







Multi-target Attachment for Surgical Instrument Tracking

Eberto Benjumea^(✉) , Juan S. Sierra , Jhacson Meza ,
and Andres G. Marrugo 

Facultad de ingeniería, Universidad Tecnológica de Bolívar, Cartagena, Colombia
ebenjumea@utb.edu.co

Abstract. The pose estimation of a surgical instrument is a common problem in the new needs of medical science. Many instrument tracking methods use markers with a known geometry that allows for solving the instrument pose as detected by a camera. However, marker occlusion happens, and it hinders correct pose estimation. In this work, we propose an adaptable multi-target attachment with ArUco markers to solve occlusion problems on tracking a medical instrument like an ultrasound probe or a scalpel. Our multi-target system allows for precise and redundant real-time pose estimation implemented in OpenCV. Encouraging results show that the multi-target device may prove useful in the clinical setting.

Keywords: Multi-target ArUco marker · Pose estimation · Optical instrument tracking

1 Introduction

Camera pose estimation and object tracking are common problems in computer vision. These tasks require a high accuracy localization of the marker's features for applications such as virtual or augmented reality [2,9], robot navigation [5,19], and even medical applications [13]. Marker-based methods have reached high popularity in recent years due to their easy usage, reliability, robustness, and high-speed detection. However, there is no general-purpose method that works efficiently for most applications.

Nowadays, it is challenging to choose from the many available solutions in terms of performance, speed, and overall accuracy [14]. Several authors have proposed different fiducial markers [7,9,10,16]. These markers have been used in augmented reality applications [1], to design wireless surgical knife attachment for medical usage [11,20], and automatic drone navigation system [17]. Nevertheless, there is still room for improvement in many of these applications.

For tracking surgical instruments, we need a clear line of sight between the marker and the camera. Due to instrument movement, marker occlusion occurs. This problem limits the physicians' freedom of movement in navigation systems

where a needle, surgical knife, or ultrasound probe must be tracked, e.g., in 3D freehand ultrasound. Occlusion is a typical limitation when optical devices are adopted, but they are often more accurate compared to electromagnetic systems [4, 12, 18]. This problem has been assessed in several works [7, 16] with partial success.

This work proposes a multi-target system for object tracking in clinical procedures under occlusion conditions. Its main novelties are the operation under occlusion and low cost, thanks to 3D printing and the use of a single camera, contrary to other proposed systems that use stereo vision [3]. Finally, this proposal is easily adaptable to any flat-faced structure as long as it contains nine markers: one source marker, four markers on the upper faces, and four markers on the lower faces. This feature allows improving the system's performance according to the type of surgical procedure: 3D reconstruction by ultrasound imaging, tool tracking, medical teaching tools, among others.

This paper is structured as follows. In Sect. 2, we explain the multi-target physical attachment with ArUco markers. Section 3 details how the pose is estimated with the multi-target system. Next, in Sect. 4, we show the proposed method's results and analysis. Finally, in Sect. 5, we draw several conclusions.

2 Multi-target Device

2.1 Physical Device

As a means to obtain a reliable detection regardless of the instrument pose or even with partial occlusions, we developed the multi-target physical attachment shown in Fig. 1. The multi-target attachment is a 3D printed solid piece designed in the computer-aided design software SolidEdge. It has a flat horizontal top face and eight diagonal faces oriented in multiples of 45 degrees. There is a unique marker on each face, as shown in Fig. 1. The bottom face is parallel to the top face and has a hole for attachment to a surgical or medical instrument.

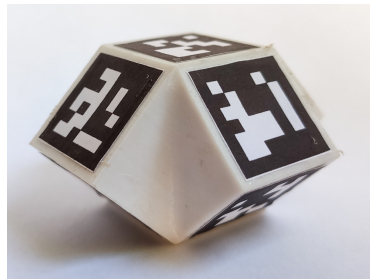


Fig. 1. Multi-target attachment.

The existence of the nine markers increases the tracking system's robustness in case of occlusion of one or more markers. This advantage comes from the fact

that each marker allows establishing the medical instrument's pose to a frame of reference (origin). We established the frame of reference on the top of the model. These coordinate transformations between markers will be explained in detail in Sect. 3.

