 0.5\text{ m}$, while large-sized parts may be larger than $1.5\text{ m} \times 1.5\text{ m}$. For smaller sized parts, all feature points are included in the camera's field of view. In this case, the robot only needs to move and observe at one position. However, for larger parts, all feature points cannot be included in one camera's field of view. To ensure that all feature points are measured, the robot needs to be moved to multiple positions and capture different images of the part through multi-views. The coordinate systems of the robotic intelligent grasping system consist of robot frame (RF), parts frame (PF), robot gripper frame (GF) and camera frame (CF).

In this system, camera is fixed on the end of robot arm and can be moved to different measurement positions along with the robot. The camera frame at the first measurement position, is called the first camera frame ($CF1$), which is taken as the reference frame. Due to the high precision of industrial machine, the parts frame (PF) can be obtained from their models in computer-aided design (CAD) software. The visual guidance system is described in detail as follows:

(1) The robot measurement path is determined by robot teaching, with which the robot brings the camera to measure the feature points on the part. Then the part pose ${}_{PF}^{CF1}T$ is determined, which is the transformation from parts frame PF to the first camera frame $CF1$.

(2) When the target is rest on the initial position, the grasping path is obtained through robot teaching. As shown in Figure 1, the transformation ${}_{PF0}^{CF1}T$, from part frame at initial position $PF0$ to the camera frame at the first measurement position $CF1$, can be given as follows:

$${}_{PF0}^{CF1}T^{-1} = {}_{CF1}^{PF0}T = {}_{GF0}^{PF0}T \times {}_{RF}^{GF0}T \times {}_{CF1}^{RF}T, \quad (1)$$

where ${}_{CF1}^{RF}T$ is the transformation from $CF1$ to RF ; ${}_{RF}^{GF0}T$ is the transformation from RF to $GF0$; ${}_{GF0}^{PF0}T$ is the transformation from $GF0$ to the initial part frame $PF0$, where $GF0$ is the robot tool frame when the robot grasp the part at the initial position.

When the robot grasps the part, the relative pose between gripper and the part is fixed, which means

$${}_{GF}^{PF}T = {}_{GF0}^{PF0}T = {}_{GF_i}^{PF_i}T. \quad (2)$$

Therefore, the transformation can be reformated as:

$${}_{PF0}^{CF1}T^{-1} = {}_{CF1}^{PF0}T = {}_{GF}^{PF}T \times {}_{RF}^{GF0}T \times {}_{CF1}^{RF}T. \quad (3)$$

When the part produces an offset relative to its initial position, the visual system needs to measure the pose of the part at current position. For small-size part and small offset, the camera can also work at the original measurement position. However, for large-size part or large offset, it does not work because current part has been out of the camera's field of view. To solve this problem, we manipulate the robot moving to several positions, in order to make the camera measure current part from different views (positions and orientations). At the i -th view, the pose, from current part to the camera ${}_{PF_i}^{CF1}T$, can be given as follow:

$${}_{PF_i}^{CF1}T^{-1} = {}_{CF1}^{PF_i}T = {}_{GF}^{PF}T \times {}_{RF}^{GF_i}T \times {}_{CF1}^{RF}T, \quad (4)$$

where ${}_{RF}^{GF_i}T$ is the transformation from RF to GF_i , and GF_i is the robot gripper frame at i -th measurement position.

Due to the high accuracy of the robot, the transform relationship between $CF1$ and RF can be regarded as fixed.

(3) In the robot control system, the pose adjustment can be realized by frame offset. The robot can accurately grasp current frame, through changing the transform relationship between RF and GF with an offset:

$${}_{Rf}^{GF_i}T = \Delta_{RF}^{GF}T \times {}_{RF}^{GF0}T, \quad (5)$$

where $\Delta_{RF}^{GF}T$ is the frame offset transformation to be solved.

(4) Combining the above equations, the frame offset transformation is estimated as:

$$\Delta_{RF}^{GF}T = {}_{GF}^{PF}T^{-1} \times {}_{PF_i}^{CF1}T^{-1} \times {}_{PF0}^{CF1}T \times {}_{GF}^{PF}T, \quad (6)$$

where ${}_{PF0}^{CF1}T$ and ${}_{PF_i}^{CF1}T$ are the pose from part frame to the camera frame at initial position and i -th measurement position.

Finally, we can calculate the base frame offset transformation matrix, with which the robot could adjust its grasping path to achieve intelligent grasping.

2.2. Hand-Eye Calibration

Hand-eye calibration is to determine the transformation relationship between the camera frame CF and the gripper frame GF . In this paper, a chessboard is used for calibration (called calibration board).

