High-accuracy 3D locators tracking in real time using monocular vision

C. Elmo Kulanesan, P. Vacher, L. Charleux, E. Roux

High accuracy 3D locators tracking in realtime using monocular vision

C. Elmo Kulanesan[a,*], P. Vacher[a], L. Charleux[a], E. Roux[a]

[a]*Université Savoie Mont Blanc Laboratoire SYMME, 74940 Annecy-le-Vieux, France*

Abstract

In the field of medical applications, precise localization of medical instruments and bone structures is crucial to ensure computer-assisted surgical interventions. In orthopedic surgery, existing devices typically rely on stereoscopic vision. Their purpose is to aid the surgeon in screw fixation of prostheses or bone removal. This article addresses the challenge of localizing a rigid object consisting of randomly arranged planar markers using a single camera. This approach is especially vital in medical situations where accurate object alignment relative to a camera is necessary at distances ranging from 80 cm to 120 cm. In addition, the size limitation of a few tens of centimeters ensures that the resulting locator does not obstruct the work area. This rigid locator consists of a solid at the surface of which a set of plane markers (ArUco) are glued. These plane markers are randomly distributed over the surface in order to systematically have a minimum of two visible markers whatever the orientation of the locator. The calibration of the locator involves finding the relative positions between the individual planar elements and is based on a bundle adjustment approach. One of the main and known difficulties associated with planar markers is the problem of pose ambiguity. To solve this problem, our method lies in the formulation of an efficient initial solution for the optimization step. After the calibration step, the reached positioning uncertainties of the locator are better than two tenth of a cubic millimetre and one tenth of a degree regardless of the orientation of the locator in space. To assess the proposed method, the locator is rigidly attached to a stylus of about twenty centimeters length. Thanks to this approach, the tip of this stylus seen by a 16.1 megapixel camera at a distance of about 1 m is localized in real time in a cube lower than 1 mm side. A surface registration application is proposed by using the stylus on an artificial scapula.

Keywords: Monocular vision, 3D locators, planar marker, 3D tracking, Aruco.

1. Introduction

The pose estimation of an object in a scene is a classical computer vision problem. It is generally seen as a least squares optimization problem called Perspective-N-Points (PNP) in which the reprojection errors of a number of points of interests (POIs) located on a target and observed on several images are minimized [1, 2, 3]. For real-time applications, the use of easy-to-detect fiducial markers provides excellent computing time performance to detect them on the images. At the same time, the use of a single camera, also referred to as monocular vision, reduces financial costs as well as the complexity of the setup. Many applications presented in the literature take place in this context ranging from positioning for augmented reality, drone navigation, gesture recognition[4, 5, 6, 7].

The fiducial markers used in the literature are mostly planar and use the 4 corners of a square as POIs. While 4 coplanars POIs are theorically sufficient for a pose estimation based on a single maker, in practice, the PNP problem becomes ill-posed. Two different solutions of pose estimation may exist, especially at large working distance and large inclination. Figure 1 presents two configurations illustrating this problem of pose ambiguity [8] in a simple case, rotation around the y axis of the marker. The first one (*configuration A*) corresponds to a non-ambiguous observation of the pose, the observed projections allow without difficulty to establish a reliable pose using the classical pose estimation algorithms. In the second configuration, the marker is far from the camera. We observe that two different orientations of the marker lead to an almost identical projection on the optical sensor. In our application, the test conditions correspond to this second configuration. In the absence of measurement noise on the detection of the four corners of the marker, the two poses are unambiguously discernible. In practice, measurement noise exists and the solution obtained by optimization does not always correspond to the real pose of the marker. The global approach proposed here consists in making the placement of the locator more reliable by increasing the number of markers visible on the images.

1.1. Related work

The related work in the field of pose estimation and localization spans various domains, including monocular and multiview camera systems as well as algorithms both with and without dedicated markers. In the realm of single-view pose estimation, the DTAM framework proposed by Newcombe et al. [9] enables real-time dense tracking and mapping, while Kendall and Cipolla [10] address uncertainty modeling in deep learning for camera relocaliza-

*Corresponding author: christian.elmo-kulanesan@univ-smb.fr

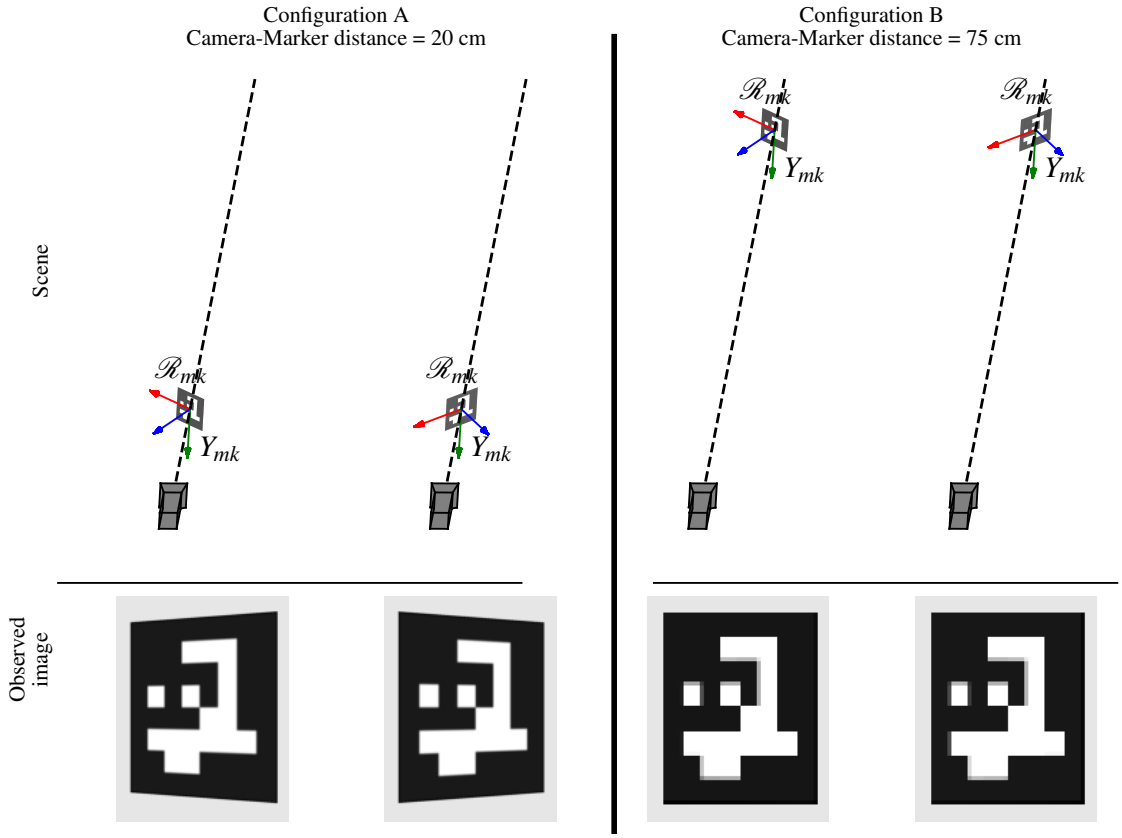

Figure 1: Illustration of the pose ambiguity phenomenon. Two configurations are presented where a marker of 12 mm side is seen by a camera whose sensor is located at two distinct distances. In this illustration, the camera has a resolution of 2048 px × 1536 px and a sensor size of 7.07 mm × 5.3 mm. The focal length is 12 mm. In each configuration, two orientations of the marker differing from a simple rotation θ around its axis Y_m by $\pm 36°$ are presented. In configuration A, the markers are positioned at a distance of 20 cm from the camera. The projections on the images clearly show the perspective effect of the markers. Visually, the vertical edges are of different lengths. In configuration B, the markers are positioned at a distance of 75 cm from the camera. The images observed for these two situations are almost identical despite the very different orientations of the two markers. A small uncertainty on the identification of the spatial position of the 4 points of this quadrilateral can lead to strongly different pose estimation of this marker at the origin of the pose ambiguity.

tion. Expanding to multiview scenarios, the seminal work by Hartley and Zisserman [11] outlines the fundamentals of multiple view geometry, and Snavely et al.'s concept of "photo tourism" [12] introduces the notion of 3D exploration through photo collections.

In the absence of dedicated markers, markerless approaches have gained prominence. Newcombe et al.'s KinectFusion [13] facilitates real-time dense surface mapping and tracking through a depth sensor, and Mur-Artal and Tardós's ORB-SLAM2 [14] presents a versatile open-source SLAM system catering to various camera setups.

Notably, optimization plays a pivotal role in refining these methods. The Levenberg-Marquardt algorithm [15] has been a cornerstone, as Levenberg's method [16] and Marquardt's algorithm [17] significantly contribute to solving nonlinear least-squares problems. Additionally, bundle adjustment, as elaborated by Triggs et al. [18], and alternative optimization techniques such as those detailed by Lepetit and Lourakis [19, 20] continue to enhance the precision and efficiency of pose and localization estimation algorithms by working on a better mathematical representation of Rodrigues' parameters.

Some studies proposed solutions to solve the pose ambiguity by using a fifth or more POIs [21, 22]. In some configurations, these approaches show low reliabilities, as demonstrated by [23].

In a noise-free configuration, this second solution exhibits a higher reprojection error than the true solution and can therefore be easily discarded [8]. In practice, it is common for the bad solution to have a lower reprojection error than the good one and makes it impossible to differentiate them directly [24]. Consequently, a single plane marker cannot always be used to estimate the pose of an object with confidence.

To avoid pose ambiguity, several markers can be positionned in a non-coplanar way on the object [25, 26]. In this paper, this set of markers stuck on a rigid object is called *locator*. A calibration is needed if the relative position of the markers is not precisely controlled. The calibration of a locator can be seen as a Structure-from-Motion (SfM) problem [12, 27] which can be solved offline to obtain an accurate model of the relative position of the fiducial markers within

the locator [28, 29].

As before, this is an ill-posed problem which requires an accurate enough starting point in order to converge towards the solution. The pose ambiguity can jeopardize the establishment of a well-suited starting point. Thus, in some applications, the markers must be placed according to a predefined pattern (e.g. on a plane grid or on the faces of a polyhedron) in order to make a direct resolution of the PNP problem possible [26, 30, 31]. In the more general case considered here where the arrangement of markers is random, a specific methods must be developed to build an initial solution taking into account the pose ambiguity.

