

Document downloaded from:

<http://hdl.handle.net/10251/151163>

This paper must be cited as:

Albiol Colomer, F.; Corbi, A.; Albiol Colomer, A. (2017). Evaluation of modern camera calibration techniques for conventional diagnostic X-ray imaging settings. *Radiological Physics and Technology*. 10(1):68-81. <https://doi.org/10.1007/s12194-016-0369-y>



The final publication is available at

<https://doi.org/10.1007/s12194-016-0369-y>

Copyright Springer-Verlag

Additional Information

Evaluation of modern camera calibration techniques for conventional diagnostic X-ray imaging settings

Francisco Albiol¹ · Alberto Corbi¹ · Alberto Albiol²

Abstract We explore three different alternatives for obtaining intrinsic and extrinsic parameters in conventional diagnostic X-ray frameworks: the direct linear transform (DLT), the Zhang method, and the Tsai approach. We analyze and describe the computational, operational, and mathematical background differences for these algorithms when they are applied to ordinary radiograph acquisition. For our study, we developed an initial 3D calibration frame with tin cross-shaped fiducials at specific locations. The three studied methods enable the derivation of projection matrices from 3D to 2D point correlations. We propose a set of metrics to compare the efficiency of each technique. One of these metrics consists of the calculation of the detector pixel density, which can be also included as part of the quality control sequence in general X-ray settings. The results show a clear superiority of the DLT approach, both in accuracy and operational suitability. We paid special attention to the Zhang calibration method. Although this technique has been extensively implemented in the field of computer vision, it has rarely been tested in depth in common radiograph production scenarios. Zhang’s approach can operate on much simpler and more affordable 2D calibration frames, which were also tested in our research. We experimentally confirm that even three or four plane-image correspondences achieve accurate focal lengths.

Keywords Conventional X-ray camera calibration · Detector resolution · Intrinsic and extrinsic parameters · Zhang’s method · Direct linear transform · Tsai’s approach

1 Introduction

Camera calibration is an important preprocessing step in computer vision applications and is significant for daily diagnostic X-ray imaging scenarios.

X-ray devices—when perceived as cameras—are composed of a Röntgen radiation source and an independent sensitive surface. Figure 1 (left) and (right) show a comparison of the methods of image reconstruction using a conventional pinhole camera device and X-ray equipment, respectively. One of the key differences between the two devices is that in the case of the X-ray equipment, a projected point Q_i is located between the anode C , which plays the role of the optical center, and the detector. The anode usually consists of a small area (actual focal spot) that increases as the X-ray tube ages. However, in this work, we assume, for the sake of simplicity, an almost null sized effective focal spot.

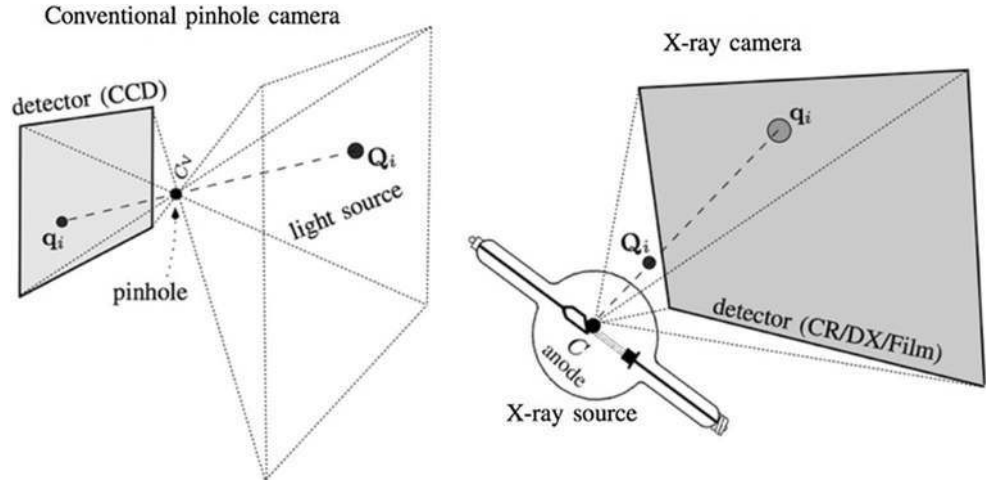
The pinhole camera representation has been applied to X-rays in many previous works [1–4]. Despite the increasing and reliable literature contributions on the subject, very few authors explicitly combine ordinary radiograph generation and modern computer vision camera calibration techniques. An adaptation of Tsai’s approach [5] was examined in the study of Miller et al. [3] to obtain measurements from planar and non-planar targets. Moura et al. [6] employed the direct linear transform (DLT) calibration process by enhancing X-ray systems with laser rangefinders. DLT is also examined by Schumann et al. [7], who used a 3D phantom and applied

✉ Alberto Corbi
alberto.corbi@ific.uv.es

¹ Instituto de Física Corpuscular, Consejo Superior de Investigaciones Científicas, Universitat de València, Valencia, Spain

² Universidad Politécnica de València, Valencia, Spain

Fig. 1 X-ray camera (*right*) vs. pinhole camera (*left*). In both devices, \mathbf{Q}_i is a 3D point whose projection in the image is \mathbf{q}_i



it to orthopedics. Some research approaches do not use a specific calibration method but specially devised techniques involving nonlinear optimizations. For instance, Selby et al. [8] apply these procedures to cylinder-shaped frames. In addition, phantom grids are investigated by Moura et al. [9] along with the minimization of the retro-projection error.

In this paper, we discuss different methods and algorithms for the geometrical calibration of X-ray systems. We pay special attention to Zhang’s method which, to our knowledge, has seldom been applied to X-ray imaging. We start this paper by summarizing how the pinhole camera representation can be applied to X-rays. Next, we analyze and compare each calibration method in different scenarios. For this purpose, we developed an improved calibration frame or device under test (DUT) and a second flat panel. The goal of these structures is to help us establish correspondences between 3D world/scene points and 2D radiographic image projections from different X-ray beam sources or frame locations (also known as poses). From these view-to-radiograph correlations, specific X-ray optical parameters can be derived. We subsequently establish several test metrics to objectively compare all techniques. Finally, we present our results and discuss the particularities of each methodology when applied in the scope of typical X-ray diagnostic protocols for patient diagnosis and object inspection.

2 Background on the geometry of X-ray imaging systems

The concept of representing X-ray imaging systems as pinhole devices requires some geometrical and operational considerations that make them distinctly different from conventional systems.

2.1 X-ray systems as pinhole cameras

As mentioned above, X-ray systems can be operatively simplified and geometrically modeled as pinhole cameras. An introduction to the subject, as applied in the field of medical imaging, can be found in Medioni et al. [10] and Bushong [11]. In a radiological device, as shown in Fig. 1 (right), the position of the pinhole is represented by the anode, which, in turn, represents the optical center C . In the case of planar X-ray imaging, C is usually a fixed and small area located in a rotating tungsten disk. Spherical aberrations, radial distortions, and skew can be ignored without loss of generality. Every \mathbf{Q}_i point is then projected to a 2D shadow in a specific coordinate \mathbf{q}_i in the sensor. In X-ray imaging, C is the origin of the photon beam, whereas in a conventional camera (Fig. 1 (left)), the source of the photon beam is the photographed object, which radiates the reflected light. This diffused light enters the camera through the pinhole and reaches the detector.

The sensor part of an X-ray setup is usually a photographic film or an array of dots (imaging plate or IP) that is sensitive to this type of radiation. Depending on how the information is read, we mainly have digital (DX) or computed (CR) radiography modalities. Both divide the detector surface into sensing pixels with resolution λ that may vary between manufacturers and clinical protocols. The parameter λ is usually independent of the orientation ($\lambda_x = \lambda_y$) and is typically referred to as the linear resolution λ_o , which is provided by the manufacturer.

2.2 World, anode, and detector reference frames

In the specific case of X-ray pinhole cameras, it is appropriate to work with different reference frames, all of which are presented in Fig. 2. The first and most intuitive one is the world coordinate system, $W = (\hat{X}_W, \hat{Y}_W, \hat{Z}_W)$, which is

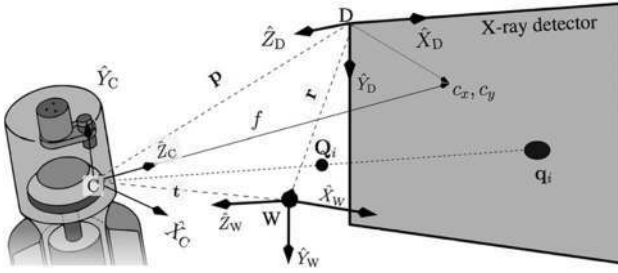


Fig. 2 World (W), detector (D), and anode (C) reference frames. A Q_i object is shown in W frame, together with its corresponding projection q_i in the D system and relative to the C reference frame

usually placed at a known 3D point in the radiographed scene, typically, over a frame or DUT. The second one is the coordinate frame attached to the detector plate itself, $D = (\hat{X}_D, \hat{Y}_D, \hat{Z}_D)$, whose origin is normally coincident with the upper leftmost pixel. The need to define the D system originates from the physical fact of working with a radiation-responsive layer that is not tied to C and from the possibility of moving the anode and/or detector with complete freedom around the object in the world coordinate system. Finally, the coordinate system $C = (\hat{X}_C, \hat{Y}_C, \hat{Z}_C)$ has its origin in C itself.

