3D measurements in conventional X-ray imaging with RGB-D sensors

Francisco Albiol[a], Alberto Corbi[a,*], Alberto Albiol[b]

[a] *Instituto de Física Corpuscular, Universitat de València, Consejo Superior de Investigaciones Científicas (Spain)*
[b] *Universidad Politècnica de València (Spain)*

Abstract

A method for deriving 3D internal information in conventional X-ray settings is presented. It is based on the combination of a pair of radiographs from a patient and it avoids the use of X-ray-opaque fiducials and external reference structures. To achieve this goal, we augment an ordinary X-ray device with a consumer RGB-D camera. The patient's rotation around the craniocaudal axis is tracked relative to this camera thanks to the depth information provided and the application of a modern surface-mapping algorithm. The measured spatial information is then translated to the reference frame of the X-ray imaging system. By using the intrinsic parameters of the diagnostic equipment, epipolar geometry, and X-ray images of the patient at different angles, 3D internal positions can be obtained. Both the RGB-D and X-ray instruments are first geometrically calibrated to find their joint spatial transformation. The proposed method is applied to three rotating phantoms. The first two consist of an anthropomorphic head and a torso, which are filled with spherical lead bearings at precise locations. The third one is made of simple foam and has metal needles of several known lengths embedded in it. The results show that it is possible to resolve anatomical positions and lengths with a millimetric level of precision. With the proposed approach, internal 3D reconstructed coordinates and distances can be provided to the physician. It also contributes to reducing the invasiveness of ordinary X-ray environments and can replace other types of clinical explorations that are mainly aimed at measuring or geometrically relating elements that are present inside the patient's body.

Keywords: X-ray, depth cameras, epipolar geometry, 3D reconstruction, movement tracking, dense surface mapping

1. Introduction

Many diagnostic protocols require several radiographs of the patient at different orientations. However, conventional and general purpose X-ray equipment usually provides geometrically uncalibrated images because the X-ray source is detached from its detector and the intrinsic and extrinsic parameters are unknown *a priori* and are mutually tied. Thus, distances in radiographs are usually very poorly estimated, and, in many cases, simple X-ray-opaque objects of everyday life (such as coins) are used as reference landmarks.

The common procedure for obtaining 3D information (locations, distances, angles, etc.) from two images j and k (also known as *projection-to-volume registration* or P2VR) requires two camera projection matrices P^j and P^k. In conventional radiography, projection matrices can be obtained by using calibration frames with N radio-opaque fiducials placed at known locations $\mathbf{Q}_i$ (with $i = 1 \cdots N$) that are projected onto each radiograph at 2D coordinates $\mathbf{q}_i^j$ and $\mathbf{q}_i^k$ for radiographs j and k, respectively. The 3D locations of these fiducials are usually expressed relative to a common *world* coordinate frame (W).

X-ray-opaque fiducials are used daily in radiology, and interesting research works on this subject can be found in the literature. For instance, Schumann et al. [1] have built and tested a special calibration object for clinical orthopedics. Structures of this type can be further customized thanks to modern 3D-printers. The problem with supporting frames is that they may be perceived as being invasive by the patient. Besides, although not a critical issue, the embedded X-ray fiducials may be projected outside the detector plate at oblique protocols or they may interfere with the image quality by creating external artifacts and adding extra Compton contribution to the radiograph produced.

P2VR in radiology has been a subject of interest for a long time. It is worth citing one of the pioneering research works carried out by Caponetti et al. [2], where a *shape-from-contour* algorithm and the back-lighting from two perpendicular projections is used, under the assumption of a parallel beam, to approximate the 3D surface of a bone.