1.2. Our contributions

In this paper, we focus on the problem of ambiguity of the pose as well as the identification of markers that are incorrectly detected or damaged but still detected (e.g., a bent marker or partially occluded marker) as shown in the Fig. 2. We propose a new method to construct a reliable estimation of the relative position of fiducial markers. This method allows the elimination of ambiguous positions as well as bad detections. The paper is organized as follows. Section 2 presents the developed method after laying the mathematical foundations. Section 3 describes a series of experiments setup to demonstrate the performance of the proposed method. Section 5 proposes a discussion and a conclusion regarding the results and the overall performance of the method.

2. Proposed method

2.1. Initial concepts and definitions

2.1.1. Definitions

In our approach, a set of planar square fiducial markers $\boldsymbol{M}_i$ with $i \in \{1, \dots, n_M\}$ are randomly glued at the surface of an object with an arbitrary geometry. It is assumed that the markers are not in a co-planar position. This object, noted $\boldsymbol{L}$, is referred to as the locator.

In the present work, ArUco markers [32] are used, but this approach could be similar to all type of square planar fiducial markers such as AprilTag [33] or ARTag [34].

A set of images $\boldsymbol{I}_j$ of the locator with $j \in \{1, \dots, n_I\}$ is created from different viewpoints with a single camera. Each marker $\boldsymbol{M}_i$ and each image $\boldsymbol{I}_j$ are associated with a reference frame. The images are processed with the ArUco method in order to detect the visible markers [35, 36]. The pixel coordinates of the 4 corners of a marker $\boldsymbol{M}_i$ visible on the image $\boldsymbol{I}_j$ are noted $\tilde{\boldsymbol{u}}_{\boldsymbol{M}_i/\mathscr{R}_I}(\boldsymbol{I}_j) \in \mathbb{R}^2 \times \mathbb{R}^4$. The sub-pixel position of these corners is refined using the "AprilTag 2 method" [37].

2.1.2. Affine transformations

When passing from one reference frame to another, the coordinates of a point undergo a rotation $\boldsymbol{r}$ followed by a translation $\boldsymbol{t}$. For the sake of simplicity, we use notations consistent with the ones used by [24]. Accordingly, such an affine transformation is noted $\gamma = (\boldsymbol{r}, \boldsymbol{t})$, with $\boldsymbol{r} = \left[r_x, r_y, r_z\right]^\mathsf{T}$, the rotation vector as defined by Rodrigues convention and $\boldsymbol{t} = \left[t_x, t_y, t_z\right]^\mathsf{T}$, the translation vector. The rotation matrix $\boldsymbol{R}$, bijectively associated to $\boldsymbol{r}$ verifies:

$$\boldsymbol{R} = \boldsymbol{I} + \bar{\boldsymbol{r}} \sin\theta + \bar{\boldsymbol{r}}^2(1 - \cos\theta) \tag{1}$$

Where $\theta = \|\boldsymbol{r}\|$ and $\bar{\boldsymbol{r}}$ is the cross product matrix:

$$\bar{\boldsymbol{r}} = \frac{1}{\theta}\begin{bmatrix} 0 & -r_z & r_y \\ r_z & 0 & -r_x \\ -r_y & r_x & 0 \end{bmatrix}, \tag{2}$$

and $\theta \in [0, 2\pi]$, due to the periodicity properties of rotations. The affine transformation matrix $\Gamma(\gamma) \in \mathbb{R}^4 \times \mathbb{R}^4$ is then defined by:

$$\Gamma(\gamma) = \left[\begin{array}{ccc|c} & \boldsymbol{R} & & \boldsymbol{t} \\ \hline 0 & 0 & 0 & 1 \end{array}\right] \tag{3}$$

Affine transformation matrices $\Gamma_a = \Gamma(\gamma_a)$ and $\Gamma_b = \Gamma(\gamma_b)$ can be composed by matrix product so that $\Gamma_{b,a} = \Gamma_b \otimes \Gamma_a$. By extension, it is assumed that the symbol $(\cdot)$ indicates the composition for γ transformations so that $\gamma_{b,a} = \gamma_b \cdot \gamma_a$.

2.1.3. Single marker pose estimation

In its own reference frame $\mathscr{R}_{\boldsymbol{M}_i}$, the spatial homogeneous coordinates of the corners ς_k with $k \in \{0, \dots, 3\}$ of a marker $\boldsymbol{M}_i$ are consolidated in matrix $\boldsymbol{C}_{\boldsymbol{M}_i/\mathscr{R}_{\boldsymbol{M}_i}}$ such as:

$$\boldsymbol{C}_{\boldsymbol{M}_i/\mathscr{R}_{\boldsymbol{M}_i}} = \frac{1}{2}\overbrace{\begin{bmatrix} -d & d & d & -d \\ -d & -d & d & d \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}}^{\begin{matrix}\varsigma_0 & \varsigma_1 & \varsigma_2 & \varsigma_3 \\ \uparrow & \uparrow & \uparrow & \uparrow\end{matrix}}, \tag{4}$$

with d the side length of each marker. These coordinates can be expressed in any reference frame $\mathscr{R}$ by their homogeneous coordinates:

$$\boldsymbol{C}_{\boldsymbol{M}_i/\mathscr{R}} = \Gamma_{\mathscr{R}\leftarrow\mathscr{R}_{\boldsymbol{M}_i}} \otimes \boldsymbol{C}_{\boldsymbol{M}_i/\boldsymbol{M}_i} = \gamma_{\mathscr{R}\leftarrow\mathscr{R}_{\boldsymbol{M}_i}} \cdot \boldsymbol{C}_{\boldsymbol{M}_i/\mathscr{R}_{\boldsymbol{M}_i}} \tag{5}$$

By composing the resulting coordinates of the corners of the marker $\boldsymbol{M}_i$ with a Ψ operator, these points are projected into the pixel space of the camera on the $\boldsymbol{I}_j$ image. Ψ is a unique affine transformation obtained by calibrating the camera sensor which integrates the change of reference frame from the camera frame $\mathscr{R}_{\boldsymbol{c}}$ to the image frame $\mathscr{R}_{\boldsymbol{I}}$ by perspective projection and a change of scale operation. The coordinates of the projected points will then be expressed as :

$$\boldsymbol{u}_{\boldsymbol{M}_i/\mathscr{R}_{\boldsymbol{I}}}(\gamma(\boldsymbol{I}_j)) = \Psi(\delta) \cdot \gamma(\boldsymbol{I}_j) \cdot \boldsymbol{C}_{\boldsymbol{M}_i/\mathscr{R}_{\boldsymbol{M}_i}} \tag{6}$$

Where $\delta = \left(f_x, f_y, c_x, c_y, k_1, \dots\right)$ is the tuple of intrinsic camera parameters, (f_x, f_y) are the focal lengths parameters, c_x, c_y the optical center parameters and $(k_1, \dots)$ the distortion parameters. The reprojection error of the marker $\boldsymbol{M}_i$ on

the image $\boldsymbol{I}_j$, as a function of $\gamma(\boldsymbol{I}_j)$, is then written:

$$e_{\boldsymbol{M}_i,\boldsymbol{I}_j}(\gamma(\boldsymbol{I}_j)) = \sum_{m=1}^{4}\sum_{l=1}^{2}\left[\Delta\boldsymbol{u}_{\boldsymbol{M}_i/\mathscr{R}_I}(\gamma(\boldsymbol{I}_j))\right]_{l,m}^2, \quad (7)$$

with l and m denote respectively the number of coordinates and the number of corners of the marker $\boldsymbol{M}_i$. $\Delta\boldsymbol{u}_{\boldsymbol{M}_i/\mathscr{R}_I} \in \mathbb{R}^2\times\mathbb{R}^4$ are the reprojection residues in pixels calculated from the difference between the measured corners $\tilde{\boldsymbol{u}}_{\boldsymbol{M}_i/\mathscr{R}_I}(\boldsymbol{I}_j)$ and the projected corners $\boldsymbol{u}_{\boldsymbol{M}_i/\mathscr{R}_I}(\gamma(\boldsymbol{I}_j))$:

$$\Delta\boldsymbol{u}_{\boldsymbol{M}_i/\mathscr{R}_I}(\gamma(\boldsymbol{I}_j)) = \boldsymbol{u}_{\boldsymbol{M}_i/\mathscr{R}_I}(\gamma(\boldsymbol{I}_j)) - \tilde{\boldsymbol{u}}_{\boldsymbol{M}_i/\mathscr{R}_I}(\boldsymbol{I}_j) \quad (8)$$

In theory, the pose of the marker $\boldsymbol{M}_i$ relative to the camera reference frame $\mathscr{R}_c$ by an image I_j can be determined by minimizing the reprojection error according to ${}^{\gamma(\boldsymbol{I}_j)}_{\mathscr{R}_c\leftarrow\mathscr{R}_{\boldsymbol{M}_i}}$:

$${}^{\gamma(\boldsymbol{I}_j)}_{\mathscr{R}_c\leftarrow\mathscr{R}_{\boldsymbol{M}_i}} = \underset{{}^{\gamma^*(\boldsymbol{I}_j)}_{\mathscr{R}_c\leftarrow\mathscr{R}_{\boldsymbol{M}_i}}\in\mathbb{R}^3\times\mathbb{R}^3}{\arg\min} e_{\boldsymbol{M}_i,\boldsymbol{I}_j}({}^{\gamma^*(\boldsymbol{I}_j)}_{\mathscr{R}_c\leftarrow\mathscr{R}_{\boldsymbol{M}_i}}) \quad (9)$$

The knowledge of the pose of each individual marker allows the locator structure to be determined. This structure is retrieved by an optimisation step called bundle adjustment described in the paragraphe 2.1.6. In practice, some pose estimation of individual marker can be wrong, due to pose ambiguity or misdetected markers.

2.1.4. Pose ambiguity

Due to the co-planarity of the corners of each individual marker, the reprojection error $e_{\boldsymbol{M}_i,\boldsymbol{I}_j}(\gamma(\boldsymbol{I}_j))$ can have two minimums [See 3, Fig. 1-3]. When two solutions exist, they are noted ${}^{\gamma(\boldsymbol{I}_j)}_{\mathscr{R}_c\leftarrow\mathscr{R}_{\boldsymbol{M}_i}},_0$ and ${}^{\gamma(\boldsymbol{I}_j)}_{\mathscr{R}_c\leftarrow\mathscr{R}_{\boldsymbol{M}_i}},_1$. One is the real solution and the other is a bad solution. The presence of measurement noise implies that in some cases, the minimum of the reprojection error does not necessarily correspond to the real solution. This problem is called *pose ambiguity* [8, 38]. In the following, we systematically calculate these two poses using the method proposed by [3]. If they are differents, both are kept and the right one will be identified afterward. We used markers of size $d = 12.41$ mm at a typical working distance of $L = 800$ to 1200 mm. In this configuration, the probability of presence of two poses is almost systematic.