Any point \mathbf{X} in the 3D space can be associated with any of the aforementioned reference frames. For instance, if \mathbf{X}_C represents the coordinates of \mathbf{X} relatively to C and \mathbf{X}_D is its representation relatively to D , then, from Fig. 2, it can be verified that:

$$\mathbf{X}_D = \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}}_{\pi \text{ rotation about } \hat{X}_C} \cdot \mathbf{X}_C + \mathbf{p} \quad (1)$$

where $\mathbf{p} = (c_x, c_y, f)$ is known as the principal point and represents the coordinates of C in D . The line that passes through C and is perpendicular to the detector plane is the principal axis and intersects the image plane at the point $(c_x, c_y, 0)$. The distance between C and the detector plane is the focal length f . The anode C can also be expressed in W coordinates by means of vector \mathbf{t} , which starts at the origin of W and ends at C .

Finally, if \mathbf{t} and \mathbf{p} are known, we can derive vector \mathbf{r} :

$$\mathbf{r} = \mathbf{p} - \mathbf{t} \quad (2)$$

which connects the W and D reference frames.

2.3 Camera calibration and projection matrices

As discussed above, a prerequisite for any application in the field of computer vision is the calibration of the camera. This step is necessary to determine the pose between the

imaging system and real-world objects. This calibration involves the calculation of five intrinsic (internal) and six extrinsic (external) parameters, which can be grouped into a 3×4 matrix, the camera projection/calibration matrix P .

Mathematically, P maps 3D points—expressed in W coordinates—to 2D points using the expression: $\hat{\mathbf{q}}_i = P \cdot \hat{\mathbf{Q}}_i$, where each $\hat{\mathbf{q}}_i$ is an image point and $\hat{\mathbf{Q}}_i$ is a W -referenced point, both expressed in homogeneous coordinates.

As mentioned above, P can be decomposed into two blocks of intrinsic (K) and extrinsic parameters. The K transformation matrix projects 3D points expressed in terms of the D reference frame to their corresponding image location on the same reference frame:

$$K = \begin{pmatrix} \alpha_x & s & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{pmatrix} = \underbrace{\begin{pmatrix} \lambda_x & 0 & 0 \\ 0 & \lambda_y & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{\lambda} \cdot \begin{pmatrix} f & \sigma & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{pmatrix} \quad (3)$$

where λ is a resolution matrix (described above) expressing the number of pixels per unit length for both x and y axes, and $\alpha_x = f\lambda_x$ and $\alpha_y = f\lambda_y$ represent the focal lengths in pixel units. Similarly, x_0 and y_0 are the counterparts of c_x and c_y in image units. The parameter s is the skewness of the camera and defines the angle between the x and y axes. However, as stated above, we assume that pixels are square, which allows us to simplify Eq. (3) as:

$$K = \begin{pmatrix} \alpha & 0 & x_0 \\ 0 & \alpha & y_0 \\ 0 & 0 & 1 \end{pmatrix} \quad (4)$$

where we have set both focal lengths to be equal to α (this can generally be assumed in the field of X-ray imaging). However, this hypothesis should be used with caution in the case of CR plates, in which the IP scanning process has its own optical oddities and limitations, as explained by Rowlands [12] and Dobbins et al. [13].

The extrinsic parameters describe a rigid transformation that maps points in space between the W and C frames. This matrix can also be decomposed into a rotation matrix R (accounting for angles $\theta_x, \theta_y, \theta_z$) and the translation vector \mathbf{t} described earlier:

$$\underbrace{\begin{bmatrix} R & \mathbf{t} \end{bmatrix}}_{\text{extrinsic matrix}} = \underbrace{\begin{pmatrix} \mathbf{I} & \mathbf{t} \\ 0 & 1 \end{pmatrix}}_{\text{3D translation}} \cdot \underbrace{\begin{pmatrix} R & 0 \\ 0 & 1 \end{pmatrix}}_{\text{3D rotation}} \quad (5)$$

A given P can be decomposed into its intrinsic/extrinsic parts using an RQ decomposition, as suggested by Hartley [14], because R is orthogonal and K has an upper triangular shape:

$$P = \underbrace{K}_{\text{intrinsic matrix}} \cdot \underbrace{[R|t]}_{\text{extrinsic matrix}} \quad (6)$$

2.4 Pose-dependent intrinsic matrices in X-ray systems

It is worth noting that in the specific scenario of X-ray imaging, a controversy may arise when characterizing K as intrinsic. In conventional pinhole cameras, the intrinsics do not change if the device is repositioned in the scene. However, in X-ray frameworks, these parameters may vary significantly between consecutive snapshots if either the sensor or the anode is shifted and/or rotated relatively to each other, as illustrated in Fig. 3. This variability originates from the fact that the detector surface and the anode are detached—and structurally independent—from each other. The term *pseudointrinsic* might be appropriate in this case.

This fact is in contrast with what occurs in conventional camera systems, where the sensor (usually a CCD/CMOS array) is architecturally fixed to the optical center. X-ray imaging involves a very interesting scenario in which the intrinsic and extrinsic parts of P are tied. Therefore, for two anode/detector locations C and C' , it is reasonable to consider that $K \neq K'$. This is the reason why the D reference frame is required when describing the geometry applied in X-ray settings (as already discussed in Sect. 2.2).

3 Outline of current calibration algorithms

Here, we outline the three methodologies that are mostly used for the estimation of P . These same approaches will be applied to X-ray imaging systems and compared in the following sections. Several authors have proposed solutions to the problem of camera calibration. Among the most popular are Tsai's algorithm, DLT, and Zhang's

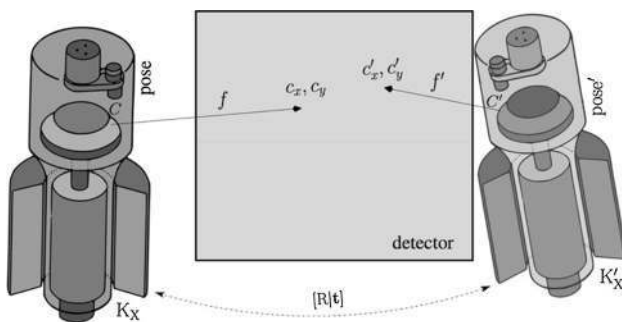


Fig. 3 Variation of intrinsic parameters (K_x to K_x') during the relocation of the anode from an original pose to a different one (the same variation would occur if the detector plate was shifted and the X-ray beam source remained anchored at a fixed location)

method, which we summarize here. Further details about the application of these methods to conventional cameras can be found in Romondino et al. [18] and Zollner et al. [19].

3.1 Tsai's method

Tsai's camera calibration method, presented by Roger Tsai [15], is one of the most famous—and probably one of the very first—modern algorithms for camera calibration. An updated description of Tsai's algorithm is provided in the works of Gupta et al. [20] and Mckerrow et al. [21].

The algorithm recovers the camera parameters using the relationship between points in a three-dimensional calibration mold and their projections in the image plane. A key aspect of this algorithm is that x_0 and y_0 (pixel coordinates of the principal point c_x, c_y) must be used as input parameters. In the case of conventional cameras, the pixel coordinates of the principal point are usually known, fixed, and provided by the detector manufacturer. Unfortunately, this approach is not applicable in X-ray systems, where the sensor and the emitter are detached from each other, as discussed above.

Tsai's method consists of two stages. The first one determines the extrinsic parameters and the focal length. This is achieved by solving a system of linear equations whose input parameters are the 3D/2D coordinates of specific points in a calibration frame, both in the image and the real world. The second phase involves a non-linear minimization process in which the radial distortion factor is determined and all other parameters are further refined.

3.2 DLT

DLT is a simple algorithm used to obtain the projection matrix, given a sufficient set of point correspondences. It was originally devised by Aziz and Karara [22] and is updated in the reference book by Hartley et al. [16].

DLT estimates P using a projective transformation, presented in Sect. 2.3, and a set of point correspondences. However, because the points are expressed in homogeneous coordinates two points are equal if their coordinates are proportional. For this reason, each point pair $\hat{\mathbf{q}}_i, \hat{\mathbf{Q}}_i$ introduces a restriction, which is better described using the vector cross product: $\hat{\mathbf{q}}_i \times P\hat{\mathbf{Q}}_i = 0$. This restriction generates two independent equations. Because the number of independent unknowns is 11 (P is determined up to a scale factor in homogeneous coordinates), a simple linear solution for P can be derived with six correspondences.

