IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received August 4, 2020, accepted September 10, 2020, date of publication September 23, 2020, date of current version October 2, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3026108

Multi-View Camera Pose Estimation for Robotic Arm Manipulation

IHTISHAM ALI[1], OLLI J. SUOMINEN[1], EMILIO RUIZ MORALES[2], AND ATANAS GOTCHEV[1]
[1]Faculty of Information Technology and Communication Sciences, Tampere University, 33720 Tampere, Finland
[2]ITER Delivery Department, Remote Handling Project Team, Fusion for Energy (F4E), 08019 Barcelona, Spain

Corresponding author: Ihtisham Ali (ihtishamaliktk@gmail.com)

This work was supported in part by the Fusion for Energy (F4E) under Contract F4E-GRT-0901, and in part by Tampere University, Finland.

ABSTRACT This article proposes a novel approach aimed at estimating the pose of a camera, affixed to a robotic manipulator, against a target object. Our approach provides a way to exploit the redundancy of the robotic arm kinematics by directly considering manipulator poses in the model formulation for camera pose estimation. We adopt a single camera multi-shot technique that minimizes the reprojection error over all the rigid poses. The results of the proposed method are compared to four other studies employing either monocular or stereo setup. The experimental results on synthetic and real data show that the proposed monocular approach achieves better and in some cases comparable results to the stereo approach. Moreover, the proposed approach is significantly more robust and precise compared to other methods.

INDEX TERMS Pose estimation, multi-view, multi-shot, machine vision, robotic arm, visual servoing.

I. INTRODUCTION

Camera pose estimation with respect to a target object/scene has been widely researched in the fields of computer and machine vision, photogrammetry and robotics. Accurate pose estimation is needed in numerous applications such as camera calibration [1], localization [2], reconstruction [3], robot visual servoing [4], and augmented reality (AR) [5]. The advances in these fields have significantly benefited users to accomplish a variety of tasks with good accuracy. Despite much progress, there is still need of improvement for application specific methods to improve accuracy and robustness. For example, an approach suited for achieving visually pleasing reconstruction might not be well suited for accurate localization.

In this study, we focus on the prerequisites of visual servoing of a robotic arm for accurate manipulation. Visual servoing uses visual information acquired from cameras to get spatial and semantic understanding of the surrounding to plan the motion of the robot. The most common applications are robotic grasping [6] and medical procedures [7]. Visual servoing depends on many independent components such as accuracy of robot positioning, hand-eye calibration, and target pose estimation. For this study, we restrict our scope to the accuracy of target pose estimation. Pose estimation of the camera against a target position/object can be achieved through various approaches that incorporate different algorithms and/or hardware configurations. Among these, monocular approaches are widely adopted for AR applications [8]. This primarily means that 6-DoF pose is obtained using a single monocular image. The depth of the object with respect to the camera can be estimated from a scaling approach by forming a geometric relationship between the camera and the known metric size of an object in view of the camera.

The generic approaches for pose estimation of a single camera with respect to the object, or vice versa, can be categorized into two groups. The first category of methods finds the solution by estimating the plane-to-view homography and then decomposing it to obtain the pose. This set of methods is known as Homography Decomposition (HD) methods [9]–[11]. Collins and Bartoli [12] proposed a method that analytically solves the problem after the homography is computed. They named their method Infinitesimal Plane-based Pose Estimation (IPPE). The underlying concept is that even when the estimated homography is noisy, it will still be close to the true transform between the image and the model plane at some regions on the plane. The method takes the points on those regions to solve for a pose using 1st order PDE. The second category of methods treats it as a rigid pose estimation problem. It uses 2D-3D point correspondence

The associate editor coordinating the review of this manuscript and approving it for publication was Pedro Neto.

for estimating the pose of the camera relative to the object. This approach is commonly known as Perspective from n points (PnP) [13]. PnP methods work by minimizing the cost function of the correspondence transfer error to estimate a rigid pose. The correspondence transfer is the error between the predicted positions of point correspondences compared with their measured positions. Collins and Bartoli [12], also makes the argument that IPPE has a deep connection with PnP problem, where the n points can be centered at infinitely small separation from each other using the estimated homography. Lu *et al.* [14], proposed a provably convergent method called RPP that iteratively solves the PnP problem. The method minimizes the collinearity error to estimate the rotation part of the pose followed by its associated translation. The method is quite efficient and usually converges in 5-10 iterations from a random geometric configuration. Schweighofer and Pinz [15], extended the work presented in [14] and introduced RPP-SP to handle ambiguous cases that results in the case of planer targets. The method first computes the pose solution in a similar way to [14] and then estimates a second pose solution by minimizing the reprojection error along 1-DoF rotation and translation at a time. The aim is to find the second local minimum if such a minimum exists. The limitations of [15] are that if the first solution is poor, then the second solution suffers as well. Moreover, it is very difficult to physically characterize the ambiguous cases since the second solution is obtained from the roots of a 4th order polynomial, where two of the roots are imaginary. Li *et al.* [16], proposed a non-iterative method that solves the PnP problem numerically in O(n) by producing subsets of three points. Each subset is then solved as a separate P3P problem. The final solution is obtained from the group of solutions that best fits the model.

Alternatively, many studies consider multi-view approaches to achieve better accuracy. In a multi-view approach, the feature points or parts of interest are observed through several views to generate a coherent and accurate model. These features can be linked across views through robust tracking and subsequently aligned through relative geometric transformation. Federico *et al.* presented a closed-form method to estimate the pose of an object from multiple views [17]. The method requires at least one point-point and two point-ray correspondences from two or more views to solve a generalized PnP problem. With the ability to efficiently and accurately match feature point across multiple views, many studies have opted for structure-from-motion (SFM) based approaches, also known as full multi-view. Daniel and Tomas proposed an SFM method that computes the rotation and translation separately for relative views [18]. The approach then optimizes the relative poses globally and evenly distributes the pose errors using bundle adjustment. Typically, approaches that opt for separate estimation of rotation and translation yield good orientation accuracy. However, the position accuracy is often compromised as the errors from rotations estimation step propagate to the translation estimation step. Nonetheless, in the case of study [18], these are compensated for in the bundle adjustment step. Collet and Srinivasa [19] introduced a modified version of full multi-view, which they termed as introspective multi-view approach. This multistep approach first estimates object and camera pose using a single-view method. Once the initial estimates are obtained, the points are clustered and the outliers from matches are removed. Finally, the poses are re-optimized in a bundle adjustment step using the filtered matches. According to the authors, the approach provides a good tradeoff between computational speed and accuracy. This study is important for our comparative analysis since it demonstrates its use for robot grasping application.