2.1.5. Bad markers and bad detections

In practice, the following 2 configurations can also lead to incorrect pose estimation illustrated in Fig. 2:

Bad marker: A damaged marker but still detected. It occures when a marker is no longer flat or if it is stained or scratched. In this case, both poses are systematically incorrect.

Bad detection: Poor marker detection typically caused by a corner being detected at the wrong position. This can happen on partially occluded markers or bad image quality. In this case, the detection results also in two incorrect poses.

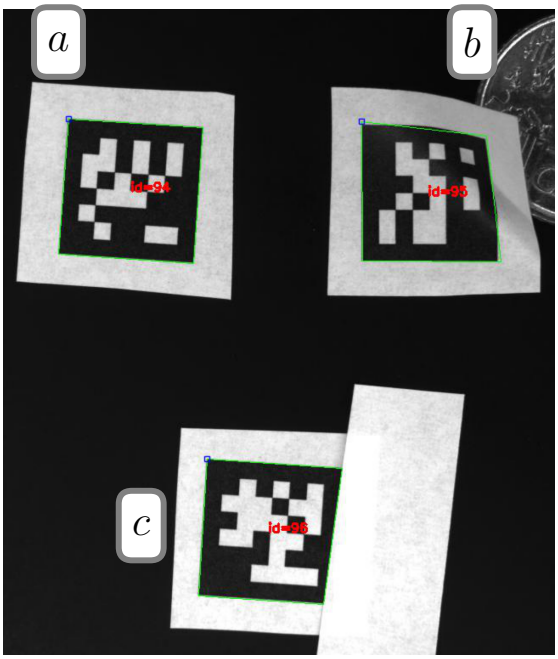

Figure 2: Illustration of two configurations that can lead to incorrect pose estimation. **(a)** Appropriate detection of the marker. **(b)** Bad marker: misdetection of the marker. In this case, it is bent but a stain could produce the same effect. The associated poses are systematically wrong. **(c)** Bad detection: a marker is detected incorrectly. In this case, the marker is partially occluded but is still detected. One of its corners is detected in the wrong position and the associated poses are erroneous.

The detections from these configurations must be eliminated before the bundle adjustment to allow convergence to the correct solution. This implies eliminating the bad detections and not considering the bad markers in the construction of the locator.

2.1.6. Bundle adjustment

The pose and structure of the locator $\boldsymbol{L}$ is determined by solving a bundle adjustment problem. Since the markers are not co-planar, the locator is not affected by pose ambiguity for a single marker. Firstly, among all the markers, a marker $\boldsymbol{M}_0$ is arbitrarily taken as reference within the locator. The structure of the locator will be fully defined by knowledge of $\boldsymbol{\gamma}_S = \left\{\gamma_{\mathscr{R}_{M_0}\leftarrow\mathscr{R}_{M_i}} | i\in\{1,\dots,n_M\}\right\}$. The pose of the locator in the image batch I_j is defined by:

$$\boldsymbol{\gamma}_I = \left\{\gamma(\boldsymbol{I}_j)_{\mathscr{R}_c\leftarrow\mathscr{R}_{M_0}} | j\in\{1,\dots,n_I\}\right\} \quad (10)$$

The reprojection error is defined as :

$$E(\boldsymbol{\gamma}_S^*,\boldsymbol{\gamma}_I^*) = \sum_{i=1}^{n_M}\sum_{j=1}^{n_I}\eta_{\boldsymbol{M}_i,\boldsymbol{I}_j}(\gamma(\boldsymbol{I}_j)_{\mathscr{R}_c\leftarrow\mathscr{R}_{M_0}}\cdot\gamma_{\mathscr{R}_{M_0}\leftarrow\mathscr{R}_{M_i}}), \quad (11)$$

where:

$$\eta_{\boldsymbol{M}_i,\boldsymbol{I}_j}(\gamma) = \begin{cases} e_{\boldsymbol{M}_i,\boldsymbol{I}_j}(\gamma), & D_{\boldsymbol{M}_i,\boldsymbol{I}_j} = 1 \\ 0, & D_{\boldsymbol{M}_i,\boldsymbol{I}_j} = 0 \end{cases}, \tag{12}$$

and D a detection mask defined as follows:

$$D_{\boldsymbol{M}_i,\boldsymbol{I}_j} = \begin{cases} 1, & \exists (i,j) \quad | \quad \boldsymbol{M}_i \subset \boldsymbol{I}_j \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

Consequently, the optimal parameters $\boldsymbol{\gamma}_S$ and $\boldsymbol{\gamma}_I$ are the solutions of:

$$\boldsymbol{\gamma}_S, \boldsymbol{\gamma}_I = \underset{\boldsymbol{\gamma}_S^*, \boldsymbol{\gamma}_I^*}{\arg\min}\, E(\boldsymbol{\gamma}_S^*, \boldsymbol{\gamma}_I^*) \tag{14}$$

The optimal result highly depends on the initial guess provided $(\boldsymbol{\gamma}_{S,0}, \boldsymbol{\gamma}_{I,0})$. We propose a method based on graph theory that is able to handle several problems prior to the bundle adjustment. It is therefore essential to eliminate bad poses beforehand by removing the pose ambiguity and by eliminating bad detections of markers.

2.2. Proposed algorithm

2.2.1. Detection graph and cycle basis

We consider a graph $\mathcal{G}_s$ whose vertices are the markers and the images. The sets of images and markers being independent, each detection $D_{\boldsymbol{M}_i,\boldsymbol{I}_j} = 1$ corresponds to an edge of this graph. A cycle basis $\{\xi_i | i \in \{1,\dots,N_c\}\}$ is then generated from this graph [39]. A basic example is shown in Fig. 3.

2.2.2. Pose classification using cyclic rotational errors

Let us consider the cycle $\xi_0 = (\boldsymbol{M}_1, \boldsymbol{I}_1, \boldsymbol{M}_2, \boldsymbol{I}_0)$ from Fig. 3 with length $N_e = 4$. For each edge of the cycle corresponding to the detection of the marker $\boldsymbol{M}_i$ on the image $\boldsymbol{I}_j$, the true affine transformation is noted $\gamma_{(\boldsymbol{I}_j)_{\mathcal{R}_c \leftarrow \mathcal{R}_{\boldsymbol{M}_i}}}$. By composing these transformations, a residual pose $\hat{\gamma}_{\xi_0} = \left(\hat{\boldsymbol{r}}_{\xi_0}, \hat{\boldsymbol{t}}_{\xi_0}\right)$ for this cycle is calculated as:

$$\begin{aligned} \hat{\gamma}_{\xi_0} = \quad & \gamma_{(\boldsymbol{I}_1)_{\mathcal{R}_{\boldsymbol{M}_2} \leftarrow \mathcal{R}_c}} \cdot \gamma_{(\boldsymbol{I}_0)_{\mathcal{R}_c \leftarrow \mathcal{R}_{\boldsymbol{M}_1}}} \\ & \cdot \gamma_{(\boldsymbol{I}_1)_{\mathcal{R}_{\boldsymbol{M}_1} \leftarrow \mathcal{R}_c}} \cdot \gamma_{(\boldsymbol{I}_0)_{\mathcal{R}_c \leftarrow \mathcal{R}_{\boldsymbol{M}_2}}} \end{aligned} \tag{15}$$

If the pose estimations were free of errors, this residual pose would be null. In practice, there is always some degree of error in the pose estimations. In addition, in case of pose ambiguity, there are two pose estimations $\gamma_{(\boldsymbol{I}_j)_{\mathcal{R}_c \leftarrow \mathcal{R}_{\boldsymbol{M}_i}}, 0}$ and $\gamma_{(\boldsymbol{I}_j)_{\mathcal{R}_c \leftarrow \mathcal{R}_{\boldsymbol{M}_i}}, 1}$. It follows that the residual affine transformation $\hat{\gamma}^k_{\xi_0}$ associated with a given cycle can be computed according to N_p combinations of estimated poses for $k \in \{1,\dots,N_p\}$ where $1 \leq N_p \leq 2^{N_e}$. *A priori*, only one of these combinations should be correct while the others should produce large residual affine transformations. Since the pose ambiguity predominantly affects the rotational component of a transformation γ, the angular residual $\hat{\theta}^k_{\xi_0} = ||\hat{\boldsymbol{r}}^k_{\xi_0}||$ is used as the sole indicator of combination quality. The following criterion is therefore used to validate the combinations associated with a cycle of length N_e:

$$\frac{\hat{\theta}^k_{\xi_0}}{N_e} \leq \hat{\theta}_c, \tag{16}$$

where $\hat{\theta}_c$ is the critical angle below which the angular error per edge is considered negligible. This angle is arbitrarily set up to 0.1° in our application.

In fact, we observe that several combinations of the same cycle may often validate this criterion. This implies that incorrect pose estimations can be involved in combinations that validate the criterion of Eq. 16. There are two main reasons for this observation. First, the presence of almost identical images within the same cycle necessarily implies that two combinations of the cycle will validate the angular criterion. Secondly, fortuitously and due to the large number of pose estimates, we sometimes observe validations of this angular criterion for wrong cycles. A statistical approach can separate the correct pose estimations from the incorrect ones.

A counter $C_{\boldsymbol{I}_j,\boldsymbol{M}_i,k}$ is associated with each pose $\gamma_{(\boldsymbol{I}_j)_{\mathcal{R}_c \leftarrow \mathcal{R}_{\boldsymbol{M}_i}}, k}$ for $k \in \{0,1\}$, which can belong to one of the following classes: bad (B), undetermined (U) and good (G). The goal of our method is to move the poses from their inital class (U) to classes (B) and (G). All the combinations of each cycle are tested using Eq. 16 and each time, one of them is validated, the counters are incremented by 1. We assume that incorrect validations of combinations involving bad poses are in the minority. We thus assume that of two poses associated with the same detection, the good one will be validated at least α times more often than the bad one. The α criterion is formalized by the following equation:

$$\alpha = \frac{\max(C_{\boldsymbol{I}_j,\boldsymbol{M}_i,0}, C_{\boldsymbol{I}_j,\boldsymbol{M}_i,1})}{\min(C_{\boldsymbol{I}_j,\boldsymbol{M}_i,0}, C_{\boldsymbol{I}_j,\boldsymbol{M}_i,1})} \tag{17}$$

To secure the validation of good poses, the value $\alpha = 1.5$ has been chosen.

Tab. 1 proposes an example of implementation of

Detection			Valid. Counter		To class	
Id.	Mark.	Img.	γ_0	γ_1	γ_0	γ_1
0	$\boldsymbol{M}_1$	$\boldsymbol{I}_0$	1	0	G	B
1	$\boldsymbol{M}_1$	$\boldsymbol{I}_1$	2	2	U	U
2	$\boldsymbol{M}_2$	$\boldsymbol{I}_0$	2	8	B	G
3	$\boldsymbol{M}_2$	$\boldsymbol{I}_1$	5		G	
4	$\boldsymbol{M}_3$	$\boldsymbol{I}_1$	0		B	
			...			

