Albiol Colomer, F.; Corbi, A.; Albiol Colomer, A. (2016). Geometrical Calibration of X-Ray Imaging With RGB Cameras for 3D Reconstruction. IEEE Transactions on Medical Imaging. 35(8):1952-1961. doi:10.1109/TMI.2016.2540929.

# Geometrical calibration of X-ray imaging with RGB cameras for 3D reconstruction

Francisco Albiol, Alberto Corbi, and Alberto Albiol

***Abstract*—We present a methodology to recover the geometrical calibration of conventional X-ray settings with the help of an ordinary video camera and visible fiducials that are present in the scene. After calibration, equivalent points of interest can be easily identifiable with the help of the epipolar geometry. The same procedure also allows the measurement of real anatomic lengths and angles and obtains accurate 3D locations from image points. Our approach completely eliminates the need for X-ray-opaque reference marks (and necessary supporting frames) which can sometimes be invasive for the patient, occlude the radiographic picture, and end up projected outside the imaging sensor area in oblique protocols. Two possible frameworks are envisioned: a spatially shifting X-ray anode around the patient/object and a moving patient that moves/rotates while the imaging system remains fixed. As a proof of concept, experiences with a device under test (DUT), an anthropomorphic phantom and a real brachytherapy session have been carried out. The results show that it is possible to identify common points with a proper level of accuracy and retrieve three-dimensional locations, lengths and shapes with a millimetric level of precision. The presented approach is simple and compatible with both current and legacy widespread diagnostic X-ray imaging deployments and it can represent a good and inexpensive alternative to other radiological modalities like CT.**

***Index Terms*—X-ray imaging, 3D reconstruction, geometric calibration, visible fiducials, camera system**

## I. Introduction

Recovering the geometrical information from multiple X-ray snapshots of a same object/patient generated from different angles and positions has become of increasing relevance in medicine, industry and surveillance. More specifically, interest has grown in relation to the identification of common points or areas of interest in several radiographs, and the derivation of useful 3D information (distances, angles, etc.) from a sparse set of images produced in conventional and primary diagnostic X-ray imaging settings. In this context, distances are usually very poorly estimated, and, in many cases, simple X-ray-opaque objects (like coins) are used as reference landmarks. In this paper, a new methodology that represents a step forward towards a better assessment of anatomical distances with standard/legacy diagnostic equipment, fiducial-less radiographs and less invasive frameworks is presented.

In computer vision, the process of retrieving 3D information from 2D bitmaps is usually referred as *image reconstruction* and can be achieved after a camera calibration phase. The geometric calibration of X-ray modalities starts with the inference of perspective projection matrices that map 3D scene points with their *projected* counterparts. Unfortunately, this process entails overcoming some obstacles, which will be addressed in this paper. Perhaps, the most important one has to do with the huge level of spatial variability present in typical X-ray imaging environments, as highlighted and summarized in Fig. 1. In these basic radiological settings, both the radiographed object and the X-ray imaging system can move with almost complete freedom around the room. These spatial changes are then usually coded as rigid transformations or translation vectors. As a collateral benefit, keeping track of the distances, positions, and orientations (*geometrical settings*) that dominate the radiographic scene can also play a role in beam equalization, and therefore, in dose control.

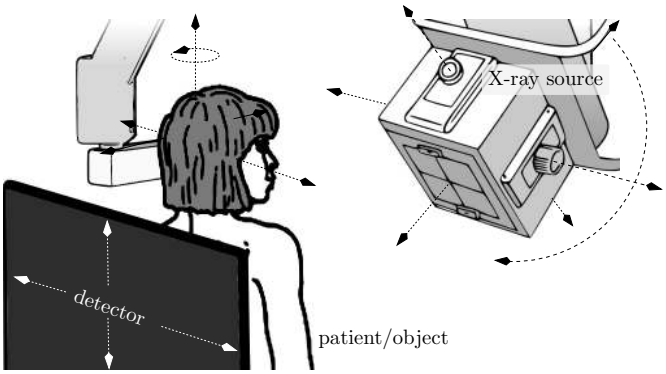


Fig. 1. Degrees of freedom found in average primary diagnostic X-ray settings: detector area can change (i.e., different CR plates) and both anode and detector can move and/or rotate, which definitely would alter camera intrinsic and extrinsic factors.

Another interesting consequence of this spatial *variability* is illustrated in Fig. 2, where a slight shift of the X-ray system (anode or detector) can drastically alter its own intrinsic parameters. This *disentanglement* between source and detector is also methodically tackled by authors in [1] in the case of mobile C-arms by introducing the concept of *virtual detector*.

In our proposal, we account for these scene alterations in basic X-ray environments with the help of a visible light camera that is *rigidly tied* to the X-ray source (Fig. 3). Alongside this goal, we also elude the use of X-ray-opaque markers for the reasons later elaborated in Section III. In that sense, we have taken a similar path to that pioneered by [2] and [3], who also combine X-ray and video modalities to achieve the registration of images produced by a (mobile) C-arm and a standard RGB camera. However, our target is quite different since our primary goal is to achieve 3D reconstruction using standard primary/diagnostic X-ray acquisition systems

Francisco Albiol is senior researcher at Instituto de Física Corpuscular (IFIC), Universitat de València, Consejo Superior de Investigaciones Científicas (Spain), e-mail: kiko@ific.uv.es.

Alberto Corbi is a PhD student at IFIC, e-mail: alberto.corbi@ific.uv.es.

Alberto Albiol is a senior researcher at Universitat Politècnica de València (Spain) e-mail: alalbiol@iteam.upv.es.

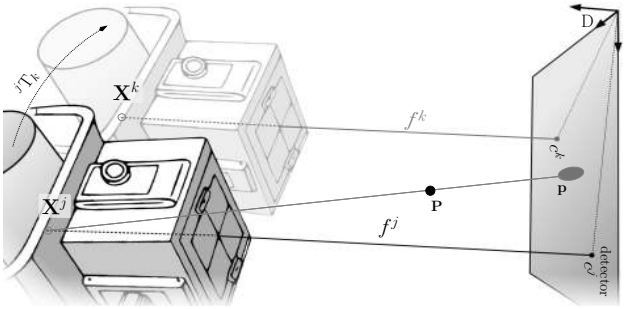


Fig. 2. Camera intrinsic parameter variation when the X-ray emitter moves.

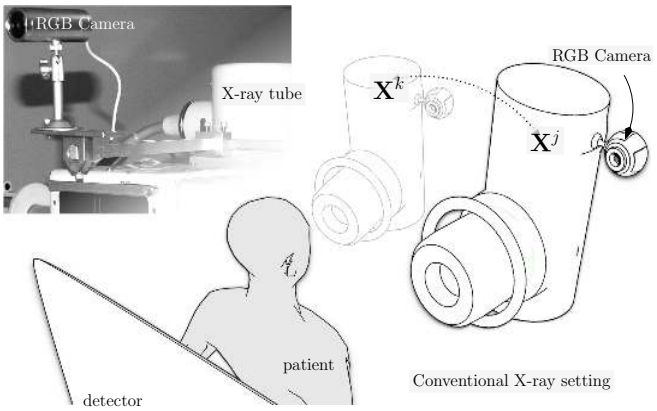


Fig. 3. Proposed X-ray acquisition system with an attached digital camera device that allows the tracking of the position/orientation of the X-ray source.

(instead of C-arms). These X-ray devices can be found in most healthcare centers worldwide, private and public, large and small. In the reconstruction process, only the aforementioned camera and a set of visual fiducial markers are used so that X-ray images do not get tainted with traces/projections from foreign objects. Our proposed method initially needs a calibration phase to retrieve the geometrical setting of both cameras. Afterwards, subsequent movements and/or rotations of the X-ray system or the examined object can be precisely tracked with the help of the RGB device. With this spatial information, X-ray projection matrices are then calculated and multiple radiographic image reconstruction can take place.

We also study two different scenarios that might be appropriate for X-ray examinations. In the first scenario, the camera system (X-ray + video) moves around the examined subject/object. Under this scheme, we successfully manage to *recreate* X-ray camera intrinsic parameters. The second scenario entails moving and/or rotating the radiographed patient (or item) while keeping both the camera system and the X-ray detector fixed. Both settings can be relevant in clinical diagnosis and/or object monitoring/scanning.