Considering that point coordinates are always measured with some error, the linear equation system used to obtain P is normally solved using the SVD method. Unfortunately,

the results of SVD depend on the origin and scale of the coordinate system in the image, which makes the algorithm unstable. To address this issue, Hartley et al. suggest a prior normalization of each $\hat{\mathbf{q}}_i$.

3.3 Zhang's method

This method, originally devised by Zhang [17], requires N_p projections of one or more planar calibration targets N_p , each with its own fiducial set $\mathbf{Q}_{i \dots N_f}$, which consists of N_f fiducial markers. Conventionally, in computer vision, $N_p = 1$; i.e., only one calibration frame is photographed/beamed N_p times at different poses ($N_p = 1 \times N_p$). However, we can also use several N_p frames, each portrayed once ($N_p = N_p \times 1$). For each projection j , a 2D homograph can be estimated from the acquired image. To compute these homographs, more than four non-collinear points are needed. Then, the Zhang method can be used to calculate a series of projective transformations $\{P_1, P_2, \dots, P_j, \dots, P_{N_p}\}$ to points $\mathbf{q}_{i \dots N_f} \times N_p$ in the N_p bitmaps up to a scale factor. It is important to note here that each projection j used in the scope of Zhang's method returns its own independent $[R|t]_j$ set. However, all such sets share the same K . This means that we can only reproject points related to each projection j and cannot locate a shared and unique world reference frame W unless one of the plane frames is radiographed at a well-known and traceable location in the 3D scene. The minimization phase helps in refining all previously derived parameters, whose value is proportional to N_p : three intrinsic (c_x, c_y, f) in addition to a total of $N_p \times (\theta_x, \theta_y, \theta_z, t_x, t_y, t_z)$ extrinsics.

3.4 Non-linear refinement

All the presented algorithms can be improved if the retrieved results are refined using a non-linear cost function $g(\dots)$. The most common such cost function is the geometric distance (or transfer error), which measures the Euclidean distance between the projection of an N_{points} set of 3D world spots \mathbf{Q}_i and their observed correspondences \mathbf{q}_i in the image:

$$g(\mathbf{q}_{1 \dots N_{\text{points}}}, \mathbf{Q}_{1 \dots N_{\text{points}}}) = \sum_i^{N_{\text{points}}} \|\mathbf{q}_i - P\mathbf{Q}_i\|^2 \quad (7)$$

In the case of the Zhang method discussed above, each plane has its own projection P_j . Each $\mathbf{Q}_{j,i}$ (i th fiducial) must be projected onto the detector as $\mathbf{q}_{j,i}$ using P_j . Each $\mathbf{Q}_{j,i}$ (fiducial i on beamed frame j) must be then projected using the corresponding P_j . In this scenario, the cost function depends on a total of $M = N_p \times N_{\text{points}}$ parameters. Equation (7) should be then rewritten as suggested in the study performed by Medioni et al. [10]:

$$g(\mathbf{q}_{1 \dots M}, \mathbf{Q}_{1 \dots M}) = \sum_{j=1}^{N_p} \sum_{i=1}^{N_{\text{points}}} \|\mathbf{q}_{j,i} - P_j \mathbf{Q}_{j,i}\|^2 \quad (8)$$

where $\mathbf{q}_{j,i}$ is the observed projection of a coplanar point i to the imaged plane j . However, if the plane projections correspond to one single physical flat frame ($N_p = 1$) with a common set of fiducial $\mathbf{Q}_1, \dots, i, \dots, N_{\text{points}}$ snapshots (as we discussed in Sect. 5.1), Eq. (8) can be simplified as:

$$g(\mathbf{q}_{1 \dots M}, \mathbf{Q}_{1 \dots N_{\text{points}}}) = \sum_{j=1}^{N_p} \sum_{i=1}^{N_{\text{points}}} \|\mathbf{q}_{j,i} - P_j \mathbf{Q}_i\|^2 \quad (9)$$

4 Methods and materials

We present here the experimental setup used to evaluate each of the camera calibration methods summarized above. Figure 4 shows the DUT used in our research with the X-ray equipment (vertical IP configuration).

4.1 X-ray imaging setup and calibration frame

To appropriately describe an X-ray diagnostic system, it is necessary to initially establish a reference frame located at the detector. The selected reference frame for the current setup is displayed in Figs. 3 and 6.

Additionally, an initial polytetrafluoroethylene 3D calibration frame or DUT, outlined in Fig. 5 (also pictured in Fig. 4), was built. The DUT had a simple rectangular shape in order to be manageable. In our setup, this DUT remained fixed and could be imaged from many locations and X-ray tube orientations.

Additionally, a set of $N_f = 13$ cross-shaped tin markers opaque to Röntgen rays were distributed on two levels of the DUT. The bottom level accommodated nine markers

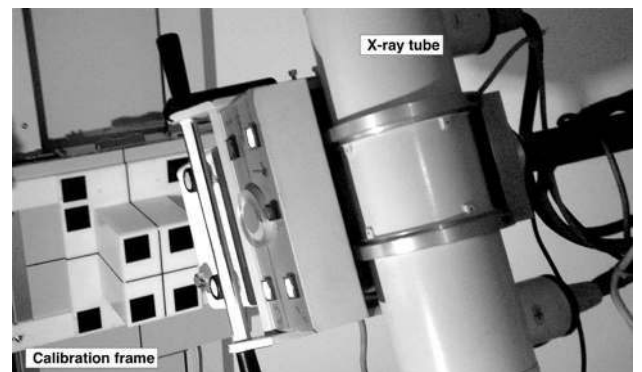
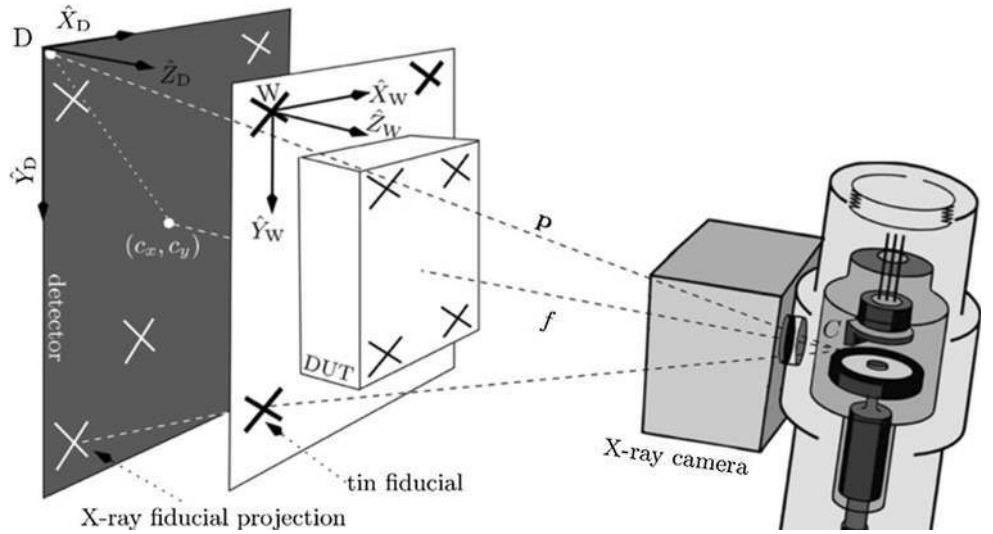


Fig. 4 Experimental setup with calibration frame and typical X-ray imaging system used in primary diagnostic procedures. Tin cross-shaped fiducials are placed beneath the square stickers for protection

Fig. 5 Schematic of the DUT, image receptor, and X-ray tube in the world (W) and detector (D) coordinate systems



and the top one only four. The DUT was also responsible for establishing the W reference frame.

Besides the N_f cross-shaped markers, our DUT also housed nine lead spherules (4 mm diameter) inside its Teflon cavity. These pellets are used when verifying the efficiency of each method and they are ignored during the calibration.

The DUT was radiographed from many angles and beam positions (57 snapshots), imitating an AP vertical examination. In our setup, the IP surface was approximately the same size as the calibration frame to capture all possible fiducial marker projections. The pixel centers of all cross-shaped traces were manually identified with the Horos DICOM Viewer (<http://horosproject.org>).

4.2 Comparison metrics

To objectively compare the three methodologies outlined in Sect. 3, several metrics were defined. Reprojection RMSE and intrinsic parameters consistency were applied on a per-projection basis. The other two metrics (detector resolution and distance to epipolars) involve the use of pairs of stereo radiographs as input [23].