For our contribution to P2VR, we present an alternative method with a special focus on its deployment in ordinary and primary diagnostic X-ray settings and the elimination of the need for calibration frames. Our approach makes use of a rigidly attached RGB-D sensor or *depth camera* that can resolve the rigid transformation that represents the patient's movement between two instants t_j and t_k and relative to a fixed X-ray system. The specific movement studied in this work is the rotation around

*Corresponding author. Email: alberto.corbi@ific.uv.es

the patient's own craniocaudal axis, from which the corresponding P^j and P^k matrices can be derived, thereby enabling the accurate distillation of 3D locations and lengths. Here, it is worth mentioning that modern depth cameras, though often seen as mere consumer products, provide high performance and good spatial resolution as Khoshelham [3, 4] demonstrates (a mean value of 1 mm in the depth direction for calibrated devices). Accuracy can be further improved with the methods described in Section 2.3.

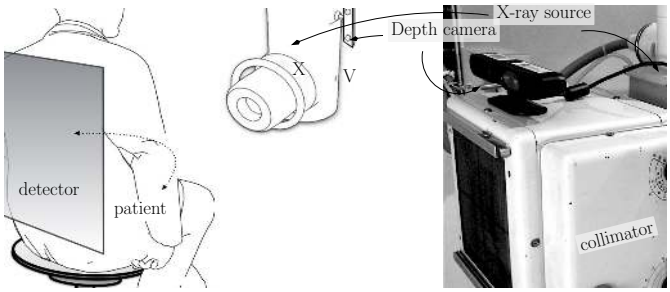

Figure 1: Depiction of the presented setup. The patient rotates in supine anteroposterior (AP) or posteroanterior (PA) position while motion is continuously tracked with the RGB-D sensor (V). This device is rigidly attached to the housing of the anode (X). The patient is radiographed at two angles relative to the vertical axis.

The augmentation of general purpose radiological modalities with other interplaying sensors has been proposed for other applications. The research carried out by Aoki et al. [5] highlights how modern consumer depth cameras have huge potential in many clinical fields, such as radiotherapy and radiography. Badal et al. [6] have engineered a dose-monitoring system that is based on the tracking of the location of patients and staff during interventional fluoroscopy sessions using depth sensors. The authors in Cook at al. [7] estimate the patient's size with the help of a Microsoft Kinect device in order to normalize the dose received, whereas Kozono K. [8] uses a similar approach to monitor the location of the patient and better assess the X-ray entrance dose. In the analysis performed by Tahavori et al. [9], a Kinect device is again used to capture the surface of the patient with the goal of detecting possible misalignments during beam radiotherapy. Three-dimensional imaging of the breast is studied in Wheat et al. [10] using a Kinect-based system, and the researchers in Bauer et al. [11] perform a similar task related to patient alignment in computerized tomographies (CT). Respiratory variations during positron emission tomographies (PET) are also determined in Noonan et al. [12].

This research shares some of the goals of the work carried out by Albiol et al. [13], who perform P2VR using videocamera-augmented X-ray equipment and visual markers. The key difference is that the methodology described in [13] is more appropriate for X-ray protocols in which the patient remains still and the imaging setup is the one that moves from location j to k.

However, despite all of these research examples on the subject of *external sensors and depth cameras in medicine and radiology*, it is very difficult to find citations about the use of RGB-D data for the determination of the patient's location/orientation relative to a conventional X-ray imaging setup (such as the one shown in Fig. 1). As the following sections demonstrate, this information can be non-invasively measured with enough precision, making a significant impact on the elaboration of the diagnosis.

2. Methods

In this section, we explain the proposed techniques based on the combination of an X-ray imaging system and a depth sensor.

2.1. System overview

Our system consists of a conventional X-ray setup and a depth sensor that is rigidly attached to the housing of the X-ray anode. During examinations, the X-ray imaging system (source/detector) and the RGB-D device both remain fixed. A RGB-D camera is comprised of depth and video components and both data streams are registered by default. The RGB-D device used measures distances by analyzing the speckle pattern of a projected infrared light.

Initially, the patient is asked to stand erect in front of the depth sensor and is rigidly rotated with the help of a swivel stool or a spinning platform. This motion is continuously tracked by the depth sensor and two radiographs are taken at different orientations j and k (instants t_j, t_k).

