High Precision Positioning with Multi-Camera Setups: Adaptive Kalman Fusion Algorithm for Fiducial Markers

Dragos Constantin Popescu [1,2,*] **, Ioan Dumitrache** [1,3] **, Simona Iuliana Caramihai** [1] **and Mihail Octavian Cernaianu** [2,*]

[1] Faculty of Automatic Control and Computers, University Politehnica of Bucharest, Splaiul Independentei No. 313, 060042 Bucharest, Romania; ioan.dumitrache@acse.pub.ro (I.D.); simona.caramihai@aii.pub.ro (S.I.C.)

[2] Extreme Light Infrastructure-Nuclear Physics (ELI-NP)/Horia Hulubei National Institute for R&D in Physics and Nuclear Engineering (IFIN-HH), Str. Reactorului No. 30, Magurele, 077125 Bucharest, Romania

[3] Department of Science and Information Technology of the Romanian Academy, Cal. Victoriei No. 125, 010071 Bucharest, Romania

* Correspondence: dragos.popescu@acse.pub.ro (D.C.P.); mihail.cernaianu@eli-np.ro (M.O.C.)

check for
updates

Abstract: The paper addresses the problem of fusing the measurements from multiple cameras in order to estimate the position of fiducial markers. The objectives are to increase the precision and to extend the working area of the system. The proposed fusion method employs an adaptive Kalman algorithm which is used for calibrating the setup of cameras as well as for estimating the pose of the marker. Special measures are taken in order to mitigate the effect of the measurement noise. The proposed method is further tested in different scenarios using a Monte Carlo simulation, whose qualitative precision results are determined and compared. The solution is designed for specific positioning and alignment tasks in physics experiments, but also, has a degree of generality that makes it suitable for a wider range of applications.

Keywords: critical infrastructures; positioning system; optical measurements; fiducial markers; adaptive kalman; measurement fusion

1. Introduction

The evolution of technology has led to an increasing demand for solving complex problems which may be viewed as attempts to control and direct system behaviours towards desired states. The inherent complexity of problems and processes requires new approaches both in system modelling and in defining the emergent interaction with a highly dynamical and sparsely defined environment.

Cognitive approaches are successfully used in contexts where the boundary between the systems and the environment is fuzzy. However, they exhibit strong interrelation and interconnection, assisted by specific perception mechanisms [1]. Advances in complex control applications can be achieved only by considering adequate design approaches for sensory systems, especially in domains like environmental applications [2], health [3], industrial control, agriculture, etc.

Although new technologies such as wireless sensor networks or Internet of Things (IoT) are providing valuable solutions for appropriate perception mechanisms, complex issues are raised with the inclusion of data fusion, reliability, flexibility, reconfigurability, and cost of the measurement system. If some of these problems can be addressed during the modelling process, there are others (e.g., sensor positioning and sensor dynamics) that have a bigger impact on the generality of the overall solution.

These issues are specific in critical infrastructures, research institutes, power systems, e-health, mining, and traffic control, where the multitude of concurrent dynamics generating a large amount of information requires adequate solving methods that cover aspects such as relevance, efficiency, and cost optimisation.

This paper follows the development of a measurement method used in a very precise and yet flexible and portable positioning system. Firstly, the method is an answer to a specific application: Automatic alignment with high precision of various instruments which have to be remotely operated in highly flexible, open, and dynamic configurations for physics experiments. However, it can be used in any application where high precision positioning over a large working area is required, where no absolute reference can be defined, or when it is necessary to simultaneously align multiple features in relative positions, with high reliability. Such applications include large scale construction sites where multiple components need to be positioned in precise locations, manufacturing facilities where large machines are assembled or spacecraft docking in zero gravity environments [4].

The article has the following structure. The specific application and context addressed in this work are presented in Section 2. The theoretical context behind the proposed algorithm is described in Section 3. It it then used to build two essential procedures for the positioning system. The precision of the proposed method is assessed in the Section 4 using a Monte Carlo simulation. The conclusions and future developments are detailed in Section 5.

2. Description of the Initial Problem

The large number of high repetition rate, ultrashort pulse, and high power laser facilities that will come online all around the world will require state-of-the-art tools to allow the harnessing of their full potential [5]. The high power lasers of the ELI-NP (Extreme Light Infrastructure-Nuclear Physics) user facility will be employed for a wide range of research topics like studies of nuclear induced reactions, dark matter search, material irradiation, or medical applications [6]. The development of the experimental setups for studying these topics requires specific instrumentation, while also having strict needs in terms of positioning and alignment, in order to ensure optimal experimental conditions.

As the setups are continuously changing, absolute position referencing is hard to achieve. This is a necessity, as multiple instruments need to be precisely positioned relative to each other during the experiment. Figure 1 displays an example of a setup for solid target alignment in which multiple optical diagnostics are positioned using motorised manipulators that have 3 to 6 degrees of freedom (DOF).
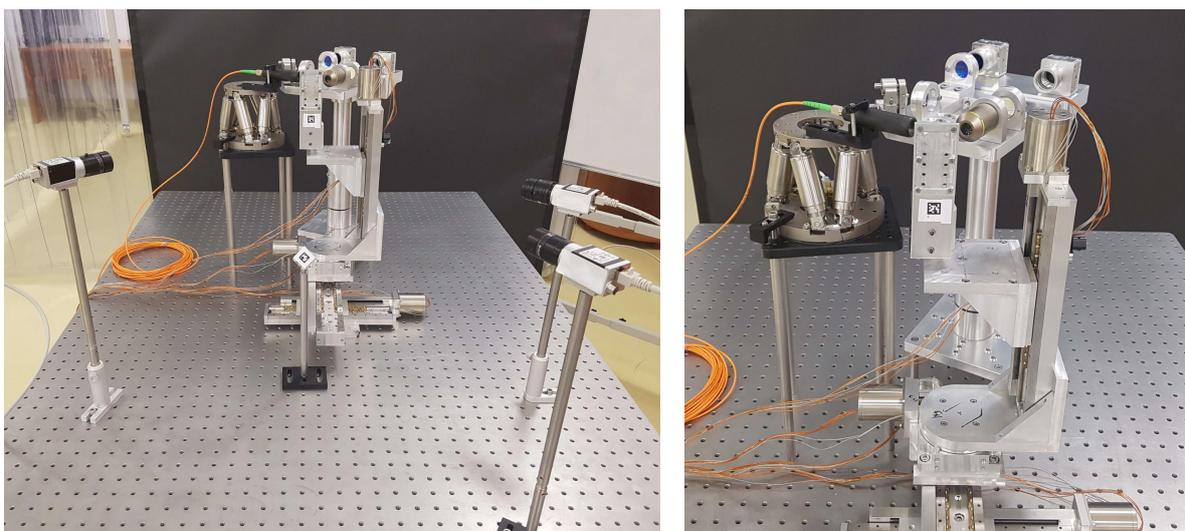

Figure 1. Solid target alignment setup example where multiple instruments are positioned using motorised manipulators.

Apart from the precision requirements, additional ones should be taken into consideration. The alignment should be done remotely because the experiments take place inside vacuum chambers and behind concrete walls used for radiation shielding. Moreover, to take advantage of the high repetition rate of the lasers and to maximise the beam-time available for the users, the positioning has to be performed in a limited amount time. In a previous work [7], we addressed this series of requirements and constraints by developing an automatic alignment system that is based on relative position measurements using imaging cameras and compact and flat fiducial markers, due to space and visibility constraints. The alignment algorithm was based on a real-time optimisation procedure which is the subject of a patent [8].

Although the fiducial markers were initially developed and used in augmented reality applications [9], due to their versatility in determining their position in a non-invasive manner (by imaging them with a camera), they were quickly adopted for a different range of applications. These include kinematic calibration and visual servoing for industrial robots [10,11], robot localisation and navigation [12–14], SLAM (simultaneous localisation and mapping) [15–17] and sensor fusion [18–20].

The motivation behind this work is to combine the measurements from multiple cameras in order to increase the working area of the system and to maximise the positioning precision. Simultaneously, the main focus is to present and test the precision of the proposed methods and algorithm, comparing the results with those recorded while using only one camera [21].