Some studies utilize multi-camera approaches to solve the pose estimation problem. Theoretically, multi-camera systems are similar to multi-view approaches for specific cases where time is not a relevant factor. Furthermore, Stereoscopic approach is a specific case of multi-camera approaches where two cameras are separated by a fixed baseline. In such a case, there must exist a considerable overlap between the views. Stereo approaches can obtain highly accurate results due to the inherent advantage of constrained two or more views. The depth estimated from stereo can be considerably more accurate compared to traditional monocular approaches. Clipp *et al.* [20], proposed an approach that estimates the pose in two steps. First, the absolute rotation and up to scale translation are estimated using a 5-point algorithm [21] in one of the cameras. The correction factor for the scale is then computed separately from an additional point correspondence in the second camera. However, the scale retrieval approach is not robust and absolute translation cannot be obtained all the time. Later Clipp *et al.* presented a modified approach that estimates the relative pose of a stereo pair by employing constraints on the feature point selection for pose estimation [22]. The pose is estimated using a selection of four feature points, where the first point is observed in all four views (both stereo-pairs). Two more points should be observed in two-views of one of the cameras (left or right), while the last point is observed in both views from the other camera. The results show improvement over a random selection of points; however, the study lacks comprehensive testing over real data. Geiger *et al.*, proposed a novel approach that generates dense 3D maps from high-resolution stereo sequences in real-time [23]. The authors claim that the presented approach achieves state-of-the-art accuracy in terms of pose estimation and its sub-sequent odometry. The method estimates pose by reprojecting the world points simultaneously on the stereo views and thereby constraining the objective function. The objective function is iteratively optimized using the Gauss-Newton method. Igor *et al.* presented a stereo approach for ego-motion estimation called SOFT [24]. The approach focuses on a careful selection of features and robust tracking for improving the overall accuracy. The author estimates the rotation with the 5- point algorithm [21] and translation with a 1-point stereo method that is iteratively optimized in both views. Raul and Juan presented a similar approach to SOFT for ego-motion estimation with slightly loose constraint on feature

selection [25]. The approach first computes the relative camera pose followed by a local bundle adjustment among a few recent poses. Later, a full bundle adjustment is performed to optimize the camera locations by minimizing the reprojection errors in all the observed views. The approaches in [24] and [25] are more suited for a large amount of data where the tradeoff is maintained between local accuracy and error distribution among all the views.

Though multi-camera approaches provide considerable advantages over monocular approaches, in many cases the additional hardware and software resources required can exceed the allocated resource budget of the task. This study is driven by the motivation to develop an accurate and robust pose estimation method for the International Thermonuclear Experimental Reactor (ITER) project using an eye-in-hand monocular approach. The goal is to perform certain tasks autonomously using a robotic arm with high precision and accuracy. In this work, we attempt to achieve comparable results to stereo approaches by proposing a multi-view monocular approach considering the case of robotic arm manipulation.

The article is organized as follows: In Section 2, we present the problem and define the preliminaries for its formulation. In Section 3, we formulate the proposed method along with other methods considered for comparative analysis. Section 4 presents the experimental setup, error metrics and experimental results using both synthetic and real datasets. Finally, Section 5 concludes the article.

II. PROBLEM FORMULATION

In this study, we attempt to elucidate the approach through geometrical relationships for thorough understanding. We adopt various notations to help us describe the problem and use them consistently throughout the study. We represent the homogeneous transformation matrix by the standard notation T and support it through various sub-indices. The sub-indices b, t, c, and w correspond to the robot base, robot tool/tool center point (TCP), camera optical center and world coordinate frames, respectively. These notations are exemplified in Fig. 1.

The TCP/end-effector pose from the base of the robot, denoted by ${}_bT^t$, is provided readily by the control system associated with the robot. Generally, the robot pose is highly accurate due to the high precision encoders used in the robotic arm at each joint. These robots, especially industrial robotic arms, are designed to perform tasks that require accuracy and high repeatability with precision around $0.1 - 0.2$ mm of the end effector's position. The transformation from the robot TCP to the camera coordinate frame ${}_tT^c$ is known as hand-eye transformation. We have discussed in detail various approaches for hand-eye and robot-world-hand-eye calibration methods in an earlier study [26]. The aforementioned article can be studied for a thorough understanding of the methods and their MATLAB implementation. For this study, we will adopt the reprojection based approach of the robot-world-hand-eye formulation to estimate ${}_tT^c$.

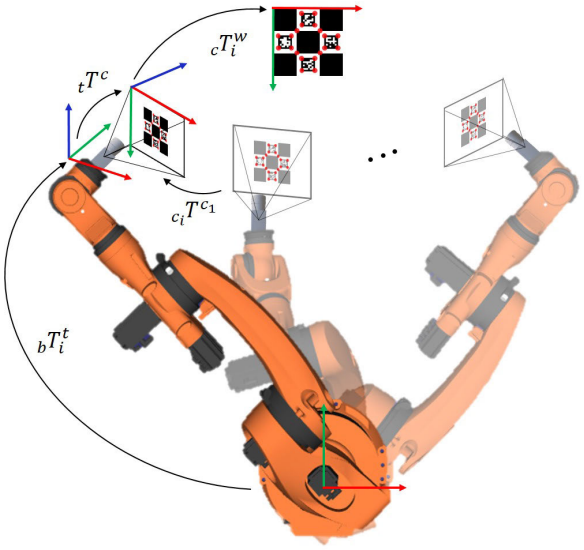

FIGURE 1. Illustration of the setup explaining the geometrical relation among various coordinate frames.

Finally, the unknown homogeneous transformation ${}_cT^w$ from camera coordinate frame to the world coordinate frame affixed to the target object needs to be estimated. The estimation of this transformation matrix defines the pose estimation problem in the described arrangement. Generally, the pose of the camera against the target object or vice versa can be computed independently of the robotic arm pose. For the case of monocular camera with one image (monocular single shot - MSS), only one camera pose ${}_cT_1^w$ exists at $i = 1$ pose. This can be estimated directly with the state-of-the-art methods mentioned in the Introduction Section. Similarly, in the case of stereo camera with one stereo image (stereo single shot - SSS), we estimate only the camera pose ${}_cT_1^w$ which should be able to validate the constrained views of the stereo image pair. The constraint between the stereo pair is a fixed transformation obtained during the calibration of the stereo camera.

In contrast, we can find the reference ${}_cT_1^w$ using multiple images from different views (monocular multi shot - MMS). For $i = (1, 2, \ldots, n)$, we estimate n ${}_cT_i^w$ transformations from camera frame, at each pose, to world frame. In the arrangement shown in Fig. 1, the camera positions at various poses of the end-effector of the robot can be considered independent camera bodies floating in space. The position and orientation of the camera at these n camera poses can be estimated and optimized freely, independent of the robot poses. The optimization is typically performed as local and/or full bundle adjustment [25], where the goal is to distribute the errors and obtain the best possible 3D reconstruction of the object/scene. Theoretically, the transformation ${}_cT_1^w$ estimated through MMS approach is more accurate as it is constrained with the help of the remaining n-1 ${}_cT_i^w$ transformations.

In this study, we propose a modified form of MMS approach where we constrain the free pose optimization of

cameras in an attempt to model the physical system more adequately. We estimate only the transformation ${}_cT_1^w$ and use the prior information (robot poses ${}_bT_i^t$ and Hand-Eye transformation ${}_tT^c$) to constrain and geometrically relate the camera views from n poses. Unlike the traditional MMS approaches, the proposed approach does not need to estimate the additional $n-1$ ${}_cT_i^w$ transformations.

III. METHODS

In this section we discuss the stereo single-shot approach presented in [23], the monocular multi-shot approach presented in [19] and our proposed monocular multi-shot approach. All these methods solve the problem of camera pose estimation in the image space. Such a method takes the world points and estimates a suitable transformation that enables us to reproject the 3D points to the image space at their corresponding views.

The stereo single-shot approach presented by Geiger *et al.* [23] was briefly discussed in Section I. We present here the mathematical relation that we use to estimate the stereo pose. The relationship is given as

$$\begin{aligned}\{q_{(c,w)},\ {}_ct^w\} &= \underset{q_{(c,w)},\ {}_ct^w}{\operatorname{argmin}} (||P^l - \Pi(K^l, [q_{(c,w)},\ {}_ct^w]_{HT}, W)||_2^2 \\ &\quad + ||P^r - \Pi(K^r, [q_{(c,w)},\ {}_ct^w]_{HT} * {}_lT^r, W)||_2^2). \quad (1)\end{aligned}$$

Here, Π is the perspective projection function that projects the 3D points $W = (X, Y, Z, 1)^T$ from world frame space to image space using the camera intrinsic (K^l and K^r) and the stereo extrinsic ${}_lT^r$. The superscript T indicates the transpose of a vector. The subscript or superscript l and r indicate the camera to which the corresponding parameters relate in the stereo pair. The cameras intrinsic and extrinsic are estimated using Zhang's stereo camera calibration approach [11]. The perspective projection yields $\tilde{x} = (\tilde{u}, \tilde{v}, 1)^T$ in the image space of the camera at the pose of interest. The reprojected points $\tilde{x}$ are compared directly against the observed/tracked 2D points (P^l and P^r) in the corresponding left and right image pair. The symbol $[\ ]_{HT}$ indicates the conversion from quaternion $q_{(c,w)}$ and translation vector ${}_ct^w$ to the homogeneous transformation matrix ${}_cT^w$. The solver minimizes the error function in quaternion representation of angles. This helps to reduce the number of unknowns from 12 to 7 parameters. We use the Levenberg–Marquardt algorithm to search for a minimum in the search space by minimizing the L2-norm ($\|\ \|_2^2$) of the residual scalar values.

