A validation strategy for a target-based vision tracking system with an industrial robot

A validation strategy for a target-based vision tracking system with an industrial robot

C Romero[1], C Naufal[1], J Meza[1], and A G Marrugo[1]

[1] Facultad de Ingeniería, Universidad Tecnológica de Bolívar, Cartagena de Indias, Colombia

E-mail: cadarome1998@hotmail.com, camilonauffalsalas@gmail.com

Abstract. Computer vision tracking systems are used in many medical and industrial applications. Understanding and modeling the tracking errors for a given system aids in the correct implementation and operation for optimal measurement results. This project aims to simulate and experimentally validate a tracking system for medical imaging. In this work, we developed a validation strategy for a target-based vision tracking system with an industrial robot. The simulation results show that the system can be accurately modeled, and the error assessment strategy is robust. Experimental verification with an EPSON C3 robot shows the reliability of the vision tracking system to obtain the target position and pose accurately. The general-purpose performance assessment strategy can be used as a vision tracking evaluation mechanism to ensure the system performance is adequate for a given application.

1. Introduction

Recent advances in computer vision have enabled modern human-computer interfaces, precise manufacturing and metrology, autonomous vehicles, among other applications [1–3]. Moreover, there are many medical imaging applications, such as free-hand ultrasound (US) [4], magnetic resonance imaging (MRI) motion artifact reduction, biomechanics [5], that require precise Three-dimensional (3D) tracking. However, the tracking error of computer vision systems is often neglected or not sufficiently evaluated.

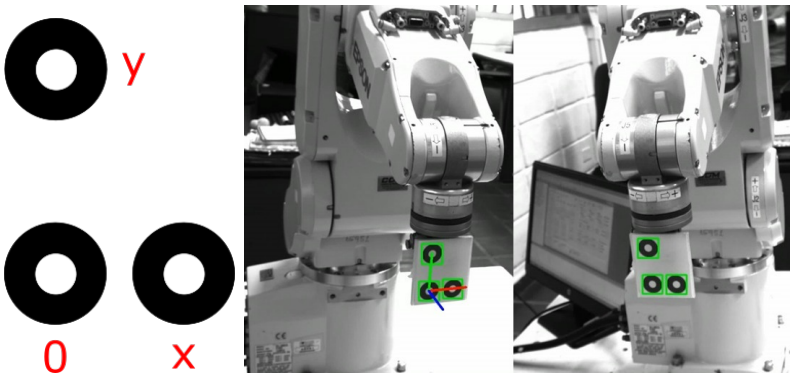
The pose and the 3D position are the relevant features for 3D tracking. In most applications, like in free-hand US, errors in the tracking of the probe ultimately lead to an inaccurate 3D representation of the object of interest [4]. Moreover, the tracking error also affects the underlying calibration [6]. For instance, Bouget, *et al.*, [7] described the challenges in surgical tool tracking and the need for data reference creation. In Ref. [8], the authors evaluated several technologies, mainly stereo-vision tracking and ultra-wideband positioning systems for construction workforce monitoring. While being an unrelated application to medical imaging, the validation and error assessment of the tracking systems is of paramount importance to determine the overall performance.

In this work, we introduce a validation strategy for assessing a 3D stereo-tracking computer vision system [9] developed for free-hand US tracking and 3D imaging. With the use of an EPSON C3 series industrial robot arm [10], reliable position and pose references are arranged to evaluate the tracking system performance. To generalize our approach, we simulate the robot and the computer vision systems, in addition to experimental laboratory results.

2. Methods

2.1. Stereo vision system

The tracking system consists of a stereo-vision system composed of two conventional cameras and a black-and-white (B&W) target with three coplanar circles shown in Figure 1(a). We attach the target to the end effector of the robot to track its position and orientation through triangulation. The tracking procedure involves estimating the centroids of the three circles in both images Figure 1(b), and matching them to obtain the 3D coordinates. The point correspondences are established through epipolar geometry constraint [9] as described in the following Figure 1.