3. Method

The basic approach behind the proposed method is to use multiple cameras in order to improve the precision of estimating the pose of fiducial markers and to extend the working area of the system. It begins with the assumption that each camera provides measurements that are erroneous and noisy. The problem can be conceived through an analogy with the GPS system where distance measurements from multiple satellites are used to estimate the position of the receiver module with high precision. The key ingredient in the GPS system is to know very precisely the position of the satellites. The proposed positioning system needs to meet the same requirement for its cameras, but wasting too much effort on this task diminishes its practicality. Consequently it is only assumed that the cameras have unknown but fixed positions.

A related approach that has similar objectives can be found in [22]. The main benefit of the method is the possibility to calibrate the setup of cameras using a 3D feature with fiducial markers having unknown configuration. However, it is not meant for estimating the pose of single markers. In order to achieve this, additional methods are required.

Our method is developed using the ArUco fiducial markers [23,24]. The ArUco library detects and estimates the pose of the marker with respect to the camera using the solvepnp algorithm in which the pinhole camera model is data fitted using a Levenberg–Marquardt non-linear optimisation procedure. The output is represented by the extrinsic camera parameters which can be expressed in terms of a homogenous transformation T_C^M (the transformation between the camera attached reference frame and the marker attached one) detailed in [7].

3.1. Adaptive Kalman Fusion Algorithm for Multiple Cameras and Fiducial Markers

The method consists of fusing noisy measurements from multiple cameras, in order to improve the precision of estimating the exact position of the fiducial marker. In the scientific literature, multiple sensor fusion and noise filtering methods have been developed over the years. Among them there is the Kalman filter [25], which is the most widely used. The noise effect is mitigated by using a dynamical model for the physical process involved and a statistical model (covariance matrix) for the noise in the measurement process.

For a discrete linear state-space dynamical model in Equation (1), the Kalman filter estimates the value of the state vector x using noisy measurements for the input u and the output y.

$$\begin{cases} x_k = A \cdot x_{k-1} + B \cdot u_{k-1} + q_{k-1} \\ y_k = C \cdot x_k + r_k \end{cases} . \tag{1}$$

In standard data fusion applications that use the Kalman filter, the pose of various objects is estimated using the input from accelerometer, gyroscope, and magnetometer sensors [26,27] and the output from distance measurement devices [28,29]. Our application makes impossible to use any type of attached sensors and, hence, we only rely on the "non-invasive" pose measurement using the solvepnp algorithm and imaging cameras.

Our approach is to employ a state-space model with free dynamics (where u is zero) and with an identity state matrix (A). In this respect, each measurement is made using only one snapshot image from all the cameras that are synchronised with the help of an electrical trigger signal. In this way, the position of the fiducial marker during the measurement is considered fixed. On the same set of images, the solvepnp algorithm is applied multiple times and, thus, the evolution of the measurement is caused only by the noise and not by the movement of the fiducial marker. Furthermore, the measurements are used to iteratively estimate the real position using the proposed algorithm.

In order to improve the results, we propose a procedure designed to update the statistical model of the noise. The Q and R covariance matrices (which are discussed below) are continuously adjusted using the newly acquired data. Any change of the noise behaviour and any existing correlations are captured and thus, the effect of the noise is mitigated more efficiently. Consequently, the proposed algorithm is an adaptive Kalman version.

The homogeneous transformation representation (T_C^M) has numerous practical benefits, especially for pose composition and inversion operations, but it is not suitable in this circumstance because it is redundant (16 numerical values for expressing a 6 DOF pose). The solution is to use an equivalent representation which is composed of the set of translation coordinates (X, Y, and Z) and the set of Euler angles (AX, AY, and AZ).

In this particular case, the structure of the state vector is the real pose of the fiducial marker defined in Equation (2), where the subscript k denotes the present discrete-time sample of the state vector.

$$x_k = \begin{bmatrix} X_k \\ Y_k \\ Z_k \\ AX_k \\ AY_k \\ AZ_k \end{bmatrix} . \tag{2}$$

The output vector is composed of the pose elements measured using all the cameras available. The structure of the output vector is defined in Equation (3), where the superscript i is indicating the index of the camera considered.

$$y_k = \begin{bmatrix} X_k^1 \\ Y_k^1 \\ Z_k^1 \\ AX_k^1 \\ AY_k^1 \\ AZ_k^1 \\ \vdots \end{bmatrix} \begin{bmatrix} \vdots \\ X_k^i \\ Y_k^i \\ Z_k^i \\ AX_k^i \\ AY_k^i \\ AZ_k^i \\ \vdots \end{bmatrix} . \tag{3}$$

The discrete-time state-space model considered is defined in Equation (4), where I_6 is the identity matrix of rank 6, q_{k-1} is a random noise signal with normal distribution (white noise) that is quantifying the false evolution induced by the noisy measurement, the H is the measurement model matrix defined in Equation (5), and r_k is the measurement noise also considered white noise.

$$\begin{cases} x_k = I_6 \cdot x_{k-1} + q_{k-1} \\ y_k = H \cdot x_k + r_k \end{cases} \tag{4}$$

$$H = \begin{bmatrix} I_6^1 \\ \vdots \\ I_6^i \\ \vdots \end{bmatrix}. \tag{5}$$

The equations that describe the Kalman filter are presented in Equation (6). The first two equations give a rough state estimate using the dynamical model while the last 5 equations are used for improving the estimate using the newly acquired output sample. P is a covariance matrix that expresses the confidence degree of the state estimation, which is updated during the algorithm iterations, Q and R are the covariance matrices associated with the noises q and r respectively, v_k is the rough estimation error (difference between the measured output and the predicted one using the first two equations), S_k is the covariance matrix associated with the predicted output, and K_k is the Kalman state update.

Predict the state

1. $\hat{x}_{k|k-1} = A \cdot \hat{x}_{k-1|k-1}$
2. $P_{k|k-1} = A \cdot P_{k-1|k-1} \cdot A^T + Q_{k-1}$

Update the prediction

3. $v_k = y_k - H_k \cdot \hat{x}_{k|k-1}$
4. $S_k = H_k \cdot P_{k|k-1} \cdot H_k^T + R_k$
5. $K_k = P_{k|k-1} \cdot H_k^T \cdot S_k^{-1}$
6. $\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k \cdot v_k$
7. $P_{k|k} = P_{k|k-1} - K_k \cdot S_k \cdot K_k^T$

$$\tag{6}$$

The filter requires good estimates for the initial state x_0 and the covariance matrices Q and R. In the proposed algorithm this is achieved using an initialisation procedure. The position of the fiducial marker is measured for N_i number of sampling times. The result is the set of samples defined in Equation (3) for $k = 1, \ldots, N_i$ (the length of the initialisation) and $i = 1, \ldots, n$ (the number of cameras available).

Averaging the samples gives a good estimate for the initial state, which is built according to Equation (7), where $E[\cdot]$ is the mean operator (expected value). The Q and R matrices are computed from the same set of samples, assuming that there is no correlation between the noises affecting the measurements.