2.2 ArUco Markers

ArUco is an open-source library for detecting squared fiducial markers in images [7, 15]. Its use allows estimating the camera pose with respect to the markers for a calibrated camera. This library is written in C++ and requires OpenCV for its implementation [7, 15]. There are some other libraries, like ArUco, used for the same function. Some of these libraries and their problems are mentioned below [6, 7, 15]: ARToolKit is prone to error and not very robust to illumination changes. ARToolKit+ is a new proprietary version of ARToolKit with a more robust error detection than ARToolKit. Also, ARTag is another proposal to prove ARToolKit, but it is a discontinued project. BinARyID does not consider the possibility of error detection and correction, and AprilTags detects and corrects errors, but its method is not adequate for a large number of markers.

ArUco library adapts to heterogeneous illumination, is highly accurate and fast at detecting markers, and is robust enough to do error detection and correction of the binary codes. These characteristics turned ArUco into the most popular system for marker detection [7, 15]. Despite that, ArUco has problems when the lighting is poor or when the image undergoes extreme motion blur [8]. Deep ChArUco is a deep convolutional neural network system trained to overcome these situations for ChArUco marker detection and pose estimation [8].

The pose estimation using the ArUco library needs a previously acquired image. It starts with an image resize in order to reduce processing time [15]. Later, image segmentation is performed with a global threshold method. A contour extraction is applied, and these contours are filtered by their shape (polygonal). With a set of resized versions of the original image, an image pyramid is created [15]. Immediately, the marker code is extracted from the image using the image pyramid, and the precise corner localization of the marker corners is estimated [15]. Finally, its pose is estimated with respect to the camera by iteratively minimizing the reprojection error of the corners [7].

The multi-target attachment has on its top faces the markers with codes: ID0, ID1, ID2, ID3, and ID4 (See Fig. 2(a)). On its bottom faces are located the ID5, ID6, ID7, and ID8 markers (See Fig. 2(b)). These last four are respectively below the ID1, ID2, ID3, and ID4 markers. This geometrical configuration allows a marker to always be visible in front of the camera in spite of the translations or rotations of the medical device. The length of the markers is 19 mm.

3 Stages of the Pose Estimation of Multi-target Attachment Using ArUco Markers

There are five stages for the pose estimation of the attachment, as Fig. 3 shows. The stages of image acquisition, marker detection, and individual marker pose

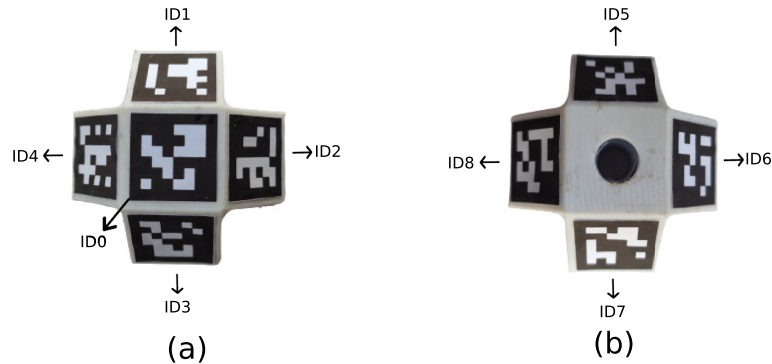


Fig. 2. Codes of the markers on the multi-target attachment. (a) Markers on the top faces, (b) Markers on the bottom diagonal faces

estimation are carried out using Python and the Open Source Computer Vision Library (OpenCV) libraries. Our novelties in the software implementation are the marker filtering and instrument pose estimation stages.

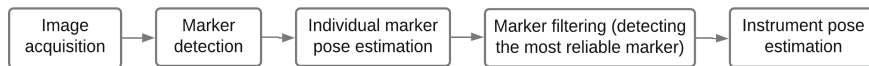


Fig. 3. Stages of the pose estimation of attachment using ArUco markers.