Table 1: Example of application of the proposed method in the case of the graph of Fig. 3. Only the poses which are in class (U) at the beginning of the current iteration are represented. The column "To class" indicates to which class the poses are moved. Detections 3 and 4 have no pose ambiguity which explains the absence of γ_1 in these two cases. We can see that except for the poses associated with detection 1, all the poses are initially in (U) are placed in (B) and (G).

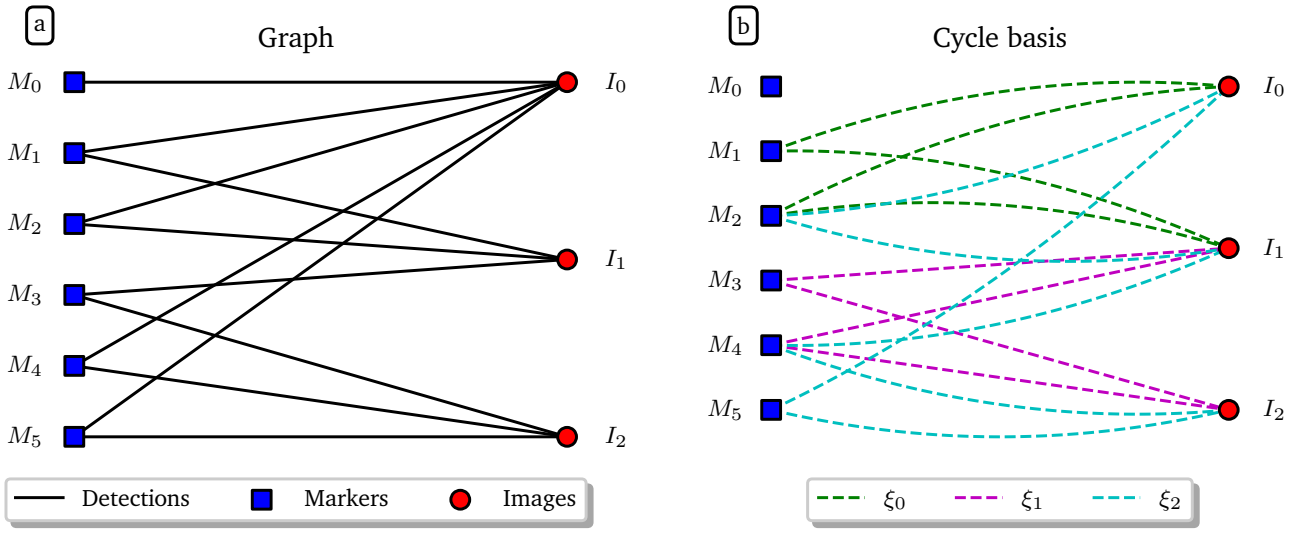

Figure 3: Simple example of construction of a graph $\mathcal{G}_s$. **a**) The sets of 3 images and 6 markers are represented. Each detection $D_{\boldsymbol{M}_i,\boldsymbol{I}_j}$ of a marker $\boldsymbol{M}_i$ on an image $\boldsymbol{I}_j$ corresponds to an edge of the graph. For example, the markers $\boldsymbol{M}_1$, $\boldsymbol{M}_2$ and $\boldsymbol{M}_3$ are here detected on the image $\boldsymbol{I}_1$. **b**) The cycle base allows to browse at least once each edge. Here, 3 cycles ξ_i are necessary to build the base. The marker $\boldsymbol{M}_0$ was observed only once, so it does not belong to any cycle and is therefore excluded from the cycle base.

this approach. The final value of this counter allows to treat each detection associated with poses belonging to (U) individually. In the absence of pose ambiguity, if $C_{\boldsymbol{I}_j,\boldsymbol{M}_i,k} > 0$, the only pose is reclassified as (G), (B) otherwise. If $C_{\boldsymbol{I}_j,\boldsymbol{M}_i,0}/C_{\boldsymbol{I}_j,\boldsymbol{M}_i,1} \geq \alpha$, the pose ${}^{\gamma(\boldsymbol{I}_j)_{\mathcal{R}_c\leftarrow\mathcal{R}_{\boldsymbol{M}_i}}}{}_{,0}$ is considered as (G) and ${}^{\gamma(\boldsymbol{I}_j)_{\mathcal{R}_c\leftarrow\mathcal{R}_{\boldsymbol{M}_i}}}{}_{,1}$ (B). Conversely, if $C_{\boldsymbol{I}_j,\boldsymbol{M}_i,1}/C_{\boldsymbol{I}_j,\boldsymbol{M}_i,0} \geq \alpha$, the pose ${}^{\gamma(\boldsymbol{I}_j)_{\mathcal{R}_c\leftarrow\mathcal{R}_{\boldsymbol{M}_i}}}{}_{,1}$ is considered (G) and ${}^{\gamma(\boldsymbol{I}_j)_{\mathcal{R}_c\leftarrow\mathcal{R}_{\boldsymbol{M}_i}}}{}_{,0}$ (B). Finally, in the remaining cases, neither pose is clearly better and both are kept in (U). The graph $\mathcal{G}_s$ is updated from the detections present in the classes (G) and (U) and a new basis cycle is generated. The $C_{\boldsymbol{I}_j,\boldsymbol{M}_i,k}$ are reset to 0 for all the remaining poses. The equation 16 is once again applied to increment the counters.
This procedure is re-iterated until the size of the (U) class does not decrease during an iteration. If the graph obtained during the iterations is disjoint then the iterations are stopped. The remaining poses in class (U) are then placed in (B) then all poses are classified as (G) or (B).
Consequently, each edge of the final graph $\mathcal{G}_s$ is associated to a single pose belonging to the class (G).

2.2.3. Formalization of the educated guess initial solution

The formulation of the initial solution requires the selection of a reference marker from which the locator structure is described. Among the markers, the one with the lowest eccentricity is chosen as the reference marker and noted $\boldsymbol{M}_0$ as in section 2.1.6. The relative pose of each marker $\boldsymbol{\gamma}_{S,0}$ and each frame $\boldsymbol{\gamma}_{I,0}$ is then estimated by looking for the shortest path between the reference marker and each other node of the simple graph $\mathcal{G}_s$ [40].

The main interest of this algorithm lies in the quality of this initial solution. Indeed, it is close enough to the optimal solution and thus represents an excellent starting point for a bundle adjustment procedure described in section 2.1.6 which allows to further refine this solution to obtain $\boldsymbol{\gamma}_S$ and $\boldsymbol{\gamma}_I$ from equation Eq. 14.

2.3. Locator pose estimation from a single image

After the proposed structural calibration, the pose of the locator can be estimated from a single image. The coordinates of the marker corners are then known relatively to the chosen reference marker $\boldsymbol{M}_0$. The locator's reference frame $\mathcal{R}_L$ is then equal to the reference frame $\mathcal{R}_{\boldsymbol{M}_0}$. Knowing the locator structure $\boldsymbol{\gamma}_S$, the coordinates of the corners of the markers are expressed in the locator's reference frame $\mathcal{R}_L$ such as:

$$\boldsymbol{C}_{\boldsymbol{M}_i/\mathcal{R}_L} = \boldsymbol{\gamma}_S \cdot \boldsymbol{C}_{\boldsymbol{M}_i/\mathcal{R}_{\boldsymbol{M}_i}} \tag{18}$$

Consequently, the coordinates of the markers in the camera reference frame $\mathcal{R}_c$ are given by the following expression:

$$\boldsymbol{C}_{\boldsymbol{M}_i/\mathcal{R}_c}(\boldsymbol{I}_j) = \gamma(\boldsymbol{I}_j)_{\mathcal{R}_c\leftarrow\mathcal{R}_L} \cdot \boldsymbol{C}_{\boldsymbol{M}_i/\mathcal{R}_L} \tag{19}$$

The coordinates of the projections of the corners of the markers visible to the camera ($D_{\boldsymbol{M}_i,\boldsymbol{I}_j} = 1$) projected onto the image $\boldsymbol{I}_j$ can be expressed in the image reference frame $\mathcal{R}_I$ as:

$$\boldsymbol{u}_{\boldsymbol{M}_i/\mathcal{R}_I}(\boldsymbol{I}_j) = \Psi(\delta) \cdot \gamma(\boldsymbol{I}_j)_{\mathcal{R}_c\leftarrow\mathcal{R}_L} \cdot \boldsymbol{C}_{\boldsymbol{M}_i/\mathcal{R}_L} \tag{20}$$

The reprojection error residuals $\Delta\boldsymbol{u}_{\boldsymbol{M}_i/\mathcal{R}_I} \in \mathbb{R}^2 \times \mathbb{R}^4$ give the difference between the projections $\boldsymbol{u}_{\boldsymbol{M}_i/\mathcal{R}_I}$ and the measured detections $\tilde{\boldsymbol{u}}_{\boldsymbol{M}_i/\mathcal{R}_I}(\boldsymbol{I}_j)$. Therefore, if at least two markers are visible, then the reprojection error of the composite marker on an image I_j can be defined as:

$$E_L(\gamma(I_j)_{\mathcal{R}_c \leftarrow \mathcal{R}_L}) = \sum_{i=1}^{n_m} \sum_{m=1}^{4} \sum_{l=1}^{2} \left(\Delta \boldsymbol{u}_{\boldsymbol{M}_i/\mathcal{R}_I}(\gamma(I_j)_{\mathcal{R}_c \leftarrow \mathcal{R}_L}) \right)^2_{l,m,i} \tag{21}$$

By optimization, the locator pose on this image is calculated by minimizing this error by a transformation $\overset{*}{\gamma}(I_j)_{\mathcal{R}_c \leftarrow \mathcal{R}_L}$ such as:

$$\gamma(I_j)_{\mathcal{R}_c \leftarrow \mathcal{R}_L} = \underset{\overset{*}{\gamma}(I_j)_{\mathcal{R}_c \leftarrow \mathcal{R}_L} \in \mathbb{R}^3 \times \mathbb{R}^3}{\arg\min} E_L(\overset{*}{\gamma}(I_j)_{\mathcal{R}_c \leftarrow \mathcal{R}_L}) \tag{22}$$