$$x_0 = \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \\ AX_0 \\ AY_0 \\ AZ_0 \end{bmatrix} = \begin{bmatrix} \underset{\substack{1 \leq k \leq N_i \\ 1 \leq i \leq n}}{E} \left[X_k^i \right] \\ \underset{\substack{1 \leq k \leq N_i \\ 1 \leq i \leq n}}{E} \left[Y_k^i \right] \\ \underset{\substack{1 \leq k \leq N_i \\ 1 \leq i \leq n}}{E} \left[Z_k^i \right] \\ \underset{\substack{1 \leq k \leq N_i \\ 1 \leq i \leq n}}{E} \left[AX_k^i \right] \\ \underset{\substack{1 \leq k \leq N_i \\ 1 \leq i \leq n}}{E} \left[AY_k^i \right] \\ \underset{\substack{1 \leq k \leq N_i \\ 1 \leq i \leq n}}{E} \left[AZ_k^i \right] \end{bmatrix} \tag{7}$$

$$Q_0 = diag\left(\underset{1 \leq i \leq n}{E} \left[\underset{1 \leq k \leq N_i}{var} \left(X_k^i \right) \right], \underset{1 \leq i \leq n}{E} \left[\underset{1 \leq k \leq N_i}{var} \left(Y_k^i \right) \right], \underset{1 \leq i \leq n}{E} \left[\underset{1 \leq k \leq N_i}{var} \left(Z_k^i \right) \right], \right.$$
$$\left. \underset{1 \leq i \leq n}{E} \left[\underset{1 \leq k \leq N_i}{var} \left(AX_k^i \right) \right], \underset{1 \leq i \leq n}{E} \left[\underset{1 \leq k \leq N_i}{var} \left(AY_k^i \right) \right], \underset{1 \leq i \leq n}{E} \left[\underset{1 \leq k \leq N_i}{var} \left(AZ_k^i \right) \right] \right) \tag{8}$$

$$R_0 = diag\left(R_0^1, \ldots, R_0^i, \ldots, R_0^n \right) \tag{9}$$

$$R_0^i = diag\left(\underset{1 \leq k \leq N_i}{var} \left(X_k^i \right), \underset{1 \leq k \leq N_i}{var} \left(Y_k^i \right), \underset{1 \leq k \leq N_i}{var} \left(Z_k^i \right), \right.$$
$$\left. \underset{1 \leq k \leq N_i}{var} \left(AX_k^i \right), \underset{1 \leq k \leq N_i}{var} \left(AY_k^i \right), \underset{1 \leq k \leq N_i}{var} \left(AZ_k^i \right) \right). \tag{10}$$

Q is a diagonal matrix defined in Equation (8) built using the mean variance of the pose elements (X, Y, Z, AX, AY, and AZ) along all the cameras. The R matrix is defined in Equation (9) where R^i is defined in Equation (10), built using the variance for each pose element measured by each camera.

After the initialisation is finished, the Kalman algorithm is iterated for N_e number of sampling times, while at each step, a new set of samples y_k in the form of Equation (3) is measured and an estimated state trajectory is built ($\hat{x}_{k|k}$ for $k = N_i + 1, \ldots, N_i + N_e$).

Newly measured system outputs can be used to improve the statistical models of the noises for increased performance. Thereby, at each iteration, every new set is added to the initialisation set and the covariance matrices Q and R are updated using Equations (11)–(13) for $k = 1, \ldots, N_e$.

$$Q_k = diag\left(\underset{1 \leq i \leq n}{E} \left[\underset{1 \leq j \leq N_i+k}{var} \left(X_j^i \right) \right], \underset{1 \leq i \leq n}{E} \left[\underset{1 \leq j \leq N_i+k}{var} \left(Y_j^i \right) \right], \underset{1 \leq i \leq n}{E} \left[\underset{1 \leq j \leq N_i+k}{var} \left(Z_j^i \right) \right], \right.$$
$$\left. \underset{1 \leq i \leq n}{E} \left[\underset{1 \leq j \leq N_i+k}{var} \left(AX_j^i \right) \right], \underset{1 \leq i \leq n}{E} \left[\underset{1 \leq j \leq N_i+k}{var} \left(AY_j^i \right) \right], \underset{1 \leq i \leq n}{E} \left[\underset{1 \leq j \leq N_i+k}{var} \left(AZ_j^i \right) \right] \right) \tag{11}$$

$$R_k = diag\left(R_k^1, \ldots, R_k^i, \ldots, R_k^n \right) \tag{12}$$

$$R_k^i = diag\left(\underset{1 \leq j \leq N_i+k}{var} \left(X_j^i \right), \underset{1 \leq j \leq N_i+k}{var} \left(Y_j^i \right), \underset{1 \leq j \leq N_i+k}{var} \left(Z_j^i \right), \right.$$
$$\left. \underset{1 \leq j \leq N_i+k}{var} \left(AX_j^i \right), \underset{1 \leq j \leq N_i+k}{var} \left(AY_j^i \right), \underset{1 \leq j \leq N_i+k}{var} \left(AZ_j^i \right) \right). \tag{13}$$

Considering that the estimated state trajectory ($\hat{x}_{k|k}$) belongs to a system with free dynamics where, in the absence of the noise effects, the state should be constant, a final estimate with a better precision can be achieved by averaging the values of the estimated state trajectory according to Equation (14). The timeline of all the procedures that are involved in the proposed algorithm is presented in Figure 2. In Algorithm 1 the proposed procedure is summarised.

$$\hat{x} = \mathop{E}_{N_i+1 \leq k \leq N_i+N_e} \left[\hat{x}_{k|k} \right] \tag{14}$$
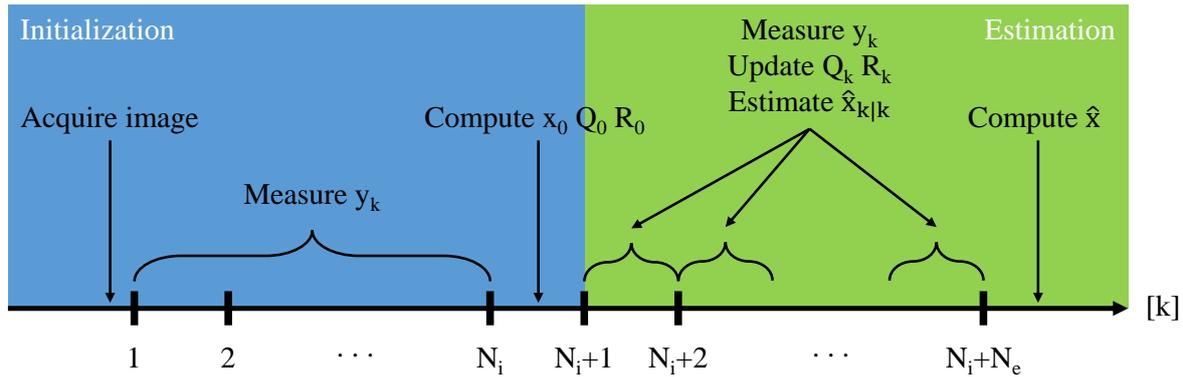
Figure 2. The sequence of procedures involved in the proposed algorithm.

Algorithm 1: The proposed algorithm for multi-camera pose fusion.

---

1. Acquire one image from all *n* available cameras;
2. Apply solvepnp algorithm on each image for N_i times and build the initialisation set from Equation (3), where $k = 1, \ldots, N_i$;
3. Compute x_0 using Equation (7);
4. Compute Q_0 using Equation (8);
5. Compute R_0 using Equations (9) and (10);
6. **for** N_e *times* **do**
7.   Compute $\hat{x}_{k|k-1}$ and $P_{k|k-1}$ using Equations (6).1 and (6).2 respectively;
8.   Apply solvepnp algorithm on each image and append a new set from Equation (3) to the initialisation set;
9.   Update Q using Equation (11);
10.   Update R using Equations (12) and (13);
11.   Compute $\hat{x}_{k|k}$ and $P_{k|k}$ using Equations (6).3–7;
12.   Add $\hat{x}_{k|k}$ to the estimated state trajectory;
13. Compute $\hat{x}$ using Equation (14);

3.2. Setup Calibration Procedure

Given a set of *n* cameras in a pre-defined fixed configuration, having unknown absolute positions, the purpose of the setup calibration procedure is to determine, in a precise manner, their relative positions. As depicted in Figure 3, the goal is to determine the set of homogenous transformations T_{Ci}^{Cj} where $i = 1, \ldots, n$ and $j = i+1, \ldots, n$.
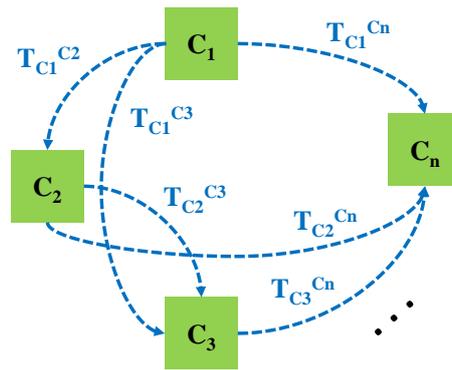
Figure 3. The calibration of a set of *n* cameras requires to determine their relative positions T_{Ci}^{Cj}.