4.2.1 Reprojection root-mean-square difference

This quantity describes how well each algorithm performs when projecting the 3D N_f points that were used in the calibration again:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N_f} \|\mathbf{q}_i - \mathbf{PQ}_i\|^2}{N_f}} \quad (10)$$

4.2.2 IP resolution estimation

The linear resolution of the IP is provided by the manufacturer ($\lambda_o = 10^4$ px/m). It can be used as a screening parameter. We can derive λ_j from each X-ray snapshot pair j with the expression:

$$\lambda_j = \left\| \frac{(\mathbf{p}_2 - \mathbf{p}_1)}{(\mathbf{t}_2 - \mathbf{t}_1)} \right\| \quad (11)$$

where \mathbf{p}_2 , \mathbf{p}_1 and \mathbf{t}_2 , \mathbf{t}_1 represent the anode coordinates in the D and W reference frames, for the X-ray source locations 1 and 2, respectively. Both \mathbf{p} and \mathbf{t} vector pairs are obtained through the RQ decomposition of each \mathbf{P}_1 and \mathbf{P}_2 that represent the camera calibration at the two stereo-anode poses:

$$\mathbf{P} \xrightarrow[\text{decomposition}]{\text{RQ}} \underbrace{\begin{bmatrix} \alpha & s & x_0 \\ 0 & \alpha & y_0 \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{p}=(x_0, y_0, \alpha)} \cdot [\mathbf{R}|\mathbf{t}] \quad (12)$$

The geometric explanation of Eq. (11) is shown in Fig. 6, where $\mathbf{p}_2 - \mathbf{p}_1$ and $\mathbf{t}_2 - \mathbf{t}_1$ represent the same spatial gap between two X-ray tube locations. The difference between $\mathbf{p}_2 - \mathbf{p}_1$ and $\mathbf{t}_2 - \mathbf{t}_1$ is that $\mathbf{p}_{1,2}$ is expressed in pixel units and $\mathbf{t}_{1,2}$ in meters. This allows us to experimentally determine the resolution (λ_j) in pixels per meter of our X-ray detector (for a given stereo snapshot j), which should be a priori similar to the factory specifications $\lambda_j \approx \lambda_o$.

Over $N_{\text{comb}} = 1540$ pair combinations from 56 radiographs were created to estimate λ_j , as shown in the example of Fig. 6 (bottom). Then, for each calibration algorithm, we obtained the overall mean value:

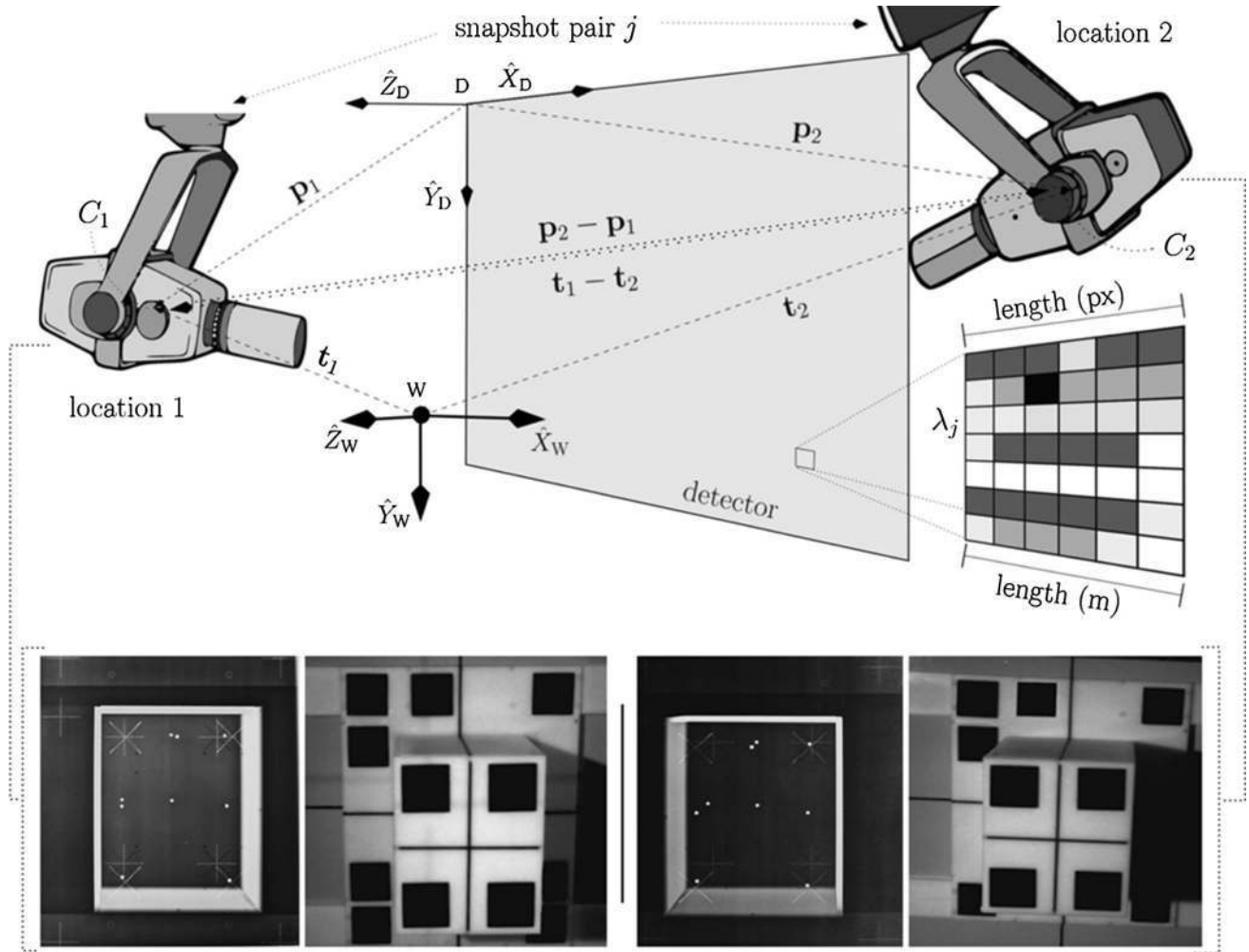


Fig. 6 Derivation of detector linear resolution. The anode is shifted between two locations C_1 and C_2 (determined by $\mathbf{p}_1, \mathbf{t}_1$ and $\mathbf{p}_2, \mathbf{t}_2$ in the D and W frames, respectively) and a radiograph of the DUT is generated. The distances $\mathbf{p}_2 - \mathbf{p}_1$ and $\mathbf{t}_2 - \mathbf{t}_1$ represent the same

spatial gap between the two X-ray tube locations and their ratio is the detector linear resolution λ_j . The *bottom row* shows an example of a snapshot pair used to derive λ_j with the corresponding photographic images of DUT

$$\lambda_{\text{method}} = \frac{\sum_{j=1}^{N_{\text{comb}}} \lambda_j}{N_{\text{comb}}} \quad (13)$$

where “method” denotes the Tsai, DLT, or Zhang methods.

4.2.3 2D distance between epipolar lines and spherules

A key step in stereo imaging entails finding point correspondences in two images. Using epipolar geometry, the search for a corresponding point can be reduced from examining an entire image to simply searching along a specific line in the image, called the epipolar line. In other words, a point in one image corresponds to a line in its stereo partner. The left camera sees a point because it is directly in line with that camera’s center of projection. However, the right camera perceives this ray

as an infinite straight segment in its image plane. The same situation also occurs in X-rays, where the beam source is shifted between two positions with a rigid transformation.

To produce these epipolars, we first compute the fundamental matrix F . This can be obtained using several means, including the interplay of the two camera projection matrices P_1 and P_2 , as described by Armangué et al. [24] and Luong et al. [25].

As mentioned in Sect. 4.1, a set of $N_s = 9$ additional lead markers were placed in the DUT plastic receptacle at specific 3D locations by means of a foam scaffold. Using the same snapshot pair analogy that we applied for the detector resolution step, we can compute the epipolar lines and their perpendicular distances $\Delta_{s,j}^{\text{method}}$ to the projections \mathbf{q}_s of each spherule (sphere) \mathbf{Q}_s for each snapshot pair j and for each calibration method. Epipolar geometry theory

supports the fact that the epipolar line \mathbf{l}_s^2 on the second image that is linked to the projection of point \mathbf{Q}_s on the first image \mathbf{q}_s^1 can be calculated with:

$$\mathbf{l}_s^2 = \mathbf{F}_{P_2}^{P_1} \cdot \mathbf{q}_s^1 \quad (14)$$

This concept is schematically summarized in Fig. 7.

The mean value of all $\Delta_{s,j}^{\text{method}}$ values is provided as the result of the Tsai ($\Delta_{s,j}^{\text{Tsai}}$) and DLT ($\Delta_{s,j}^{\text{DLT}}$) calibration techniques:

$$\Delta_{s,j}^{\text{method}} = \frac{\sum_{j=1}^{N_{\text{comb}}} \sum_{s=1}^{N_s} \Delta_{s,j}}{N_s \times N_{\text{comb}}} \quad (15)$$

When deriving $\Delta_{s,j}^{\text{Zhang}}$ and considering Eq. (12), P1 and P2 should represent the corresponding projection matrices on the same Zhang plane (#1, ..., #13 in Fig. 8) containing the given projected spherule from each stereo location.