(a) (b)

Figure 1. (a) The target is composed of three co-planar circles used for the position and orientation tracking. (b) The target attached to the robot arm for the experimental validation, as seen from the left and the right cameras.

We assume each camera follows a pinhole camera model [11]. Thus, a point X in 3D is imaged at x_1 and x_2, in camera 1 and cameras 2, respectively, as shown in Figure 2(a). Note that X and the camera centers O_1 and O_2 form a plane which intersects the image plane of camera 2 forming a line l_2. This geometrical constraint is useful since the algorithm does not need to search for the corresponding point of x_1 in the entire image plane 2, but the search is restricted to the line l_2.

The fundamental matrix $\mathbf{F}$ encapsulates the epipolar geometry between two views with the expression shown in Equation (1).

$$x_1^{\mathsf{T}} \mathbf{F} x_2 = 0. \tag{1}$$

For each point x_1 in the first view, we evaluate the epipolar constraint from Equation (1) with each center x_2. Thus, the true correspondence of x_1 is the point x_2 that gives the lowest epipolar constraint value.

2.2. Triangulation

With the correct center matches, we obtain the 3D coordinates of the three points through triangulation. From the pinhole camera model depicted in Figure 2(b), we can describe the projection of a 3D point in homogeneous coordinates in the image plane based on extrinsic and intrinsic parameters matrix of the camera through Equation (2).

$$s\mathbf{x} = \mathbf{K}\, {}^{C}\mathbf{M}_{W}\, \mathbf{X}^{W}, \tag{2}$$

where $\mathbf{X}^W = [X,\ Y,\ Z,\ 1]^\mathsf{T}$ is a 3D point given in a world coordinate system, s is a scale factor, $\mathbf{x} = [x,\ y,\ 1]^\mathsf{T}$ is the homogeneous 2D coordinates of the projected 3D point, $\mathbf{K}$ is the matrix built from the intrinsic parameters of the camera as focal length, screw factor and main camera point as show in Equation (3), and ${}^{C}\mathbf{M}_W = [\mathbf{R} \mid \mathbf{t}]$ is the extrinsic matrix that describes the orientation $\mathbf{R}$ and position $\mathbf{t}$ of the world frame relative to the camera frame. The matrix $\mathbf{P} = \mathbf{K}\,{}^{C}\mathbf{M}_W$ is referred to as the camera matrix or the projection matrix of the camera.

$$\mathbf{K} = \begin{bmatrix} f_x & \gamma & x_0 \\ 0 & f_y & y_0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{3}$$

The intrinsic and extrinsic parameters of the two cameras are obtained with the standard camera calibration procedure [12], which enables the estimation of the 3D coordinates of a point given its image point matches. From Equation (2), We have a set of six equations and we need to solve for $\mathbf{X}$ where s_1 and s_2 scale factors are also unknowns in Equation (4). This is achieved through the homogeneous solution or direct linear transform to this problem [11].

$$s_1\mathbf{x}_1 = \mathbf{P}_1\mathbf{X}\ \ ,\ \ s_2\mathbf{x}_2 = \mathbf{P}_2\mathbf{X}, \tag{4}$$

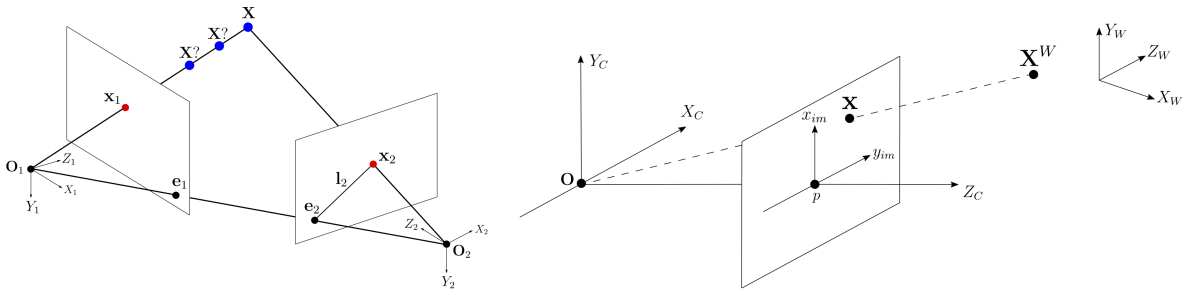


Figure 2. (a) Epipolar geometry. (b) Pinhole camera model.