This requires the use of a precision gauge, which can be manufactured in the form of a cube of fiducial markers like in [21] or in any 3D shape where the markers can be viewed all around. The gauge needs to be manufactured or calibrated with increased precision. It can be seen as the absolute precision reference used to calibrate the entire positioning system. In Figure 4, a conceptual diagram of a gauge containing *m* fiducial markers is presented. The set of homogenous transformations between all the markers (T_{Mi}^{Mj} where $i = 1, \ldots, m$ and $j = i + 1, \ldots, m$) is precisely known.
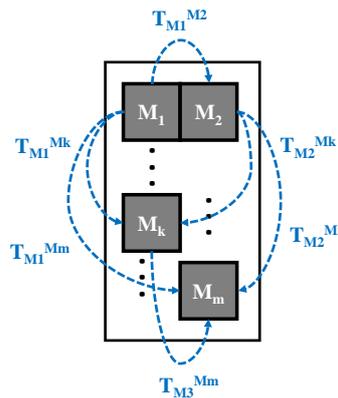
Figure 4. A setup of cameras is calibrated using a precision gauge. It is a manufactured feature having *m* fiducial markers where all their relative positions are precisely known.

The camera setup calibration procedure is done simultaneously for all the camera pairs C_i, C_j where $i = 1, \ldots, n$ and $j = i + 1, \ldots, n$, with the goal of estimating the homogenous transformation T_{Ci}^{Cj} (as presented in Figure 5). It starts with placing the precision gauge inside the environment. Depending on the orientation, each camera sees a different portion of the gauge, i.e., from all *n* fiducial markers, C_i can measure the position of n_i markers and C_j can measure only n_j, where $n_i \leq n$ and $n_j \leq n$. It is preferable that at least one fiducial marker M_k is seen by both cameras, otherwise, it can be virtually determined using the gauge transformations.

The proposed algorithm estimates the T_{Ci}^{Mk} homogenous transformation using the noisy measurements from all n_i markers using the C_i camera. This transformation is expressed multiple times in terms of T_{Ci}^{Ml} pose measurement (where $l = 1, \ldots, n_i$) using Equation (15). The T_{Ml}^{Mk} transformation is precisely known from the gauge.

$$T_{Ci}^{Mk} \approx T_{Ci}^{\hat{M}k}{}_l = T_{Ci}^{Ml} \cdot T_{Ml}^{Mk} \tag{15}$$

The resulting $T_{Ci}^{\hat{M}k}{}_l$ pose is converted to translations and Euler angles which are used for building the output measurement vector of Equation (3).
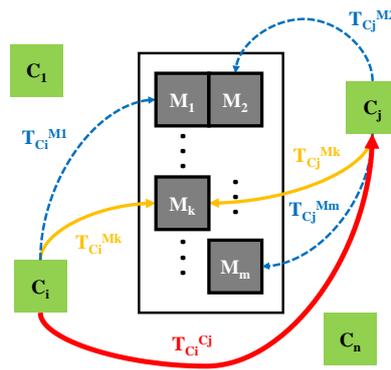
Figure 5. The process of calibrating all the pairs of n cameras, which illustrates all the homogenous transformations involved in finding the relative position of C_i and C_j cameras.

The proposed algorithm is further applied and the result is an estimation of T_{Ci}^{Mk} having a higher degree of precision and being closer to the real value. For the C_j camera, T_{Cj}^{Mk} is estimated in a similar manner. Consequently, T_{Ci}^{Cj} transformation is computed using Equation (16).

$$T_{Ci}^{Cj} = T_{Ci}^{Mk} \cdot \left(T_{Mk}^{Cj} \right)^{-1} \tag{16}$$

3.3. Position Estimation Procedure

Having a calibrated camera setup, the aim of this procedure is to estimate the position of a fiducial marker attached to a specific instrument, according to the application.

In Figure 6, the conceptual diagram of this procedure is depicted. Depending on the configuration, the M marker can be seen only by a number of n_m out of n cameras ($n_m \le n$). An arbitrary camera (C_r) is chosen, which is considered the positioning reference.
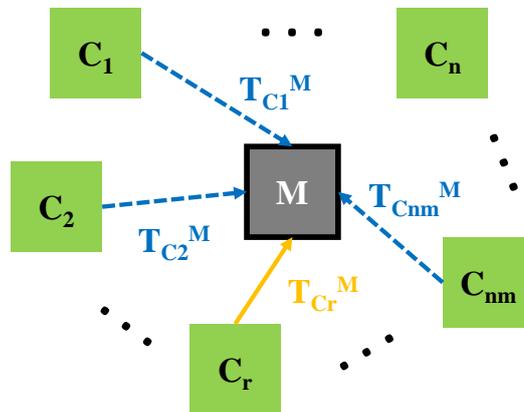
Figure 6. The homogenous transformations involved in estimating the position of a fiducial marker placed in the environment of a calibrated camera setup.

The proposed algorithm is estimating the T_{Cr}^{M} homogenous transformation using the noisy measurements from each of the n_m cameras. In a similar manner to the setup calibration procedure, the T_{Cr}^{M} transformation is expressed multiple times in terms of T_{Ck}^{M} pose measurement (where $k = 1, \ldots, n_m$) using Equation (17). The T_{Cr}^{Ck} is precisely known from the setup calibration.

$$T_{Cr}^{M} \approx \hat{T}_{Crk}^{M} = T_{Cr}^{Ck} \cdot T_{Ck}^{M}. \tag{17}$$

The resulting $T_{Crk}^{\hat{M}}$ pose is converted to translations and Euler angles and used for building the output measurement vector of Equation (3). The proposed algorithm is further applied and the result is an estimation of T_{Cr}^{M} which has a higher degree of precision and is closer to the real value.

This approach can be used in a similar manner for measuring the relative position of multiple fiducial markers in the environment, which is very useful in alignment tasks as presented in [7].

4. Simulation Results

4.1. Monte Carlo Setup

There are two factors that contribute and affect the precision of estimating the pose of the fiducial marker. First, there are the physical and environment-related aspects, which include the optical specifications of the imaging system (the sensor and optical resolution, the focal length, the depth of field, and the field of view), the environment illumination conditions (how strong, uniform, and consistent the lighting is) and if the cameras are optimally positioned so as to achieve good viewing angles and maintain consistent accuracy along the working area. Secondly, there are factors related to the algorithms regarding how precise the pose can be estimated and how much the effect of the noise can be mitigated during the data fusion. In this paper we decouple the two types of factors and only consider the behaviour of our proposed method, so as to have a first qualitative assessment regarding the precision.

Consequently, we developed a simulation environment that aims to replicate how our method performs in a real setup, considering the noisy pose estimations it receives from the solvepnp algorithm. For an increased confidence in the results, we adopted a Monte Carlo approach in which we statistically analysed how the noise is propagated through our method and how the precision is affected. The simulation is implemented in MATLAB where positions of multiple cameras, precision gauges, and fiducial markers are virtually defined. In order to simulate the noisy input from the solvepnp, each position (that is supposed to be measured in the real environment) is disturbed with additive random noise having normal distribution. The noise is configured considering the precision limits we determined experimentally for one camera in our previous work [21]: for X and Y, 75 μm, for Z, 300 μm, and for AX, AY and AZ, 0.02°. The mean value of the noise is 0 while the standard deviation (σ) was configured in such a way that 95.45% of the noise values are within the experimental limits (inside $[-2\sigma, 2\sigma]$).

The simulation puts in place three scenarios which are additionally used to assess the different contributions between the number of cameras and the number of the fiducial markers from the gauge:

1. 3 cameras, a precision gauge with 5 markers, and one marker whose position must be estimated;
2. 5 cameras, a precision gauge with 3 markers, and one marker whose position must be estimated;
3. 5 cameras, a precision gauge with 5 markers, and one marker whose position must be estimated.