By using the patient's displacement relative to the depth sensor in combination with epipolar geometry, P^j and P^k can be updated. However, before proceeding, the dual imaging system must first be geometrically calibrated.

2.2. System calibration

The goal of the calibration phase is to obtain both the intrinsic parameters of the X-ray setting and the geometrical relation between the two imaging systems which remain invariant during the examination. The calibration frame shown in Fig. 3 was designed to do this. It accommodates 13 copper, cross-shaped markers $\mathbf{Q}_i^{\text{cal}}$ (with $i = 1 \cdots 13$) that are opaque to the Roentgen radiation. It also contains a matching number of visible markers that can be easily detected using the incorporated video camera of the depth sensor. Both fiducial types are made coincident to ease calculations. Also, the frame defines the origin and orientation of the W coordinate frame. The 3D coordinates of all of the calibration markers are known by construction and referenced relative to W.

In this article, rigid transformations are expressed with the nomenclature ${}^{\text{dest}}\mathrm{T}_{\text{orig}}$, that is, how points in the *origin* reference frame are translated to the *destination* coordinate system. Three other coordinate systems are defined, which are shown in Fig. 2. The first one is the

X coordinate system, whose origin is at the X-ray anode and has one axis that is orthogonal to the detector plate. The second one is the camera coordinate system (V), whose origin is the optical center of the built-in video camera. V also defines the origin of the depth data given that both streams (video and depth) are registered. Finally, an object-dependent coordinate frame W′ is used for testing purposes, as discussed in Section 2.5. All of the coordinate systems (X, V, W, and W′) can be related using rigid transformations, namely, ${}^{\mathrm{V}}\mathrm{T}_{\mathrm{W}}$, ${}^{\mathrm{X}}\mathrm{T}_{\mathrm{W}}$, ${}^{\mathrm{X}}\mathrm{T}_{\mathrm{V}}$, and ${}^{\mathrm{W}}\mathrm{T}_{\mathrm{W}'}$.

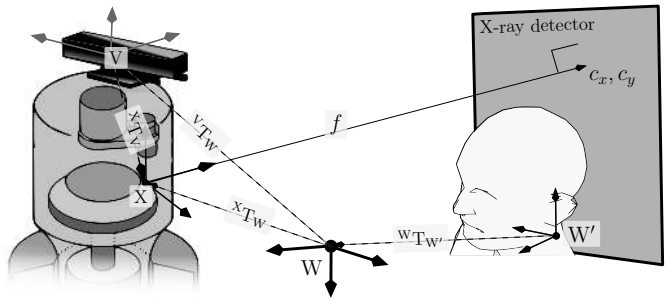

Figure 2: Coordinate frames and geometrical relations used in the proposed imaging system. W is a common coordinate frame whose origin is a known position in the room. W′ has its origin inside the object under study and is linked to it. V and X are the proper coordinate frames of the anode and depth camera, respectively.

The calibration process can be summarized as follows:

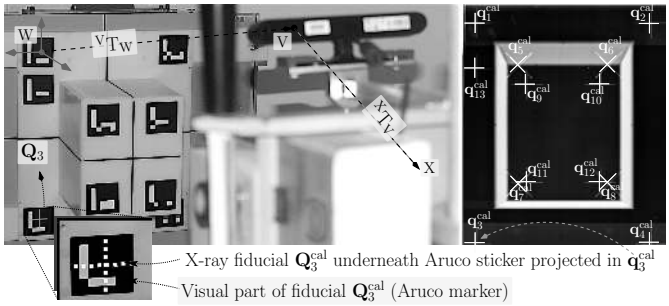

Figure 3: Left: information gathered during calibration: a photograph and a radiograph of the frame. Both images enable the derivation of the transformation that connects both imaging systems (${}^{\mathrm{X}}\mathrm{T}_{\mathrm{V}}$). Visible fiducials help find the relation between the world and the depth camera (${}^{\mathrm{V}}\mathrm{T}_{\mathrm{W}}$). Right: projections of the cross-shaped X-ray markers (hidden beneath the visual fiducials) in the radiograph.