2.3. Stereo camera calibration

The calibration of the stereo-vision system consists of estimating the intrinsic parameters of both cameras using the method proposed by Zhang [12] using a B&W checkerboard arbitrarily placed at different positions.

For the extrinsic parameters, we align the world coordinate system with the camera 1 coordinate system. In this way, the extrinsic matrix for camera 1, ${}^{C_1}\mathbf{M}_W = [\mathbf{I} \mid \mathbf{0}]$, where $\mathbf{I}$ is the 3×3 identity matrix and $\mathbf{0}$ is a 3×1 zero vector. Then, we estimate the extrinsic matrix of camera 2, ${}^{C_2}\mathbf{M}_W = [\mathbf{R} \mid \mathbf{t}]$, where $\mathbf{R}$ and $\mathbf{t}$ represents the position and orientation of the (world) camera 1 relative to the camera 2 frame which are estimated based on the correspondence of points and internal parameters previously estimated. Furthermore, the camera calibration also consists of estimating the lens distortion coefficients not considered in the standard pinhole camera model [12].

2.4. Position and orientation tracking of the target

With the reconstructed centers of each circle, we estimate the position and orientation of the target coordinate system. This target frame is defined based on the labels of each circle from Figure 1(a), where the center of the 0-label circle represents the origin of the frame, and the x-label and y-label centers are the directions of the x- and y-axis, respectively.

The 3D position of the center labeled as 0, represents the translation vector $\mathbf{t}$ of the target frame relative to the stereo-vision system, and the orientation is given by $\mathbf{R} = [\hat{\mathbf{x}}\ \hat{\mathbf{y}}\ \hat{\mathbf{z}}]$ where $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ are the unit vectors of the x- and y-axis, and $\hat{\mathbf{z}} = \hat{\mathbf{x}} \times \hat{\mathbf{y}}$.

2.5. Error evaluation

To evaluate the performance of the tracking system, we locate our target with the end effector of the arm in two different positions and evaluate the displacement and rotation of the target between both locations. As for performance metrics, we use the root mean square (RMS) error shown in Equation (5) which gives us an average measure of the precision of the system. The mean absolute error (MAE) depicted in Equation (6) to verify how far is our measured data with the system from the reference value, and the ℓ_∞ norm of the error as worst system accuracy performance through Equation (7).

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{n}(\hat{p}_i - p)^2}, \tag{5}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{n}|\hat{p}_i - p|\,, \tag{6}$$

$$\|e\|_\infty = \max_{1\leq i\leq n}|\hat{p}_i - p|\,, \tag{7}$$

where $\hat{p}_i$ is the measured variable (displacement or rotation), and p is the reference value. Note that all these metrics are measured between two different target locations or image frames. That is, $\hat{p}_i$ is the displacement or rotation measured between two different frames.

2.6. Building a position and orientation reference

To evaluate the performance of the vision tracking system, we compare the target position and pose estimation from the tracking system with those obtained from an industrial Epson C3 robot arm (Epson Robots, CA, USA) with 6 degrees of freedom. The arm is a high speed, and a high precision industrial robot. The kinematic description of the robot is carried out using the standard Denavit-Hartenberg matrix [13]. This procedure makes it possible to relate the last link of the kinematic chain to the base. Also, we used the ARTE Library [14] for performing the visualization of the links and for positioning the target.

3. Experiments and results

We designed the validation strategy through simulation using the Peter Corke Robotics Toolbox [14], which allows for kinematic and dynamic analysis of the robotic manipulator. For the stereo vision system, we used the Peter Corke computer vision library [15], which allows setting up the cameras arbitrarily in space and simulating the target imaging. In the simulation, the robotic arm moved the target at predefined positions and orientations, which served as reference values, and the cameras imaged the target points with noise added to the pixel coordinates. The error metrics were computed as described in the previous section. Finally, the same experiments were carried out in the real setting with the EPSON C3 robot.