The results are compared against estimating the pose using only one camera in the same environment. The Monte Carlo simulation performs 5000 iterations where, in each run and for each of the scenarios, the setup is calibrated using the gauge. The calibration is further used to estimate the pose of the fiducial marker which is compared with the true, predefined one. The N_i and N_e parameters are configured to 20 and 50 respectively. In a real application, the choice of these two parameters is a matter of cost optimisation, considering the computational effort, the resources available, and the required measuring frequency.

Compared with a real setup, we adopted one simplifying hypothesis. The angle of incidence of the marker relative to the camera, as we showed in [21], has an influence on the precision. Close to normal angles of incidence tend to bring more noise in the estimation. In this study, we only consider that the precision of the solvepnp algorithm to be consistent, regardless of the angle. However, in other respects, the simulation is considering the worst case scenario because of the following arguments:

- Noises from different cameras and from different elements of the pose (*X*, *Y*, *Z*, *AX*, *AY*, and *AZ*) are considered completely uncorrelated. In real circumstances this might not be the case (e.g., the noise induced by the environment illumination which affects all the measurements in a similar manner) thus, any relaxed conditions are contributing to an increased precision of the estimation. This additional information is harnessed using the update procedure for *Q* and *R* covariance matrices (which in this case would no longer be diagonal);
- The precision of one camera measurement along the *Z* axis is 5 times lower compared with the *X* and *Y* axis. In the simulation this is taken as it is, but in a real situation, this effect can be diminished by placing the camera setup in an optimal configuration. For instance, the measurement along the *Z* axis from one camera can be replaced by measurements from two cameras placed in lateral positions, for which the *Z* axis motion is decomposed in *X* and *Y* components that have a higher precision.

4.2. Results

Figures 7–12 presents the simulation results for estimating *X*, *Y*, and *Z* coordinates and *AX*, *AY*, and *AZ* orientation angles, which are given in terms of the probability density function (PDF) and the standard deviation (SD) of the error. The first plot from each Figure depicts the estimation error achieved using only one camera. The following three plots give the results achieved using the setups from each of the above-mentioned scenarios.

The results show that the proposed algorithm achieves an increase in precision which is close to an order of magnitude. It can also be observed that it is of greater importance to have more fiducial markers in the precision gauge instead of more cameras. In scenario #3, a slight decrease of precision is experienced in comparison to the #1 scenario. This is to be expected as each added camera is an additional noise source. However, the benefit of achieving a larger working area is far more important. In Table 1 the results are summarised and compared with regards to the limits of the variation interval $[-2\sigma, 2\sigma]$ where it is expected that 95.45% of the errors will occur. This can be considered the precision that the positioning system is expected to achieve when using th proposed method.
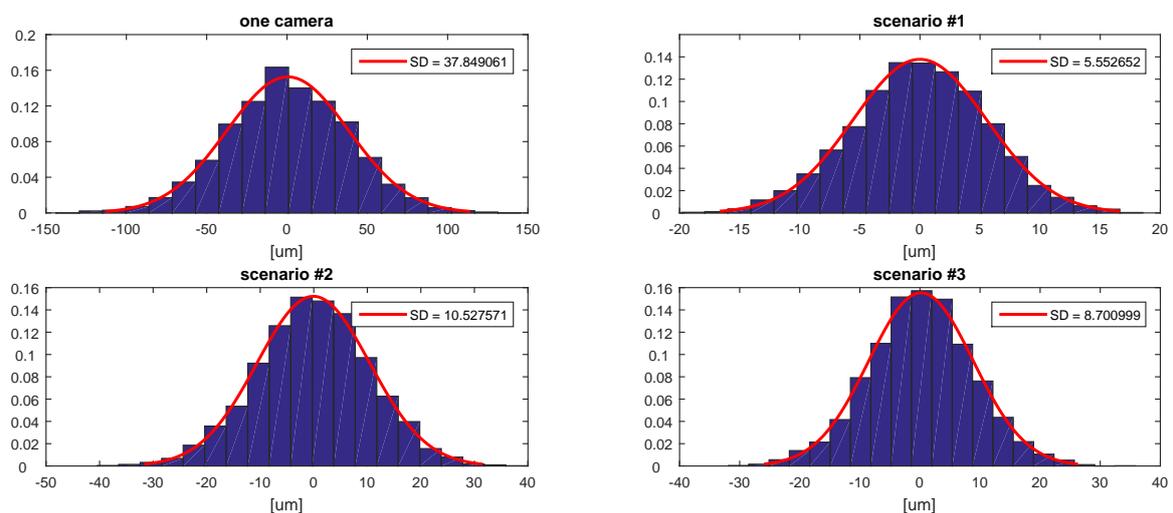

Figure 7. The probability density function (PDF) and the standard deviation (SD) of the error for estimating the *X* coordinate.
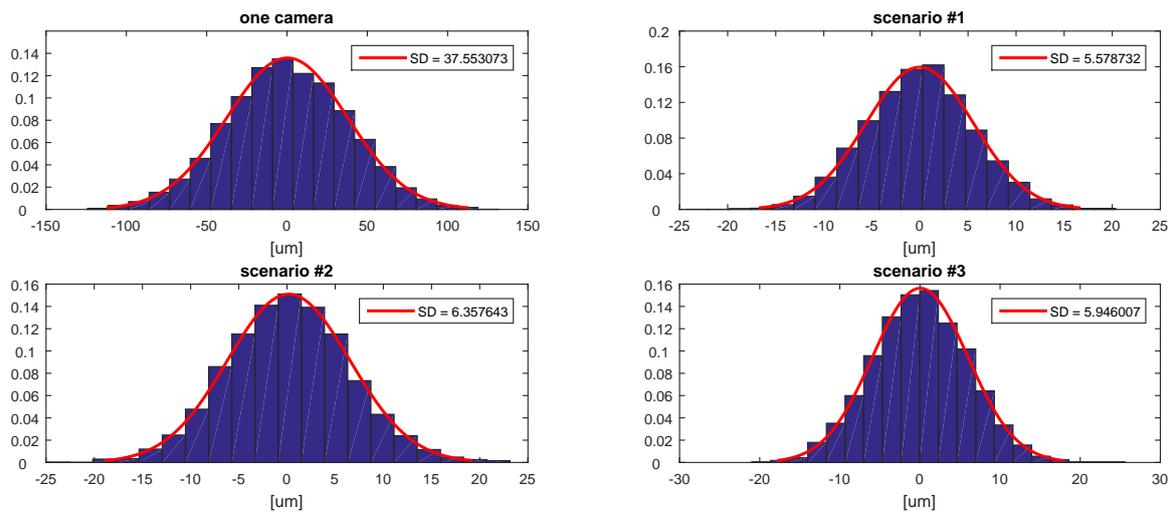
Figure 8. The probability density function (PDF) and the standard deviation (SD) of the error for estimating the *Y* coordinate.
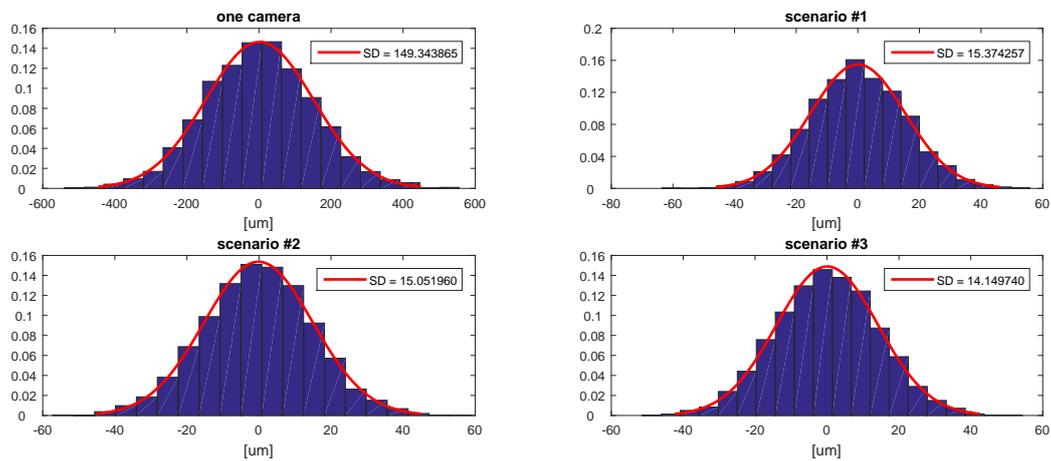