3. Experiments and results

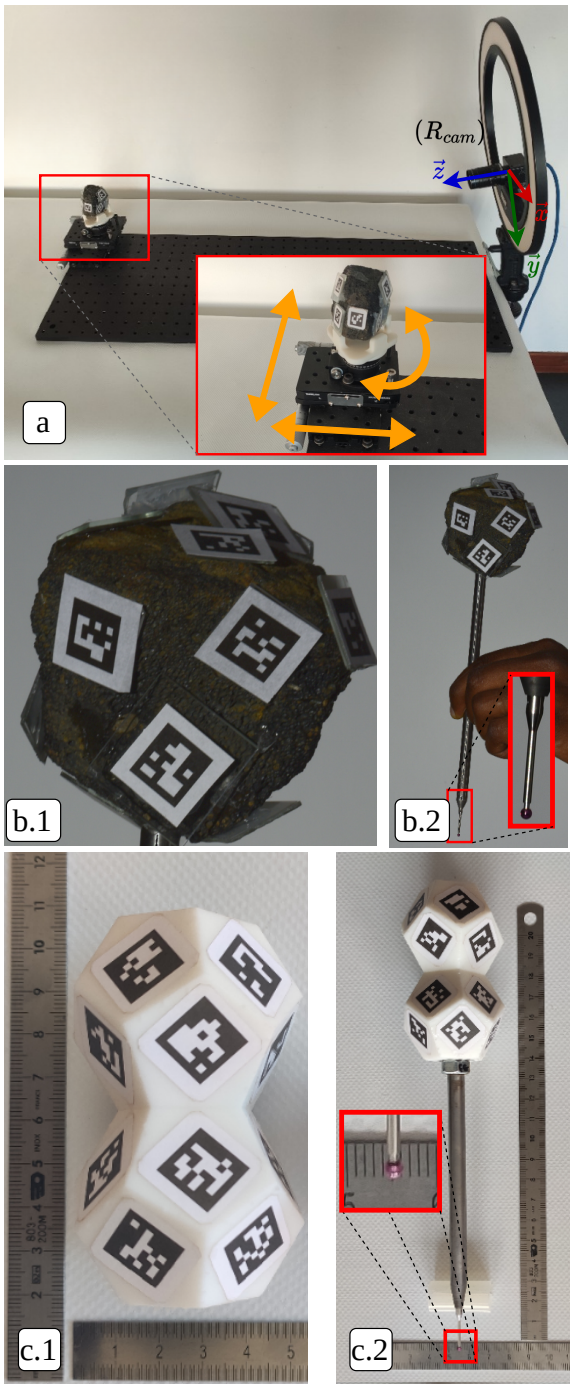

Figure 4: Experimental setup equipments. **a**) Camera, ring light and micrometric stage. The camera reference frame is represented. The micrometric stage allows to perform rotations and translations close to camera x and z axis. **b.1**) Locator made from stone, ArUco were glued on glass. **b.2**) Stylus made from stoned-based locator. **c.1**) The double-dodecahedron used had the dimensions 9 cm × 7 cm × 7 cm. ArUco markers glued on 21 faces of the solid. **c.2**) Stylus using the double-decahedron. The spherical contact probe is shown.

This section presents the experimental validation of the proposed method. A first part is focused on the calibration and measurement uncertainties of locators made from two different shapes of the host objects. A second one concern the use of these locators as stylus where their respective tip are tracked. The last part is dedicated to the surface registration on an artificial scapula with points digitized by one of the presented styli. All tests were performed on a laptop using an Intel Core i9-9880H processor with 32 GB of RAM and Ubuntu 18.04 as the operating system. An SVS-VISTEK EXO542 MU3 camera with a resolution of 5320×3032 pixels, 23 fps max and a 16 mm focal length PENTAX TV lens were used. The scene is lit by a ring light. The intrinsic camera parameters were identified by the chessboard corner method [41]. An overview of the experimental setup is provided in Fig. 4.a. The working distance (distance camera-object) is in the range of 800 mm to 1200 mm. The aperture of the lens diaphragm is set to obtain a depth of field of about 400 mm.

3.1. Calibration and measurement uncertainties of locators

The first proposed locator is composed by 12 mm sided 6 × 6 ArUco markers placed on the surface of a stone of approximate dimension of 70 × 70 × 60 mm^3 with a random shape, see Fig. 4.b.1. The ArUco markers are stuck on glass plates to guarantee their flatness, themselves glued on the stone. These markers were placed randomly on the surface of the object, simply avoiding placing them in a co-planar orientation. Depending on the geometry of the host object, the number of markers visible to the camera can vary. We propose to analyze another geometry in the form of a double-dodecahedron, see Fig. 4.c.1. This geometry offers the advantage of presenting a minimum of 6 ArUco markers whatever the camera's observation point.

As shown in Fig. 5.a, a continuous recording during about 20 s of freehand manipulations of the stylus is performed with great variety of poses. This provided a set of 423 frames. Attention was paid to avoid motion blur, overexposure and to respect the camera depth of field.

The proposed algorithm is applied to calibrate the locator made from a stone. The structure of the locator is calculated from the individual poses of each of the ArUco markers on the set of images (2525 detections lead to 2 × 2525 poses due to the pose ambiguity). The Fig. 5.b represents the graph which connects all the images to the 15 markers glued on the locator. Fig. 5.c shows the corner positions detected by the AprilTag 2 fiducial marker detection method [42]. The position of these same corners determined by projection with our method alone and after the bundle adjustment are represented jointly on these 4 images. Some marker's detections, although visible on some images, are excluded by our method because they don't reach the criteria defined by Eq.16 and Eq. 17 . Figure 5.d presents the starting point given by our method and the optimized state. The mean and standard deviation values of these reprojection errors along x and y axes are respectively : $\bar{X}_x = 3.73e{-}4$ pixel, $\sigma_x = 1.55e{-}1$ pixel; $\bar{X}_y = -1.36e{-}3$ pixel, $\sigma_y = 1.61e{-}1$ pixel. To sum up, 98% of the points have a reprojection errors contained in a square box of 1 pixel side. The points outside this area are due to

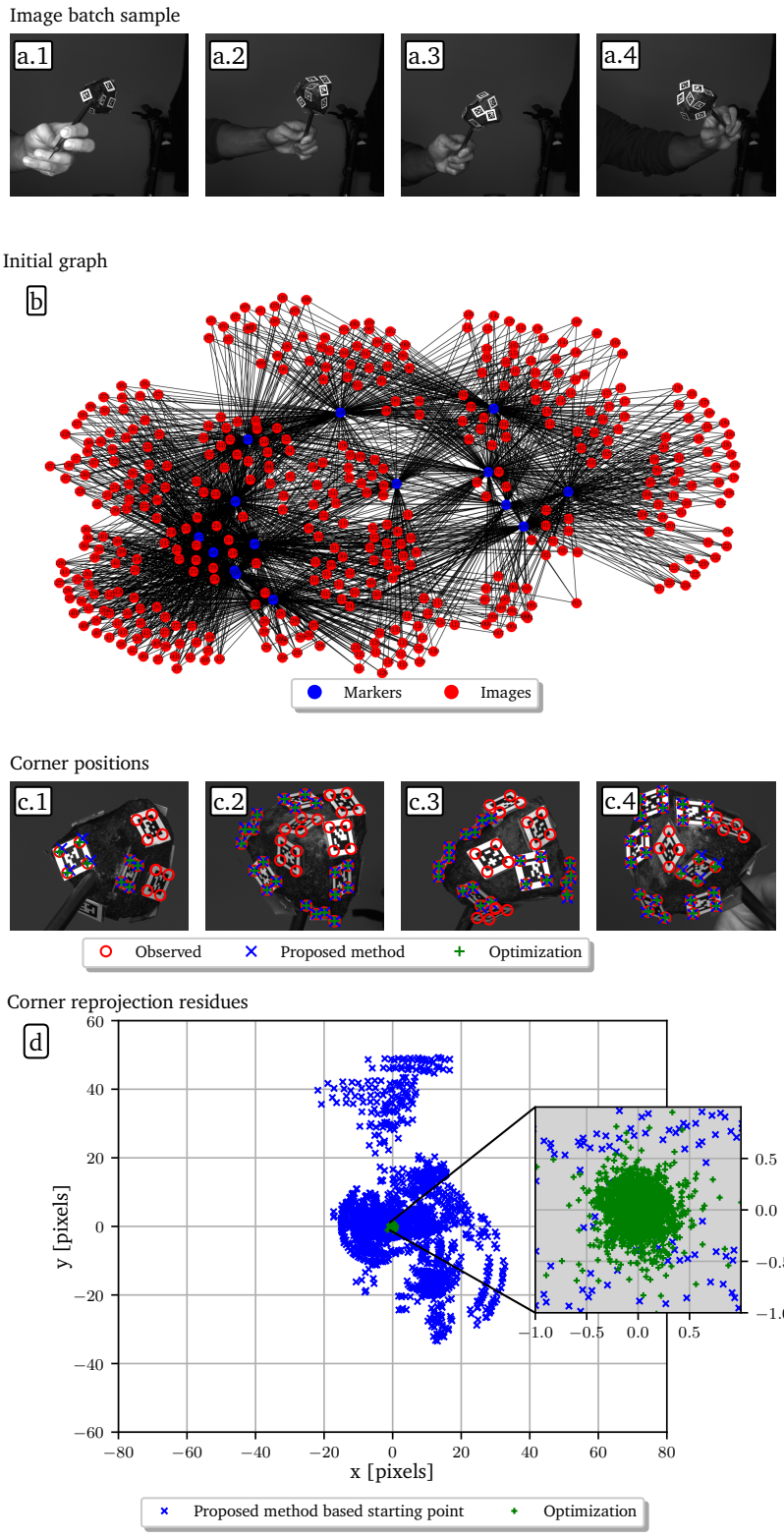

Figure 5: Workflow for the calibration of a rigid body. **a**) Sample of 4 calibration images of the locator object. The stylus was moved manually in front of the camera during image acquisition. All faces of the stone were presented so that all markers are considered during calibration. **b**) Initial graph obtained during the calibration of the locator object, red and blue nodes denote images and markers respectively. The edges represent the detections. **c**) Positions of each corner of each marker on the images. In red, the positions obtained using the AprilTag 2 fiduciary detection method [42]. In blue, the positions obtained after projection of the poses estimated by our method. And finally, in green, the positions obtained after the least squares optimization using the blue positions as initial solutions. **d**) Residuals of reprojection errors of the markers corners from initial solution (blue) and the results of least-square optimization (green) on the all image batch. The majority of the points **(**98%**)** after optimization presents an absolute value reprojection error lower than 0.5 pixel.

markers strongly tilted to the camera's optical axis (approximately greater than 60°). Keeping or removing them only marginally affects the result of the optimization. In practice, the calibration time varies between 3 and 6 minutes depending on the number of images recorded and the number of present markers.