3.1. Simulation tests

The validation strategy was implemented in MATLAB to perform an error and uncertainty analysis based on the relative positions of the target as detected from the vision tracking system and the robotic manipulator reference. Figure 3 shows the simulated tracking system, where the target is represented by three spheres with coplanar centers. The simulation was carried out with

the following parameters: the distance between the robot base and the stereo vision system was 1.27 m, and the camera base-line distance was 1 m. The distance between the 0-label and the x-label centers is 0.2 m, and the distance between centers labeled as 0 and y is 0.4 m. The camera settings were: sensor size 10 mm $\times$ 10 mm, $f = 20$ mm, and 1024 $\times$ 1024 pixel resolution. Extrinsic parameters of both cameras (positions and orientations) are given with respect to the robot base coordinate system, which in the case of simulations is the world coordinate frame. In this way, 3D points obtained through triangulation with the stereo-vision system are in the base robot frame. For each pose, the target points were imaged by the two cameras. The subpixel positions of the targets were corrupted with white Gaussian noise ($\mathcal{N}(0, 0.02)$) to produce realistic results. We evaluate the simulated setup in terms of displacement and rotation.

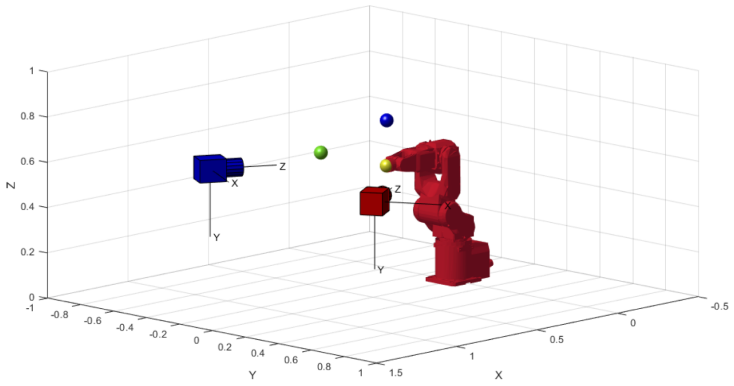


Figure 3. Simulated system with the robot arm, the stereo-vision system and our target represented by three spheres with coplanar centers.

For the displacement evaluation, the end effector of the robot is moved a fixed distance of Δx in 12 different positions. The vision system measured the displacement of the end effector by tracking the target between the 11 consecutive frames. The tracking consists of reconstructing each circle center of the target for each frame and measuring the Euclidean distance between the 3D points in consecutive frames. From this procedure, 11 displacement measurements are obtained for each target circle; *i.e.*, for the three target circles, there are a total of 33 measurements. In Table 1, we show the error evaluation of the estimated displacements for each target circle for different Δx: 10 mm, 20 mm, 30 mm, and 40 mm. Note that the errors are relatively independent of the displacement magnitude Δx, which is expected for a robust tracking system. The simulation also allowed us to validate the performance assessment script.

For the rotation evaluation, the joint of the end effector is rotated a fixed $\Delta\theta$ angle of 5°, 10°, 15° and 20°. In this way, the points of the target were rotated in the same plane in all the frames. A total of 11 different frames were acquired for each $\Delta\theta$ value; that is, we calculate a total of 10 angles between consecutive frames. The rotation estimation tracking consists of calculating the rotation matrix of the target coordinate system relative to the stereo-vision system in each frame. Then, between two consecutive frames, we estimate the equivalent rotation angle using Equation (8).

$$\theta = \arccos\left[\frac{\mathrm{Tr}(\mathbf{R}_{12}) - 1}{2}\right], \tag{8}$$

where $\mathbf{R}_{12} = \mathbf{R}_1^{\mathsf{T}}\mathbf{R}_2$, is the equivalent rotation matrix between frame 1 and frame 2, that is, the rotation of the target coordinate system between frame 1 and frame 2. Error evaluation of the measurement for each $\Delta\theta$ are reported in Table 2. As in case of translation, we can note that errors are relatively independent of the rotation magnitude $\Delta\theta$.