3.1 Image Acquisition, Marker Detection and Individual Marker Pose Estimation

First, we record a video of the surgical instrument. The proposed method was tested offline, and a real-time implementation is currently being developed. All the frames of the video are read. Each frame is processed for the detection of ArUco markers. The system can detect one or several markers in a frame. In this stage, we obtain the ID number and the corners of the markers in individual marker pose estimation. This data is taken for the individual pose estimation of the markers. For each detected marker, the pose is given in a rotation vector and a translation vector. These vectors are transformed in the extrinsic matrix of the marker.

3.2 Marker Filtering and Multi-target Pose Estimation

Prior camera calibration is required to obtain the intrinsic matrix of the camera. Since we established the frame of reference (origin) on the top of the model, it is necessary to calculate the relative position matrices between the origin marker

and the other markers one by one (Fig. 4 shows some of these transformations). Therefore, we must estimate the pose of the origin marker and the pose of the other markers and formulate the following equation

$${}^C H = {}^C_M H \times {}^M_O H, \quad (1)$$

where ${}^C_M H$ is the pose matrix of any target, ${}^C_O H$ is the pose matrix of the ID0 target, and ${}^M_O H$ is the relative position matrix of any target with respect to ID0 marker. From this equation, using matrix algebra, we obtain

$${}^M_O H = [{}^C_M H]^{-1} \times {}^C_O H, \quad (2)$$

which allows us to calculate the relative position matrix of any marker.

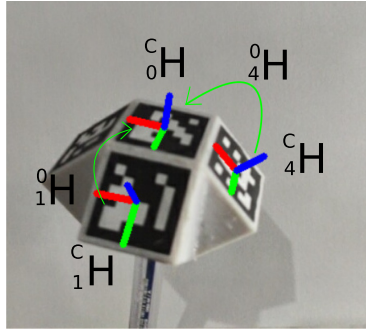


Fig. 4. Matrices of relative positions.

Nevertheless, the estimation of relative position matrices based on a single image causes errors in the instrument's pose estimation. This problem is shown in the experiments and results section. Accordingly, it is necessary to estimate these matrices with many representative images of each pair of targets' possible poses. For this, we take the poses of each pair of markers, ID0 and IDX (IDX is the marker which we want to know its relative position matrix), in each image. These values are stored into two matrices which we stack the poses of ID0 and IDX on all processed images. Therefore, Eq. (2) turns into the following,

$$\begin{bmatrix} {}^C_O H_1 \\ {}^C_O H_2 \\ {}^C_O H_3 \\ \cdot \\ \cdot \\ \cdot \\ {}^C_O H_n \end{bmatrix} = \begin{bmatrix} {}^C_M H_1 \\ {}^C_M H_2 \\ {}^C_M H_3 \\ \cdot \\ \cdot \\ \cdot \\ {}^C_M H_n \end{bmatrix} \times {}^M_O H, \quad (3)$$

where n is the number of processed images, and the problem is solved by least-squares. The relative position matrices of the markers on the bottom faces are indirectly estimated through the top face's relative position matrices. The previously explained process is applied to these matrices, but instead of ID0 as a reference frame, we use the corresponding top face to each bottom marker. Equation (3) describes this estimation. Later, the matrix with respect to ID0 is calculated by Eq. (4).

$${}_{0}^{Bottom\ marker} H = {}_{Top\ marker}^{Top\ marker} H \times {}_{Top\ marker}^{Bottom\ marker} H. \quad (4)$$

The camera center coordinates (CC) project in the coordinates system of the target ID0 through each detected marker's coordinates system. The resulting coordinates are averaged, and the marker with the closest coordinates is selected to estimate the surgical device's pose. We multiply this matrix with its corresponding relative position matrix to the origin marker (top marker with ID0), as shown in Eq. (1). The result is the pose of the surgical instrument. Finally, to visualize the result on the image, we use the following equation for the complete perspective projection, given by

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = K \times {}_O^C H \times {}^O P, \quad (5)$$

where K is the intrinsic camera matrix and ${}^O P$ are the axes points of the target in homogeneous coordinates. Then, the origin coordinates x_{im} and y_{im} on the image are given by $x_{im} = x_1/x_3$ and $y_{im} = x_2/x_3$.