As shown in Figure 2, the chessboard is placed in the camera's field of view, and the transformation relationship ${}_{BF}^{CF}T$ between calibration board frame (BF) and camera frame CF can be obtained by detecting the corner points on the calibration board. By obtaining the position of the corner points of the calibration board in RF , the transformation relationship ${}_{BF}^{RF}T$ between RF and calibration board frame can be calculated. Through the robot teaching device, the robot pose at current position can be achieved. Then the transformation ${}_{CF}^{GF}T$ can be calculated as follows:

$${}_{CF}^{GF}T = {}_{GF}^{RF}T^{-1} \times {}_{BF}^{RF}T \times {}_{BF}^{CF}T^{-1}. \quad (7)$$

Of course, we can also adopt other hand-eye calibration methods, such as eye-on-hand calibration method, in which the robot drives the camera to move and takes images of calibration board at different positions and angles. In another way, the relationship between camera, calibration board and the robot is established through high-precision measurement equipments (such as theodolite, laser tracker).

Figure 1. The coordinate systems of the robotic intelligent grasping system consist of robot frame (RF), parts frame (PF), robot gripper frame (GF) and camera frame (CF).

Figure 2. Hand-eye calibration is to determine the transformation relationship between CF and GF .

2.3. Multi-View-Based Pose Estimation

The multi-view-based 6-DOF pose estimation method is proposed for the pose estimation of large-size parts. In this situation, the feature points have a wider distribution and the camera can not see all the feature points in one single view. The camera, along with the robot, has to move to several different views to capture all the feature points. The framework of the multi-view-based 6D pose estimation method can be seen in Figure 3. The method begins with feature points extraction and stereo construction from each view. The final 6D pose is estimated after coordinate transformation, initial value estimation and nonlinear optimization.

Figure 3. The framework of the multi-view-based 6D pose estimation method.

In each view, we extract the image points by image processing methods. First, automatic image thresholding [30] is performed to separate the pixels into foreground and background. Second, the candidate objects (ellipses) are selected with their areas and shapes. Third, the coordinates of feature points are computed using Hough transform [31]. The corresponding 3D coordinates of these feature points can be achieved through stereo construction.

Based on camera perspective projection model, the relationship between the coordinate of the camera points P_C and the image points P_I are given as follows:

$$\begin{cases} Z_l P_{Il} = A_l P_{Cl} \\ Z_r P_{Ir} = A_r P_{Cr}, \end{cases} \quad (8)$$

where Z_l and Z_r are the depth factors. $P_{Cl} = [x_{Cl}, y_{Cl}, z_{Cl}, 1]^T$ and $P_{Cr} = [x_{Cr}, y_{Cr}, z_{Cr}, 1]^T$ are the homogeneous coordinates of a spatial point in left camera frame and right camera frame separately. $P_{Il} = [u_l, v_l, 1]^T$ and $P_{Ir} = [u_r, v_r, 1]^T$ are the corresponding coordinates in image frame. A_l and A_r are the left and right camera intrinsic parameters, which can be calculated from the camera calibration method [32], which are formulated as:

$$A_l = \begin{bmatrix} f_{xl} & 0 & u_l & 0 \\ 0 & f_{yl} & v_l & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad A_r = \begin{bmatrix} f_{xr} & 0 & u_r & 0 \\ 0 & f_{yr} & v_r & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

The coordinates P_{Cl} and P_{Cr} have the relationship:

$$P_{Cr} = {}_l^r T \times P_{Cl}, \quad (9)$$

where ${}_l^r T = \begin{bmatrix} {}_l^r R & {}_l^r t \\ 0 & 1 \end{bmatrix}$ is the transformation from left camera frame to right camera frame, which can be computed by stereo camera calibration.

Assuming that there are $M (M \geq 1)$ views and n_i feature points for each view. Therefore, the total number of feature points for all views is

$$N = \sum_{i=1}^M n_i$$

Let P_j^i , $1 \leq j \leq n_i, 1 \leq i \leq M$ represents the coordinate of j -th feature point at i -th view. We need to transform coordinates of feature points according to the transformation between $CF1$ and CFi . The transformation ${}_{CF1}^{CFi}T$ can be expressed as follows:

$${}_{CF1}^{CFi}T = {}_{CFi}^{GF_i}T^{-1} \times {}_{GF_1}^{GF_i}T \times {}_{CF1}^{GF_1}T = {}_{CFi}^{GF_i}T^{-1} \times {}_{GF_i}^{RF}T^{-1} \times {}_{GF_1}^{RF}T \times {}_{CF1}^{GF_1}T, \quad (10)$$

where ${}_{CFi}^{GF_i}T$ is the transformation from CFi to GF_i .