Figure 9. The probability density function (PDF) and the standard deviation (SD) of the error for estimating the *Z* coordinate.
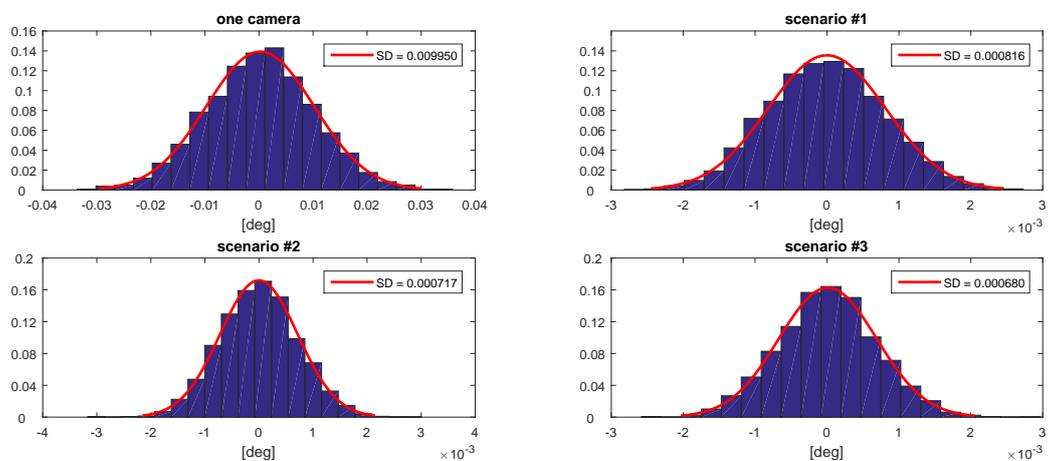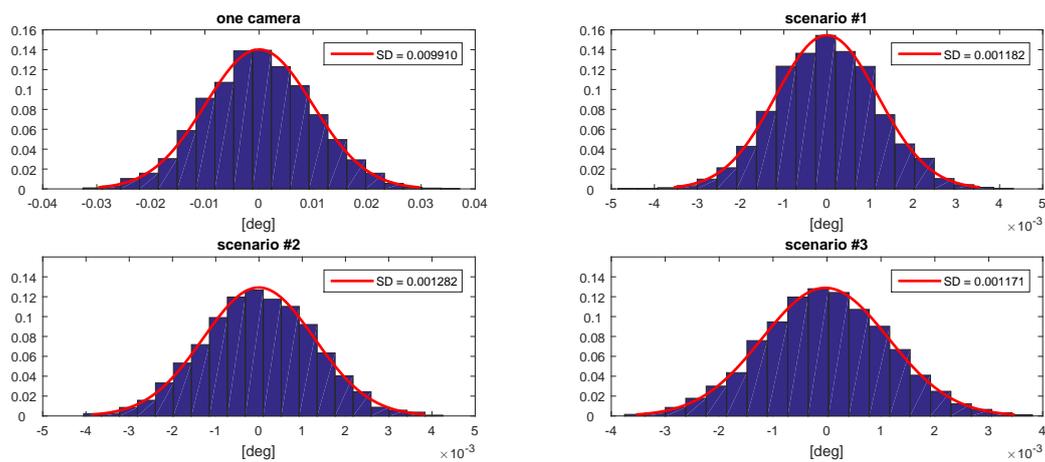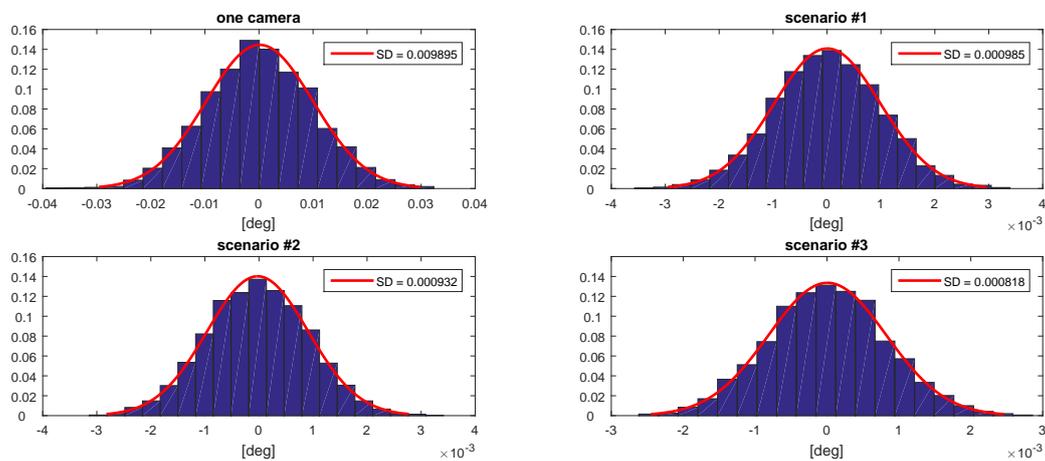

Figure 10. The probability density function (PDF) and the standard deviation (SD) of the error for estimating the *AX* angle.

Figure 11. The probability density function (PDF) and the standard deviation (SD) of the error for estimating the *AY* angle.

Figure 12. The probability density function (PDF) and the standard deviation (SD) of the error for estimating the *AZ* angle.

Table 1. The simulation results for estimating the pose using only one camera and multiple cameras in three scenarios. The value is the boundary of the error variation interval (2σ).

Pose Element	UM	One Camera	Scenario #1	#2	#3
X	(μm)	75.68	11.1	21.04	17.4
Y	(μm)	75.1	11.14	12.7	11.88
Z	(μm)	298.68	30.74	30.01	28.28
AX	(deg)	0.019	0.0016	0.0014	0.0013
AY	(deg)	0.019	0.0022	0.0024	0.0022
AZ	(deg)	0.019	0.0019	0.0018	0.0016

In order to further emphasise the simulation results achieved by the proposed method, Figures 13–18 depict a set of examples from the 3rd scenario which show how the estimation of the state vector elements is evolving. Each of the figures presents: The noisy measurement from the five cameras, the value of the state vector computed after the initialisation procedure, the evolution of the state vector during the Kalman estimation, and the final state estimation which falls close to the true value. Although the noise in the measurements is amplified because the reconstruction of the

position of the marker in different cameras, the evolution of the Kalman estimation shows a clear noise damping effect.
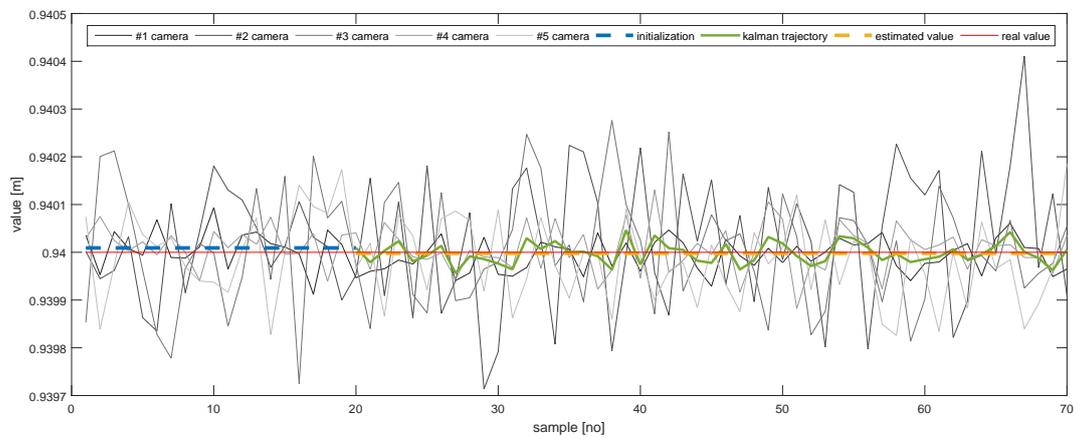