The second method is a MMS approach known as Introspective Multiview Approach [19]. The method first extracts feature points from the scene and estimates a camera pose for each view using a single-view method. The points are then clustered, filtered, and matched across the views. Finally, the individual poses are re-optimized in a bundle adjustment step using the filtered matches/clusters. The study presents two mathematical relationships for solving the problem; one is based on reprojection and the other is based on back-projection. It then argues that both relations are equivalent in Euclidean space and one may use either of the approaches. The reprojection based approach is generally preferred since it is invariant to projective transformations, while the back-projection does not provide useful information in projective space [19]. The authors opted for back-projection based approach to extend their implementation to be used with other sensors e.g. LIDAR data. However, we use the relationship provided for reprojection based approach as it concurs to the approach we have adopted throughout this study. In the original implementation, feature points from multiple objects in the scene were extracted and multiple hypotheses are generated; one for each cluster of points tracked across views. Since we are using one target pattern, the formulation can be simplified to a single hypothesis optimization problem. In line with the argument in Problem Formulation Section, we estimate and re-optimize n ${}_cT_i^w$ transformations in this MMS approach. The mathematical relationship is given as

$$\{h^*\} = \underset{{}_cT_i^w}{\operatorname{argmin}} \sum_{i=1}^{n} \delta_j^i ||P_1^i - \Pi(K,\ {}_cT_i^w, W)||_2^2. \quad (2)$$

where $h^* = \{h_1^*, h_2^*, ..h_n^*\}$ indicates the set of optimal hypotheses and δ_j^i is a logical operator that switches to 1 when P_1^i has points in a cluster and 0 otherwise. We have fixed the lower subscript to 1 since we assume one cluster of points i.e. the target pattern.

In the proposed method, we use data from n poses and explicitly take the robot poses into consideration. The primary difference between our proposed approach and [19] is that we recommend introducing the robotic arm transformations in the optimization step to constrain the model and minimize the number of unknow transformations from n ${}_cT_i^w$ to ${}_cT_1^w$.

From Fig. 1, we can form the following relationship among n manipulator poses.

$$\begin{aligned}{}_bT_1^t\ {}_tT^c\ {}_cT_1^w &= {}_bT_2^t\ {}_tT^c\ {}_cT_2^w \\ &= {}_bT_3^t\ {}_tT^c\ {}_cT_3^w \\ &\vdots \\ &= {}_bT_n^t\ {}_tT^c\ {}_cT_n^w. \quad (3)\end{aligned}$$

During the estimation step, we optimize only for one homogeneous transformation ${}_cT^w$ that transforms a point from camera frame position in the first/reference view to the fixed object/world coordinate frame. Hence, we curtail the geometric relationship in (3) and accumulate the transformations from world frame to the camera frames at all poses except the reference pose. The resultant transformation $\overline{T}_i$ transforms the 3D world points from object/world coordinate frame, through the first reference pose, to the camera frames at the remaining $n-1$ poses.

$$\overline{T}_i = {}_cT^t\ {}_tT_i^b\ {}_bT_1^t\ {}_tT^c\ {}_cT^w. \quad (4)$$

Since we use quaternion and translation vector representation during optimization, we re-write (4) as

$$\overline{T}_i = {}_tT^{c^{-1}}\ {}_tT_i^b\ \ {}_tT_1^{b^{-1}}\ {}_tT^c\ \ [q_{(c,w)}\ ,\ {}_ct^w]_{HT}. \tag{5}$$

We can now estimate ${}_cT^w$ by optimizing the following expression

$$\{q_{(c,w)}\ ,\ {}_ct^w\} = \underset{q_{(c,w)}\ ,\ {}_ct^w}{\operatorname{argmin}} \sum_{i=1}^{n} ||P_i - \Pi(K, \overline{T}_i, W)||_2^2. \tag{6}$$

Many studies suggest optimizing the camera intrinsic parameters along the solution estimation to achieve better results [27]. On the other hand, Koide and Menegatti [28] contend this argument as an overfitting problem. The rationale that [28] provides is that upon optimizing the intrinsic parameters for the reprojection error, the model overfits to the given samples. This will yield poor results for all error metrics other than the reprojection error and carry the estimate away from the true solution. We observed a similar response while optimizing for ${}_tT^c$. As mentioned before, this transformation is obtained from the robot-world-hand-eye calibration. The calibration is performed on a significantly higher number of poses (10-20) compared to the number of poses used for object pose estimation (3-5). Due to fewer poses, the result deteriorates and the errors propagate to the final solution. Based on the presented argument and experimental results, we opted for a single camera intrinsic and robot-world-hand-eye calibration.

It is noteworthy that the proposed approach is not invariant to the choice of initial estimates for the solver. However, we have successfully constrained the number of unknown parameters to just 7, which improves the convergence of the solution, even with a rough initial estimate. The initial estimates for [19] and our proposed method are obtained from MATLAB's implementation of Zhang's method [11] for camera calibration and monocular pose estimation.

IV. EXPERIMENTAL SETUP AND RESULTS

To assess the performance of our proposed method against other studies, we carry out tests on simulated data with synthetic images and real data. The motivation for using simulated data is to check the actual response of the method against actual ground truth. In contrast, the real data is used to assess the performance of the methods in a real working environment where perturbations in the data are higher and the ground truth is always an approximation.

When estimating the object pose by relaying information through the image space, the selection of feature points for tracking plays a significant role in the overall accuracy of the system. Many studies prefer a markerless approach similar to SFM approaches [18], [29]–[31] to make the system independent of special fiducial patterns. These approaches use feature point correspondences from the feature-rich scene and track them through the views. This is immensely useful for the case where the environment is unregulated and the use of markers is difficult. However, the drawback of such an approach is that the feature correspondence step is prone to outliers. Even in the presence of powerful consensus generator algorithms such as RANSAC [32], the approach inherits additional errors in the form of weak feature correspondence due to tracking or matching. As a result, the accuracy is always an approximation of what it can be in the presence of specialized markers. The study aims to develop an accurate pose estimation method for robotic arm manipulation, where the environment is moderately regulated. It is to our advantage to use specialized markers.

For this study, we use classical checkerboard and ChArUco diamond marker [33], for simplicity referred hereafter to as *diamond marker*. The diamond marker consists of 3×3 squares with 4 ArUco markers placed inside the white squares. This pattern and its detection approach are more robust compared to the use of only markers and compact compared to the original ChArUco pattern.

In this section, we provide a quantitative analysis of the proposed method against four other state-of-the-art methods. Moreover, we discuss the error metrics used to assess the performance of each method. Among the methods used for comparison, IPPE [12] and Zhang [11] are based on monocular single shot (MSS) approaches. Collet and Srinivasa [19] is a monocular multi-shot/multi-view (MMS) approach, while Geiger *et al.* [23] requires a single shot from stereo (SSS) camera pair. We have not considered a comparison with a stereo multi-shot approach as it is redundant for this task and is in conflict with the aim of this study i.e. improving results with reduced hardware. Finally, we discuss and report the experimental results of all these methods for each test case.