Before depicting the aforementioned scenarios of application, we first review how X-ray images (generated by the most widely used radiological modalities) are commonly calibrated. Then, a methodology to compute X-ray projection matrices from ordinary color or grayscale images is summarized, followed by a portrayal of the calibration stage. Subsequently, we discuss some practical epipolar geometry tools and how they can be useful for diagnosis in plain X-ray imaging. Finally, a few tests carried out with phantoms (geometrical and anthropomorphic) and with a real brachytherapy treated patient are presented and their results are discussed.

## II. Related work

X-ray calibration techniques and X-ray image registration procedures commonly use calibrated C-arms and CT scanners [4], which involve, of course, having access to such radiological devices. Tomosynthesis [5] devices are more affordable and are already widely used in digital screening mammography. However, their application in everyday X-ray examinations, where legacy hardware is the common denominator, may require the complete renewal of the imaging set.

In order to achieve the same goal with less expensive, more accessible/widespread radiological tools and lower dose levels, several approaches have been (or are currently being) explored. All these efforts involve the use of an external and adjacent device that interplays with the X-ray imaging apparatus. For instance, there is a trend in research focused on using depth and time-of-flight cameras in order to reconstruct 3D data and 3D models of objects being radiographed or scanned. This 3D data is then combined with X-ray images to obtain different and meaningful information. As an example, the approach followed by [6] combines 3D modeling with X-ray images in order to three-dimensionally locate and define the shape and silhouette of hidden objects inside boxes, which can have very interesting applications in surveillance and QA processes. Another precedent can be found in the study carried out by [7], who estimates patient's size, volume and appropriate dose with the help of a Microsoft Kinect™ device. Other attempts like the one described by [8] try to rebuild the X-ray system extrinsic parameters with the help of a distance meter device (in this case, a laser rangefinder located close to the X-ray emitter). The problem is further simplified by the authors of [9], who require each X-ray source and sensor to be placed at known locations. Conventional approaches use special X-ray calibration structures that accommodate fiducials that leave visible *shadows* in the radiographic image. These foreign frames remain present during each snapshot (either attached to the patient [10] or to the X-ray system) *contaminating* the acquired radiographs with their own projected traces, introducing artifacts and invading the patient's space.

As stated in the introduction, there also exist recent and laudable research efforts around interplaying video information with X-ray imaging. In [11], [12] and [13] researchers calibrate, model and study the clinical and surgical applications of camera augmented mobile C-arms, which also involve the precise registration of visible and X-ray images. Close to this work, we find that carried out by [14], which highlights the contribution of external cameras to radiation exposure reduction and surgery planning improvement. Radiograph mosaicking is also a subject of interest. In this direction, we find interesting citing the work performed by researchers in [15] who focus on accurate X-ray image stitching (also in C-arm modalities). Their goal requires a pure rotation around the X-ray source center which is accomplished with the help of visual

fiducials and a video camera that contribute to estimating the translational part of the motion so that it can be later compensated. Similarly to this approach, we also derive the translational motion of the X-ray source with the help of visual markers and a RGB camera. However, two key differences arise. First, one of the main goals of [15] is the registration of the X-ray and RGB modalities. For this reason a set of mirrors is used so that both optical centers are made coincident. In our case this is not necessary because we do not perform image registration. Second, the X-ray source in [15] undergoes a pure rotational motion so that it is possible obtain a parallax-free mosaic of both imaging systems. Regarding our goal, we are also interested in the translational part of the beam origin in order to derive 3D information from plain radiographs, as later tackled in Section V-B.

Beyond image-to-image registration, we find increasing interest in relating bitmap content (radiographs, CT/MR slices and video) to volumes. The authors of [16] and [17] present a concise review of state-of-the-art around the topic applied to minimally invasive therapy and image guided interventions. The reconstruction of 3D structures from sets of 2D X-ray projections is studied in [18] and [19] with the help of custom designed phantoms. However, none of these research efforts explore the event of an alteration of the geometrical setting. Researchers in [20] and [21] try to tackle this problem by using image similarity measures (entropy, intensity, gradient, patterns, etc.) which can be used (with some difficulty) without the need of a phantom or an *ad hoc* calibration phase.

In addition, the literature on the combination of visual fiducials with radiology, medicine, and surgical environments is steadily growing. The works in [22], [23], [24] and [25] are just a few examples of applications of the realm of augmented reality (AR) in hospital and clinical environments.

## III. Derivation of X-ray projection matrices from visual information

Unlike other setups that can retrieve 3D information (like C-arm), ours relies on standard clinical X-ray imaging systems, like those comprised of a 4 degrees of freedom (DOF) *movable* X-ray source and a 1 or 2-DOF vertical/horizontal sliding imaging plate or IP (which receives and integrates the emitter radiation). During examinations, the patient is placed vertically next to this detector (*standing erect*) or horizontally over it (*supine anteroposterior projection*). In these radiological settings, there is no way to know with enough precision, the beam source spatial position relative to the detector or, more generally, to a coordinate reference frame (later referred to as *world*). Therefore it is not aprioristically possible to find projection matrices and achieve geometric reconstruction.

However, X-ray imaging devices have been commonly modeled as pinhole cameras, which enables the application of projective geometry. The research efforts carried out by [26], [27], [9] and [28] represent a few examples, and a more succinct introduction on the subject applied to the context of medical imaging can be found in Section 2.2.1 of [29] and chapter 20 of [30]. As stated in our introduction, in order to combine several X-ray images from the same object/patient at each different geometrical setting $j$, we need associated camera projection matrices $\mathrm{P}^j$, $\mathrm{P}^k$, etc. Each $\mathrm{P}^j$ relates 3D points ${}^{\mathrm{W}}\mathbf{Q}_i$ in a coordinate reference frame called *world* (W) with their 2D *observed* projections $\mathbf{q}_i^j$ on each radiographic image $j$. Using projective geometry, we can write:

$$\hat{\mathbf{q}}_i^j = \mathrm{P}^j \cdot {}^{\mathrm{W}}\hat{\mathbf{Q}} \tag{1}$$

where ${}^{\mathrm{W}}\hat{\mathbf{Q}}_i$ and $\hat{\mathbf{q}}_i$ are the homogeneous coordinates of ${}^{\mathrm{W}}\mathbf{Q}_i$ and $\mathbf{q}_i^j$, respectively. With a RQ decomposition, $\mathrm{P}^j$ can be expressed as:

$$\mathrm{P}^j = \mathrm{K} \cdot {}^{\mathrm{X}^j}\mathrm{T}_{\mathrm{W}} \tag{2}$$

where $\mathbf{K}$ is a $3 \times 3$ upper triangular matrix that contains the intrinsic parameters of the X-ray system (for a given geometrical setting *j*) and ${}^{\mathrm{X}^j}\mathrm{T}_{\mathrm{W}}$ is a rigid transformation that translates 3D homogeneous points relative to W to coordinates of the X-ray camera ($\mathrm{X}^j$), whose reference frame is centered at the radiation emitting anode (focal point) and one of its axes is orthogonal to the X-ray detector plane.

As already highlighted in Section II, conventionally in X-ray imaging, projection matrices are obtained with the help of a calibration frame equipped with fiducials $\mathbf{Q}_i$ that are opaque to the Roentgen radiation which are then projected to $\mathbf{q}_i$ spots in the image. This frame is placed around or over the examined object/patient. Combinations of $\mathbf{Q}^j{}_i$, $\mathbf{q}_i^j$ pairs are then fed into a calibration algorithm such as Direct Linear Transform or DLT (introduced by [31] and very succinctly described in chapter 4.1 of [32]) and projection matrices can be then derived.

The problem with such radiation-opaque fiducials is that they usually leave visible traces, artifacts and extra Compton contribution in the radiographic image that can seriously obstruct the analysis and/or diagnosis process. They also can be easily projected outside the imaging sensor area when using oblique protocols, like those exemplified in [33] and in Fig. 4, which can harden their application. Last but not least, essential supporting frames in which they are usually accommodated, can be perceived as invasive by the patient.

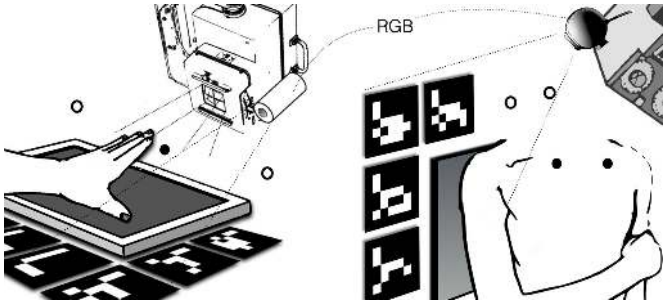


Fig. 4. Examples of lateral and oblique X-ray protocols whose projections (i.e., imaged fiducial set of black and white circles) and principal point may lie outside the final image (i.e., white circles). However the system can still be geometrically tracked through visible information and RGB cameras.