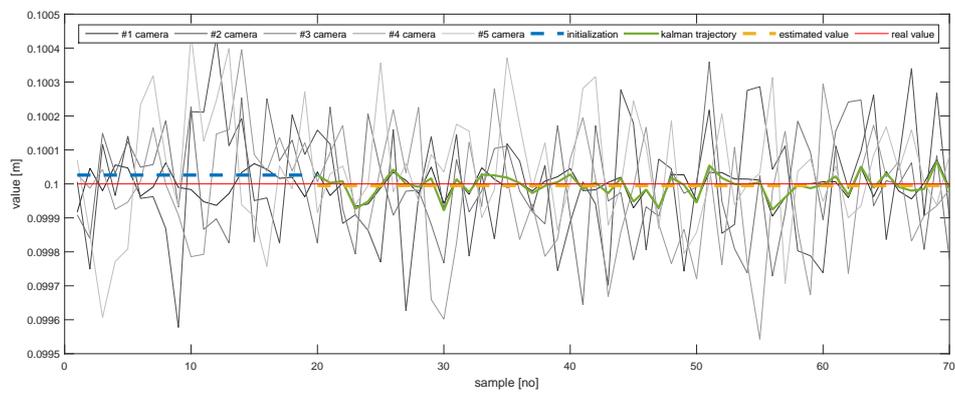**Figure 13.** The evolution of the estimation of X.

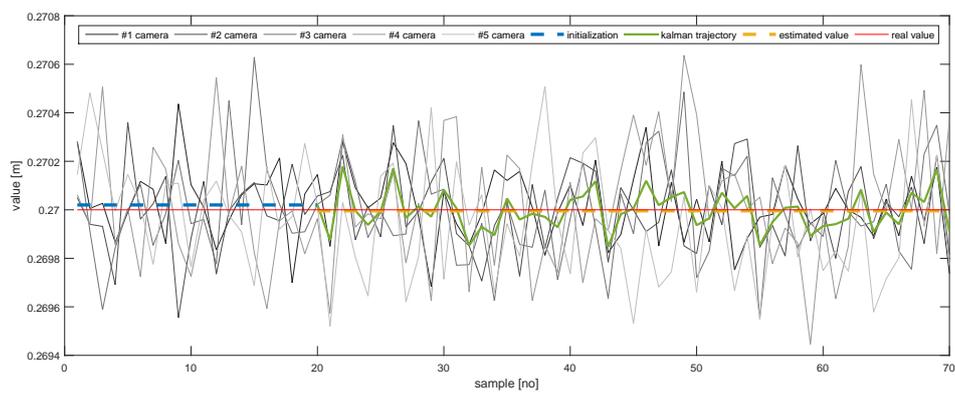**Figure 14.** The evolution of the estimation of Y.

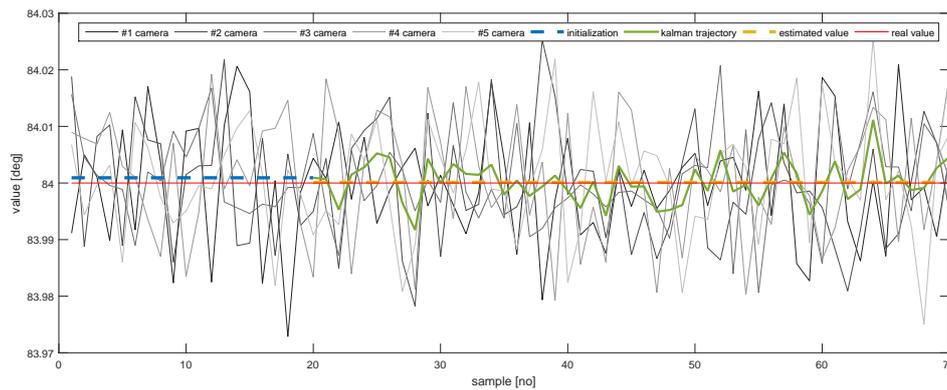**Figure 15.** The evolution of the estimation of Z.
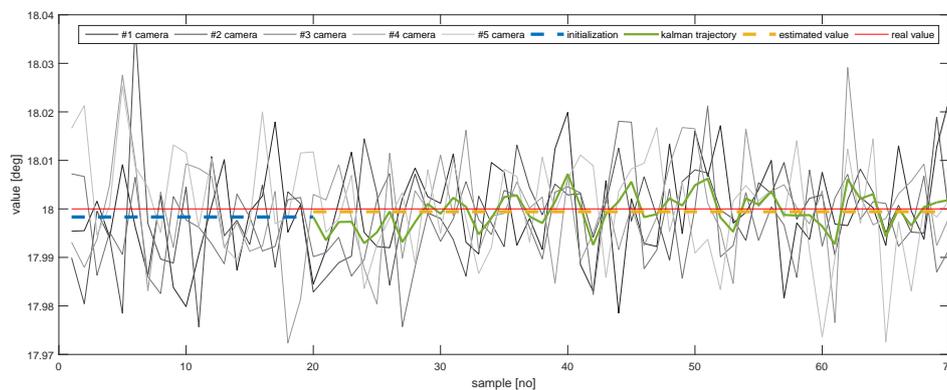
Figure 16. The evolution of the estimation of *AX*.

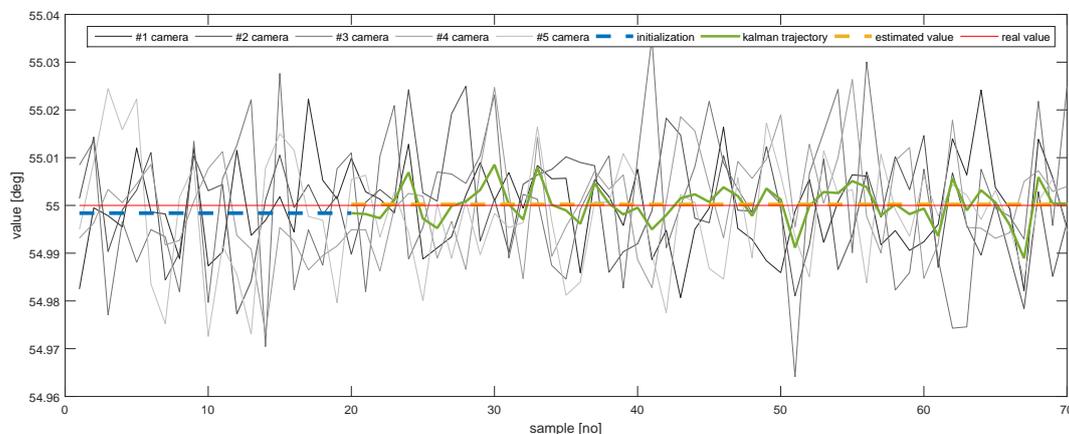**Figure 17.** The evolution of the estimation of *AY*.

Figure 18. The evolution of the estimation of *AZ*.

5. Conclusions

With respect to the considered case study, the results show that the proposed method and algorithm have managed to successfully meet the objectives. The working area could be increased in accordance with the number of cameras in the setup. This is a decision-making procedure that needs to take into account the cost relative to the working area and the precision required. For our simulation scenarios, the precision increase was close to an order of magnitude, which was around 10–15 μm for *X* and *Y* coordinates, 30 μm for *Z* and 0.002° for *AX*, *AY*, and *AZ* orientation angles.

The cost is an important benefit of the system compared to other solutions like laser trackers which tend to be extremely expensive. In addition to that, our proposed method could achieve relative and simultaneous positioning of multiple fiducial markers, which supports the development of advanced applications.

Future work will include a complete analysis of the method in a real environment where all physical and algorithm-related factors are considered, and a precision comparison against other methods presented in the literature.

References

1. Albus, J.S.; Barbera, A.J. RCS: A cognitive architecture for intelligent multi-agent systems. *IFAC Proc. Vol.* **2004**, *37*, 1–11. [CrossRef]

2. Prasanna, S.; Rao, S. An overview of wireless sensor networks applications and security. *Int. J. Soft Comput. Eng.* **2012**, *2*.