A. ERROR METRICS

To estimate the error in the computed pose against the ground truth poses, we use absolute rotation error (deg), absolute translation error (mm), and reprojection error (px). The absolute errors require that the ground truth poses are known. In the case of synthetic data, the ground truth poses are exactly known. In the case of real data, the approximates of ground truth are found through dedicated manual steps as explained in later sections. The absolute rotation error e_{aR} and the absolute translation error e_{at} are given as follows

$$e_{aR} = ||\ angle\ ({}_bR^{w^{-1}}\ {}_bR_{gt}^w)\ ||_2^2, \tag{7}$$

$$e_{at} = ||\ {}_bt_{gt}^w\ -\ {}_bt^w\ ||_2^2. \tag{8}$$

Here the angle() represents the conversion from a rotation matrix to axis-angle for simpler user interpretation. ${}_bR^w$ and ${}_bt^w$ are the rotation and translation from the base of the robot to the world frame.

The final metric that we use, is the reprojection root mean squared error (rrmse). This metric is measured in pixels and is a good way to assess the quality of the results in image space. Moreover, it consolidates the absolute error metrics since the proposed model of reprojection error back-projects the 3D points onto the images by first transforming them through the robotic arm. Hence, each pose has to be accurate and in

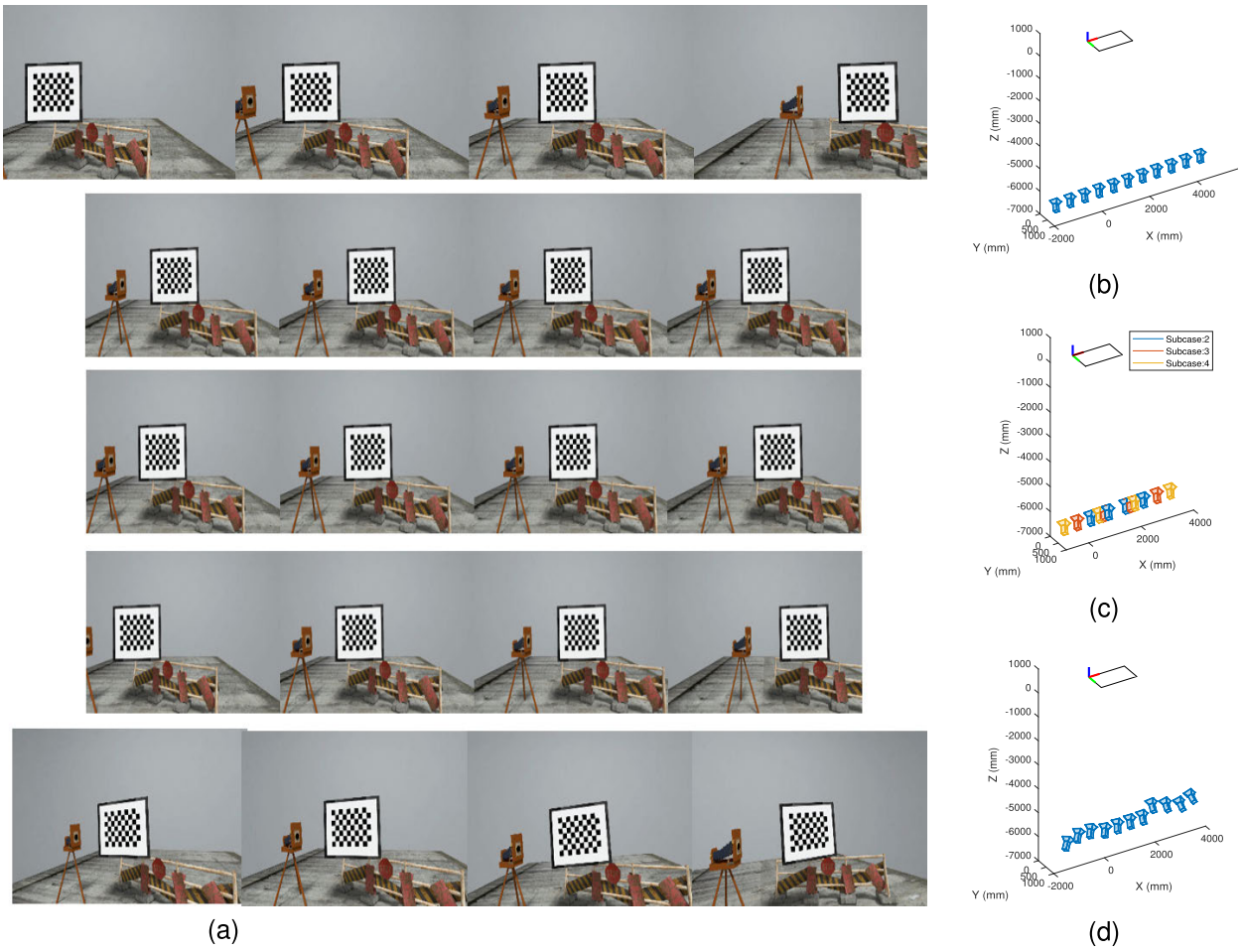

FIGURE 2. Example of the rendered images from the simulated cases and the 6DoF poses of the camera. (a) The first row of images exemplifies case 1 with varying number of poses with motion along one axis. Second to the fourth row include extracts from case 2 (subcase 2-4) with varying inter-pose distance. The last row of images exhibits case 3 where all axes are excited during camera motion and rotation (b-d) Camera poses against the target pattern for each synthetic case shown in (a).

agreement to the overall geometric relationship for the 3D points to back-project precisely onto these different views from corresponding robot poses. The transformation is the same as used in (6) and shown in (4). The reprojection error is given as

$$e_{rrmse} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} ||\overline{P}_i - \Pi(K, \overline{T}_i, W)||_2^2}. \tag{9}$$

B. TESTS ON SYNTHETIC DATA

A significant advantage of using simulated data is the availability of exact ground truth information. In the case of simulated data, we have the ground truth robot poses hand-to-eye transformation and the camera to world object transformation. In real cases, the ground truth hand to eye transformation is not available as it is not feasible to estimate the exact location of the optical frame in a physical setup. Moreover, any ground truth robot pose and the camera pose is only the best possible approximation of the actual information. For the simulated case, we generated high-resolution synthetic images instead of simulated points, as shown in Fig. 2a. These synthetic images were generated using Blender, a 3D computer graphics software. It should be noted that the tripod in the images is part of the scene and is not to be confused with the virtual camera that is capturing the scene. To simplify the experimentation, we assume that the camera and the robot TCP position is the same for the simulated test cases. This means the virtual position of the camera is the position of the robot TCP. Then the homogeneous transformation from hand-to-eye constitutes of just rotation. This rotation is the result of the transformation between the Blender world frame and Blender camera frame.