In this case, a subset of 43 X-ray images and $N_{\text{comb}} = 903$ snapshot pair combinations were used.

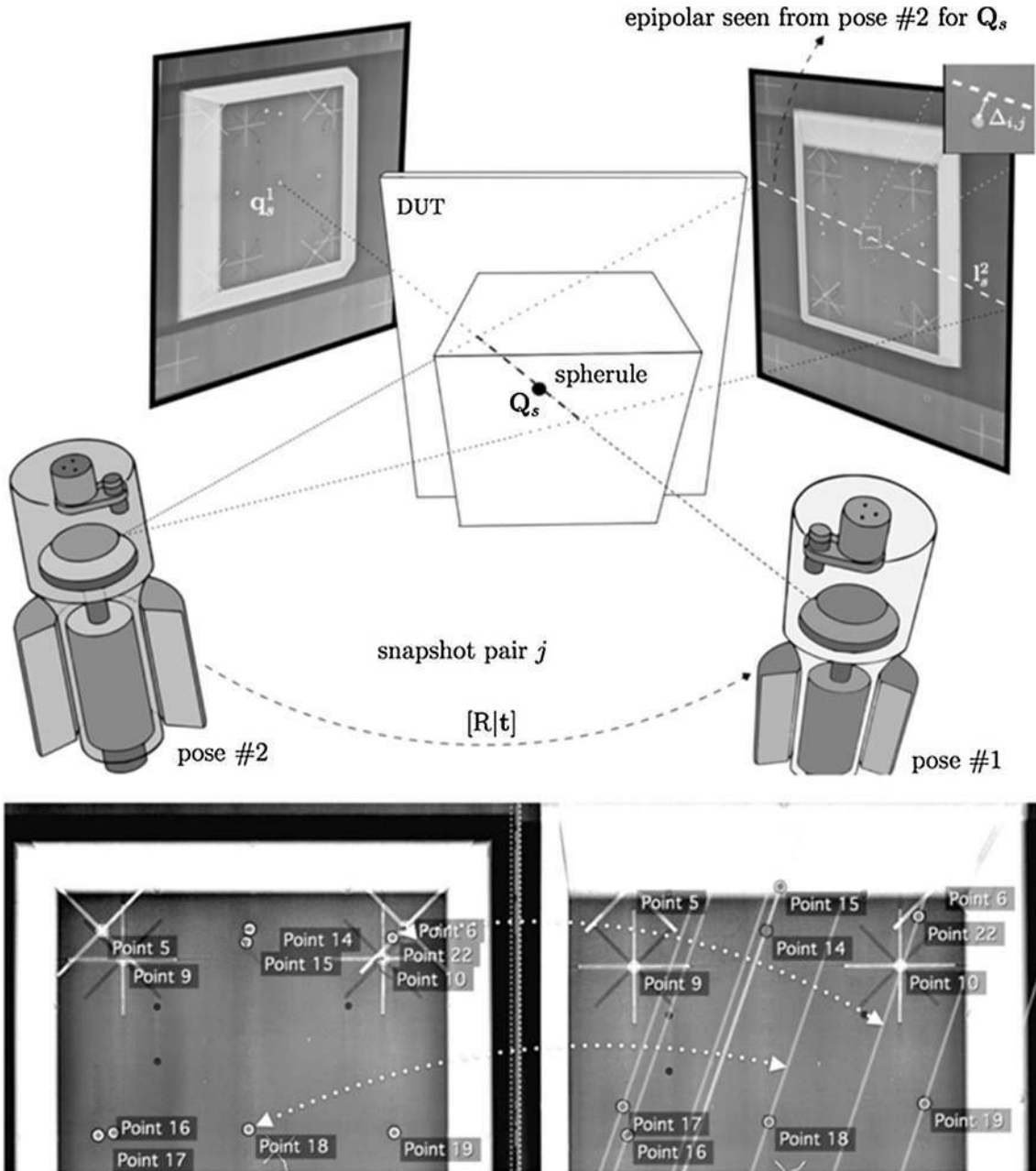


Fig. 7 Epipolar geometry in stereo X-ray imaging and, at the *bottom*, an example of two stereo snapshot pairs and the *epipolar line* from each lead spherule, as seen from a paired anode location

Figure 7 (bottom) shows some examples of these stereo pairs and epipolars.

4.2.4 Consistent intrinsic parameters

The final test involves investigating whether the intrinsic parameters calculated at the RQ decomposition stage are physically consistent, i.e., if they represent cohesive physical dimensions and distances between the X-ray emitter and the detector. More specifically, we focus on the focal length f of each X-ray system pose. We then examine whether we can derive a coherent spatial distance between the anode and the IP. For this purpose, the focal distance obtained with the DLT approach is considered as a near-true distance and the Tsai and Zhang methodologies are compared against DLT (Δf_{DLT}). Because DLT almost always delivers consistent intrinsic sections, it seems reasonable to compare the other two methodologies against it.

This metric is not intended to serve as a comparison point for the three calibration methods but as a final health check of the consistency of each calculated P . A projection matrix can perform very well in all three aforementioned metrics. However, this does not necessarily mean that the intrinsics (i.e., f) are well derived or that they produce well-formed and physically relevant K_{Tsai} and K_{Zhang} matrices.

4.3 Practical considerations

The most straightforward method is DLT, whose application starts with a set of correspondences between world points N_f and image points; a linear solution can be then computed. Tsai is rather similar but requires an initial guess of x_0 and y_0 (usually the image center), which must later be fine tuned with a least-squares calculation stage.

The Zhang method differs significantly from the other two, as it operates over 2D frames and their planar projections. In a manner similar to the DLT and Tsai methods, in the Zhang approach, we initially assumed that the accuracy of the camera calibration process is proportional to the number of 3D-2D point correspondences, which implies more planar projections (as discussed in Sect. 3.3). In this context, we initially considered a total of $N_p = 13$ virtual planes using all possible fiducial marker combinations in the DUT. These ideal planes are shown in Fig. 8a and a sample subset of their X-ray projections is displayed in Fig. 8b. Each of these planes contains four fiducial points. Planes #11 and #12 can be expanded with the additional antisymmetric copper landmark (whose X-ray projection is labeled as point 13 in Fig. 8b, c).

The lead spherules described in Sect. 4.1 were positioned as coplanarly as possible to some of the built Zhang planes #1–#10 shown in Fig. 8a; however, some pellets were omitted from the calculations if their geometrical

distance from a given plane exceeded 5 mm. In the case of the Zhang method, every plane is related to its own projection matrix $P_{1\dots 13}$ and, thus, only the 3D traits that lie on that plane can be projected. This feature has been considered when applying the comparison metric described in Sect. 4.2.3.

After some preliminary tests, we observed that the use of all 13 planes for the calibration produced unstable results in many cases. A more careful investigation showed that fluctuating outcomes were obtained when some of the virtual planes were almost perpendicular to the image surface and/or anode. When our DUT was radiographed, these steep/oblique planes projected an almost negligible area. Figure 8c shows some examples of this type of problematic traces.

For this reason, we devised a pre-step to the Zhang algorithm that easily discards planes that generate unstable results based on their projected area, expressed in number of pixels. Section 5 explores this topic.