3. Apolle, R.; Brückner, S.; Frosch, S.; Rehm, M.; Thiele, J.; Valentini, C.; Lohaus, F.; Babatz, J.; Aust, D.E.; Hampe, J.; et al. Utility of fiducial markers for target positioning in proton radiotherapy of oesophageal carcinoma. *Radiother. Oncol.* **2019**, *133*, 28–34, doi:10.1016/j.radonc.2018.12.016. [CrossRef]

4. Fiala, M. Artag fiducial marker system applied to vision based spacecraft docking. In Proceedings of the International Conference on Intelligent Robots and Systems (IROS) 2005 Workshop on Robot Vision for Space Applications, Edmonton, AB, Canada, 2 August 2005; pp. 35–40.

5. Gales, S.; Tanaka, K.A.; Balabanski, D.L.; Negoita, F.; Stutman, D.; Tesileanu, O.; Ur, C.A.; Ursescu, D.; Andrei, I.; Ataman, S.; et al. The extreme light infrastructure-nuclear physics (ELI-NP) facility: New horizons in physics with 10 PW ultra-intense lasers and 20 MeV brilliant gamma beams. *Rep. Prog. Phys. Phys. Soc.* **2018**, *81*, 094301. [CrossRef] [PubMed]

6. Tanaka, K.A.; Spohr, K.M.; Balabanski, D.L.; Balascuta, S.; Capponi, L.; Cernaianu, M.O.; Cuciuc, M.; Cucoanes, A.; Dancus, I.; Dhal, A.; et al. Current status and highlights of the ELI-NP research program. *Matter Radiat. Extrem.* **2020**, *5*, 024402, doi:10.1063/1.5093535. [CrossRef]

7. Popescu, D.C.; Cernaianu, M.O.; Dumitrache, I. Automatic rough alignment for key components in laser driven experiments using fiducial markers. *J. Phys. Conf. Ser.* **2018**, *1079*, 012013, doi:10.1088/1742-6596/1079/1/012013. [CrossRef]

8. Popescu, D.C.; Cernăianu, M.O. *Method and System for Automatic Positioning of Elements for Configurations of Experiments with High-Power Lasers*; Romanian patent RO132327B1; Library Catalog: ESpacenet; OSIM: Bucharest, Romania, 2019.

9. Fiala, M. ARTag, a fiducial marker system using digital techniques. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; Volume 2, pp. 590–596, ISSN: 1063-6919, doi:10.1109/CVPR.2005.74. [CrossRef]

10. Boby, R.A.; Saha, S.K. Single image based camera calibration and pose estimation of the end-effector of a robot. In Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA), Stockholm, Sweden, 16–21 May 2016; pp. 2435–2440. [CrossRef]

11. Cai, C.; Dean-León, E.; Somani, N.; Knoll, A. 6D image-based visual servoing for robot manipulators with uncalibrated stereo cameras. In Proceedings of the 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, Chicago, IL, USA, 14–18 September 2014; pp. 736–742, ISSN: 2153-0866, doi:10.1109/IROS.2014.6942640. [CrossRef]

12. Chen, D.; Peng, Z.; Ling, X. A low-cost localization system based on artificial landmarks with two degree of freedom platform camera. In Proceedings of the 2014 IEEE International Conference on Robotics and Biomimetics (ROBIO 2014), Bali, Indonesia, 5–10 December 2014; pp. 625–630. [CrossRef]

13. Babinec, A.; Jurišica, L.; Hubinský, P.; Duchoň, F. Visual Localization of Mobile Robot Using Artificial Markers. *Procedia Eng.* **2014**, *96*, 1–9. [CrossRef]

14. Bi, S.; Yang, D.; Cai, Y. Automatic Calibration of Odometry and Robot Extrinsic Parameters Using Multi-Composite-Targets for a Differential-Drive Robot with a Camera. *Sensors* **2018**, *18*, 3097, doi:10.3390/s18093097. [CrossRef]

15. Houben, S.; Droeschel, D.; Behnke, S. Joint 3D laser and visual fiducial marker based SLAM for a micro aerial vehicle. In Proceedings of the 2016 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI), Baden-Baden, Germany, 19–21 September 2016; pp. 609–614. [CrossRef]

16. Muñoz-Salinas, R.; Marín-Jimenez, M.J.; Medina-Carnicer, R. SPM-SLAM: Simultaneous localization and mapping with squared planar markers. *Pattern Recognit.* **2019**, *86*, 156–171. [CrossRef]

17. Muñoz-Salinas, R.; Medina-Carnicer, R. UcoSLAM: Simultaneous localization and mapping by fusion of keypoints and squared planar markers. *Pattern Recognit.* **2020**, *101*, 107193. [CrossRef]

18. Xing, B.; Zhu, Q.; Pan, F.; Feng, X. Marker-Based Multi-Sensor Fusion Indoor Localization System for Micro Air Vehicles. *Sensors* **2018**, *18*, 1706, doi:10.3390/s18061706. [CrossRef] [PubMed]

19. Poulose, A.; Han, D.S. Hybrid Indoor Localization Using IMU Sensors and Smartphone Camera. *Sensors* **2019**, *19*, 5084, doi:10.3390/s19235084. [CrossRef] [PubMed]

20. Wu, Y.; Niu, X.; Du, J.; Chang, L.; Tang, H.; Zhang, H. Artificial Marker and MEMS IMU-Based Pose Estimation Method to Meet Multirotor UAV Landing Requirements. *Sensors* **2019**, *19*, 5428, doi:10.3390/s19245428. [CrossRef] [PubMed]

21. Popescu, D.C.; Cernaianu, M.O.; Ghenuche, P.; Dumitrache, I. An assessment on the accuracy of high precision 3D positioning using planar fiducial markers. In Proceedings of the 2017 21st International Conference on System Theory, Control and Computing (ICSTCC), Sinaia, Romania, 19–21 October 2017; pp. 471–476. [CrossRef]

22. Sarmadi, H.; Muñoz-Salinas, R.; Berbís, M.A.; Medina-Carnicer, R. Simultaneous Multi-View Camera Pose Estimation and Object Tracking With Squared Planar Markers. *IEEE Access* **2019**, *7*, 22927–22940, doi:10.1109/ACCESS.2019.2896648. [CrossRef]

23. Garrido-Jurado, S.; Muñoz-Salinas, R.; Madrid-Cuevas, F.J.; Marín-Jiménez, M.J. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognit.* **2014**, *47*, 2280–2292. [CrossRef]

24. Garrido-Jurado, S.; Muñoz-Salinas, R.; Madrid-Cuevas, F.J.; Medina-Carnicer, R. Generation of fiducial marker dictionaries using Mixed Integer Linear Programming. *Pattern Recognit.* **2016**, *51*, 481–491. [CrossRef]

25. Kalman, R.E. A New Approach to Linear Filtering and Prediction Problems. *J. Basic Eng.* **1960**, *82*, 35–45, doi:10.1115/1.3662552. [CrossRef]

26. Hao, Y.; Xu, A.; Sui, X.; Wang, Y. A Modified Extended Kalman Filter for a Two-Antenna GPS/INS Vehicular Navigation System. *Sensors* **2018**, *18*, 3809, doi:10.3390/s18113809. [CrossRef]

27. Hosseinyalamdary, S. Deep Kalman Filter: Simultaneous Multi-Sensor Integration and Modelling; A GNSS/IMU Case Study. *Sensors* **2018**, *18*, 1316, doi:10.3390/s18051316. [CrossRef]

28. Bischoff, O.; Wang, X.; Heidmann, N.; Laur, R.; Paul, S. Implementation of an ultrasonic distance measuring system with kalman filtering in wireless sensor networks for transport logistics. *Procedia Eng.* **2010**, *5*, 196–199. [CrossRef]

29. Li, S.E.; Li, G.; Yu, J.; Liu, C.; Cheng, B.; Wang, J.; Li, K. Kalman filter-based tracking of moving objects using linear ultrasonic sensor array for road vehicles. *Mech. Syst. Signal Process.* **2018**, *98*, 173–189. [CrossRef]