Using equation 22, the pose of the locator obtained after calibration can be calculated from an image in lower than 0.1 second, examples are shown in figure 6. We propose to qualify the measurement uncertainties in translation and in rotation of the resulting locator.

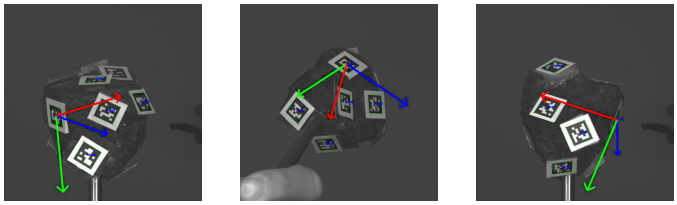

Figure 6: Different poses of the locator. The pose of the locator can be calculated in real time regardless of its orientation if a minimum of two markers are visible. The reference frame associated to the locator correspond to center of the chosen reference marker $\boldsymbol{M}_0$. Note that even if $\boldsymbol{M}_0$ is not visible on an image, the pose of the locator can be still calculated and drawn using the other markers detections.

→ Translation uncertainties

From a reference position noted $\boldsymbol{Q}_0$, two successive translations of target distance $\boldsymbol{L}_t = 25$ mm are performed on the locator using a micrometric translation stage. The directions of these two translations are orthogonal to each other. The first translation moves the locator from the position $\boldsymbol{Q}_0$ to position $\boldsymbol{Q}_1$ along a direction close to the camera optical axis (z axis). The second translation brings the locator from $\boldsymbol{Q}_1$ to $\boldsymbol{Q}_2$ with a direction close to the camera x axis. For each position, $N_i = 10$ images are recorded. We calculate the associated Euclidean distance noted $\boldsymbol{L}$ between the different positions. Note that the distances are calculated from all the combinations of images between the two positions, so there are $N_i^2 = 100$ distances calculated between each position. The whole procedure is repeated for 3 orientations of the locator and the associated distances are noted $\boldsymbol{L}_l$, with $l \in \{0,1,2\}$. We impose that the visible markers between the 3 orientations are different.

The results of these tests are shown in Figure 7. For all the movements made along the camera pseudo z axis, an average distance of 25.08 mm is recorded. The associated standard deviation dispersion is 0.06 mm. Similarly, the measurements taken on the pseudo x camera axis give an average of 25.01 mm and a standard deviation dispersion of 0.03 mm. According to these measurements, a positioning error of less than one millimeter is possible. As expected, the standard deviation dispersion is higher for the movements along the camera optical than the orthogonal direction.

→ Rotation uncertainties

A qualification of the measurement uncertainties in rotation of the locator is proposed. From an initial orientation, successive imposed rotations of $\theta_t = 45°$ are performed on the

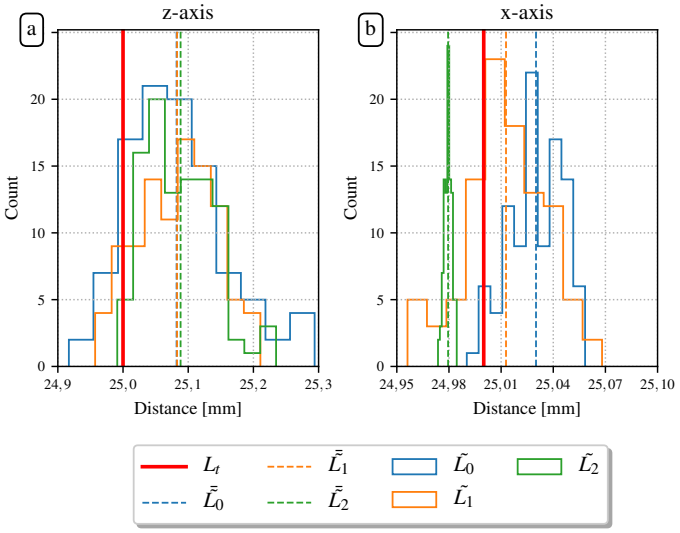

Figure 7: Translation uncertainties of the locator for $l \in \{0,1,2\}$ given orientations. **a**) Distribution of the distances $\boldsymbol{L}_l$ calculated between positions $\boldsymbol{Q}_0$ and $\boldsymbol{Q}_1$, close to camera z axis direction. **b**) Distribution of the distances $\boldsymbol{L}_l$ calculated between positions $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$, close to camera x axis direction.

locator using a rotation plate. A total of $N_r = 7$ rotations are thus imposed. For each orientation, $N_i = 10$ images are taken. Between each rotation, the pose of the locator is estimated and the value of the rotation angle θ performed is then deduced.

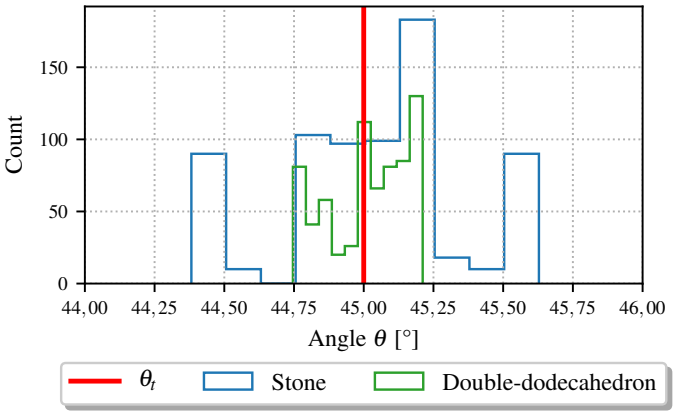

Figure 8: Rotational uncertainties of the stone locator. From a initial orientation, 7 rotations of $\theta_t = 45°$ are imposed.

The results of these tests are given in Fig. 8. For the imposed angles θ, a standard deviation of $\sigma_\theta = 0.33°$ and $\sigma_\theta = 0.14°$ respectively for the stone and the double-dodecahedron locators.

We applied the same translations and rotations tests to the double-dodecahedron locator geometry. The test results for translational uncertainties along the camera z-axis and rotational uncertainties from the two types of locator geometries are given in Table 2.

There is a significant improvement in the results with the use of the double-dodecahedron, since the number of visible markers is systematically greater.

Host object	$\lvert\bar{d_z} - d_t\rvert$	σ_{d_z}	d_t**	$\lvert\bar{\theta} - \theta_t\rvert$	σ_θ	θ_t**
Stone	0.080	0.060	25	0.041	0.330	45
Dd*	0.002	0.041		0.010	0.140	
Unit	[mm]			[°]		

Table 2: Accuracy uncertainties in translation and rotation of the double-dodecahedron shape locator compared to the locator made from a stone.*Double-dodecahedron, **Target values.

Host object	σ_x	σ_y	σ_z
Stone	0.120	0.040	0.240
Dd*	0.096	0.039	0.302
Unit	[mm]		

Table 3: Comparison of the stability of tip point identification for the stone and the double-dodecahedron locators. The results are obtained by calculating the standard deviations of the coordinates of the tip point in the camera frame. *Double-dodecahedron.

3.2. Stylus tracking

A steel rod of length 200 mm with a spherical contact probe (Ø2 mm) was embedded in the stone and the double-dodecahedron locators see Fig. 4.b-c.2.

The position of the stylus tip was identified by using movements in which the tip was kept in contact with a fixed conical surface relative to the camera. The performed movements are random combinations of rotations including rotations of 360° around the axis of revolution of the stylus to ensure observation of all markers. Twenty images are recorded as illustrated by figure 9.

The center of the tip is determined by looking for a unique point $\boldsymbol{P}_{A/\mathscr{R}_L}$ in the locator frame of reference $\mathscr{R}_L$ that would also be stationary in the camera's reference frame $\mathscr{R}_c$ during the movements performed. This condition can be expressed as the minimization of the variance of the coordinates of point $\boldsymbol{P}_{A/\mathscr{R}_c}$ in the camera frame:

$$\boldsymbol{P}_{A/\mathscr{R}_L} = \underset{\overset{*}{\boldsymbol{P}}_{A/\mathscr{R}_L}\in\mathbb{R}^3}{\arg\min}(Var(\left\{\gamma(I_j)_{\mathscr{R}_c\leftarrow\mathscr{R}_L}\cdot\overset{*}{\boldsymbol{P}}_{A/\mathscr{R}_L}\right\})) \tag{23}$$

The coordinates of stylus tip's center are thus known in the locator reference frame.

To quantify uncertainties in the position of the tip of this stylus, we reuse the same types of movements as those applied when the tip identification. These motions are performed via a micrometric translation stage for three positions $\boldsymbol{Q}_0$, $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ of the tip relatively to the camera. The displacement between $\boldsymbol{Q}_0$ and $\boldsymbol{Q}_1$ corresponds to a translation of $d_t = 10$ mm in a direction close to the optical axis camera (z axis). The second displacement $\boldsymbol{Q}_1$ to $\boldsymbol{Q}_2$ corresponds to a translation of in a direction close to the x axis camera with the same d_t value. For each position, 100 images are recorded.

In order to identify the stability of the process, the standards deviations of the tip coordinates are calculated for the $\boldsymbol{Q}_0$ position. The results are shown in the following table 3.

Then, we calculate the distances using the Euclidean norm separating the positions $\boldsymbol{Q}_0$ and $\boldsymbol{Q}_1$ noted d_{z_i} and the positions $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ noted d_{x_j}. The results are shown in figure 10 for the stone locator.

It is thus possible to identify by monocular vision the position of the tip of a stylus observed at a distance of 1 metre with a standard deviation of less than 0.2 mm and a mean deviation of less than 0.1 mm.

The comparison of the uncertainties related to the two styli is given in table Tab. 4.

Host object	$\lvert\bar{d_{z_i}} - d_t\rvert$	$\sigma_{d_{z_i}}$	$\lvert\bar{d_{x_i}} - d_t\rvert$	$\sigma_{d_{x_i}}$	d_t**
Stone	0.063	0.344	0.058	0.169	10
Dd*	0.081	0.193	0.022	0.053	
Unit	[mm]				

Table 4: Comparison of the uncertainties on the position of the tip of the styli between the localizers made from a stone and a double-decahedron. *Double-dodecahedron, **Target distance value.

3.3. Surface registration on an artificial scapula

In the context of surgical navigation, we need to identify the position of a bone structure relative to the camera by simply observing a locator rigidly attached to the bone structure. To illustrate our approach, we propose to find the position of an artificial scapula in relation to the camera using a surface registration. The shape of the used locators were double-dodecahedron geometry.