In this context, the following sections describe in detail how each $\mathrm{P}^j$, for each different position *j* of the X-ray emitter (or patient), can be obtained using just the help of an interplaying

visible light camera that has a broader field of view and whose fiducials are transparent to the radiograph production process.

### A. Coordinate systems

Although some coordinate systems have already been briefly mentioned, here we review in depth the ones that are used in our proposal. These are graphically summarized in Fig. 5 where, for clarity, we assume the camera system is located at a given position $j$ (i.e., $\mathrm{X}^j$ and $\mathrm{V}^j$).

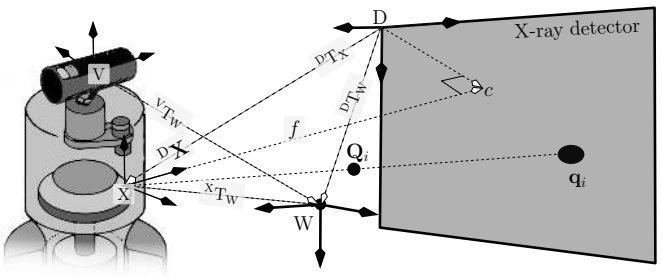


Fig. 5. Coordinate frames available in an X-ray setting. The point ${}^{\mathrm{D}}\mathbf{X} = (c_x, c_y, f)$ represents the anode location in detector coordinates. ${}^{\mathrm{D}}\mathrm{T}_{\mathrm{X}}$ is the rigid transformation that connects the detector and anode reference frames.

- W, or the world coordinate system, whose origin is a 3D spot in the scene.
- D, the reference system of the X-ray detector, whose origin is usually situated at the upper-leftmost pixel with one of its axes (*Z*) being orthogonal to the detector plane.
- X, the X-ray coordinate system, whose origin is the beam source and which also has one axis (i.e., z) orthogonal to the detector by definition. Note that there is a different $\mathrm{X}^j$ coordinate system for each spatial position of the X-ray emitter but all of them have their z-axis orthogonal to the detector plane.
- V, the video camera coordinate system. Since it is rigidly attached to the X-ray emitter, there will also be a different $\mathrm{V}^j$ for each location of the anode.

Fig. 5 shows that the relation between reference frames D and X is always a $\pi$ rotation around the horizontal axis ($x$) of D and a translation in the case of X-ray cameras. This relation can be coded as a rigid transformation expressing how to translate points (in homogeneous coordinates) from X to D as follows:

$$ {}^{\mathrm{D}}\hat{\mathrm{T}}_{\mathrm{X}} = \begin{bmatrix} \mathrm{R}_x(\pi) & {}^{\mathrm{D}}\mathbf{X} \\ \mathbf{0} & 1 \end{bmatrix} = \begin{pmatrix} 1 & 0 & 0 & c_x \\ 0 & -1 & 0 & c_y \\ 0 & 0 & -1 & f \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{3} $$

where ${}^{\mathrm{D}}\mathbf{X}$ is the beam origin of the X-ray expressed in D coordinates. The mark ($\hat{}$) over ${}^{\mathrm{D}}\hat{\mathrm{T}}_{\mathrm{X}}$ indicates that an extra row $(0, 0, 0, 1)$ has been appended to the transformation matrix, which allows us to work in homogeneous coordinates.

### B. Intrinsic and extrinsic parameters in X-ray cameras

In a radiological device that is modeled as a pinhole camera, the beam origin is represented by the tube anode, which also plays the role of the optical center. If there are no lenses present, we can ignore spherical aberrations, radial distortions, and skew ($s = 0$), w.l.o.g. Using perspective projection equations, it is possible to transform 3D points in X coordinates to 2D homogeneous radiographic image dots:

$$ \hat{\mathbf{p}}_i = \mathrm{K} \cdot {}^{\mathrm{X}}\mathbf{Q}_i \tag{4} $$

The intrinsic parameters matrix K can be decomposed as:

$$ \mathrm{K} = \begin{pmatrix} \alpha_x & 0 & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{pmatrix} = \underbrace{\begin{pmatrix} \lambda_x & 0 & 0 \\ 0 & \lambda_y & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{\lambda} \cdot \begin{pmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{pmatrix} \tag{5} $$

where $c_x$, $c_y$ are the coordinates of the so-called *principal point* and $f$ is the focal length, which perpendicularly connects the anode and the detector plane. These components are expressed in *pixels*, but they can be transcribed to spatial dimensions (meters) if multiplied by a known resolution $\lambda$, which is provided by the detector manufacturer. It is conventionally assumed that $\lambda$ is axis-independent, i.e., $\lambda_x = \lambda_y$.

Fig. 5 and Eq. (5) show that the components of the optical center relative to D are coincident with the intrinsics:

$$ {}^{\mathrm{D}}\mathbf{X} = (c_x, c_y, f) \tag{6} $$

Eqs. (5) and (6) reveal that the intrinsic parameters of any X-ray system will change if either the sensor or the emitter are moved, as already highlighted in Section I. This variability lies in the fact that the detector surface and the radiation source are detached from each other. This peculiarity vividly contrasts with video camera systems, where the sensor (typically a CCD/CMOS array) stays fixed relative to the optical center.

The extrinsic parameters can be written as a rigid transformation that relates points in word coordinates (W) and X-ray camera coordinates (X) and can be derived from ${}^{\mathrm{D}}\mathbf{X}$ with the following expression:

$$ {}^{\mathrm{X}}\hat{\mathrm{T}}_{\mathrm{W}} = \left({}^{\mathrm{D}}\hat{\mathrm{T}}_{\mathrm{X}}\right)^{-1} \cdot {}^{\mathrm{D}}\hat{\mathrm{T}}_{\mathrm{W}} \tag{7} $$

where ${}^{\mathrm{D}}\mathrm{T}_{\mathrm{X}}$ can be obtained using Eq. (3) and ${}^{\mathrm{D}}\mathrm{T}_{\mathrm{W}}$ is one of the results of the initial calibration process (described in Section IV) that maps points from W to D.

### C. Scene tracking with visible fiducials

From Eqs. (5) and (7), we conclude that the spatial location of the X-ray optical center ${}^{\mathrm{D}}\mathbf{X}$ is essential for the calculation of the intrinsic and extrinsic parameters of the X-ray system. In our proposal, after a single calibration (itemized in Section IV), ${}^{\mathrm{D}}\mathbf{X}$ is tracked with the help of visible fiducial markers and an ordinary RGB camera. An initial optical calibration of this device is necessary, which provides its invariant intrinsics and contributes to minimizing distortions produced by lenses.

Extrinsic parameters ${}^{\mathrm{V}}\mathrm{T}_{\mathrm{W}}$, which connect points in W coordinates to the V reference frame, can be determined with the help of visible fiducials with known 3D coordinates. Some examples of other commonly used visible fiducial markers are

ReacTIVision [34], Intersense [35], BinARyID [36], ARTag [37] and Cantag [38]. In this work, we use the Aruco markers described in [39]. The Aruco framework was chosen over other alternatives because of its robustness against noise and vertex jitter. There are 1024 different instances of this fiducial system that can be easily detected in real time. Figs. 6 and 11 show a few examples of these instances. Aruco also comes with a companion C++ library and utilities that allow a straight and fast integration in our workflow. An interesting aspect from this fiducial system is that the *camera pose* (location and orientation) can be determined with just one Aruco marker with enough fidelity, if necessary. This feature makes the system very robust to occlusions made by scene objects. However, in a normal setting many Aruco markers are detected and the camera pose is retrieved with high precision.

Since the coordinates of the anode relative to the external camera (${}^{\mathrm{V}}\mathbf{X}$) are constant (obtained in the calibration stage discussed in Section IV) and given that both imaging systems are rigidly tied, we are allowed to write:

$$ {}^{\mathrm{D}}\hat{\mathbf{X}} = {}^{\mathrm{D}}\hat{\mathrm{T}}_{\mathrm{W}} \cdot \left( {}^{\mathrm{V}}\hat{\mathrm{T}}_{\mathrm{W}} \right)^{-1} . {}^{\mathrm{V}}\hat{\mathbf{X}} \tag{8} $$

Together with Eq. (7), we can compute the X-ray intrinsic and extrinsic parameters of the X-ray imaging system.