Because the camera is fixed on the end of the robot, the transformation from camera frame to robot frame is determined, which means ${}_{CFi}^{GF_i}T = {}_{CF1}^{GF_1}T = {}_{CF}^{GF}T$. In this way, the above equation can be reformulated as:

$${}_{CF1}^{CFi}T = {}_{CF}^{GF}T^{-1} \times {}_{GF_i}^{RF}T^{-1} \times {}_{GF_1}^{RF}T \times {}_{CF}^{GF}T, \quad (11)$$

where ${}_{GF_i}^{RF}T$ is the transformation from GF_i to RF for i -th view, which can be obtained directly through robot controller.

Using the calculated transformation ${}_{CF1}^{CFi}T$, the coordinates of different views can be transformed to the first view:

$${}^TP_j^i = {}_{CF1}^{CFi}T^{-1} \times P_j^i, \quad (12)$$

where ${}^TP_j^i$ is transformed coordinates in the first view.

When all the feature points are transformed to the first view, we can calculate an initial transformation from the part frame PF to the camera frame CF , which will be used in the following nonlinear optimization.

In this section, the 6-degree-of-freedom pose of the part is obtained by nonlinear optimization based on the initial iteration value. As shown in Figure 4, the transformation relationship between PF and CFi can be given as follows:

$${}_{PF}^{CFi}T = {}_{CF1}^{CFi}T \times {}_{PF}^{CF1}T. \quad (13)$$

Figure 4. The transformation relationships between PF and $CF1$ is known, and transformation relationships between $CF1$ and CFi can be obtained. The transformation relationships between PF and CFi can be calculated.

If the pose of the part relative to $CF1$ is known, the theoretical corresponding coordinate of feature points in image frame can be calculated based on camera perspective projection model. For i -th view, the theoretical corresponding coordinate of feature points in image frame can be calculated as follows:

$$\begin{aligned} l_i \begin{bmatrix} \hat{u}_j^i \\ \hat{v}_j^i \\ 1 \end{bmatrix} &= A_l \times_{PF}^{CFi} T \times P_j^i \\ &= A_l \times_{CF1}^{CFi} T \times_{PF}^{CF1} T \times P_j^i. \end{aligned} \quad (14)$$

Finally, the 6-DOF pose of the part relative to the first view can be estimated by minimizing the following objective function:

$$\hat{T} = \arg \min_{PF} \sum_{i=1}^M \sum_{j=1}^{n_i} \left[(u_j^i - \hat{u}_j^i)^2 + (v_j^i - \hat{v}_j^i)^2 \right], \quad (15)$$

where (u_j^i, v_j^i) the coordinates in image frame of j -th feature point at i -th view, and $(\hat{u}_j^i, \hat{v}_j^i)$ is the corresponding coordinates transformed to the image frame. Especially, when the number of views is 1 ($M = 1$), the objective function becomes:

$$\hat{T} = \arg \min_{PF} \sum_{j=1}^N \left[(u_j - \hat{u}_j)^2 + (v_j - \hat{v}_j)^2 \right], \quad (16)$$

which is the traditional form of single-view pose estimation. In this sense, our method has universal applicability, and single-view pose estimation can be considered as a special form of our proposed multi-view-based pose estimation.

Equation (15) is derived from multi-view-based pose estimation problem of a moved camera. The formula is the sum of distance between the coordinates of the image point under current view and the coordinates of the projected image point transformed into this view. The transformation, which transform the image point from one view to another, plays an important role. In this sense, as long as the transformation relationship between the camera coordinate systems can be obtained, the formula can estimate the pose, whether it is one camera with multi-views, or one-shot with multi-cameras. From this side, it also illustrates the general adaptability and application potential of the formula we derived.

3. Experiments and Discussion

In this paper, we propose a multi-view-based pose estimation method and apply it in industrial manufacturing. In this experimental section, we verify the effectiveness of the proposed method and evaluate its role in actual production line from three aspects. First, simulation analysis is carried out to analyse its performance with different factors. Second, a verification system is built in laboratory, in order to evaluate the proposed visual guidance system. Third, the intelligent grasping system is applied into automobile manufacturing, which has played an important role in the actual production line.