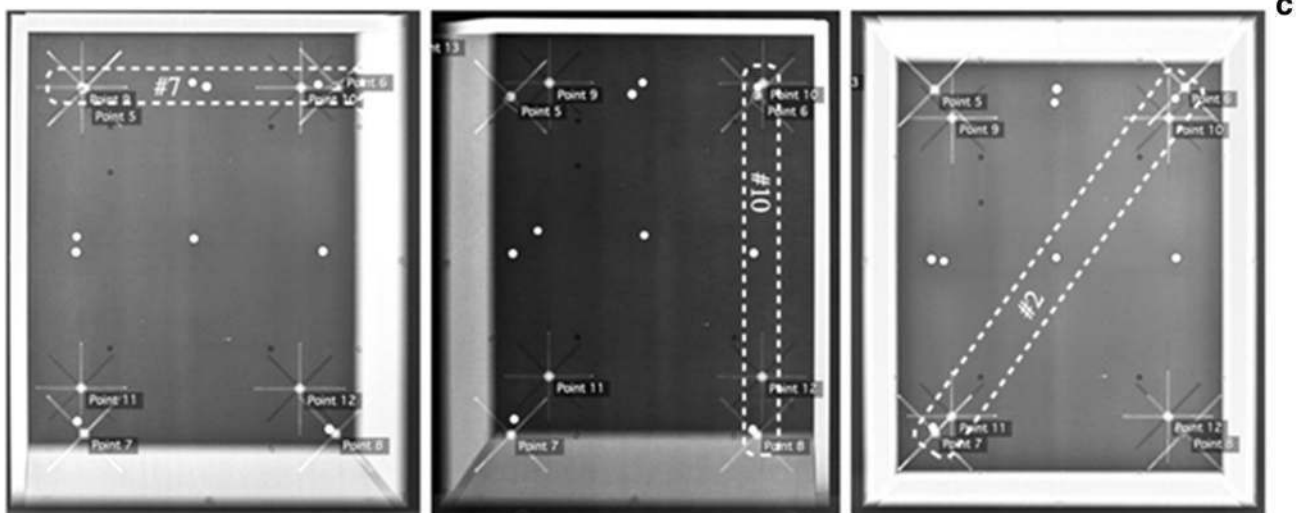
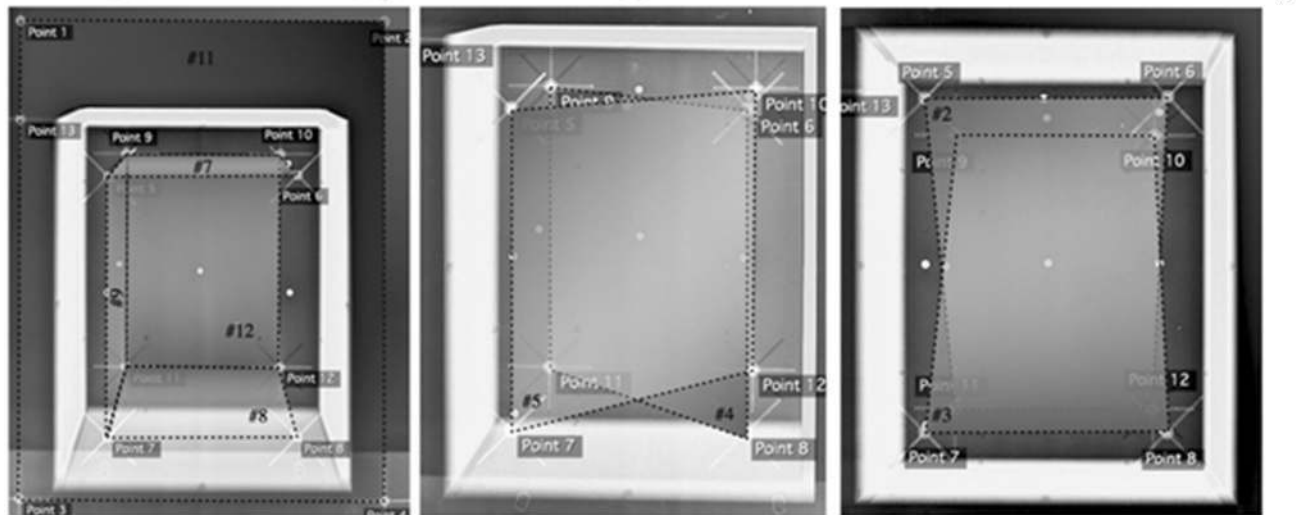
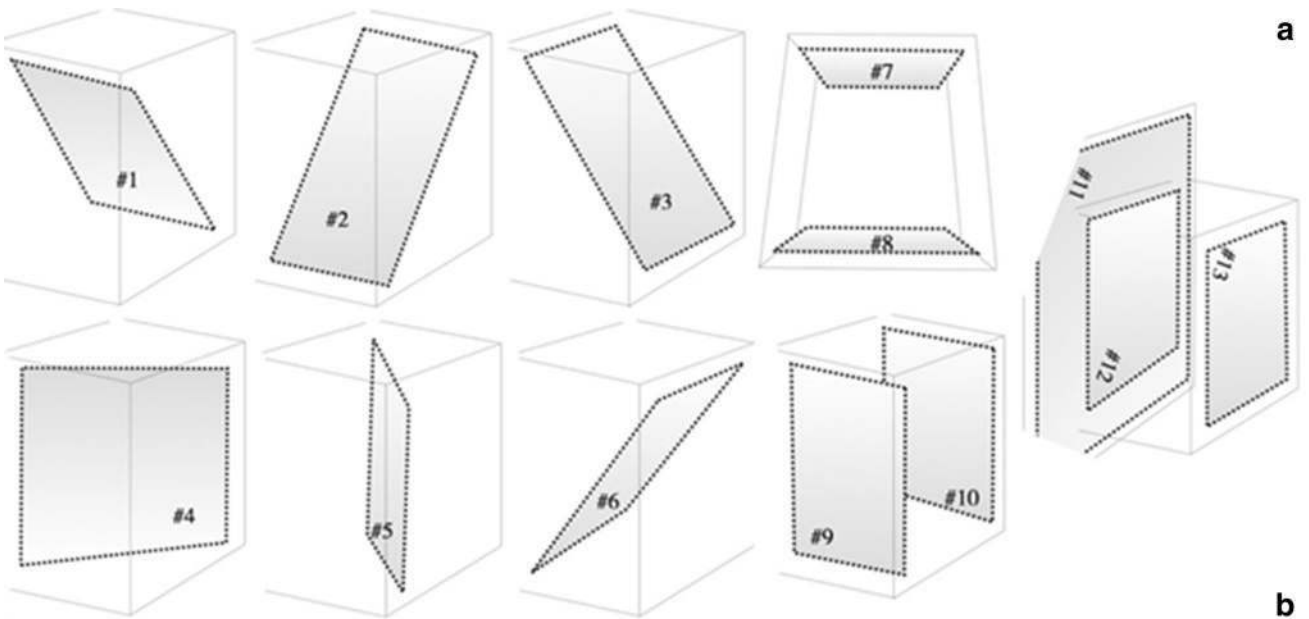
Furthermore, to apply the test metric described in Sect. 4.2.2, we must select the projection matrix P_{11} , which is linked to plane #11 (Fig. 8a), in each stereo snapshot. This plane contains the fiducial cross marker (upper left corner) that defines a shared W reference frame. When the matrix P_{11} that is linked to this plane is RQ-decomposed, it reveals the translation vector \mathbf{t} that allows us to derive the detector resolution λ_j for a specific radiograph stereo pair j using Eq. (11).

As discussed in Sect. 3, all calibration methods allow further optimization through non-linear minimization. We have chosen the geometric distance as the non-linear cost function. All necessary calculations, such as the aforementioned minimizations, SVD, and optimizations, were performed with the GNU Octave software package (<https://www.gnu.org/software/octave>).

5 Results and discussion

Table 1 compares all calibration algorithms using the metrics introduced in Sect. 4.2. We provide the results before and after the refinement process (the latter denoted with the suffix “/R”). As discussed in Sect. 4.2, σ is the reprojection error—calculated using Eq. (10)—of the fiducial markers considered for the calibration. The metric Δ is the mean distance from each spherule image location to the epipolar of that same spherule—obtained with (13)—viewed from another paired anode. Δ is calculated using Eq. (15). It should be noted that in the case of the Zhang algorithm, each spherule is projected using the specific P_j that is related to the plane that fully contains it.

The mean resolution λ of the detector is estimated in pixels/meters and using Eq. (11). λ is then compared with



◀**Fig. 8 a** The 13 virtual internal planar structures inside the 3D DUT used for the Zhang calibration method. **b** X-ray projections of virtual Zhang planes built with fiducials present in the DUT. **c** X-ray projections of some problematic Zhang virtual planes (inside the *dotted line*) that affect the results

the factory specifications λ_o for each calibration method. We note here that in the Zhang method, we used the projection matrix P_{11} for plane #11 and its translation vector \mathbf{t} to compute λ . We also focus on the quality of the focal length f provided by the Tsai and Zhang methods by comparing it with that derived with DLT (Δf_{DLT}).

In general, all methods produce good results and show applicability to X-ray calibration. σ is very low, λ is almost identical to λ_o , and Δ is approximately zero.

DLT achieves the best results, even without a minimization phase. This indicates that this time-consuming stage can be omitted. In contrast, the Zhang and Tsai methods require a non-linear minimization step to deliver accurate results. Although Tsai’s approach is quite precise and outputs even better results when refined through a minimization process, it must be initialized with the principal point coordinates, which are a priori unknown in X-ray environments. If the X-ray emitter is positioned approximately perpendicularly to the examined object, as in setup a) in Fig. 10, x_0 and y_0 can be assigned as the coordinates of the image center and still obtain satisfactory results. However, at poses approaching the limits of the X-ray system, this central landmark can have a coordinate that is distant from the center, as shown in the example of Fig. 3.

As detailed in Sect. 4.3, in the case of Zhang’s method, some degenerate planes needed to be rejected using the simple technique of requiring a minimum projected area belonging to each virtual plane. In this method, the most accurate results are achieved when a minimal area of 10^3 px is set as a threshold for each plane. However, these results are not as accurate as those obtained with DLT/Tsai. This may be explained by the fact that planes must be populated with a denser fiducial grid, as observed in Sánchez et al. [26].