In the example presented here, a digital scapula model (Fig. 11.a) was obtained from a CT scan of a patient. This STL model was used to reproduce an artificial copy of the scapula by photopolymerisation using the *Form 3+* printer from the manufacturer *Formlabs*. In this test, we propose to use two locators, one attached to the stylus and the second attached to the artificial scapula whom references frames are respectively $\mathscr{R}_{sty}$ and $\mathscr{R}_{scp}$. The coordinates of the points of the numerical model of the scapula are known in a reference frame noted $\mathscr{R}_{stl}$. Using the stylus, scans of certain areas of the artifical scapula are made during which $N_i = 90$ images are taken (Fig. 11.b).

For each image $j \in \{1,\ldots,N_i\}$, the centre of the spherical tip of the stylus noted $Pc_{j/\mathscr{R}_{scp}}$ is known in the reference frame $\mathscr{R}_{scp}$ by the following equation:

$$Pc_{j/\mathscr{R}_{scp}} = \gamma(I_j)_{\mathscr{R}_{scp}\leftarrow\mathscr{R}_c}\cdot\gamma(I_j)_{\mathscr{R}_c\leftarrow\mathscr{R}_{sty}}\cdot Pc_{/\mathscr{R}_{sty}} \tag{24}$$

Using the proximity.signed_distance() f function from *Trimesh* module (Python3), we can calculate the signed distance, noted $\boldsymbol{ls}$, between any point expressed in $\mathscr{R}_{stl}$ reference and the surface $\boldsymbol{S}$ forming the numerical model. The

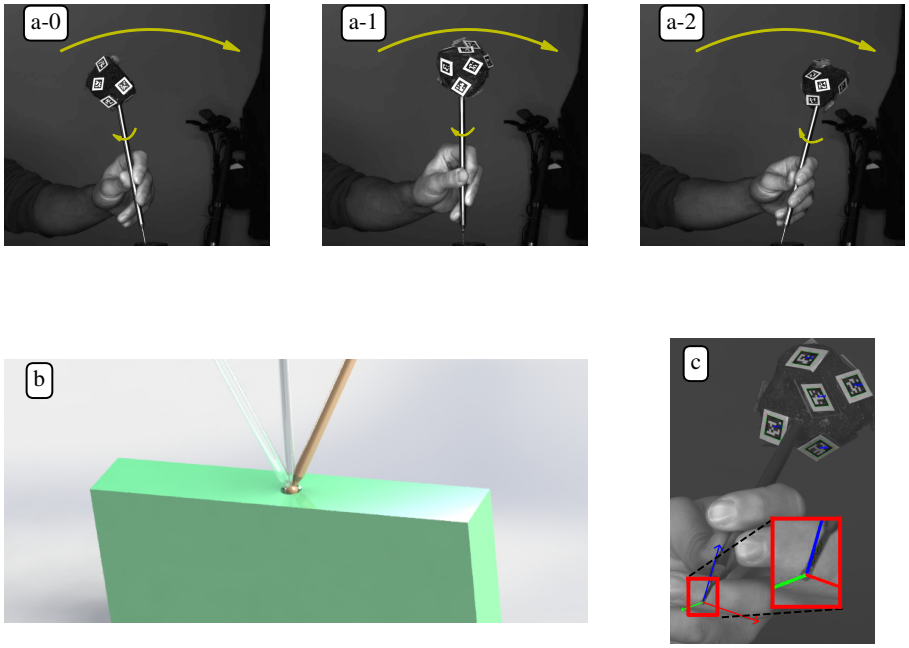

Figure 9: Stylus tip identification. **a**) Sample calibration images of the locator tip. **b**) Illustration of the movements performed during image capture, the tip of the stylus is held in a conical notch. The operator manipulates the stylus while keeping contact with the notch. At the same time, the operator rotates the stylus along the axis of the stainless steel rod to present all the faces of the locator object. The range of motion of the locator is about ±100 mm around the vertical position. **c** Photography of the tip of the stylus. The origin of the drawn reference frame is positioned on the tip of the stylus.

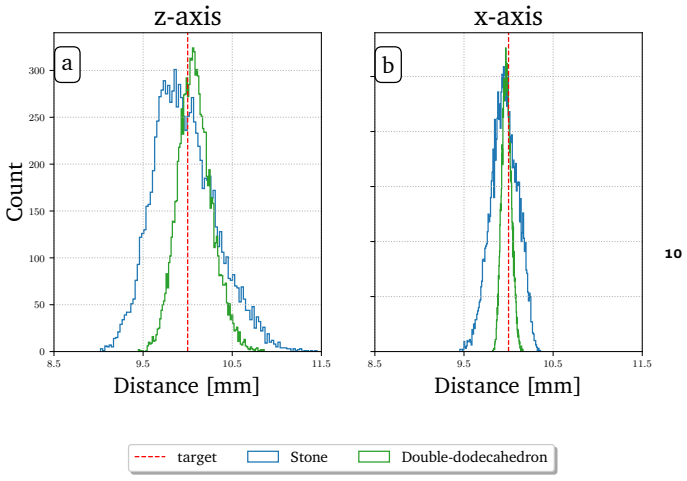

Figure 10: Styli-tip distance of the locators dispersion where target value is 10 mm. **a**) Distribution of distances d_{z_i} measured between positions $\boldsymbol{P}_{0/\mathscr{R}_c}$ and $\boldsymbol{P}_{1/\mathscr{R}_c}$ along an axis close to the z-axis camera. **b**) Distribution of distances d_{x_i} measured between positions $\boldsymbol{P}_{1/\mathscr{R}_c}$ and $\boldsymbol{P}_{2/\mathscr{R}_c}$ along an axis close to the x-axis camera.

sign of $\boldsymbol{ls}$ is negative when the considered point is outside the surface $\boldsymbol{S}$.

During scanning, the actual probe point is shifted by the radius of the contact sphere of radius $r = 1$ mm. Taking into account the offset, we can write the following optimisation equation to identify the rigid transformation $\gamma_{\mathscr{R}_{stl}\leftarrow\mathscr{R}_{scp}}$ between the digital model and the scapula locator as :

$$\gamma_{\mathscr{R}_{scp}\leftarrow\mathscr{R}_{stl}} = \underset{\overset{*}{\gamma}_{\mathscr{R}_{scp}\leftarrow\mathscr{R}_{stl}}}{\arg\min} \sum_{j=1}^{Ni} \left[f\left(\overset{*}{\gamma}_{\mathscr{R}_{stl}\leftarrow\mathscr{R}_{scp}} \cdot Pc_{j/\mathscr{R}_{scp}}, \boldsymbol{S} \right) + r \right]^2 \tag{25}$$

Then the coordinates of the points $Pc_{j/\mathscr{R}_{stl}}$ can then be expressed in the reference frame $\mathscr{R}_{stl}$ according to the following equation:

$$Pc_{j/\mathscr{R}_{stl}} = \gamma_{\mathscr{R}_{stl}\leftarrow\mathscr{R}_{scp}} \cdot Pc_{j/\mathscr{R}_{scp}} \tag{26}$$

We have repositioned these points in the stl reference frame to represent them on the surface of the numerical model (Fig. 11.c). We evaluated the solution found by calculating the obtained signed distance $\boldsymbol{ls}$ between the points $Pc_{j/\mathscr{R}_{stl}}$ and the surface $\boldsymbol{S}$ forming the numerical model.

The results obtained are presented in figure 12. The mean value of the signed distance $\overline{ls_j + r} = 0.011$ mm is observed with a standard deviation $\sigma_{ls_j} = 0.26$ mm.

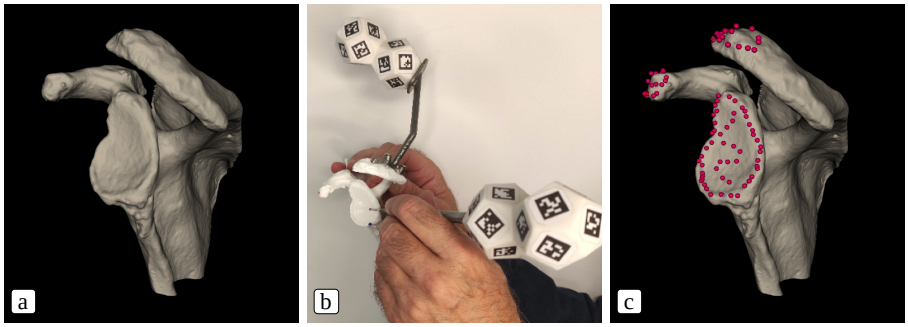

Figure 11: **a**) Numerical model of the scapula from CT-scan of cadaveric scapula. **b**) Scannning of the artificial scapula. One of the double-dodecahedron shape locator is rigidly attached to the artificial scapula. The scapula and pen are manipulated freehand. The digitalisation of the points is made using the stylus with a second double-dodecahedron shape locator. **c**) Registration of the scanned points on the artificial scapula with the digital model of the scapula.

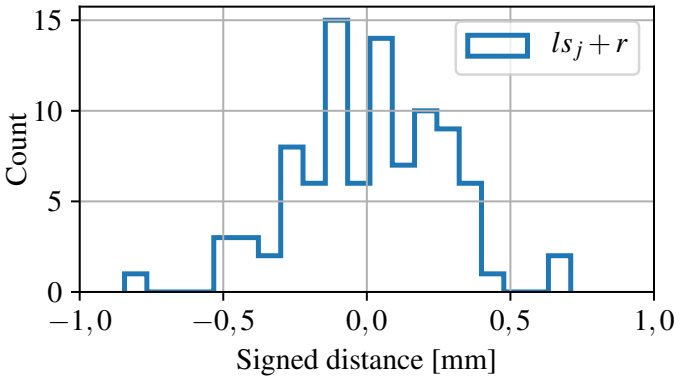

Figure 12: Distribution of the signed distance $ls_j + r$ between the measured points $Pc_{j/\mathcal{R}_{stl}}$ and the surface $\boldsymbol{S}$ of the digital scapula model taking into account the offset of the contact sphere radius r.

4. Discussion

Based on the aforementioned results, we evaluate the quality of the following aspects: calibration of locator, the enhanced value associated with the geometry of the double-dodecahedron, positioning of the tip of the stylus, and surface registration for the presented medical application.

A calibration step allow to find the relative positions of the planar markers stuck on these objects by simply observing images of them manipulated freehand in front of the camera. This calibration method is based on graph theory. After optimization, we found the structure of the locator. The reprojection errors observed on the images were mostly contained in a square of 1 pixel.

To test the found solution, controlled movements were imposed on the locators alone. The analysis of the uncertainties for these movements showed that the geometry of double-dodecahedron with at least 6 planar markers performs best. For this locator, we obtain uncertainties of the order of 41 µm in translation along the camera optical axis (the most critical axis in monocular vision) and uncertainties in rotation of 0.14°.