4 Experiments and Results

We use a monocular system composed of a camera with a resolution of 640×480 pixels, the multi-target attachment, and a personal computer. As explained above, the software was implemented using Python and OpenCV. The camera calibration parameters are the following intrinsic values (K) and distortion coefficients ($dist$)

$$K = \begin{bmatrix} 491.0607 & 0 & 319.8941 \\ 0 & 492.8343 & 238.0494 \\ 0 & 0 & 1 \end{bmatrix},$$

$$dist = [0.0534, -0.2042, -0.00136, 0.0009, 0.1593].$$

The experiments were carried out with a distance of approximately 50 cm between the camera and the multi-target attachment. And the surgical instrument moved, like an ultrasound probe, in a volume of 4000 cm^3 composed of $20 \text{ cm} \times 20 \text{ cm} \times 10 \text{ cm}$. The video duration 40 s with a rate of 20 frames/seconds.

To test the accuracy of the proposed device and system, we carried out a graphical experiment. This consisted of drawing the target ID0: under the given pose estimation directly by the ArUco library and under the proposed method. The estimation of the relative position matrices with one and several images was analyzed in this test. Figure 5(a) shows two axes. The axes on the top face of the attachment were drawn with the given pose matrix by ArUco library. The axes which are floating were drawn with the proposed method but with the relative positions matrix based on only a single image. Figure 5b show us the same situation but with the relative positions matrix based on a group of twenty images. We can see that the two axes are almost superimposed. The multi-target attachment and its system obtained good results in the graphic experiment. Likewise, this test demonstrated the need to estimate the relative positions matrices with several images.

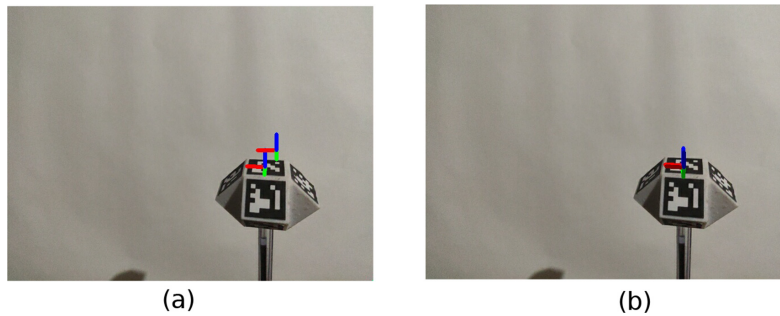


Fig. 5. Result of the graphic experiment to pose estimation of the surgical instrument. *a)* With relative position matrix based in a single image. *b)* With relative position matrix based in several images.

To present the system operation, we show the OpenCV-obtained poses for one of the multiple acquired images. The image showed in Fig. 6a was processed, and the ID0, ID1, and ID4 markers were detected, and their poses were calculated, obtaining the following data (See Fig. 6(b))

$$ID0 \text{ marker pose} = \begin{bmatrix} -0.9689 & -0.1715 & 0.1785 & -0.0144 \\ -0.2355 & 0.4176 & -0.8776 & 0.0122 \\ 0.0759 & -0.8923 & -0.4450 & 0.1383 \end{bmatrix},$$

$$ID1 \text{ marker pose} = \begin{bmatrix} -0.9737 & -0.2177 & -0.067 & -0.0185 \\ -0.1464 & 0.8235 & -0.5480 & 0.0271 \\ 0.1745 & -0.5238 & -0.8338 & 0.1161 \end{bmatrix},$$

$$ID4 \text{ marker pose} = \begin{bmatrix} -0.5916 & -0.1598 & 0.7903 & 0.0054 \\ -0.7611 & 0.4339 & -0.4820 & 0.0254 \\ -0.2659 & -0.8867 & -0.3783 & 0.1459 \end{bmatrix}.$$