3.1. Simulation Analysis

The accuracy of multi-view-based pose estimation method proposed in this paper is affected by many factors, such as the detection accuracy of feature points, the number of feature points and the number of views. We first design the simulation experiments with generated data, in order to further study the influence of these factors on the final results of multi-view-based pose estimation.

- Influence of feature point detection accuracy on multi-view-based pose estimation

We set the number of feature points to 4 and the number of multi-views is 2. The 3D coordinates of each feature point in the workpiece coordinate system are adopted according to the CAD model of an automobile front floor. The ground-truth 2D coordinates in image frame is generated through camera perspective model with a ground-truth pose value. Then, based on the ground-truth coordinates, Gaussian noises with different levels are added to each feature point. The pose is estimated using the proposed multi-view-based pose estimation method. Finally, the deviations of rotation and translation are calculated, comparing the estimated pose and the ground-truth value. For each noise level, 2000 tests are conducted, and the average error is obtained. The result is shown in Figure 5:

Figure 5. Influence of feature point detection accuracy on multi-view-based pose estimation: (a) rotation error; (b) translation error.

It can be seen from the results that, as the noise level gradually increases, the deviation of the pose estimation also increases, and is approximately linear. When the noise level is 1 pixel, the rotation error is about 0.1° , with the translation error about 1 mm. This shows that the estimated pose is influenced by feature point detection accuracy. In practice, We should reduce the detection error of feature points by improving image quality or adopting special image processing algorithms

- Influence of number of feature points on multi-view-based pose estimation

In order to determine the influence of number of feature points on multi-view-based pose estimation, we set the number of feature points growth from 4 to 20. In each group, the number of multi-views is fixed at 2 and the Gaussian noise is set to 1 pixel. Then we randomly generate a corresponding number of feature points within a certain range. In each test, we perform 2000 times with the added Gaussian noise level, while we calculate the average error of rotation and translation between the estimated pose and the ground-truth value. The result is shown in Figure 6:

As shown in the simulation results, when the number of feature points increases, the pose deviation decreases rapidly; when the number of feature points increases to a certain number (>16), the deviation of the pose estimation tends to be stable and basically does not change. Experiments show that increasing the number of feature points within a certain range can improve the accuracy of pose estimation. Therefore, we should select as many reliable feature points as possible.

Figure 6. Influence of number of feature points on multi-view-based pose estimation: (a) rotation error; (b) translation error.

- Influence of number of multi-views on multi-view-based pose estimation

In order to study the influence of number of multi-views on multi-view-based pose estimation, we conduct three different simulation experiments. One is to not consider the hand-eye calibration error (zero), the other two are to consider the hand-eye calibration errors. In this situation, we first add a smaller hand-eye calibration error (hand-eye calibration error: low-level), and then a larger hand-eye calibration error (hand-eye calibration error: high-level) is added between different views. In the simulation, the number of views is increased from 1 to 20, the number of feature points is 18, and Gaussian noise level is 1 pixel. Simulation experiments are conducted 2000 times with the noise level at each number of multi-views, for three different hand-eye calibration error (zero, low-level, high-level). The average error of rotation and translation between the calculated pose and the ground-truth value is analyzed. The final result is shown in Figure 7.

Figure 7. Influence of number of multi-views on multi-view-based pose estimation: (a) rotation error; (b) translation error.

As shown in Figure 7, when the hand-eye calibration error is zero, the deviation of pose estimation does not increase with the growth of multi-views. Actually, the deviation basically remains unchanged, and is only affected by the accuracy of the feature point detection. However, when different levels of hand-eye calibration are introduced: As the number of multi-views increases, the pose estimation

deviation gradually grows. The larger the hand-eye calibration error it has, the larger the estimated deviation it gains. This is because the error of hand-eye calibration will affect the solution of the objective function, resulting in deviation of pose estimation. However, when the views increase to a certain number (≥ 6), as the number of multi-views increases, the pose deviation tends to stabilize. This shows that the effect of hand-eye calibration errors is random. When there are many views, the offsets of the hand-eye calibration errors between different views are random and will be eliminated by an average way, to some extent. The simulation results show that when there are hand-eye calibration errors, the increase in the number of multi-views will affect pose estimation accuracy, resulting in poor accuracy. In practice, when facing large-size target, we have to use multi-views to estimate its pose. At this time, it is necessary to control the error of hand-eye calibration as much as possible. Therefore, when applied, on the premise of being able to estimate the pose of the object, we should select as few views as possible. In particular, after the number of multi-views is increased to more than 6, the effect of increasing the number of multi-views on pose estimation accuracy tends to be stable, the translation error is less than 1.8 mm, and the rotation error is less than 0.14° . The proposed algorithm still has high accuracy.