Figure 9 provides an optimized visualization of the influence of the minimum required projected area in Zhang’s algorithm. The figure shows the mean difference (Δf_{DLT}) between the focal lengths f_{Zhang} derived with Zhang’s method and those obtained using DLT (f_{DLT}).

From these plots, we can infer that the best performance seems to be achieved with the combination of four large projected planes: the three widest parallel ones (#11, #12, and #13 in Fig. 8a) and an additional fourth plane. In other words, the Zhang method can be improved by adding more planes, but the selection of planes according to their projected area is also of great importance. Too many perpendicular planes can interfere with the calibration process. This outcome agrees with the experimental results obtained by Zhang et al. [17]. Additionally, we observed that Zhang’s algorithm delivered better results when the DUT was radiographed from wider angles (configuration b in Fig. 10) that originated from larger projected zones.

Zhang’s method was evaluated in its original form for computer vision using a single 2D calibration frame ($N_p = 1 \times N_p$) projected on alternating poses (discussed in Sect. 3.3). This parallel study is detailed below.

5.1 Alternative Zhang calibration experiment with a 2D calibration frame

As a supplementary experiment, we radiographed a separate 2D square DUT (equipped with 16 radiation-opaque fiducials and four additional ones to break the symmetry) at 27 arbitrary orientations and distances (Figs. 11, 12). We designed a special holding structure consisting of an adjustable support system with a panoramic head that allowed us to freely move and re-orientate this structure before radiography.

The focal length f_{Zhang} was estimated by means of this device and the X-ray imaging system remained stationary throughout the measurements. We calculated the mean

Table 1 Test metrics and results of the three calibration methods

| | | Tsai | DLT | Zhang | Tsai/R | DLT/R | Zhang/R |
|-------------------------|------|--------|------|--------|--------|-------|---------|
| σ | px | 36 | 7 | 270 | 7 | 7 | 19 |
| std(σ) | | 15 | 5 | 136 | 5 | 5 | 11 |
| Δ | Px | 15 | 3 | 69 | 3 | 3 | 12 |
| std(Δ) | | 19 | 6 | 49 | 7 | 6 | 8 |
| λ | px/m | 10,071 | 9993 | 10,114 | 9991 | 9991 | 10,040 |
| std(λ) | | 287 | 192 | 237 | 181 | 190 | 278 |
| Δf_{DLT} | m | 0.10 | | 0.12 | 0.00 | | 0.01 |
| std(Δf_{DLT}) | | 0.06 | | 0.02 | 0.00 | | 0.06 |

The suffix “/R” denotes non-linear least-squares refinement

focal length (Fig. 12c) for all plane combinations (e.g., 2300 combinations for three planes, 12,650 for four planes, and 53,130 for five planes).

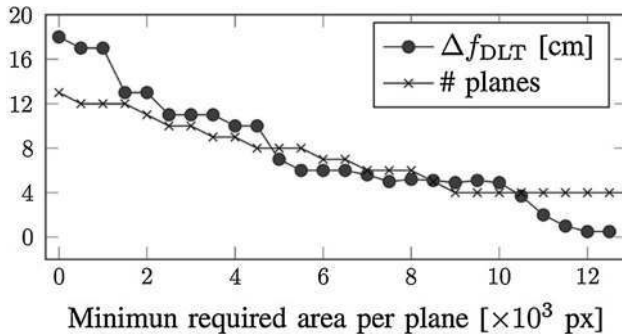


Fig. 9 Mean difference between the focal lengths calculated with the DLT and Zhang methods according to the minimum number of pixels required in each virtual Zhang plane inside the DUT

Fig. 10 Most effective geometrical configuration of the X-ray system for Zhang's approach (b) and optimal frontal placement for Tsai's algorithm (a)

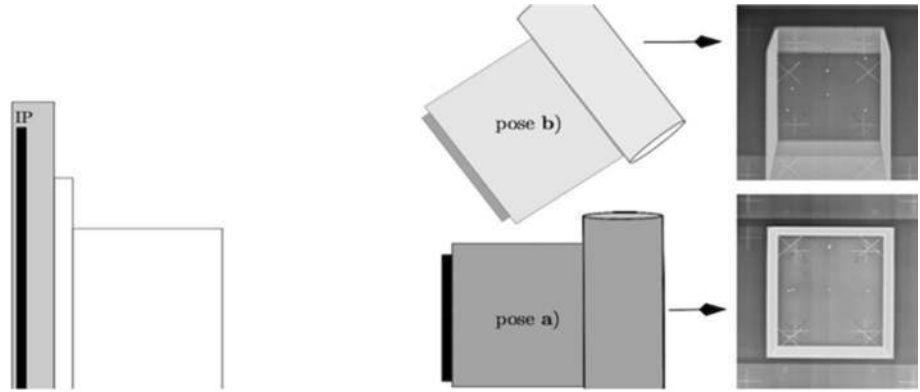
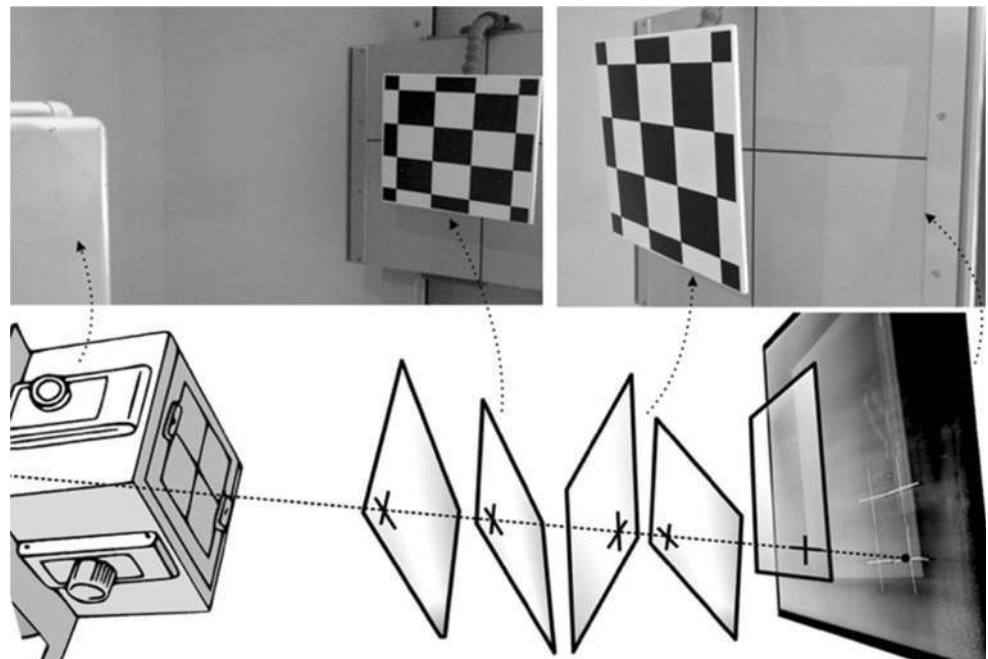


Fig. 11 Visual representation of the experimental setup examined with the Zhang method. A single calibration flat panel is adjusted over a strand of flexible plastic segments, which allows perfect positioning. The panel is equipped with 16 hidden X-ray-opaque markers and is then irradiated at random poses to derive f_{Zhang}



These results clearly demonstrate how Zhang's method, combined with a simple 2D frame, can provide a good alternative for the calibration of any X-ray device. As expected, calculations involving a greater number of projected planes produced a more accurate (or a convergence towards a stable) focal length. Nevertheless, some combinations of even three or four planes generated a very precise solution. As an example, Fig. 12a shows a combination of only three planes, from which we can already derive the same f_{Zhang} that is achieved with the remaining 22 projections.