To study the effect of various parameters, we set up three test cases for the simulated data based experimentation. The excerpts from these cases are shown in Fig. 2. The results from these simulated data aids in selecting suitable parameters to use for real data acquisition and testing. Moreover, we induce visual noise to the points detected for pose estimation. The noise was introduced to study how well the methods can converge towards an accurate solution in the presence of uncertainties. The generated noise has a Gaussian distribution

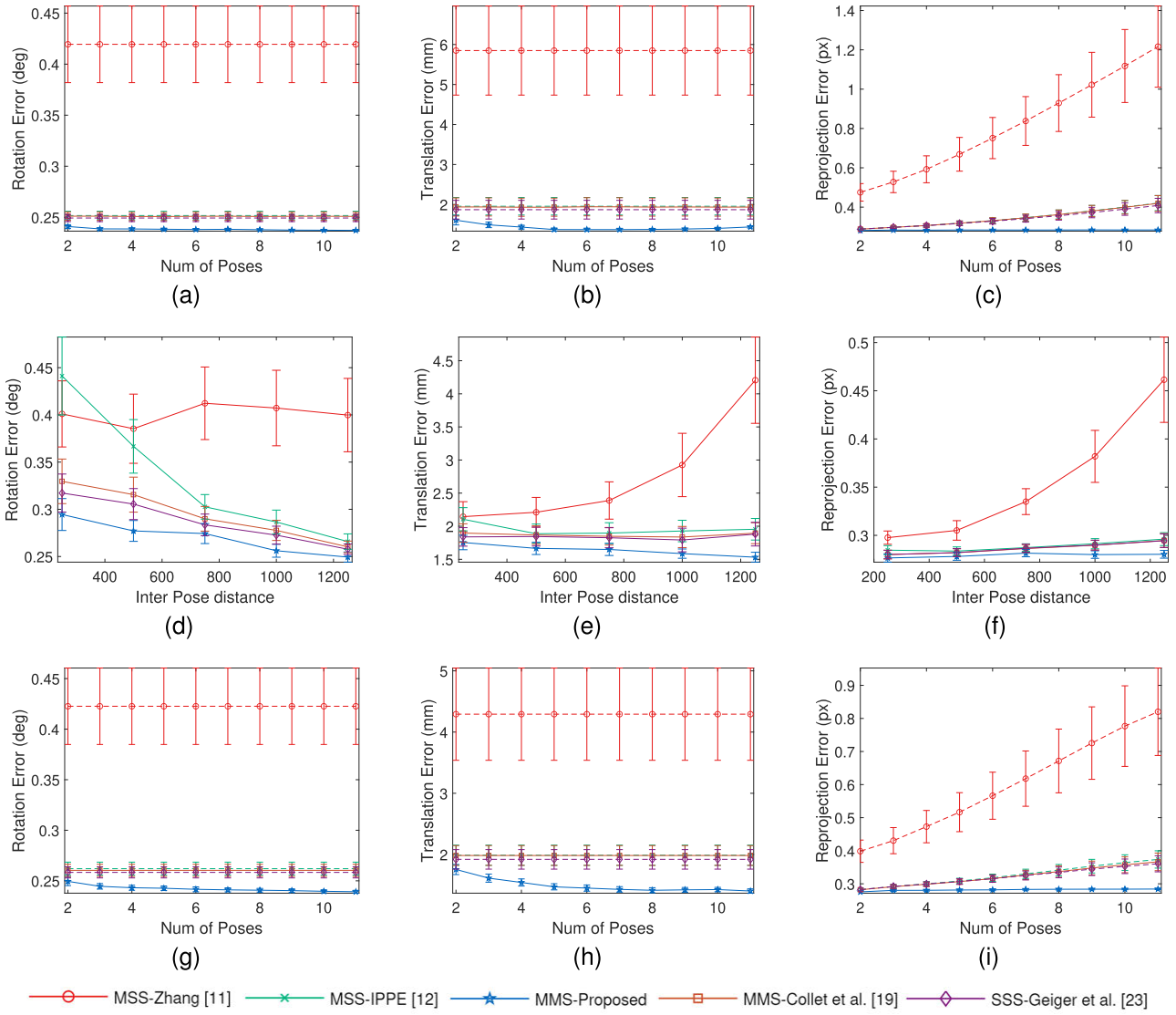

FIGURE 3. **Metric error results for the synthetic data with added Gaussian noise to images (a-c) Results for case 1 of the synthetic data where the camera moves along one axis (d-f) Results for case 2 of the synthetic data with varying inter-pose distance and one axis motion (g-i) Results for case 3 of the synthetic data where all axes are excited during camera motion and rotation. The dashed line exhibit extrapolation from the first pose estimate for the case of monocular methods.**

with a mean of 0.5 and standard deviation of 0.5. To avoid a biased result due to the addition of noise to synthetic data, we repeat the experiment for 50 iterations. In each iteration, we introduce the same level of noise, randomly generated, with a Gaussian distribution. Finally, the mean performance over these 50 iterations is considered a stable response of the corresponding method over the given data.

In the first case, we study the effect of varying the number of poses while moving the camera only in one axis. We choose the horizontal axis. Few images from this case are shown in the first row of Fig. 2a, where the camera moves in one axis only. The camera pose distribution against the calibration pattern for case 1 can be observed from Fig. 2b. We analyze the response of the methods when we increase the number of poses from where the object is viewed. The response of the methods can be observed in Fig. 3a, 3b and 3c. The rotation and translation error show slight improvement especially in the case of the proposed method. It is noteworthy, that the confidence interval of the proposed method is the smallest, which correlates to good precision over varying noise. On the contrary, Zhang [11] show significant deviation from its mean results. The rising trend in the reprojection error can be explained by the fact that the single shot (MSS and SSS) estimates the pose from one image only. The estimate might be accurate for that specific viewpoint, however, its global accuracy is poor as we attempt to use that geometrical information to transform and reproject the points onto other poses. Moreover, Collets and Srinivasa [19] shows a similar increase in the reprojection error, even though, it is based on a multi-shot based approach similar to the proposed method. IPPE [12] shows significantly better results despite being a single-shot approach. The proposed method begins

to reach the minimum error using 5 unique poses for image acquisition. It is noteworthy, that the rotation and translation errors are constant over the increasing number of poses for the single-shot methods. This is because they are extrapolated from the first/reference view for which the camera pose is estimated against the target pattern. The extrapolated dashed line is intended to assist readers in visually comparing the single-shot methods with multi-shot methods along the increasing number of poses. In contrast, the reprojection error for the single-shot methods is not constant over the varying number of poses. This is because the mean reprojection error is estimated over all the views for all methods by using the estimated first/reference pose and the ground truth poses of the relative views. This is done to ensure that the estimated pose from a method is not the result of a local minimum rather the solution is globally consistent and accurate.

In the second case, we attempt to examine the effect of interpose distance on object pose estimation. We fix the number of images for estimating the pose so that the only varying parameter is the interpose distance. The second case has further six sub-sequences where each sequence has four images. In each of the sub-cases, the interpose distance is varied. In Fig. 2a, we show three sub-sequences (of case 2) in row 2, 3 and 4, where the interpose distance is 400, 600 and 800mm, respectively. We visualized the camera pose distribution against the calibration pattern for the aforementioned sub-cases of case 2 in Fig. 2c. The response of the methods on the data from case 2 can be observed in Fig. 3d, 3e and 3f. A noticeable change can be observed between the results of case 1 and case 2. The responses on the case 2 are more sharply varying especially for the single-shot approaches. This is because the camera pose for the reference image (first image) changes as we increase the pose distance from the middle of the scene. In case 1, we started from one side of the scene and moved the camera along the horizontal axis by adding more frames. As a result, the reference frame always remained the same. In contrast, the reference pose/view point in case 2 changes as we move further from the center of the scene. Pose estimates of the same object from different view points may incorporate different levels of uncertainties. As a consequence, we observe in Fig. 3 (d-f) that the estimates between two consecutive data points exhibit a sharp change in response as we vary the interpose distance. This effect strongly points toward the data dependency of many single-camera methods. This data dependency results in the form of imprecise solutions. Here, the primary factor causing this dependency is variation in the poses chosen for calibration, however, such an effect may also be observed due to the model of the robot, and how the robot is mounted, which may introduce new errors. Nonetheless, the proposed method yields the best result over varying interpose distance followed by Geiger *et al.* stereo based approach [23]. The overall trend shows that increasing the interpose distance improves the accuracy of the estimate, with the exception for Zhang's response [11]. Moreover, MMS and SSS approaches exhibit more stable response compares to MSS approaches.