1. The calibration frame is introduced in the scene and a photograph (with the RGB-D camera) and a radiograph are generated from it (Fig. 3-right). The structure is then removed and is no longer necessary during the examination of the patient.
2. An initial X-ray projection matrix $\mathrm{P}^{\mathrm{cal}}$ is built using the Direct Linear Transform (DLT) algorithm [14], the 3D location $\mathbf{Q}_i^{\mathrm{cal}}$ of each radio-opaque marker (relative to W), and the coordinates (manually outlined) of the corresponding 2D projection $\mathbf{q}_i^{\mathrm{cal}}$ in the radiograph obtained in the previous step.
3. The intrinsic (K) and extrinsic (${}^{\mathrm{X}}\mathrm{T}_{\mathrm{W}}$) parameters of the X-ray system are extracted from $\mathrm{P}^{\mathrm{cal}}$ using the RQ decomposition: $\mathrm{P}^{\mathrm{cal}} = \mathrm{K} \cdot {}^{\mathrm{X}}\mathrm{T}_{\mathrm{W}}$. In ordinary cameras, intrinsic parameters are fixed. However, in X-ray imaging, detector and source are decoupled, which entails that K will be altered if either the imaging plate or anode are shifted. Conventionally, the elements of K contain the principal point c_x, c_y and the focal length f, which perpendicularly connects the anode and the detector. The extrinsic part relates 3D coordinates between W and X frames.
4. The position and orientation of the video camera (and depth sensor) relative to the world coordinate system (${}^{\mathrm{V}}\mathrm{T}_{\mathrm{W}}$) are obtained by using the photograph of the calibration frame (and detected visual markers) taken in the first step. In this work, we use an automatic process for visual marker identification based on the Aruco library, which is described in Garrido et al. [15]. This framework was chosen over other alternatives because of its robustness against vertex jitter.
5. Finally, the relation between the X-ray and RGB-D coordinate system can be obtained as:

$$ {}^{\mathrm{X}}\mathrm{T}_{\mathrm{V}} = {}^{\mathrm{X}}\mathrm{T}_{\mathrm{W}} \cdot \left({}^{\mathrm{V}}\mathrm{T}_{\mathrm{W}}\right)^{-1} \quad (1) $$

This transformation remains constant as long as the relative position between devices remains fixed. With Eq. (1), the derived spatial information can be translated to the reference frame of the X-ray imaging system. The part $\left({}^{\mathrm{V}}\mathrm{T}_{\mathrm{W}}\right)^{-1}$ is the *pose* of the depth camera (i.e., its position/orientation relative to W).

2.3. Estimation of motion and projection matrices

Once the system is calibrated, the patient enters the scene and comfortably sits or stands on a rotating platform. At this initial instant t_j, the first radiograph is captured and then the platform starts spinning gently while depth images are continuously generated. During this process, the patient should remain still to ensure that a motion be as close to rigid as possible. The captured depth data (also known as *point cloud*) is continuously analyzed until the patient reaches a second orientation at t_k. At this instant, the second radiograph is taken and the process ends.

In this work, we use the KinectFusion technology developed by Newcombe et al. [16] to estimate the rigid motion of a patient ${}^{\mathrm{V}^k}\mathrm{T}_{\mathrm{V}^j}$ between consecutive X-ray snapshots obtained at t_j and t_k. This algorithm was originally designed to reconstruct 3D scenes robustly by moving the Microsoft Kinect sensor around an object or person and performing a Dense Surface Mapping (DSM), which is described by Tong et al. [17]. KinectFusion also adds extra

accuracy to the 3D derived geometry. For instance, Meister et al. [18] report having achieved a 2 mm precision for the Euclidean error of a scanned 40 cm high human-like statue, with the camera/object distance being ∼1 m. This error decreases if the tracked object is bigger because more points contribute to the estimation of the motion. Apart from the mentioned surface, KinectFusion also refines the location/orientation of the sensor relative to W (i.e., its pose).