In our study, the accuracy of the determination of the mean f_{Zhang} was tested by including and omitting some oblique Zhang planes. It is important to identify the conflicting Zhang planes (Fig. 12b) with a very small projected area as well as non-conflicting planes (Fig. 12a). Removing the conflicting planes from the

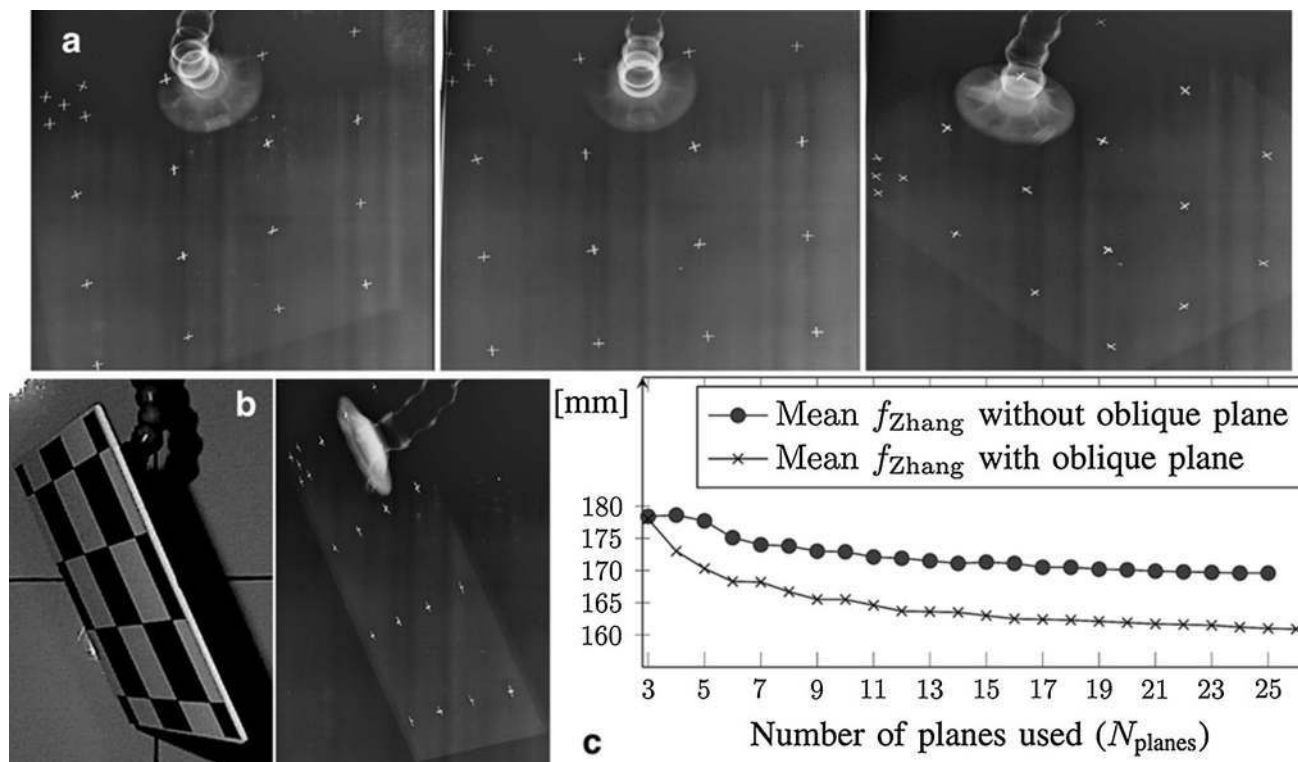


Fig. 12 **a** Combination of three random Zhang planes that produce a very accurate focal length. **b** Problematic configuration of a Zhang plane. **c** Convergence of the mean f_{Zhang} determined using the random

calculations accelerated the stable and consistent calibration, as shown in Fig. 12c.

6 Conclusions

We have applied three well-known camera calibration methods to primary diagnostic X-ray environments. All calibration methods share the same principle: determining a projection matrix that connects 3D points in the scene with their projections in the image. For this study, we developed a 3D calibration frame fitted with fiducial markers and a set of metrics.

The main and most evident outcome is that all three methods can perform well with these types of X-ray settings. In other words, when using a similar calibration frame to ours, all three methodologies deliver functional projection matrices and accurate intrinsics before or after non-linear least-squares refinement. However, each technique exhibits some significant operational differences and particularities that have also been analyzed.

We have paid special attention to the Zhang method, which is frequently used for computer vision but not in regular X-ray settings, and demonstrated that it can be a reasonable alternative to heavier 3D phantoms. Within a Zhang framework, three radiographs from a single planar

positioning of DUT (with and without oblique planes) towards the stable mean focal length due to an increase in the number of planes used

device can suffice; however, very oblique projections can significantly alter the calculations and should be previously identified and discarded. From the observations reported in this paper, fixed 3D calibration structures are preferable when working under the DLT or Tsai schemes.

Finally, we have proposed an innovative method to derive the detector pixel density from pairs of X-ray snapshots. This procedure might be suitable for the QA of clinical environments for verifying detector and contrast sensor resolution against original factory specifications.

Acknowledgments This work was carried out with the support of Information Storage S.L., University of Valencia (Grant #CPI-15-170), CSD2007-00042 Consolider Ingenio CPAN (Grant #CPAN13-TR01), Spanish Ministry of Industry, Energy and Tourism (Grant #TSI-100101-2013-019), IFIC (Severo Ochoa Centre of Excellence #SEV-2014-0398), and Dr. Bellot's medical clinic.

Compliance with ethical standards

Conflict of interest We have nothing to declare for this study.

References

1. Selby BP, Sakas G, Groch W-D, Stilla U. Patient positioning with X-ray detector self-calibration for image guided therapy. *Aust Phys Eng Sci Med.* 2011;34:391–400.

2. Markelj P, Likar B. Registration of 3D and 2D medical images. PhD Thesis, University of Ljubljana; 2010.
3. Miller T, Quintana E. Stereo X-ray system calibration for three-dimensional measurements. Springer, 2014. pp. 201–207.
4. Rougé A, Picard C, Ponchut C, Troussset Y. Geometrical calibration of X-ray imaging chains for three-dimensional reconstruction. *Comput Med Imaging Graph.* 1993; 295–300.
5. Trucco E, Verri A. Introductory techniques for 3-D computer vision. Prentice Hall Englewood Cliffs, 1998.
6. Moura DC, Barbosa JG, Reis AM, Tavares JMRS. A flexible approach for the calibration of biplanar radiography of the spine on conventional radiological systems. *Comput Model Eng Sci.* 2010; 115–137.
7. Schumann S, Thelen B, Ballestra S, Nolte L-P, Buchler P, Zheng G. X-ray image calibration and its application to clinical orthopedics. *Med Eng Phys.* 2014;36:968–74.
8. Selby B, Sakas G, Walter S, Stilla U. Geometry calibration for X-ray equipment in radiation treatment devices. 2007. pp. 968–974.
9. de Moura DC, Barbosa JMG, da Silva Tavares JMR, Reis A. Calibration of bi-planar radiography with minimal phantoms. In: *Symposium on Informatics Engineering.* 2008. pp. 1–10.
10. Medioni G, Kang SB. *Emerging topics in computer vision.* Prentice Hall. 2004.
11. Bushong S. *Radiologic science for technologists: physics, biology, and protection.* Elsevier. 2012.
12. Rowlands JA. The physics of computed radiography. *Phys Med Biol.* 2002;47:123–66.
13. Dobbins JT, Ergun DL, Rutz L, Hinshaw DA, Blume H, Clark DC. DQE(f) of four generations of computed radiography acquisition devices. *Med Phys.* 1995;22:1581–93.
14. Hartley R. Self-calibration from multiple views with a rotating camera. In: *European Conference on Computer Vision.* 1994. pp. 471–478.
15. Tsai R. A versatile camera calibration technique for high accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE J Robot Autom.* 1985;3(4):323–44.
16. Hartley R, Zisserman A. *Multiple view geometry in computer vision.* Cambridge University Press. 2004.
17. Zhang Z. A flexible new technique for camera calibration. *IEEE Trans Pattern Anal Mach Intell.* 2000;22:1330–4.
18. Remondino F, Fraser C. Digital camera calibration methods: considerations and comparisons. *Symposium Image Eng Vis Metrol.* 2006;36:266–72.
19. Zollner H, Sablatnig R. Comparison of methods for geometric camera calibration using planar calibration targets. In: *Digital Imaging in Media and Education.* 2004. pp. 237–244.
20. Gupta M, Upadhyay S, Nagawat AK. Camera calibration technique using Tsai's algorithm. In: *International Journal of Enterprise Computing and Business Systems.* 2011.
21. Mckerrow PJ, Abrantes J. Problems encountered in the implementation of Tsai's algorithm for camera calibration. In: *Australasian Conference on Robotics and Automation.* 2002. pp. 66–70.
22. Aziz AYI, Karara HM. DLT into object space coordinates in close-range photogrammetry. In: *Symposium on Close-Range Photogrammetry.* 1971. pp. 1–18.
23. Albiol AJ, Albiol A, Fernandez A, Rodrigo G, Albiol F. Dispositivo y procedimiento de obtención de imágenes densitométricas de objetos mediante combinación de sistemas radiológicos y cámaras de profundidad. Consejo Superior de Investigaciones Científicas. 2012.
24. Armangué X, Salvi J. Overall view regarding fundamental matrix estimation. *Image Vis Comput.* 2003;21:205–20.
25. Luong Q-T, Faugeras O. The fundamental matrix: theory, algorithms, and stability analysis. *Int J Comput Vision.* 1996;17:43–75.
26. Sánchez JA, Destefanis EA, Canali LR. Plane-based camera calibration without direct optimization algorithms. *IV Jornadas Argentinas de Robótica.* 2006.