From the above three simulations, the standard deviations of rotation error and translation error are about 0.01° and 0.1 mm, for all the conditions (noise level, number of feature points, number of views), which also shows our method has good stability.

Our method is aimed at the pose estimation problem of large size targets and its application in intelligent manufacturing. The similar method is in Reference [29], in which a monocular-based 6-DOF pose estimation technology is proposed for robotic intelligent grasping systems. It can estimate the pose of large-size parts through the camera movement. To verify the adaptability of our proposed method, we follow the settings and data in Reference [29], moving without rotation, and use the same comprehensive angle and position error calculated as:

$$\begin{cases} dR = \sqrt{\Delta\alpha^2 + \Delta\beta^2 + \Delta\gamma^2} \\ dT = \sqrt{\Delta X^2 + \Delta Y^2 + \Delta Z^2}. \end{cases} \quad (17)$$

The estimation errors of the two methods are compared in Figure 8. We can see that there is no significant difference between Liu's method and our proposed method, when limiting the camera movement to translation instead of rotation. However, In industrial applications, the parts have different situations and the camera should observe them from various views. If the camera movement is restricted to just translation, not rotation, the effectiveness of pose estimation method would be greatly reduced. Therefore, in the following experiments, only our proposed method can be applied, because the movements have both rotation and translation in automobile manufacturing.

Figure 8. The comprehensive rotation and position errors of two methods: (a) rotation error; (b) translation error.

3.2. Experimental Validation in Laboratory

In order to verify the effectiveness of the proposed multi-view-based pose estimation method, a robotic intelligent grasping system is setup and experimental validation is performed. As shown in Figure 9, the experimental setup consists of an industrial robot (Fanuc R-2000iA), two industrial cameras (Basler acA2040-35gc). Two cameras are mounted on the fixture of robot. A front floor of automobile is used as the test part. The floor part is a large-size automobile part and the cameras have to move to two different positions to cover it (Figure 10). There are four feature points, two of which are measured in each view (Figure 11).

The accuracy validation experiments are carried out on the experimental setup in Figure 9. After the hand-eye calibration, The front floor part is placed at several arbitrary positions. In each position (named as “Index”), the following two steps are carried out. First, the pose of the deviated floor part relative to the initial pose is measured by robot teaching (ground-truth value). Second, the robot brings the camera to capture the feature points through two views, and the pose is calculated with the proposed method (seen as computed value).

Figure 9. Experimental setup for robot intelligent grasping system. Robot at reset position.

(a) robot at the first view

(b) robot at the second view

Figure 10. Experimental setup for robot intelligent grasping system: (a) Robot at the first view; (b) Robot at the second view.

(a) left and right images at the first view

(b) left and right images at the second view

Figure 11. Two images (left camera and right camera) are captured at the first view (first row) and the second view (second row). Feature points 2 and 3 are measured at first view. Feature points 1 and 4 are measured at second view: (a) left image and right images at the first view; (b) left and right images at the second view.

As shown in Table 1, the part (front floor) is conducted at 9 different positions (from Index 1 to 9). Compared with robot teaching, the root mean squared error for the angle and position is about 0.123° and 0.731 mm; furthermore, the standard deviations are about 0.053° and 0.220 mm. Due to the consistency error of the automobile body and parts, when grasping this workpiece, the required accuracy is usually 0.2° and 1.0 mm [13], which is also the usual accuracy of robot manual teaching. Therefore, it has reached the accuracy requirements of robot intelligent grasping system [29]. The robot grasping achieves good performance using the pose estimated by the proposed method.

3.3. Applications in Production Line

On this basis, the vision-based grasping system has been used in automobile intelligent manufacturing by reconfiguring the traditional production line. Through the integration with production line (see Figure 12), production-manufacturing can be accomplished intelligently, quickly and accurately with the grasping system.

Table 1. Pose estimation errors (rotation and translation) compared between proposed method and robot teaching (seen as ground-truth).

Index	Rotation Error (Degree)			Translation Error (mm)		
	α	β	γ	X	Y	Z
1	0.31	−0.11	0.06	0.64	−0.68	0.47
2	0.26	−0.23	0.08	0.57	−0.37	−0.11
3	0.29	−0.31	−0.04	0.65	−0.16	−0.14
4	0.17	−0.02	0.03	0.19	−0.62	0.06
5	−0.30	0.03	−0.16	0.27	−0.05	−0.30
6	0.23	0.26	−0.07	0.24	−0.59	−0.51
7	0.10	0.15	−0.09	0.39	0.09	0.07
8	−0.08	−0.34	−0.13	0.12	−0.42	0.43
9	0.30	0.14	−0.11	−0.27	0.44	0.42

(a) picture of the first station

(b) picture of the second station

Figure 12. Two different stations in the automobile production line with intelligent grasping system: (a) the first station; (b) the second station.