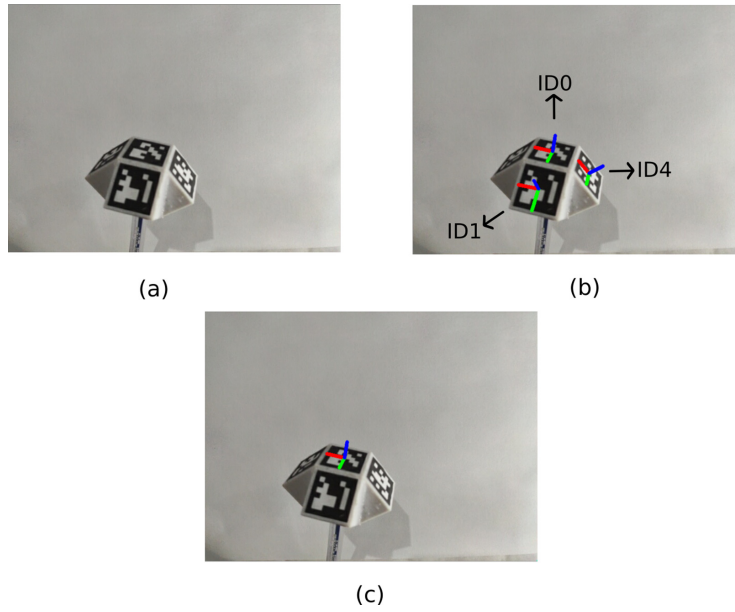


Fig. 6. Result of pose estimation of the surgical instrument. *a)* Input image. *b)* Axes plotted with the pose of each marker. *c)* Axes plotted with the pose of frame of reference.

Based on the target selection metric, the algorithm chooses the ID1 marker to estimate the pose of the frame of reference. Therefore, with the relative position matrix of the marker (See Eq. (6)), it uses Eq. (1) to obtain the optimal pose (See Eq. (7)). In Fig. 6c, we show the origin of the image coordinates.

$$ID1 \text{ relative position matrix} = \begin{bmatrix} 0.9912 & -0.04989 & -0.12295 & 0.002 \\ -0.0229 & 0.8487 & -0.5285 & -0.0248 \\ 0.1306 & 0.5266 & 0.8400 & -0.0106 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (6)$$

$$Pose \text{ of the surgical instrument} = \begin{bmatrix} 0.9912 & -0.04989 & -0.12295 & 0.002 \\ -0.0229 & 0.8487 & -0.5285 & -0.0248 \\ 0.1306 & 0.5266 & 0.8400 & -0.0106 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (7)$$

In the future, an application of this multi-target attachment would be a robust and low-cost 3D free-hand ultrasound [13]. In the case showed in [13], we can change the three circular markers by our proposal and locate the camera any position with line of sight. Also, we would use an industrial robot to obtain a quantitative validation of the proposed method and errors of pose estimation [14].

We described a performance problem for pose estimation under poor lighting and fast motion of the experiments' attachment. We will focus our efforts to solve this situation, possibly using Deep ChArUco [8].

5 Conclusion

We implemented an offline system for pose estimation of a surgical instrument using ArUco markers. The developed system estimates the pose in the world coordinates even when some markers are occluded due to the device's movement. However, the existence of nine markers in the attachment guarantees the detection of at least one marker, and ultimately, the instrument pose. Future work involves exploiting the redundancy for improving pose estimation in challenging medical environments.

Acknowledgement. E. Benjumea thanks MinCiencias and Sistema General de Regalías (Programa de Becas de Excelencia) for a PhD scholarship. J. Sierra and J. Meza thank Universidad Tecnológica de Bolívar (UTB) for a post-graduate scholarship. J. Meza acknowledges support from MinCiencias, and MinSalud for a “Joven Talento” scholarship.