Table 1. Error evaluation in the target displacements between consecutive frames.

	$\Delta x = 10$ mm	$\Delta x = 20$ mm	$\Delta x = 30$ mm	$\Delta x = 40$ mm
MAE (mm)	0.09953	0.1522	0.09425	0.06422
RMS (mm)	0.12583	0.18817	0.12146	0.08296
ℓ_∞ norm (mm)	0.34109	0.42050	0.26157	0.15252

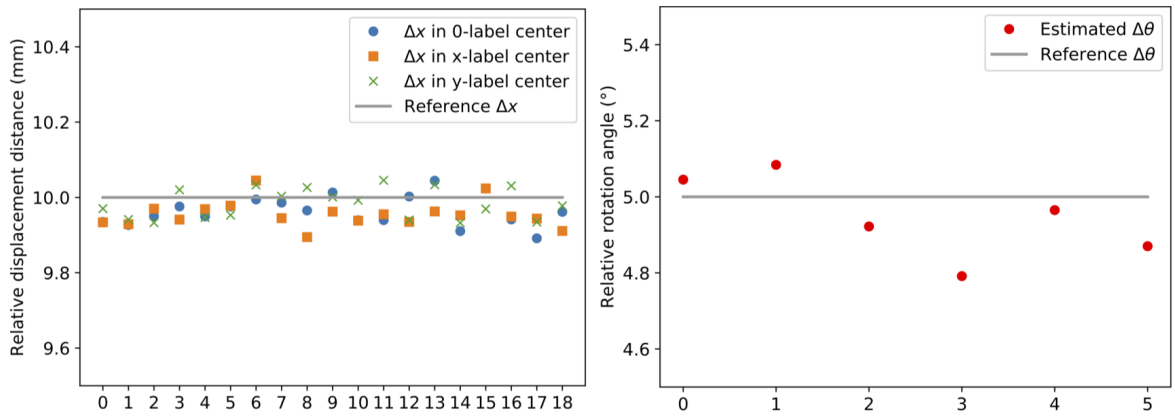
Table 2. Error evaluation in the target coordinate system rotation between consecutive frames.

	$\Delta\theta = 5°$	$\Delta\theta = 10°$	$\Delta\theta = 15°$	$\Delta\theta = 20°$
MAE (°)	0.00290	0.00196	0.00230	0.00367
RMS (°)	0.00334	0.00229	0.00278	0.00417
ℓ_∞ norm (°)	0.00541	0.00371	0.00476	0.00673

3.2. Experimental tests

In our experimental setup, we use a robot arm EPSON C3, two monochromatic complementary metal oxide semiconductor (CMOS) cameras (Basler acA1300-200um; 1280 x 1024; 203 fps), two objectives (Computar M0814-MP2; 8mm; F1.4) and the B&W target shown in Figure 1(a), which we attach to the end effector of the EPSON arm, as shown in Figure 1(b). We calibrate our stereo-vision system with the procedure described in section 2.3. We carried out two experiments to evaluate the performance of the tracking method with displacements and rotations.

In the first experiment, the end effector is moved at a fixed distance of $\Delta x = 10$ mm. The distance of each center of the three circles was measured between successive frames. The target was captured in a total of 20 different positions, which gives us 19 Δx values estimated for each center, *i.e.*, a total of 57 Δx measurements including for the three target circles. Figure 4(a) shows the estimated relative displacement of each circle center between two consecutive frames. The results can be compared with the reference displacement line of 10 mm. The error evaluation of values obtained in this experiment are reported in Table 3. An RMS error of 0.05078 mm and a maximum error of 0.10857 mm confirms that vision tracking system is a reliable method for displacement tracking.