To quantitatively evaluate its effectiveness and measurement accuracy of the intelligent system, we conduct a verification experiment on the automobile production line. In this experiment, the vehicle workpieces (floor panel and bonnet in the test) to be operated are fixed on the tooling which is made up of multiple fixtures. Each fixture can be controlled to move forward or backward, up or down. By adjusting the states of all the fixtures, the workpieces can be rotated and translated. The state parameters of all fixtures can be directly read by the tooling controller and transformed into the 6-DOF pose of the workpiece fixed on it. In the experiment, the pose parameters of the tooling are recorded as the ground-truth value P_{true} . The intelligent grasping system measures the current workpiece and calculates the position and orientation of the workpiece which is relative to the camera. Then the relative pose is transformed to the end of the welding robot through the external parameters obtained by calibration, so as to guide the welding robot to a specific position accurately and complete the welding operation. The pose parameters measured by the camera are denoted as calculated value P_{comp} , so the measurement error of the system is $P_{err} = P_{true} - P_{comp}$.

In the applications, two different stations are selected for measurement. In the first station, there are four cameras, fixed on the upper side (two cameras) and back side (two cameras) of the workpiece. In the second station, there are three cameras, fixed on upper side (one camera) and back side (two cameras) of the workpiece. The measurement error of the two stations is shown in Tables 2 and 3.

There are 9 sets of measurements in the first station. The average rotation error ($^{\circ}$) is $(0.20, 0.04, -0.05) \pm (0.29, 0.25, 0.06)$, and the average translation error (mm) is $(0.24, 0.23, 0.06) \pm (0.39, 0.25, 0.31)$. Similarly, in the second station there are 15 sets of measurements. The average rotation error ($^{\circ}$) is $(0.33, -0.05, -0.05) \pm (0.34, 0.21, 0.09)$, and the average translation error (mm) is $(0.31, -0.26, 0.04) \pm (0.29, 0.37, 0.35)$. From the two Tables (Tables 2 and 3), we can see that the vision measurement is accurate. What is more, the time consuming of each measurement is less than 50 ms. The system has been applied in actual vehicle production line and the result of long-term operation shows that it can correctly and quickly guide the welding robot arm to complete the welding operation and meet the production requirements of the vehicle production line.

Table 2. The measurement error of workpiece in the first station.

Index	Rotation Error (Degree)			Translation Error (mm)		
	α	β	γ	X	Y	Z
1	0.61	−0.11	0.06	0.64	−0.68	0.47
2	0.56	−0.23	0.08	0.57	−0.37	−0.11
3	0.69	−0.31	−0.04	0.65	−0.16	−0.14
4	0.37	−0.02	0.03	0.19	−0.62	0.06
5	−0.30	0.03	−0.16	0.27	−0.05	−0.30
6	0.23	0.26	−0.07	0.24	−0.59	−0.51
7	0.60	0.15	−0.09	0.39	0.09	0.07
8	−0.08	−0.34	−0.13	0.12	−0.42	0.43
9	0.30	0.14	−0.11	−0.27	0.44	0.42

Table 3. The measurement error of workpiece in the second station.

Index	Rotation Error (Degree)			Translation Error (mm)		
	α	β	γ	X	Y	Z
1	0.42	−0.00	0.02	0.80	−0.25	−0.38
2	−0.07	0.03	0.07	0.37	−0.22	0.17
3	0.19	−0.09	−0.09	0.15	0.45	0.17
4	−0.34	−0.09	−0.01	0.28	0.32	0.39
5	0.39	0.16	−0.11	0.56	0.11	0.01
6	−0.18	−0.21	−0.11	0.47	0.07	0.02
7	0.43	0.31	0.00	0.38	0.35	−0.08
8	0.27	0.24	−0.05	0.13	0.20	0.57
9	0.39	−0.50	−0.01	0.29	0.26	0.18
10	0.34	0.23	−0.02	−0.55	0.47	0.22
11	0.34	0.46	−0.09	0.07	0.42	0.35
12	0.42	−0.25	−0.01	0.74	0.18	−0.51
13	0.41	0.27	−0.14	−0.19	0.73	0.05
14	0.37	0.10	−0.08	0.52	0.11	−0.46
15	−0.39	−0.08	−0.16	−0.43	0.25	0.21