For the specific needs of shoulder surgery, a probing device is necessary. A stylus of approximately 20 cm in length consisting of a spherical tip and a double dodecahedron locator is proposed. Tests from paragraph 3.2 have shown that it is possible to track the centre of this sphere with an uncertainty of about 0.2 mm.

This stylus was used to perform a surface registration between an artificial scapula and its digital model based on the scanning of a restricted area of this scapula (glenoid cavity). The deviations between the set of scanned points and the mesh of the digital model are all less than 1 mm This accuracy is fully compatible with the expectations of orthopaedic shoulder surgery. The results obtained from cadaveric shoulders will be presented in a forthcoming publication. The calibration time of the locators and the identification of the stylus tip do not impact the operating time since these steps can be performed preoperatively. However, the palpation of the glenoid cavity and the surface registration are included in the operating time (about 2 minutes in total).

Improvement perspectives are possible for this monocular localization. Currently, the calibration of the locator with a hundred images in input takes about 1 minute. The cycle basis obtained from the graph tool proposes short cycles but also longer cycles, the latter penalizing the processing time. A reflection to limit this type of long cycle has to be conducted. More generally, structural optimizations of the algorithm should allow an improvement of the processing speed using parallelization approaches.

5. Conclusion

In this work, a method to track in real time an object postionned at about 1 meter from the camera by monocular vision was proposed. From a random shape object (stone) or from a more controlled geometry (double-dodecahedron), 3D locators based on a cluster of planar markers were analysed.

The resulting locators have translation uncertainties on the order of one-tenth of a millimeter and orientation uncertainties on the order of one-hundredth of a degree. We have also shown that these locators can be used for surface registration with sub-millimetre uncertainties. These uncertainties are reached whatever the visible plane markers of this locator. These results, along with the locators' geometric versatility, show great potential for their implementation in orthopedic surgical navigation applications. In this application, locators can be customized for use with medical tools such as probing styluses, drills, surgical saws, implants, or prostheses.

The freedom offered by this approach allows to imagine applications for other types of surgeries but also for various fields of applications requiring real-time sub-milimetric localizations between different objects.

6. Acknowledgements

This work was financed by the European program INTERREG France-Switzerland.

References

[1] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981. 1

[2] S. Li, C. Xu, and M. Xie, "A robust o (n) solution to the perspective-n-point problem," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 7, pp. 1444–1450, 2012. 1

[3] G. Schweighofer and A. Pinz, "Robust Pose Estimation from a Planar Target," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 28, pp. 2024–2030, Dec. 2006. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence. 1, 4

[4] J. Wang, C. Sadler, C. F. Montoya, and J. C. Liu, "Optimizing Ground Vehicle Tracking Using Unmanned Aerial Vehicle and Embedded Apriltag Design," in *2016 International Conference on Computational Science and Computational Intelligence (CSCI)*, pp. 739–744, Dec. 2016. 1

[5] M. F. Sani and G. Karimian, "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)*, (Kuching), pp. 102–107, IEEE, Nov. 2017. 1

[6] H. Ishiyama and S. Kurabayashi, "Monochrome glove: A robust real-time hand gesture recognition method by using a fabric glove with design of structured markers," in *2016 IEEE Virtual Reality (VR)*, pp. 187–188, Mar. 2016. ISSN: 2375-5334. 1

[7] X. Zhang, S. Fronz, and N. Navab, "Visual marker detection and decoding in AR systems: a comparative study," in *Proceedings. International Symposium on Mixed and Augmented Reality*, pp. 97–106, Oct. 2002. 1

[8] D. Oberkampf, D. F. DeMenthon, and L. S. Davis, "Iterative Pose Estimation Using Coplanar Feature Points," *Computer Vision and Image Understanding*, vol. 63, pp. 495–511, May 1996. 1, 2, 4

[9] R. A. Newcombe, S. J. Lovegrove, and A. J. Davison, "Dtam: Dense tracking and mapping in real-time," in *2011 international conference on computer vision*, pp. 2320–2327, IEEE, 2011. 1

[10] A. Kendall and R. Cipolla, "Modelling uncertainty in deep learning for camera relocalization," in *2016 IEEE international conference on Robotics and Automation (ICRA)*, pp. 4762–4769, IEEE, 2016. 1

[11] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003. 2

[12] N. Snavely, S. M. Seitz, and R. Szeliski, "Photo tourism: exploring photo collections in 3d," in *ACM siggraph 2006 papers*, pp. 835–846, 2006. 2

[13] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohi, J. Shotton, S. Hodges, and A. Fitzgibbon, "Kinectfusion: Real-time dense surface mapping and tracking," in *2011 10th IEEE international symposium on mixed and augmented reality*, pp. 127–136, Ieee, 2011. 2

[14] R. Mur-Artal and J. D. Tardós, "Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras," *IEEE transactions on robotics*, vol. 33, no. 5, pp. 1255–1262, 2017. 2

[15] J. J. Moré, "The levenberg-marquardt algorithm: implementation and theory," in *Numerical analysis: proceedings of the biennial Conference held at Dundee, June 28–July 1, 1977*, pp. 105–116, Springer, 2006. 2

[16] K. Levenberg, "A method for the solution of certain non-linear problems in least squares," *Quarterly of applied mathematics*, vol. 2, no. 2, pp. 164–168, 1944. 2

[17] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *Journal of the society for Industrial and Applied Mathematics*, vol. 11, no. 2, pp. 431–441, 1963. 2

[18] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, "Bundle adjustment—a modern synthesis," in *Vision Algorithms: Theory and Practice: International Workshop on Vision Algorithms Corfu, Greece, September 21–22, 1999 Proceedings*, pp. 298–372, Springer, 2000. 2

[19] V. Lepetit, F. Moreno-Noguer, and P. Fua, "Ep n p: An accurate o (n) solution to the p n p problem," *International journal of computer vision*, vol. 81, pp. 155–166, 2009. 2

[20] M. Lourakis and G. Terzakis, "A globally optimal method for the pnp problem with mrp rotation parameterization," in *2020 25th International Conference on Pattern Recognition (ICPR)*, p. 3058–3063, Jan 2021. 2

[21] D. Nister, "An efficient solution to the five-point relative pose problem," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, p. 756–770, Jun 2004. 2

[22] D. Barath, M. Polic, W. Förstner, T. Sattler, and Z. Kukelova, "Making a ngeecoomrertersypocnodmenpcuetsawtioonrk in camera," p. 17. 2

[23] S. S ˇegvic, G. Schweighofer, and A. Pinz, "Performance evaluation of the five-point relative pose with emphasis on planar scenes1," p. 8. 2

[24] R. Muñoz-Salinas, M. J. Marín-Jimenez, and R. Medina-Carnicer, "SPM-SLAM: Simultaneous localization and mapping with squared planar markers," *Pattern Recognition*, vol. 86, pp. 156–171, Feb. 2019. 2, 3

[25] S. Garrido-Jurado, R. Muñoz-Salinas, F. Madrid-Cuevas, and M. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognition*, vol. 47, pp. 2280–2292, June 2014. 2

[26] P.-C. Wu, R. Wang, K. Kin, C. Twigg, S. Han, M.-H. Yang, and S.-Y. Chien, "DodecaPen: Accurate 6DoF Tracking of a Passive Stylus," in *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, UIST '17, (Québec City, QC, Canada), pp. 365–374, Association for Computing Machinery, Oct. 2017. 2, 3

[27] C. Wu, "Towards linear-time incremental structure from motion," in *2013 International Conference on 3D Vision-3DV 2013*, pp. 127–134, IEEE, 2013. 2

[28] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, "Bundle Adjustment — A Modern Synthesis," in *Vision Algorithms: Theory and Practice* (G. Goos, J. Hartmanis, J. van Leeuwen, B. Triggs, A. Zisserman, and R. Szeliski, eds.), vol. 1883, pp. 298–372, Berlin, Heidelberg: Springer Berlin Heidelberg, 2000. Series Title: Lecture Notes in Computer Science. 3

[29] J. L. Schonberger and J.-M. Frahm, "Structure-From-Motion Revisited," pp. 4104–4113, 2016. 3

[30] D. Kotake, S. Uchiyama, and H. Yamamoto, "A marker calibration method utilizing a priori knowledge on marker arrangement," in *Third IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 89–98, Nov. 2004. 3

[31] H. Kato, M. Billinghurst, I. Poupyrev, K. Imamoto, and K. Tachibana, "Virtual object manipulation on a table-top ar environment," in *in ISAR 2000: Proceedings of the IEEE and ACM International Symposium on Augmented Reality*, pp. 111–119, 2000. 3

[32] R. Munoz-Salinas, "Aruco: a minimal library for augmented reality applications based on opencv," *Universidad de Córdoba*, vol. 386, 2012. 3

[33] E. Olson, "Apriltag: A robust and flexible visual fiducial system," in *2011 IEEE International Conference on Robotics and Automation*, pp. 3400–3407, IEEE, 2011. 3

[34] M. Fiala, "ARTag, a Fiducial Marker System Using Digital Techniques," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 2, (San Diego, CA, USA), pp. 590–596, IEEE, 2005. 3

[35] F. J. Romero-Ramirez, R. Muñoz-Salinas, and R. Medina-Carnicer, "Speeded up detection of squared fiducial markers," *Image and Vision Computing*, vol. 76, pp. 38–47, Aug. 2018. 3

[36] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and R. Medina-Carnicer, "Generation of fiducial marker dictionaries using Mixed Integer Linear Programming," *Pattern Recognition*, vol. 51, pp. 481–491, Mar. 2016. 3

[37] J. Wang and E. Olson, "AprilTag 2: Efficient and robust fiducial detection," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4193–4198, Oct. 2016. ISSN: 2153-0866. 3

[38] T. Collins and A. Bartoli, "Infinitesimal Plane-Based Pose Estimation," *Int J Comput Vis*, vol. 109, pp. 252–286, Sept. 2014. 4

[39] K. Paton, "An algorithm for finding a fundamental set of cycles of a graph," *Commun. ACM*, vol. 12, pp. 514–518, Sept. 1969. 5

[40] E. W. Dijkstra *et al.*, "A note on two problems in connexion with graphs," *Numerische mathematik*, vol. 1, no. 1, pp. 269–271, 1959. 6

[41] A. De la Escalera and J. M. Armingol, "Automatic chessboard detection for intrinsic and extrinsic camera parameter calibration," *Sensors*, vol. 10, no. 3, pp. 2027–2044, 2010. 8

[42] J. Wang and E. Olson, "Apriltag 2: Efficient and robust fiducial detection," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4193–4198, IEEE, 2016. 8, 9