### D. Application scenarios

We envision two possible operation scenarios (Fig. 7).

*1) Moving camera scenario :* This scenario is the most intuitive from the point of view of stereo and computer vision and is undoubtedly more suitable if the injury being examined prevents the patient from moving safely. In this setting, the X-ray source is placed at different locations and orientations relative to W, and an X-ray image of the patient (who stands still) is generated at each position *j*. Using visual fiducials and the process described in Section III-C, we can determine the location of the beam source ${}^{\mathrm{D}}\mathbf{X}^{j}$ for each setting $j$.

Note that, in this scenario, the intrinsic part ($\mathrm{K}^{j}$) of $\mathrm{P}^{j}$ changes when the X-ray emitter is moved. Nevertheless, $\mathrm{K}^{j}$ can be easily updated using Eqs. (5) and (6) together with the relation ${}^{\mathrm{D}}\mathbf{X}^{j}$ for each setting $j$. Finally, the projection matrix for each geometrical setting can be obtained as:

$$ \mathrm{P}^{j} = \mathrm{K}^{j} \cdot {}^{\mathrm{X}^{j}}\mathrm{T}_{\mathrm{W}} \tag{9} $$

where ${}^{\mathrm{X}^{j}}\mathrm{T}_{\mathrm{W}}$ is the outcome of Eq. (7).

*2) Moving patient scenario:* In this setting, the X-ray emitter remains fixed and the object/patient rotates and/or moves. A different radiograph is produced at each patient/object position $j$. If the patient is able to move him/herself and/or his/her examined anatomic part, this scenario might turn out interesting because it allows larger geometrical changes (it also depends on the examination protocol being applied). Visual markers have to be rigidly fixed somehow to the patient as in the example shown in Fig. 7-right. This scenario might also be interesting for object inspection, i.e., boxes, packages, etc., that can easily be rotated and where markers can be easily stuck over flat surfaces.

The first key difference when compared with the previous setup is that, in this case, intrinsic parameters remain constant ($\mathrm{K}^{j} = \mathrm{K}$) because the position of the X-ray emitter (obtained during calibration) also remains fixed (i.e., ${}^{\mathrm{D}}\mathbf{X}^{j} = {}^{\mathrm{D}}\mathbf{X}$).

The second key difference is that what is tracked is not the absolute position of the X-ray emitter relative to W but the relative location between a new world reference system $\mathrm{W}'$ (attached to the patient) and the X-ray emitter. The 3D points $\mathbf{Q}_i$ of Eq.1 are then expressed in $\mathrm{W}'$ coordinates (i.e., ${}^{\mathrm{W}'}\mathbf{Q}_i$).

Using the visual tracking described in Section III-C, we measure ${}^{\mathrm{V}^{j}}\mathrm{T}_{\mathrm{W}'}$, which connects points in the $\mathrm{W}'$ reference system and visible light camera coordinates for each patient location/orientation *j*. With this information, the projection matrix in this scenario can be expressed as:

$$ \mathrm{P}^{j} = \mathrm{K} \cdot {}^{\mathrm{X}}\mathrm{T}_{\mathrm{W}} . \left( {}^{\mathrm{V}^{j}}\mathrm{T}_{\mathrm{W}} \right)^{-1} \cdot {}^{\mathrm{V}^{j}}\mathrm{T}_{\mathrm{W}'} \tag{10} $$

where K, ${}^{\mathrm{X}}\mathrm{T}_{\mathrm{W}}$ and ${}^{\mathrm{V}^{j}}\mathrm{T}_{\mathrm{W}}$ are just results of the initial calibration (presented in Section IV). Note that ${}^{\mathrm{V}^{j}}\mathrm{T}_{\mathrm{W}}$ remains constant for each patient position because the visible light camera remains fixed relative to the original W.

## IV. Calibration phase

The goal of the calibration step is to obtain the necessary (and invariant) geometrical relations connecting scene elements, imaging systems and reference frames to each other and to W. To that end, the Teflon structure shown in Fig. 6, was designed. Once this information is obtained, the calibration frame can be removed from the scene.

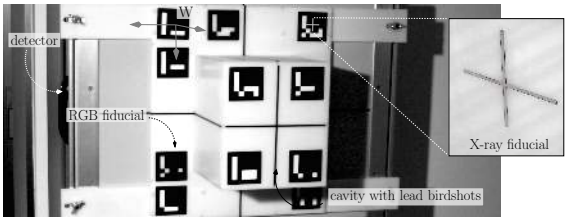


Fig. 6. Calibration frame with X-ray and visible Aruco fiducials.

It accommodates, at two different planes, 13 tin/copper, cross-shaped markers which are opaque to X-rays. It also contains 12 visible Aruco fiducials that can be easily detected using the visible light camera. The coordinate frame W is centered in one of the fiducials, and the 3D coordinates of all of the fiducials (visible and opaque to radiation) are known relative to it after a careful construction process. The calibration phase can be summarized in the following steps:

1) The calibration frame is introduced in the scene, and a radiograph and a RGB or grayscale image are generated from it (similar to those shown in Fig. 11).
2) ${}^{\mathrm{V}}\mathrm{T}_{\mathrm{W}}$ is calculated with the acquired photograph and the methodology in Section III-C. This transformation is used in both application scenarios in Eqs. 8 and 10.
3) An initial X-ray projection matrix P is computed with the DLT algorithm and combinations of 3D locations of X-ray-opaque fiducials and their corresponding 2D projections on a calibration X-ray instance.

4) Intrinsic K and extrinsic ${}^{\mathrm{X}}\mathrm{T}_{\mathrm{W}}$ parts are extracted from the RQ decomposition of P. Matrix K (unaltered) is used in the moving patient scenario, specifically in Eq. (10).
5) Vector ${}^{\mathrm{D}}\mathbf{X}$ and matrix ${}^{\mathrm{D}}\mathrm{T}_{\mathrm{X}}$ are rebuilt from K using Eqs. (5) and (3) together with the detector resolution provided by the manufacturer. Finally, the rigid transformation ${}^{\mathrm{D}}\mathrm{T}_{\mathrm{W}}$ used in Eq. (8) is computed.
6) The relative position between the X-ray emitter and the camera ${}^{\mathrm{V}}\mathbf{X}$ is also obtained. This relation remains invariant and is later applied in Eq. (8).

$$ {}^{\mathrm{V}}\mathbf{X} = {}^{\mathrm{V}}\mathrm{T}_{\mathrm{W}} \cdot \left({}^{\mathrm{D}}\mathrm{T}_{\mathrm{W}}\right)^{-1} \cdot {}^{\mathrm{D}}\mathbf{X} \tag{11} $$

7) When all the aforementioned relations have been collected the calibration frame is no longer necessary.

## V. 3D INFORMATION FROM PROJECTION MATRICES

In this section, we examine how our technique can be used in real application scenarios and how 3D information from correlated X-ray images can be extracted. From this point onwards, projection matrices are obtained with the help of visible fiducials as explained in Section III. Fig. 7 shows some examples of this operation stage.

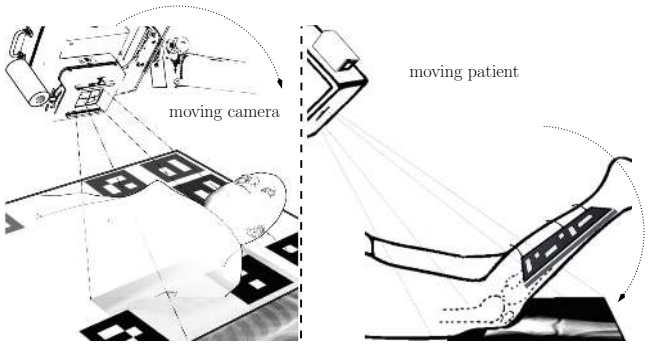


Fig. 7. Applications of the procedures detailed in this text. Depending on the scenario of application, visible fiducials can be placed over the detector, a wall, or the examination table (left), or they can be tied to the patient or even stuck over scanned items (right).