The final test case of the simulated data focuses on studying the impact of position and orientation change in more than one axis. The dominant motion is the same as in case 1. However, small position and orientation changes are also introduced in other axes as well. Few images from this case are shown in the last row of Fig. 2a and the camera poses are given in Fig. 2d. An apparent change between case 1 and case 3 can be observed in the images and the camera poses in the form of change in yaw angle. All other movements are minute and cannot be observed from the images. The response of the methods on the data from case 3 can be observed from Fig. 3g, 3h and 3i. The results follow the trend of case 1, where IPPE [12], Collet and Srinivasa [19], and Geiger *et al.* [23] show almost similar responses with Geiger *et al.* [23] method yielding the lowest errors among them. Zhang [11] shows the largest error while the proposed method yields the best results on all the error metrics. It is noteworthy that the mean errors for case 3 are marginally lower than the errors in case 1. This exemplifies that it is important to excite motion and rotation around all axes to yield better results.

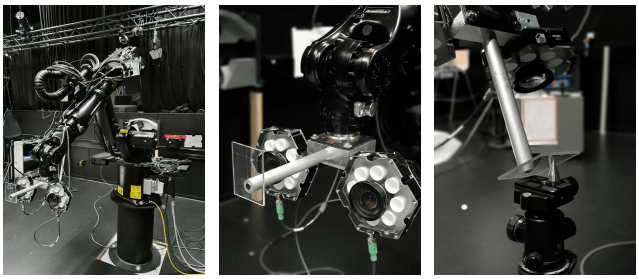

FIGURE 4. An example of the experimental setup (a) KUKA KR16L6-2 robotic manipulator used for recording data (b) Close up of the adaptor with the tool, stereo camera pair and lights affixed to the manipulator using customized hardware (c) A snapshot from the tool 4 point XYZ-calibration step.

C. TESTS ON REAL DATA

We further study the performance of the methods using real data. The real data is acquired using industrial-grade equipment for high accuracy. The experimental setup is shown in Fig. 4. A custom adaptor was designed to fix two Basler acA1920-50gc cameras to KUKA KR16L6 -2 serial 6-DoF robot arm, as shown in Fig. 4b. We use 6mm lens with each camera. The stereo pair has a baseline of 14 cm. Moreover, we use dedicated lamps to uniformly light the target. The use of these lamps is not mandatory; however, they are convenient in maintaining a uniform brightness irrespective of the room lighting condition. The adaptor not only houses the cameras but also holds a custom tool. The tool is an aluminum bar with a Polycarbonate sheet at the end. A cross-hair marker is drawn on this sheet. The purpose of this tool is to manually measure the position and orientation of the target object as accurately as possible. The intersection of the cross-hair marker helps to pinpoint the position while the planer surface of tool sheet aids in measuring the orientation of the planer target. Since the study focuses on accurately estimating the

TABLE 1. Comparative results using checker board as target object.

Methods	Abs. Rotation Error, $\mu_R(deg)$	Abs. Translation Error, $\mu_t(mm)$	Abs. Reprojection Error, $\mu_{re}\ (px)$	Rotation std. dev., $\sigma_R\ (deg)$	Translation std. dev., $\sigma_t\ (mm)$	Reprojection std. dev., $\sigma_{re}\ (px)$
MSS-Zhang [11]	1.9546	3.4217	1.3194	0.036274	0.19947	0.21217
MSS-IPPE [12]	1.9586	3.6378	1.3345	0.047422	0.2165	0.22098
SSS-Geiger et al. [23]	1.9325	**2.931**	1.1621	**0.027213**	0.22121	0.26702
MMS-Collet et al. [19]	1.9498	3.4847	1.3306	0.03223	0.20967	0.2204
MMS-Proposed	**1.6796**	3.1285	**1.0733**	0.045619	**0.15001**	**0.12447**

pose of the camera against the target without target handling; the tool effectively fulfills the purpose. The tool is calibrated for the robotic arm using KUKA's XYZ 4-point method for position and ABC 2-point method for orientation calibration as illustrated in Fig. 4c.

We utilize the tool for both initial ground truth measurement and evaluation of the estimated pose. The ground truth is measured by manually aligning the tool marker on the target object and recording the robot pose information. The estimated poses from the experiments are then compared to this recorded pose.

However, we observed a marginal instability at the base of the robotic arm used for experimentation. The base instability is observed near the maximum reach of the arm inside the workspace. As a result, the manual measurement of the ground truth has small uncertainties (around 2mm). Therefore, we refer to it as the desired pose instead of the ground truth in this work. Nonetheless, the evaluation of our estimates against the desired pose provides us with invaluable results for relative comparison of the methods under consideration in this study.

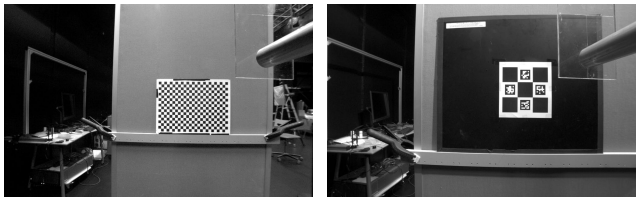

FIGURE 5. Example of the captured images using checkerboard and diamond marker as target objects.

We perform two sets of experiments using different patterns as the target objects, shown in Fig. 5. The first set of experiments uses a checkerboard of size 18×25, while the other uses a diamond marker. Based on our earlier results from synthetic data, we can infer that 5 poses are sufficient for a multishot approach to converge to a stable solution. The distance of the target object from the tool is kept around 0.76 meters since the reach of the robot with the custom tool is approximately 2.15 meters. It is not possible to manually measure the desired pose on the target object for evaluation after 2.15 meters with the current setup. Each experiment is repeated multiple times from randomly initialed pose with varying additional poses. At each pose, the stereo pair takes images of the target object. We can easily see the distribution of the poses in 3D space in relation to the calibration pattern for the real test cases from Fig. 6.

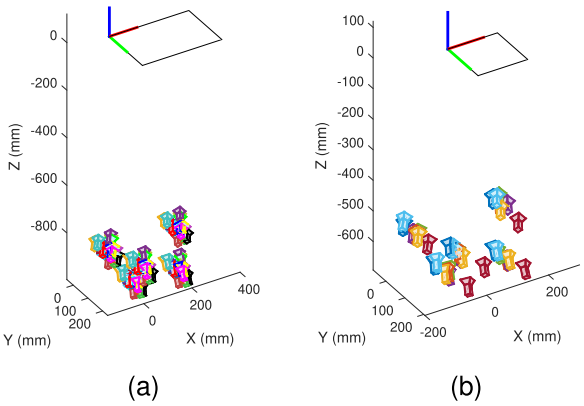

FIGURE 6. Camera poses against target pattern for multiple set of experiments where each color represent a unique set of experiment (a) Poses for the checkerboard (b)Poses for the diamond marker.

The experimental results for the tests using checkerboard as the target pattern are shown in Table 1 and Fig. 7. We compute the arithmetic mean (μ) and standard deviation (σ) of the corresponding errors from all the test iterations. The tabulated results show that the proposed method yields the least absolute rotation error (μ_R) and absolute reprojection error (μ_{re}). The least absolute translation error (μ_t) is obtained by the stereo approach in [23], however, the proposed approach yields a comparative result with the second-best translation estimate. The results obtained for μ_R, μ_t,and μ_{re} using [11], [12] and [19] are quite similar for the given set of experiments.

In addition to accuracy, the system must be consistent in realizing its accuracy over varying data samples. If a method achieves good results only half of the time then the system is not robust and requires improvement in precision. We also tabulate the standard deviation of the estimates for quantitative analysis of the robustness of the methods. It can be observed from Table 1, that the proposed method yields significantly lower deviations over translation (σ_t) and reprojection (σ_{re}) estimates. The least standard deviation for rotation (σ_R) is achieved by SSS-Gieger *et al.* [23]. A comparable result is obtained by the proposed method for σ_R

TABLE 2. Comparative results using diamond marker as target object.