4. Conclusions

In this paper, a multi-view-based pose estimation method is proposed, while an intelligent grasping system is established and applied in intelligent manufacturing. This method is designed for production lines with intelligent requirements, capable of measuring zero-offset poses and adjusting the robot path to complete grasping parts. The industrial camera is fixed on the end of the robot. The robot brings the camera to measure the feature points on the parts in different views. This characteristic makes the proposed method applicable to the posture measurement of large-size parts, and can also solve the posture estimation to a certain extent. In addition, this paper proposes a hand-eye calibration method, which can quickly calibrate the transformation relationship between the camera coordinate system and the robot system. Simulation experiments and actual applications have proved the effectiveness and accuracy of the proposed method. It is hoped that the multi-view-based pose

estimation method and the established intelligent grasping system can be applied to more intelligent manufacturing scenarios.

Author Contributions: Conceptualization, H.Y. and F.W.; methodology, H.Y. and P.J.; writing—original draft preparation, H.Y.; writing—review and editing, P.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Major Science and Technology Projects of China (No. 2019ZX01008103) and Program of Introducing Talents of Discipline to University (B13043).

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

References

1. Rad, M.; Lepetit, V. Bb8: A scalable, accurate, robust to partial occlusion method for predicting the 3d poses of challenging objects without using depth. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 3828–3836.
2. Teng, X.; Yu, Q.; Luo, J.; Wang, G.; Zhang, X. Aircraft Pose Estimation Based on Geometry Structure Features and Line Correspondences. *Sensors* **2019**, *19*, 2165. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Cui, Z.; Liu, Y.; Ren, F. Homography-based traffic sign localisation and pose estimation from image sequence. *IET Image Process.* **2019**, *13*, 2829–2839. [\[CrossRef\]](#)
4. Su, Y.; Rambach, J.; Minaskan, N.; Lesur, P.; Pagani, A.; Stricker, D. Deep Multi-state Object Pose Estimation for Augmented Reality Assembly. In Proceedings of the 2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), Beijing, China, 10–18 October 2019; pp. 222–227.
5. Citraro, L.; Márquez-Neila, P.; Savarè, S.; Jayaram, V.; Dubout, C.; Renaut, F.; Hasfura, A.; Shitrit, H.B.; Fua, P. Real-time camera pose estimation for sports fields. *Mach. Vis. Appl.* **2020**, *31*, 1–13. [\[CrossRef\]](#)
6. Ren, X.; Luo, J.; Solowjow, E.; Ojea, J.A.; Gupta, A.; Tamar, A.; Abbeel, P. Domain randomization for active pose estimation. In Proceedings of the 2019 International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 20–24 May 2019; pp. 7228–7234.
7. Gao, Q.; Liu, J.; Ju, Z.; Zhang, X. Dual-hand detection for human-robot interaction by a parallel network based on hand detection and body pose estimation. *IEEE Trans. Ind. Electron.* **2019**, *66*, 9663–9672. [\[CrossRef\]](#)
8. Liu, Z.; Li, J.; Liu, X. Novel Functionalized BN Nanosheets/Epoxy Composites with Advanced Thermal Conductivity and Mechanical Properties. *ACS Appl. Mater. Interfaces* **2020**, *12*, 6503–6515. [\[CrossRef\]](#)
9. Li, J.; Zhang, H.; Zhou, C.; Chen, Z.; Chen, X.; Long, Z.; Liu, X.; Zhu, W. A Multiparameter Numerical Modeling and Simulation of the Dipping Process in Microelectronics Packaging. *IEEE Trans. Ind. Inform.* **2018**, *15*, 3808–3820. [\[CrossRef\]](#)
10. Li, J.; Zhang, X.; Zhou, C.; Zheng, J.; Ge, D.; Zhu, W. New applications of an automated system for high-power LEDs. *IEEE ASME Trans. Mechatron.* **2015**, *21*, 1035–1042. [\[CrossRef\]](#)
11. Wang, Z.; Fan, J.; Jing, F.; Liu, Z.; Tan, M. A pose estimation system based on deep neural network and ICP registration for robotic spray painting application. *Int. J. Adv. Manuf. Technol.* **2019**, *104*, 285–299. [\[CrossRef\]](#)
12. Liu, Q.; Mo, Y.; Mo, X.; Lv, C.; Mihankhah, E.; Wang, D. Secure pose estimation for autonomous vehicles under cyber attacks. In Proceedings of the 2019 IEEE Intelligent Vehicles Symposium (IV), Paris, France, 9–12 June 2019; pp. 1583–1588.
13. Li, R.; Qiao, H. A Survey of Methods and Strategies for High-Precision Robotic Grasping and Assembly Tasks—Some New Trends. *IEEE ASME Trans. Mechatron.* **2019**, *24*, 2718–2732. [\[CrossRef\]](#)
14. Norman, A.R.; Schönberg, A.; Gorch, I.A.; Schmitt, R. Validation of iGPS as an external measurement system for cooperative robot positioning. *Int. J. Adv. Manuf. Technol.* **2013**, *64*, 427–446. [\[CrossRef\]](#)
15. van Diggelen, F. Indoor GPS theory & implementation. In Proceedings of the 2002 IEEE Position Location and Navigation Symposium, Palms Springs, CA, USA, 15–18 April 2002; pp. 240–247.
16. Xie, Y.; Lin, J.; Yang, L.; Guo, Y.; Zhao, Z. A new single-station wMPS measurement method with distance measurement. In *AOPC 2015: Optical Test, Measurement, and Equipment*; International Society for Optics and Photonics: Washington, DC, USA, 2015; Volume 9677, p. 96771Z.
17. Xiong, Z.; Zhu, J.; Zhao, Z.; Yang, X.; Ye, S. Workspace measuring and positioning system based on rotating laser planes. *Mechanics* **2012**, *18*, 94–98. [\[CrossRef\]](#)