References

1. Avola, D., Cinque, L., Foresti, G.L., Mercuri, C., Pannone, D.: A practical framework for the development of augmented reality applications by using aruco markers. In: International Conference on Pattern Recognition Applications and Methods, vol. 2, pp. 645–654. SCITEPRESS (2016)
2. Azuma, R.T.: A survey of augmented reality. *Presence: Teleoperators/Virtual Environment*. **6**(4), 355–385 (1997)
3. Bootsma, G.J., Siewerdsen, J.H., Daly, M.J., Jaffray, D.A.: Initial investigation of an automatic registration algorithm for surgical navigation. In: 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 3638–3642 (2008). <https://doi.org/10.1109/IEMBS.2008.4649996>
4. Colley, E., Carroll, J., Thomas, S., Varcoe, R.L., Simmons, A., Barber, T.: A methodology for non-invasive 3-D surveillance of arteriovenous fistulae using free-hand ultrasound. *IEEE Trans. Biomed. Eng.* **65**(8), 1885–1891 (2017)
5. DeSouza, G.N., Kak, A.C.: Vision for mobile robot navigation: a survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(2), 237–267 (2002)
6. Garrido-Jurado, S., Muñoz Salinas, R., Madrid-Cuevas, F., Medina-Carnicer, R.: Generation of fiducial marker dictionaries using mixed integer linear programming. *Pattern Recogn.* **51**(C), 481–491 (2016). <https://doi.org/10.1016/j.patcog.2015.09.023>
7. 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 Recogn.* **47**(6), 2280–2292 (2014)
8. Hu, D., DeTone, D., Malisiewicz, T.: Deep charuco: dark charuco marker pose estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2019

9. Kato, H., Billinghamurst, M.: Marker tracking and HMD calibration for a video-based augmented reality conferencing system. In: Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR 1999), pp. 85–94. IEEE (1999)
10. Knyaz, V.A.: The development of new coded targets for automated point identification and non-contact 3D surface measurements. *IAPRS* **5**, 80–85 (1998)
11. Koeda, M., Yano, D., Shintaku, N., Onishi, K., Noborio, H.: Development of wireless surgical knife attachment with proximity indicators using aruco marker. In: Kurosu, M. (ed.) *HCI 2018*. LNCS, vol. 10902, pp. 14–26. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-91244-8_2
12. Mercier, L., Langø, T., Lindseth, F., Collins, D.L.: A review of calibration techniques for freehand 3-D ultrasound systems. *Ultrasound Med. Biol.* **31**(4), 449–471 (2005)
13. Meza, J., Simarra, P., Contreras-Ojeda, S., Romero, L.A., Contreras Ortiz, S.H., Arámbula Cosío, F., Marrugo, A.G.: A low-cost multi-modal medical imaging system with fringe projection profilometry and 3D freehand ultrasound. *Proc. SPIE* **11330**, 1133004 (2020)
14. Romero, C., Naufal, C., Meza, J., Marrugo, A.G.: A validation strategy for a target-based vision tracking system with an industrial robot. *J. Phys.: Conf. Ser.* **1547**, 012018 (2020). <https://doi.org/10.1088/1742-6596/1547/1/012018>
15. Romero-Ramirez, F., Muñoz-Salinas, R., Medina-Carnicer, R.: Speeded up detection of squared fiducial markers. *Image Vision Comput.* **76** (2018). <https://doi.org/10.1016/j.imavis.2018.05.004>
16. Romero-Ramirez, F.J., Muñoz-Salinas, R., Medina-Carnicer, R.: Fractal markers: a new approach for long-range marker pose estimation under occlusion. *IEEE Access* **7**, 169908–169919 (2019)
17. Sani, M.F., Karimian, G.: Automatic navigation and landing of an indoor AR. Drone quadrotor using arUco marker and inertial sensors. In: 2017 International Conference on Computer and Drone Applications (IConDA), pp. 102–107. IEEE (2017)
18. Treece, G.M., Gee, A.H., Prager, R.W., Cash, C.J., Berman, L.H.: High-definition freehand 3-D ultrasound. *Ultrasound Med. Biol.* **29**(4), 529–546 (2003)
19. Wang, J., Olson, E.: Apriltag 2: efficient and robust fiducial detection. In: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 4193–4198. IEEE (2016)
20. Yano, D., Koeda, M., Onishi, K., Noborio, H.: Development of a surgical knife attachment with proximity indicators. In: Marcus, A., Wang, W. (eds.) *DUXU 2017*. LNCS, vol. 10289, pp. 608–618. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-58637-3_48