Methods	Abs. Rotation Error, $\mu_R(deg)$	Abs. Translation Error, $\mu_t(mm)$	Abs. Reprojection Error, $\mu_{re}\ (px)$	Rotation std. dev., $\sigma_R\ (deg)$	Translation std. dev., $\sigma_t\ (mm)$	Reprojection std. dev., $\sigma_{re}\ (px)$
MSS-Zhang [11]	2.508	4.0691	2.3638	0.92636	1.0597	0.20078
MSS-IPPE [12]	2.3373	4.1017	2.2399	0.52881	0.47025	0.18418
SSS-Geiger et al. [23]	2.3903	4.2761	2.1124	0.20142	0.30155	**0.13553**
MMS-Collet et al. [19]	2.3095	4.0959	2.2096	0.25623	0.50482	0.17876
MMS-Proposed	**2.1837**	**4.0628**	**2.105**	**0.14443**	**0.21076**	0.15846

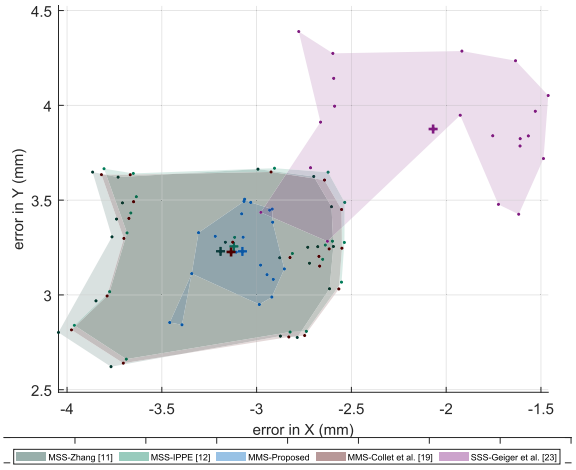

FIGURE 7. A 2D precision plot for the set of experiments using a checkerboard target. Each data point is the error of the estimate against the desired pose without the measure from the axis containing the depth information. The corresponding information from each method is shown in a unique color. The mean of each set of data points is presented by the + symbol and the region encompassing the scattered estimates shows the spread of estimates from respective methods.

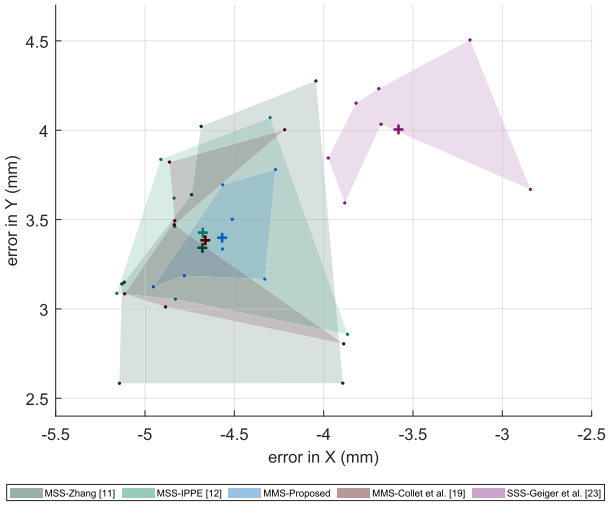

FIGURE 8. A 2D precision plot for the set of experiments using a diamond marker as target object. Each data point is the error of the estimate against the desired pose without the measure from the axis containing the depth information. The corresponding information from each method is shown in a unique color. The mean of each set of data points is presented by the + symbol and the region encompassing the scattered estimates shows the spread of estimates from respective methods.

with a significantly small deviation value. To better illustrate the effect of the standard deviation, we plot the translation estimates error without its depth dimension in Fig. 7. The mean of the data points is presented by the + symbol and the region encompassing the scattered estimates is shown in a unique color. It can be easily observed from the plot that the proposed method shows the most consistent results compared to other aforementioned methods.

We can see from the results that an offset is observed as the estimated errors lie in the quadrant formed by the negative X-axis and positive Y-axis. All the estimate errors lie within this quadrant, which is not common for a natural distribution of error. The offset indicates an uncertainty produced by a more direct cause, which is the instability of the robot base, as mentioned earlier. The effect is more apparent as the tool moves away from the robot base causing a marginal flex. However, the comparative performance of the methods under consideration and their statistical analysis are not directly affected by this problem.

The results for the second set of experiments using the diamond marker as the target object are presented in a similar structure in Table 2 and Fig. 8. The tabulated results show that the proposed method achieved the best results over all metrics except for σ_{re}, where it yields a result comparable to the stereo approach. SSS-Gieger *et al.* [23] shows a comparable error distribution to the proposed method.

The monocular single-shot approach by Zhang [11] seems to have the least consistent performance. The standard deviation of the estimate errors (σ_R, σ_t, and σ_{re}) is the highest in this case. The MSS-IPPE [12] performs comparatively better among the MSS approaches and yields a comparative result to the monocular multi-shot approach by Collins and Bartoli [12].

Moreover, it can be noted that the overall error of this set of experiments is larger by some factor compared to the results obtained for the experiments using the checkerboard. This might be due to the reason that we extract 408 corner points from a checkerboard of size 18 $\times$ 25. On the other hand, we use only 20 points from the diamond marker as illustrated in Fig. 1. The number of points is almost 20 times less for the case of a diamond marker with comparatively small spatial distribution to the checkerboard. In our opinion, the increase

in error for such a situation is in accordance with the stated reason.

The distribution of error estimates can be observed in Fig. 8 for the case of the diamond marker. As before, the proposed approach yields the most consistent result with its estimates being more precise and uniformly distributed.

V. CONCLUSION

In this article, we proposed a monocular multi-shot approach to estimate the 6-DoF pose of the camera against a planar target (object). The proposed approach models the geometric relation among various coordinate systems and explicitly incorporates the robotic manipulator poses into the formulation. It uses a non-linear optimizer to iteratively minimize the reprojection error based cost function. The experimental results were compared to four other existing studies, which included two monocular single shot, one monocular multi-shot, and one stereo approach. The tests were performed on both simulated data with synthetic images and real data. Two target patterns were considered for real data testing. Our method demonstrates significant improvement and robustness on many metrics in various test cases against other methods. In addition to improved accuracy, our approach achieves the most precise results.

ACKNOWLEDGMENT

The results are intended to be integrated in advanced camera-based systems attached to robotic manipulators in order to achieve complex remote handling for maintenance and operation in a safe and efficient way.

This article reflect the views of the authors. F4E and Tampere University (TAU) cannot be held responsible for any use which may be made of the information contained herein.

REFERENCES

[1] A. Richardson, J. Strom, and E. Olson, "AprilCal: Assisted and repeatable camera calibration," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Nov. 2013, pp. 1814–1821.

[2] J. A. Castellanos and J. D. Tardos, *Mobile Robot Localization Map Building: A Multisensor Fusion Approach*. Boston, MA, USA: Springer, 1999.

[3] I. Ali, O. Suominen, and A. Gotchev, "Discrimination of active dynamic objects in stereo-based visual SLAM," *Electron. Imag.*, vol. 2018, no. 13, pp. 1–6, 2018.

[4] W. Pan, M. Lyu, K.-S. Hwang, M.-Y. Ju, and H. Shi, "A neuro-fuzzy visual servoing controller for an articulated manipulator," *IEEE Access*, vol. 6, pp. 3346–3357, 2018.

[5] S. Dong, A. H. Behzadan, F. Chen, and V. R. Kamat, "Collaborative visualization of engineering processes using tabletop augmented reality," *Adv. Eng. Softw.*, vol. 55, pp. 45–55, Jan. 2013.