### A. Epipolar lines between radiographic images

A key step in stereo imaging (including diagnostic X-ray imaging) involves finding point correspondences in two stereo images. Using epipolar geometry, the search for a corresponding point $\mathbf{q}_i^j$, which is initially observable in image $j$, can be reduced to a *search through a line* on the second radiograph *k*. The line on this second image is called the *epipolar line*. This simple technique can dramatically contribute to resolving ambiguities when two points of interest lie very close to each other in one image. This could create some difficulty in distinguishing between them. However, they are easily recognizable in another projection (i.e., a *paired* radiograph), where they can be more efficiently matched by a corresponding epipolar line.

Given two paired X-ray images along with their corresponding projection matrices $\mathrm{P}^j$ and $\mathrm{P}^k$, we can compute the Fundamental Matrix F presented in chapter 9 of [32] which enables the mapping of any observed point $\mathbf{q}_i^j$ on the first X-ray image $j$ to an *artificial* infinite epipolar $\mathsf{l}_i^k$ in image $k$.

This epipolar line can be *bounded* or shortened if the rough dimensions of the object under examination (e.g., chest/patient thickness) are taken into account. This restriction simplifies the search on the second image to a much shorter segment ($\updownarrow_i^k$) instead of the infinite $\mathsf{l}_i^k$. In order to obtain the limits of the segment for a particular point $\mathbf{q}_i^j$, we begin by *back-projecting* it into a 3D infinite ray $\overline{\mathsf{Q}_i^j}(\zeta)$:

$$ \overline{\mathsf{Q}}_i^j(\zeta) = \left(\mathrm{P}^j\right)^+ \cdot \hat{\mathbf{q}}_i^j + \zeta \mathbf{X}^j \tag{12} $$

where $\mathbf{X}^j$ is the anode location in W coordinates and $\zeta$ is a scalar that parametrizes the ray that passes over the point $\mathbf{q}_i^j$, continues towards $\mathbf{Q}_i$, and finally reaches the anode $\mathbf{X}^j$ (see Section 6.2.2 of [32] for details). Next, the boundaries $\mathsf{Q'}_i$ and $\mathsf{Q''}_i$ of the segment $\overline{\mathsf{Q}}_i$ that are coherent with the rough thickness of the studied object/person are iteratively matched. Finally, these two confining 3D points are *reprojected* on the second X-ray image $k$:

$$ \left.\begin{array}{l} \hat{\mathsf{q}}'^k_i = \mathrm{P}^k \cdot \hat{\mathsf{Q}}'_i \\ \hat{\mathsf{q}}''^k_i = \mathrm{P}^k \cdot \hat{\mathsf{Q}}''_i \end{array}\right\} \updownarrow_i^k = \left(\mathsf{q}'^k_i, \mathsf{q}''^k_i\right) \tag{13} $$

and a *bounded* $\updownarrow_i^k$ between them can be plotted (Fig. 8).

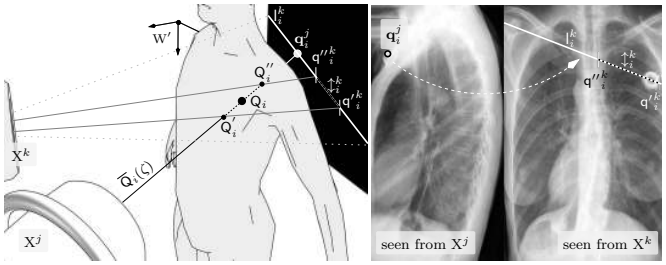


Fig. 8. Point reconstruction using two X-ray images with epipolars.

Real examples of bounded epipolars are shown in Fig. 13 for the moving camera scenario.

### B. 3D reconstruction from two image pairs

This process is also known as *projection-to-volume* reconstruction or *projective reconstruction* and is described in Section 12.2 of [32] and in [40]. It enables the determination of the 3D location of an observed point $\mathbf{q}_i$ in two images ($j$ and $k$). Given two projection matrices $\mathrm{P}^j$ and $\mathrm{P}^k$ and using Eq. (1), we can write:

$$ \hat{\mathsf{q}}_i^j = \mathrm{P}^j \cdot \hat{\mathbf{Q}}_i \qquad \hat{\mathsf{q}}_i^k = \mathrm{P}^k \cdot \hat{\mathbf{Q}}_i \tag{14} $$

Since we are working in homogeneous coordinates, the equivalence between two points has to be expressed using the cross product:

$$ \begin{array}{l} \hat{\mathbf{q}}_i^j \times \hat{\mathsf{q}}_i^j = \hat{\mathbf{q}}_i^j \times \mathrm{P}^j \cdot \hat{\mathbf{Q}}_i = \mathbf{0} \\ \hat{\mathbf{q}}_i^k \times \hat{\mathsf{q}}_i^k = \hat{\mathbf{q}}_i^k \times \mathrm{P}^k \cdot \hat{\mathbf{Q}}_i = \mathbf{0} \end{array} \tag{15} $$

Each of these expressions determines two linearly independent equations that can be written in the form of a linear system. When solved through a single value decomposition method (SVD), we can derive the 3D location of a specific point $\hat{\mathbf{Q}}_i$ observed in the two images.

## VI. Tests and results

In this section, the set of techniques described in Sections III to V are tested. In the first group of experiments, we make use of the calibration frame presented in Section IV as a device under test (DUT). Then, we perform similar research on an arm-shaped anthropomorphic phantom, which allows us to show the same procedures with a more realistic target. Finally, we evaluate our methodology in a real radiotherapy environment. Regarding the first two tests, results are provided for both the DUT and the phantom, and in each case, for the moving camera/patient scenarios.

### A. Tests on the calibration frame acting as DUT

In this part, the calibration frame will play the role of a DUT. Graphical representations of the two application scenarios are shown in Figs. 9 and 10. Within this test, the X-ray fiducials used during calibration are completely ignored. This means that each $\mathrm{P}^j$ is now derived using only visual information. In order to evaluate the algorithm, we placed nine 2 mm radius lead spherules or bearings at different well-known positions inside the calibration frame (see Fig. 12-left).

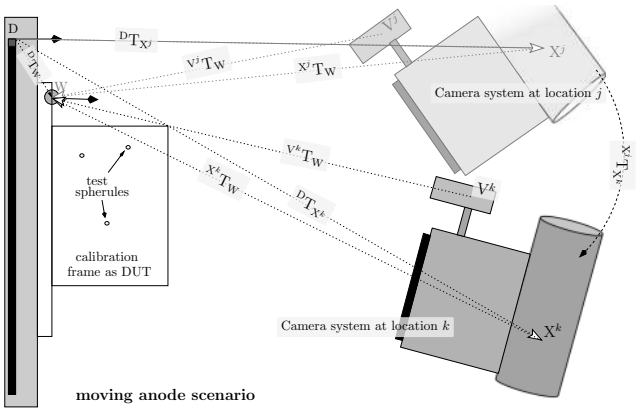


Fig. 9. Moving camera system scenario consisting of a moving anode. D, X and V stand for the coordinate frames described in Section III-A. The rigid transformations ${}^{\mathrm{D}}\mathrm{T}_{\mathrm{X}^j}$ and ${}^{\mathrm{D}}\mathrm{T}_{\mathrm{X}^k}$ are estimated from visual fiducials only detected at each camera location $\mathrm{V}_j$ and $\mathrm{V}_k$ as explained in Section III-D1. X-ray projection matrices $\mathrm{P}^j$, $\mathrm{P}^k$, etc., are derived with Eq. (9).

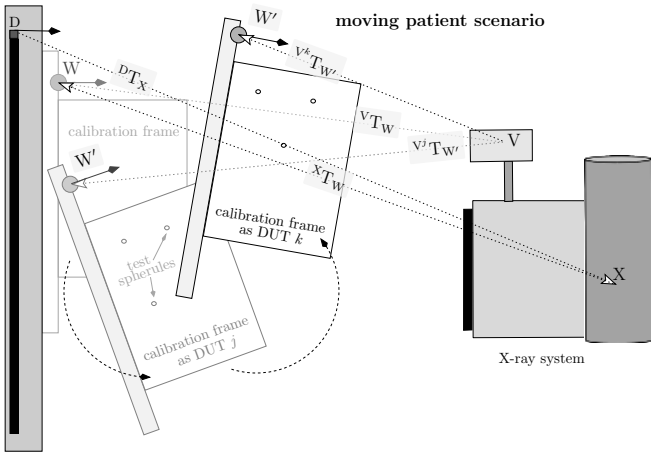


Fig. 10. Moving patient application scenario with the calibration frame now acting as a DUT. The DUT is shifted to other positions while the anode remains fixed. Intrinsics remain constant (i.e., $\mathrm{K}^j = \mathrm{K}^k = \mathrm{K}$) and each DUT position is tracked using visual information. X-ray projection matrices $\mathrm{P}^j$, $\mathrm{P}^k$, etc., are obtained with Eq. (10) and as discussed in Section III-D1.