(a) (b)

Figure 4. Experimental results. (a) Estimated displacement of each center for all the 19 consecutive frames. (b) Estimated rotation of the target coordinate system for all the six consecutive frames.

In the second experiment, we evaluate our system in terms of rotations. The end effector is rotated $\Delta\theta = 5°$ between two frames. We evaluate the rotation of the target coordinate system

drawn in the image of Figure 1(b) using Equation (8) described in the simulation experiments. A total of seven different frames were acquired, which give us six $\Delta\theta$ values. The estimated relative rotation angles between each consecutive frames are shown in Figure 4(b), which can be compared with the reference rotation angle. The error evaluation of values obtained in this experiment are reported in Table 3. The RMS error of 0.11311° and the maximum error of 0.20856° show the reliability of the tracking system to obtain the target pose and relative rotations.

Table 3. Error evaluation of the two real experiments.

Experiment	MAE	RMS	ℓ_∞ norm
Displacement (mm)	0.04461	0.05078	0.10857
Rotation	0.09676°	0.11311°	0.20856°

4. Conclusions

Reliable target tracking in computer vision requires a well-defined target that is easily detectable. To test the performance of a stereo-vision tracking system, we simulated the positioning of a target with a robotic arm and the tracking of the target with the vision system. This approach enabled us to obtain a general-purpose performance assessment strategy to evaluate a vision tracking system. Encouraging experimental results with a robotic arm show the validity of the proposed approach. Moreover, this work could provide a means for reliable 3D free-hand ultrasound imaging. Future work involves assessing the performance of the vision tracking system in real-time.

Acknowledgments

This work has been partly funded by Universidad Tecnológica de Bolívar, Cartagena de Indias, Colombia, projects C2018P005 and C2018P018. J. Meza thanks Universidad Tecnológica de Bolívar for a post-graduate scholarship.

References

[1] Voulodimos A, Doulamis N, Doulamis A and Protopapadakis E 2018 *Computational Intelligence and Neuroscience* **2018(7068349)** 1
[2] de Oliveira Baldner F, Costa P B, Gomes J F S and Leta F R 2020 A review on computer vision applied to mechanical tests in search for better accuracy *Advances in Visualization and Optimization Techniques for Multidisciplinary Research* (Singapore: Springer) pp 265–281
[3] Janai J, Güney F, Behl A and Geiger A 2019 *ArXiv* **abs/1704.05519v2** 1
[4] Mozaffari M H and Lee W S 2017 *Ultrasound in Medicine and Biology* **43(10)** 1
[5] Ferber R, Osis S T, Hicks J L and Delp S L 2016 *Journal of Biomechanics* **49(16)** 3759
[6] Cenni F, Monari D, Desloovere K, Aertbeliën E, Schless S H and Bruyninckx H 2016 *Computer Methods and Programs in Biomedicine* **136(C)** 179
[7] Bouget D, Allan M, Stoyanov D and Jannin P 2017 *Medical Image Analysis* **35** 633
[8] Yang J, Cheng T, Teizer J, Vela P A and Shi Z 2011 *Advanced Engineering Informatics* **25(4)** 736
[9] Meza J, Simarra P, Contreras-Ojeda S, Romero L A, Contreras-Ortiz S H, Arambula Cosio F and Marrugo A G 2019 A low-cost multi-modal medical imaging system with fringe projection profilometry and 3d freehand ultrasound *15th International Symposium on Medical Information Processing and Analysis* vol 11330 (Colombia: International Society for Optics and Photonics)
[10] El-Gohary M and McNames J 2015 *IEEE Transactions on Biomedical Engineering* **62(7)** 1759
[11] Hartley R and Zisserman A 2003 *Multiple View Geometry in Computer Vision* (United States of America: Cambridge University Press)
[12] Zhang Z 2000 *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**
[13] Craig J J 2009 *Introduction to Robotics: Mechanics and Control, 3/E* (India: Pearson Education)
[14] Corke P 2007 *IEEE Robotics & Automation Magazine* **14(4)** 16
[15] Corke P I 2005 *IEEE Robotics & Automation Magazine* **12(4)** 16