18. Ersü, E.; Wienand, S. Vision system for robot guidance and quality measurement systems in automotive industry. *Ind. Robot Int. J.* **1995**, *22*, 26–29. [[CrossRef](#)]
19. Pérez, L.; Rodríguez, Í.; Rodríguez, N.; Usamentiaga, R.; García, D.F. Robot guidance using machine vision techniques in industrial environments: A comparative review. *Sensors* **2016**, *16*, 335. [[CrossRef](#)] [[PubMed](#)]
20. Guo, S.; Diao, Q.; Xi, F. Vision based navigation for Omni-directional mobile industrial robot. *Procedia Comp. Sci.* **2017**, *105*, 20–26. [[CrossRef](#)]
21. Zhu, W.; Mei, B.; Yan, G.; Ke, Y. Measurement error analysis and accuracy enhancement of 2D vision system for robotic drilling. *Robot. Comput. Integr. Manuf.* **2014**, *30*, 160–171. [[CrossRef](#)]
22. Michalos, G.; Makris, S.; Eytan, A.; Matthaiakis, S.; Chryssolouris, G. Robot path correction using stereo vision system. *Procedia CIRP* **2012**, *3*, 352–357. [[CrossRef](#)]
23. Agarwal, A.; Triggs, B. Recovering 3D human pose from monocular images. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *28*, 44–58. [[CrossRef](#)]
24. Lim, K.B.; Xiao, Y. Virtual stereovision system: New understanding on single-lens stereovision using a biprism. *J. Electron. Imaging* **2005**, *14*, 043020. [[CrossRef](#)]
25. Zhang, S.; Huang, P.S. Novel method for structured light system calibration. *Opt. Eng.* **2006**, *45*, 083601.
26. Liu, B.; Zhang, F.; Qu, X. A method for improving the pose accuracy of a robot manipulator based on multi-sensor combined measurement and data fusion. *Sensors* **2015**, *15*, 7933–7952. [[CrossRef](#)]
27. Zhang, Z. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1330–1334. [[CrossRef](#)]
28. Usamentiaga, R.; Molleda, J.; Garcia, D.F. Structured-light sensor using two laser stripes for 3D reconstruction without vibrations. *Sensors* **2014**, *14*, 20041–20063. [[CrossRef](#)] [[PubMed](#)]
29. Liu, T.; Guo, Y.; Yang, S.; Yin, S.; Zhu, J. Monocular-Based 6-Degree of Freedom Pose Estimation Technology for Robotic Intelligent Grasping Systems. *Sensors* **2017**, *17*, 334. [[CrossRef](#)] [[PubMed](#)]
30. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 62–66. [[CrossRef](#)]
31. Fernandes, L.A.; Oliveira, M.M. Real-time line detection through an improved Hough transform voting scheme. *Pattern Recognit.* **2008**, *41*, 299–314. [[CrossRef](#)]
32. Zhang, Z. Flexible camera calibration by viewing a plane from unknown orientations. In Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 20–27 September 1999; Volume 1, pp. 666–673.

© 2020 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/>).