The DUT was radiographed ten times for each of the application scenarios. Fig. 11 shows a few examples of the captured images. We established several evaluation metrics in order to check the goodness of the results. All metrics make a distance comparison between the *real location* of the lead spherules and the *estimated location* predicted by using the generated projection matrices.

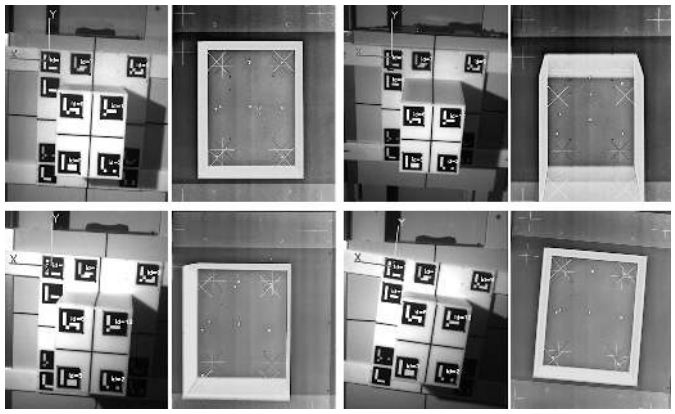

Fig. 11. Moving camera (top) and moving patient (bottom) scenarios.

*1) The mean distance between* generated *projections of known landmark locations and their observed image coordinates:* This is the most basic quality test for the projection matrices. It consists of measuring the mean 2D image distance between the *observed projection* $\mathbf{q}_i$ (on the image $i$) of each of the 9 spherules ($\mathbf{Q}_i$) and their *computed projection* $\mathrm{q}_i$, estimated with Eq. (1). The results (highlighted in Fig. 12-right) show this mean distance is equal to 12 px (with a dispersion of 8 px) for the moving camera scenario and 8 px (with a dispersion of 6 px) for the patient moving scenario. As expected, the difference in accuracy between the two scenarios is not significant since the geometry involved in both problems is very similar.

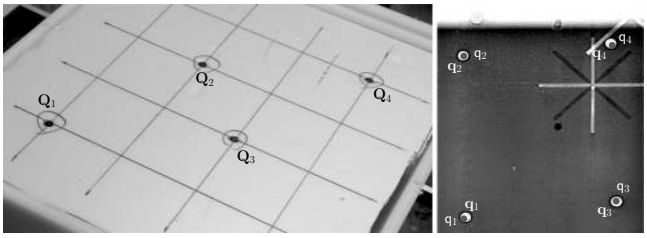


Fig. 12. (Left) Lead spherules inside the calibration frame (acting now as a DUT) placed at 3D known locations $\mathbf{Q}_i$. (Right) An example of the level of agreement between some lead spherule projection centers $\mathbf{q}_i$ (white dots) and their predicted image location $\mathrm{q}_i$ (gray dots) by using the estimated projection matrix $\mathrm{P}^j$. The cross-shaped mark at the top-right corner corresponds to one of the fiducial set that previously played a role during calibration (Fig. 6) but is now ignored during the whole testing phase.

*2) The distance between epipolars and projections of spherules:* The proposed metric consists of computing the

mean 2D distance (for all possible radiograph combinations) between each lead spherule, whose observed projection is $\mathbf{q}_i^k$, and its corresponding (and calculated) bounded epipolar line $\updownarrow_i^k$. Again this mean distance for both moving camera and moving patient scenarios were obtained.

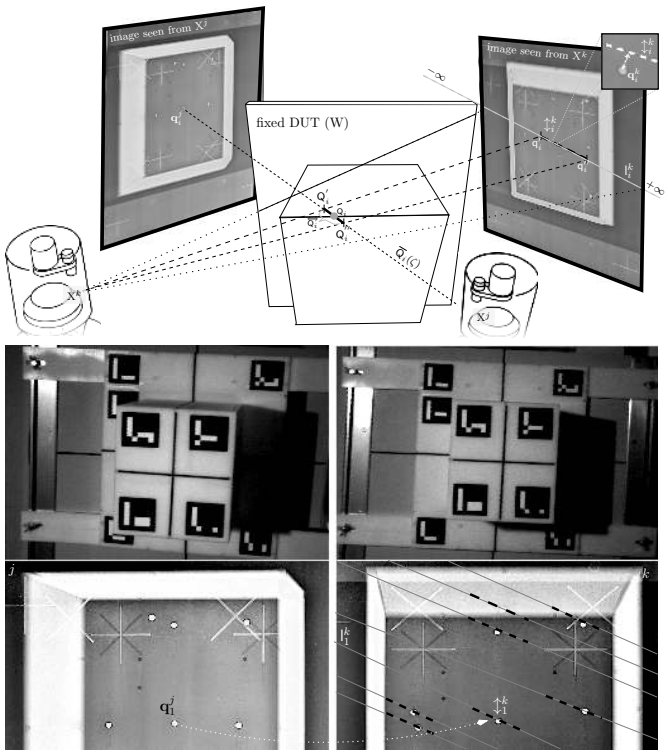


Fig. 13. (Top) Schematic representation of the geometrical relations described in Section V-A for a moving camera scenario (detector appears twice and in two positions for the sake of clarity). The segment $\overline{\mathsf{Q}_i}$ roughly spans the mean length of the calibration frame ($|\mathsf{Q}\text{''}_i - \mathsf{Q}\text{'}_i| \approx 0.24\,\mathrm{m}$). The projection of this segment in the X-ray image is $\updownarrow_i^k$. (Bottom) Example of resolved bounded epipolars in a moving patient scenario.

Epipolars were bounded assuming an average DUT thickness of 24 cm. Fig. 13 illustrates this test and some of its results, which confirm a mean distance of 13 px with a dispersion of 8 px for the camera moving scenario and a mean distance of 10 px with a dispersion of 6 px for the patient moving scenario. Again the differences between scenarios is not significant.

*3) The distance between back-projections and real 3D locations:* Since the 3D location of the spherules is very well known (Fig. 12-left), we test the accuracy of the algorithm described in Section V-B by calculating the mean value of the distances between the known 3D locations $\mathbf{Q}_i$ and predicted ones $\mathsf{Q}_i$ using all possible radiograph combinations and all spherules. Outcomes indicate a mean offset of 2 mm (deviation of 2 mm) when the spherules are located 2 meters away from the anode (approximately) and a mean baseline of 45 cm (displacement of X-ray anode between radiographs).

### *B. Tests on anthropomorphic phantom*

In order to check our procedures on a more real target, we use an arm-shaped anthropomorphic phantom from Life/form® that includes a splinter fracture (Fig. 14).

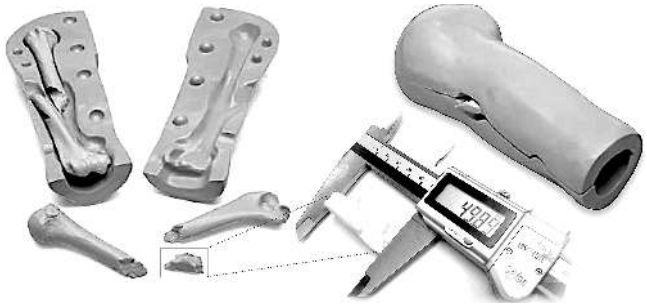

Fig. 14. Anthropomorphic arm-shaped phantom with a 5 cm splinter fracture.

With this phantom, we applied the techniques described in Section VI-A3 in order to compute the splinter length from X-ray image pairs and to compare it with the real length ($\sim$ 5 cm). Obviously, in cases with severely injured patients, the moving camera scenario (immobilized patient) might be more appropriate. However, we also analyzed the moving patient scenario for the sake of completeness. Both of them are graphically summarized in Fig. 15.