[6] L. Minati, N. Yoshimura, and Y. Koike, "Hybrid control of a vision-guided robot arm by EOG, EMG, EEG biosignals and head movement acquired via a consumer-grade wearable device," *IEEE Access*, vol. 4, pp. 9528–9541, 2016.

[7] M. Zhou, X. Hao, A. Eslami, K. Huang, C. Cai, C. P. Lohmann, N. Navab, A. Knoll, and M. A. Nasseri, "6DOF needle pose estimation for robot-assisted vitreoretinal surgery," *IEEE Access*, vol. 7, pp. 63113–63122, 2019.

[8] V. Lepetit, L. Vacchetti, D. Thalmann, and P. Fua, "Fully automated and stable registration for augmented reality applications," in *Proc. 2nd IEEE ACM Int. Symp. Mixed Augmented Reality*, Oct. 2003, pp. 93–102.

[9] T. Collins, J.-D. Durou, P. Gurdjos, and A. Bartoli, "Singleview perspective shape-from-texture with focal length estimation: A piecewise affine approach," in *Proc. 5th Int. Symp. 3D Data Process., Vis. Transmiss.* Paris, France: Espace Saint Martin, May 2010.

[10] P. Sturm, "Algorithms for plane-based pose estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 2000, pp. 706–711.

[11] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.

[12] T. Collins and A. Bartoli, "Infinitesimal plane-based pose estimation," *Int. J. Comput. Vis.*, vol. 109, no. 3, pp. 252–286, Sep. 2014.

[13] M. A. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[14] C.-P. Lu, G. D. Hager, and E. Mjolsness, "Fast and globally convergent pose estimation from video images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 6, pp. 610–622, Jun. 2000.

[15] G. Schweighofer and A. Pinz, "Robust pose estimation from a planar target," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2024–2030, Dec. 2006.

[16] S. Li, C. Xu, and M. Xie, "A robust O(n) solution to the Perspective-n-Point problem," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1444–1450, Jul. 2012.

[17] F. Camposeco, T. Sattler, and M. Pollefeys, "Minimal solvers for generalized pose and scale estimation from two rays and one point," in *Proc. 14th Eur. Conf. Comput. Vis. (ECCV)*. Amsterdam, The Netherlands: Springer, Oct. 2016, pp. 202–218.

[18] D. Martinec and T. Pajdla, "Robust rotation and translation estimation in multiview reconstruction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[19] A. Collet and S. S. Srinivasa, "Efficient multi-view object recognition and full pose estimation," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2010, pp. 2050–2055.

[20] B. Clipp, J.-H. Kim, J.-M. Frahm, M. Pollefeys, and R. Hartley, "Robust 6DOF motion estimation for non-overlapping, multi-camera systems," in *Proc. IEEE Workshop Appl. Comput. Vis.*, Jan. 2008, pp. 1–8.

[21] D. Nister, "An efficient solution to the five-point relative pose problem," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2, Jun. 2003, p. 195.

[22] B. Clipp, C. Zach, J.-M. Frahm, and M. Pollefeys, "A new minimal solution to the relative pose of a calibrated stereo camera with small field of view overlap," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 1725–1732.

[23] A. Geiger, J. Ziegler, and C. Stiller, "StereoScan: Dense 3d reconstruction in real-time," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2011, pp. 963–968.

[24] I. Cvišić, J. Ćesić, I. Marković, and I. Petrović, "SOFT-SLAM: Computationally efficient stereo visual simultaneous localization and mapping for autonomous unmanned aerial vehicles," *J. Field Robot.*, vol. 35, no. 4, pp. 578–595, 2018.

[25] R. Mur-Artal and J. D. Tardos, "ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras," *IEEE Trans. Robot.*, vol. 33, no. 5, pp. 1255–1262, Oct. 2017.

[26] I. Ali, O. Suominen, A. Gotchev, and E. R. Morales, "Methods for simultaneous robot-world-hand–eye calibration: A comparative study," *Sensors*, vol. 19, no. 12, p. 2837, 2019.

[27] A. Tabb and K. M. A. Yousef, "Solving the robot-world hand-eye (s) calibration problem with iterative methods," *Mach. Vis. Appl.*, vol. 28, nos. 5–6, pp. 569–590, 2017.

[28] K. Koide and E. Menegatti, "General hand–eye calibration based on reprojection error minimization," *IEEE Robot. Automat. Lett.*, vol. 4, no. 2, pp. 1021–1028, Apr. 2019.

[29] T. Nöll, A. Pagani, and D. Stricker, "Markerless camera pose estimation-an overview," in *Proc. Vis. Large Unstructured Data Sets-Appl. Geospatial Planning, Model. Eng. (IRTG)*. Wadern, Germany: Schloss Dagstuhl-Leibniz-Zentrum füer Informatik, 2011, pp. 55–129.

[30] P. Vicente, L. Jamone, and A. Bernardino, "Towards markerless visual servoing of grasping tasks for humanoid robots," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2017, pp. 3811–3816.

[31] Y.-M. Wei, L. Kang, B. Yang, and L.-D. Wu, "Applications of structure from motion: A survey," *J. Zhejiang Univ. Sci. C*, vol. 14, no. 7, pp. 486–494, Jul. 2013.

[32] K. G. Derpanis, "Overview of the RANSAC algorithm," *Image Rochester NY*, vol. 4, no. 1, pp. 2–3, 2010.

[33] A. Kaehler and G. Bradski, *Learning OpenCV 3: Computer Vision in C++ With the OpenCV Library*. Sebastopol, CA, USA: O'Reilly Media, Inc, 2016.

IHTISHAM ALI received the B.Sc. degree in mechatronics engineering from the University of Engineering and Technology, Pakistan, in 2014, and the M.Sc. degree in automation engineering from Tampere University, Finland, in 2017. He is currently a Ph.D. Researcher with the 3D Media Group, Tampere University. He has worked on several industrial projects pertaining to machine automation using visual cues. His research interests include computer vision and robotics, specifically object pose estimation, 3-D reconstruction, and visual SLAM.

OLLI J. SUOMINEN received the B.Sc. and M.Sc.(Tech.) degrees in information technology, with a major in signal processing, from the Tampere University of Technology (TUT), in 2011 and 2012, respectively, where he is currently pursuing the Ph.D. degree with the Laboratory of Signal Processing, 3D Media Group. He also manages the construction and development of the Centre for Immersive Visual Technologies. After starting the B.Sc. thesis using only one camera (depth image-based rendering) and the M.Sc. thesis using two cameras (stereo depth estimation), he has scaled up to 40 cameras for the Ph.D. with research interests in multicamera systems, 3-D reconstruction, multimodal sensor fusion, SLAM, and light field capture. He focuses on applications in heavy mobile work machines, leading several industry driven research projects, and developing relations with the industry.

EMILIO RUIZ MORALES received the M.Sc. degree in electro-mechanical engineering and telecommunications from the École Polytechnique, Université Libre de Bruxelles, in 1990. He is currently the Project Manager for remote handling control systems of the ITER Project at the EU Fusion for Energy Agency. He has dedicated his career to the design and development of robotics control systems and advanced robotics applications in the fields of remote handling, nuclear, and surgical robotics.

ATANAS GOTCHEV received the M.Sc. degrees in radio and television engineering, in 1990, and in applied mathematics, in 1992, the Ph.D. degree in telecommunications from the Technical University of Sofia, in 1996, and the D.Sc.(Tech.) degree in information technologies from the Tampere University of Technology, in 2003. He is currently a Professor of Signal Processing and the Director of the Centre for Immersive Visual Technologies, Tampere University. His recent work concentrates on the algorithms for multisensor 3-D scene capture, transform-domain light-field reconstruction, and Fourier analysis of 3-D displays.