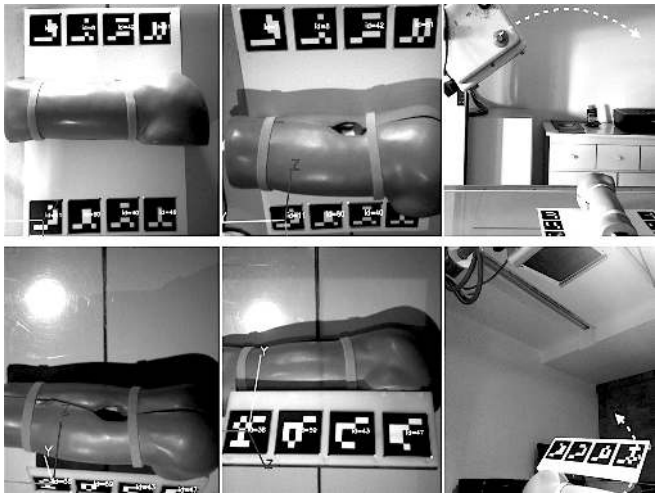

Fig. 15. Anthropomorphic phantom tested for the camera moving scenario (top) and patient moving scenario (bottom). In the camera moving scenario, a board with visible fiducials at known 3D coordinates relative to W is used and several radiographs are generated at different angles/locations. In the patient moving scenario, the board of visible fiducials is rigidly tied to the object under examination, which arbitrarily moves and rotates while the X-ray imaging system remains fixed.

Combining all available image pairs (similar to those in Fig. 16-bottom) we obtained a mean length of 4.9 cm with a deviation of 0.1 cm for the fixed arm setting. In the case of the moving phantom, we retrieved a mean distance of 5.0 cm with a deviation of 0.2 cm.

Bounded epipolars are drawn in Fig. 16-top using the algorithm described in Section VI-A2. For this experiment, several lead spherules were added as we did with the DUT in Section VI-A. The distances between computed epipolars and real observed projections was under 15 px for the camera moving scenario and under 12 px for the phantom moving scenario. This result eases a good identification of hidden facets which are missing or indistinguishable in one image, but that are quite evident in a second radiograph.

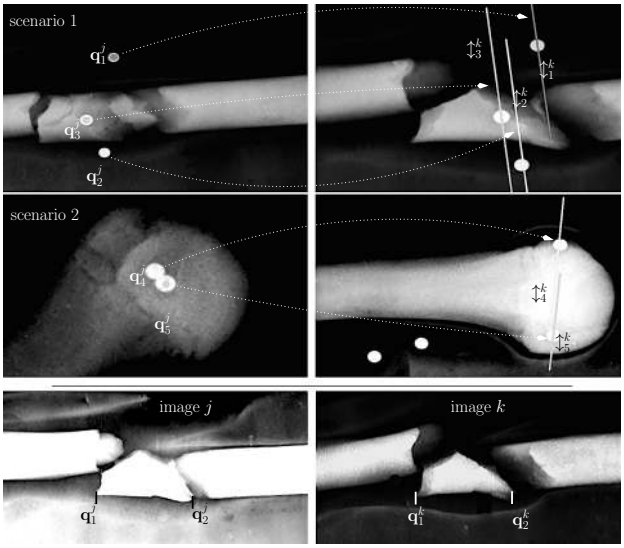


Fig. 16. (Top) Some examples of bounded epipolar lines between stereo X-ray snapshots. Points 4 and 5 lie very close to each other and are difficult to discern in the left image. However, they can be clearly differentiated in the stereo pair on the right. (Bottom) Two radiographs (produced in a moving camera setting) of our anthropomorphic phantom. Two points have been manually selected on each image at both ends of the splinter. With this information, we can estimate $|\mathsf{Q}_2 - \mathsf{Q}_1|$ with the mathematical tools described in Section V-B.

### C. 3D reconstruction of brachytherapy (cervix) applicators

We have also applied the technique described in Section V-B to the reconstruction of a tandem applicator for cervix tumor treatment used in the FDA-cleared planning system presented in [41]. This specific brachytherapy system makes use of two radiographs (AP and LAT) produced at two respective imaging system locations (Fig. 17). The patient's pelvis is imaged in these two configurations, together with a surrounding and permanent frame (equipped with conventional X-ray-opaque fiducials). With both stereo radiographs and the mathematical

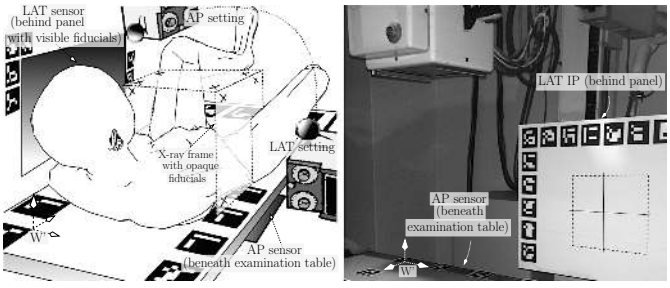


Fig. 17. Proposed and tested brachytherapy scenario using conventional X-ray imaging and visible cameras. (Left) Patient lies comfortably in an AP supine position over a conventional X-ray examination table in a typical examination room (appropriately conditioned for this type of radiotherapy) during treatment. We show the original X-ray-opaque fiducials supporting frame (eventually removable with our enhancements). The patient is radiographed from a frontal and lateral view using two imaging sensors. (Right) Real treatment setup where the sensor for the AP snapshot is under the examination table and the LAT one is behind the Aruco fiducials frame.

tools described in Section V, the system allows oncologists to graphically link the 2D coordinates of the radioactive seeds between both X-ray instances thanks to the constraints imposed by epipolar lines. It can also resolve the correct 3D location, orientation and size of the tandem applicator inside the patient during each radiotherapy session. Initially the position of the first radiation source is calculated and from there, the location of the remaining seeds is derived by construction. These coordinates, which are definitely extremely critical, define the spatial aspect of a radiotherapy treatment of this type.

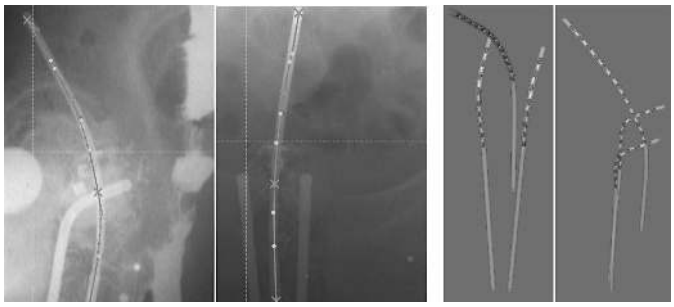

Fig. 18. (Left) LAT and AP radiographs from the patient's treated area where bounded epipolars have contributed to link applicator points in both images. (Right) Two images of the 3D reconstructed brachytherapy tandem applicator. Only visual information has been taken into account for calculations.

We have updated this FDA-cleared brachytherapy scenario with our visual marker detection procedure, eliminating the need for the aforementioned frame around the patient. In order to verify our enhancements, two simple visual marker sets at each orientation have been added to the original system.

Fig. 18-left shows two images produced at the two different X-ray imaging system locations. Original X-ray fiducials (and required supporting frame) remained present during the experiment (and visible in the X-ray images) but were left out of calculations and only used for verification of our methodology. This upgraded brachytherapy system works in the moving camera framework tackled in Section III-D2, where spatial points are expressed relative to a single reference frame $\mathrm{W}'$ (the relative transformation between lateral and frontal Aruco fiducial marker sets is known by design but it can be easily derived since they are automatically detected).

We manage to reconstruct the tandem applicator (Fig. 18-right) and the 3D locations of each radioactive source. We also achieve the same functionality and precision related to seed identification between stereo radiographs with the help of the bounded epipolars introduced in Section V-A. The deviations from the results obtained with the original system (based on X-ray fiducials) are negligible. The tested modifications could allow the complete removal of such surrounding reference frames over patients, resulting in less invasive radiotherapy scenarios and clearer X-ray images, widening the options available for choosing a treatment site.

## VII. Conclusions

We have introduced an innovative methodology for determining the geometrical setting in standard X-ray imaging systems. This information is essential for 3D reconstruction using several X-ray images, which could be a relevant tool for diagnosis and object inspection. Contrary to the usual

approach, ours is not based on the projection of external opaque fiducials on each radiograph. Instead, we account for all the necessary spatial information and modifications thanks to an external camera and visible markers present in the scene. Our technique enables the removal of foreign reference marks and interposed frames, contributing to reducing the complexity and invasiveness of the X-ray diagnostic process. The same technique also allows the combination of X-ray images produced at acuter angles/protocols, where conventional fiducials would be projected outside the radiographic plane. Two application scenarios have been discussed, each one involving different geometrical, optical, and mathematical challenges and application suitability. The accuracy attained in both scenarios is equivalent and the choice of one or the other will depend on the final application. The *moving patient* approach may contribute to enhance X-ray based object scanning frameworks by for instance, providing the exact 3D location and size of items masked in boxes, packages, etc.

The experiments on two phantoms (anthropomorphic and boxlike) show that it is possible to estimate epipolar lines in a second image from given points in a first one with a very good level of precision. This simple but effective subtlety can help radiologists correlate points/areas of interest between different radiographs. The determination of real anatomic lengths is also achieved with great fidelity and a real successful example involving the reconstruction of a tandem applicator for cervix cancer has been shown. Finally, 3D reconstruction using ordinary X-ray images represents an elegant alternative to invasive techniques like CT and similar radiological equipment.

## 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) as well as with the support of the Spanish Ministry of Industry, Energy and Tourism (grant #TSI-100101-2013-019).

## REFERENCES

[1] N. Navab and M. Mitschke, "US patent 6236704: Method and apparatus using a virtual detector for three-dimensional reconstruction from X-ray images," 2001.

[2] N. Navab, A. Bani-Hashemi, and M. Mitschke, "Merging visible and invisible: two camera-augmented mobile C-arm (CAMC) applications," in *2nd IEEE/ACM International Workshop on Augmented Reality*, 1999.

[3] M. Mitschke and N. Navab, "Recovering the X-ray projection geometry for three-dimensional tomographic reconstruction with additional sensors: Attached camera versus external navigation system," *Med. Image Analysis*, 2003.

[4] M. J. Daly, J. H. Siewerdsen, Y. B. Cho, D. A. Jaffray, and J. C. Irish, "Geometric calibration of a mobile C-arm for intraoperative cone-beam ct." *Med Phys*, 2008.

[5] D. G. Grant, "Tomosynthesis: a three-dimensional radiographic imaging technique," *IEEE Transactions on Biomedical Engineering*, 1972.

[6] P. Taddei and V. Sequeira, "X-ray and 3D data fusion for 3D reconstruction of closed receptacle contents," in *Conf. on 3D Vision*, 2013.

[7] T. S. Cook, G. Couch, T. J. Couch, W. Kim, and W. Boonn, "Using the Microsoft Kinect for patient size estimation and radiation dose normalization." *J Digit Imaging*, 2013.

[8] D. C. Moura, J. G. Barbosa, M. R. Tavares, João, and A. M. Reis, "Calibration of bi-planar radiography with a rangefinder and a small calibration object," 2008.

[9] T. J. Miller and E. C. Quintana, "Stereo X-Ray System Calibration for 3D Measurements," in *Advancement of Optical Methods in Experimental Mechanics*, H. Jin, C. Sciammarella, S. Yoshida, and L. Lamberti, Eds. Springer, 2014.

[10] S. Schumann, B. Thelen, S. Ballestra, L.-P. Nolte, P. Büchler, and G. Zheng, "X-ray image calibration and its application to clinical orthopedics," *Medical Engineering & Physics*, 2014.

[11] N. Navab, S.-M. Heining, and J. Traub, "Camera augmented mobile C-arm (CAMC): calibration, accuracy study, and clinical applications," *IEEE Transactions on Medical Imaging*, 2010.

[12] L. Wang, R. Zou, S. Weidert, J. Landes, E. Euler, D. Burschka, and N. Navab, "Modeling kinematics of mobile C-arm and operating table as an integrated six degrees of freedom imaging system," in *Computer Aided Medical Procedures*, T. U. München, Ed., 2010.

[13] X. Chen, H. Naik, L. Wang, N. Navab, and P. Fallavollita, "Video guided calibration of an augmented reality mobile C-arm," in *IJCARS*, 2014.

[14] S. Nicolau, P. Y. Lee, H. S. Wu, and M. H. Huang, "Fusion of c-arm x-ray image on video view to reduce radiation exposure and improve orthopedic surgery planning : first in-vivo evaluation," in *CARS*, 2011.

[15] L. Wang, J. Traub, S. Weidert, S. M. Heining, E. Euler, and N. Navab, "Parallax-free intra-operative X-ray image stitching," *Med. Image Analysis*, 2010.

[16] R. Liao, L. Zhang, Y. Sun, S. Miao, and C. Chefd'hotel, "A review of recent advances in registration techniques applied to minimally invasive therapy," *IEEE Transactions on Multimedia*, 2013.

[17] P. Markelj, D. Tomaževič, B. Likar, and F. Pernuš, "A review of 3D/2D registration methods for image-guided interventions," *Med. Image Analysis*, 2012.

[18] A. Rougee, C. L. Picard, Y. L. Trousset, and C. Ponchut, "Geometrical calibration for 3d x-ray imaging," in *Medical Imaging*, 1993.

[19] B. E. Claus, "Geometry calibration phantom design for 3D imaging," in *Medical Imaging*, 2006.

[20] G. Penney, J. Weese, J. Little, P. Desmedt, D. L. G. Hill, and D. Hawkes, "A comparison of similarity measures for use in 2-d-3-d medical image registration," *IEEE Transactions on Medical Imaging*, 1998.

[21] S. Aouadi and L. Sarry, "Accurate and precise 2d-3d registration based on X-ray intensity," *Computer Vision and Image Understanding*, 2008.

[22] J. de Siebenthal and F. Langlotz, "Use of a new tracking system based on artoolkit for a surgical simulator: accuracy test and overall evaluation," in *Augmented Reality Toolkit, 1st IEEE International Workshop*, 2002.

[23] J. Fischer, M. Neff, D. Freudenstein, and D. Bartz, "Medical augmented reality based on commercial image guided surgery," in *10th Eurographics Conf. on Virtual Environments*, 2004.

[24] T. Hachaj and M. Ogiela, "Augmented reality interface for visualization of volumetric medical data," in *Image Processing and Communications Challenges 2*, R. Choraś, Ed. Springer Berlin Heidelberg, 2010.

[25] J. B. A. Maintz and M. A. Viergever, "A survey of medical image registration," *Med. Image Analysis*, 1998.

[26] B. P. Selby, G. Sakas, W.-D. Groch, and U. Stilla, "Patient positioning with X-ray detector self-calibration for image guided therapy." *Australas Phys Eng Sci Med*, 2011.

[27] P. Markelj and B. Likar, *Registration of Three-dimensional and Two-dimensional Medical Images: PhD Thesis*. P. Markelj, 2010.

[28] A. Rougé, C. Picard, C. Ponchut, and Y. Trousset, "Geometrical calibration of X-ray imaging chains for 3D reconstruction," *Computerized Medical Imaging and Graphics*, 1993.

[29] G. Medioni and S. B. Kang, *Emerging Topics in Computer Vision*. Prentice Hall PTR, 2004.

[30] S. Bushong, *Radiologic Science for Technologists: Physics, Biology, and Protection*. Elsevier - Health Sciences Division, 2012.

[31] A. Y. I. Aziz and H. M. Karara, "Direct linear transformation into object space coordinates in close-range photogrammetry," in *Proc. of the Symposium on Close-Range Photogrammetry*, 1971.

[32] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004.

[33] *X-ray Patient Positioning Manual*. Aunt Minnie, 2008.

[34] M. Kaltenbrunner and R. Bencina, "reacTIVision: a computer-vision framework for table-based tangible interaction," in *Int. Conf. on Tangible and Embedded Interaction*, 2007.

[35] L. Naimark and E. Foxlin, "Fiducial detection system," 2007, US Patent 7,231,063.

[36] B. Fröhlich, R. Blach, and R. van Liere, "A lightweight id-based extension for marker tracking systems," 2007.

[37] M. Fiala, "Artag, a fiducial marker system using digital techniques," in *IEEE Computer Society Conf. on Computer Vision*, 2005.

[38] A. C. Rice, A. R. Beresford, and R. K. Harle, "Cantag: an open source software toolkit for designing and deploying marker-based vision systems," in *Int. Conf. on Pervasive Computing*, 2006.
[39] 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*, 2014.
[40] R. Hartley, "Projective reconstruction and invariants from multiple images," *Pattern Analysis and Machine Intelligence*, 1994.
[41] U.S. Food & Drug Administration, "Premarket notification on SIMUPLAN Treatment Planning System (KO30821)," 2003.