Alternatively, in our devised setup, the patient rotates relative to a fixed Kinect-like device. This is not a problem for the KinectFusion algorithm because it only needs the relative motion between the patient and the depth sensor. This situation can be alternatively understood as a *moving virtual camera system* around the fixed W reference but with the key advantage of preserving the intrinsic parameters (K) of the X-ray equipment and the transformation (${}^{\mathrm{X}}\mathrm{T}_{\mathrm{V}}$) between the X-ray equipment and the depth camera, which are obtained during calibration. The only significant difference with respect to the default KinectFusion application scenario is that, in our *virtual camera* approach, the background information has to be subtracted from the point dataset frame by frame. To achieve this, any of the methods already applied by many Kinect-based applications, such as the gesture recognition studied by Biswas and Basu [19], can be used. For example, a depth snapshot of the room can be obtained in advance and kept for background removal. A length/depth threshold, beyond which point clouds are no longer considered, can be established as well.

The X-ray projection matrices P^j and P^k for instants t_j and t_k can be calculated as:

$$\begin{aligned} \mathrm{P}^j &= \mathrm{K} \cdot \left\{ {}^{\mathrm{X}^j}\mathrm{T}_{\mathrm{V}^j} \cdot {}^{\mathrm{V}^j}\mathrm{T}_{\mathrm{W}} \right\} \\ \mathrm{P}^k &= \mathrm{K} \cdot \left\{ {}^{\mathrm{X}^k}\mathrm{T}_{\mathrm{V}^k} \cdot {}^{\mathrm{V}^k}\mathrm{T}_{\mathrm{V}^j} \cdot {}^{\mathrm{V}^j}\mathrm{T}_{\mathrm{W}} \right\} \end{aligned} \qquad (2)$$

where ${}^{\mathrm{X}^j}\mathrm{T}_{\mathrm{V}^j}$ and ${}^{\mathrm{X}^j}\mathrm{T}_{\mathrm{V}^j}$ are the result of Eq. (1) and are constant and equal to ${}^{\mathrm{X}}\mathrm{T}_{\mathrm{V}}$. The parts inside the brackets in Eq. (2) correspond to the composition of the rigid transformations involved and represent the extrinsic parameters of each projection matrix. The patient's motion and the corresponding angular span around the vertical axis (included in the term ${}^{\mathrm{V}^k}\mathrm{T}_{\mathrm{V}^j}$) between t_j and t_k should be chosen following medical criteria. It is also recommended that each radiograph have a diagnostic meaning of its own. An important practical question is how the patient's breathing might affect the estimation of the rigid motion. In a real scenario, these motions are unnoticeable under a hospital gown. However, the experiment described in Section 2.5 shows that their influence is negligible even for a naked body. It is, however, convenient for the X-ray images to be taken with a similar state of the lungs in order to minimize organ displacements and provide more precise 3D information. Additionally, X-ray images and motion information should be in sync, but this is normally the default case given that radiographs are produced by DICOM compliant hardware and depth cameras also append time information (i.e., t_j and t_k). Altogether, the process summarized here should not take more than a few seconds in a fully-digital X-ray setting or about a minute in a CR-equipped setup, representing minor annoyances for the patient.

Finally, the evaluation of the motion performed by KinectFusion is more appropriate for relatively large body parts, such as the torso, but it can also work with smaller ones such as the head.

2.4. 3D point reconstruction

Given the two matrices P^j and P^k from Eq. (2) and two observed landmarks $\mathbf{q}_a^j$ and $\mathbf{q}_a^k$ in images j and k, respectively, the 3D location Q_a (derived) of the imaged point $\mathbf{Q}_a$ (ground-truth) can be determined. In our implementation, the radiologist should be the one that manually locates $\mathbf{q}_a^j$ and $\mathbf{q}_a^k$, assisted by the automatic drawing of epipolar lines (like l_a^k and l_b^k shown in Fig. 5). It is therefore important for these traces to be visible in both radiographs. As Hartley and Zisserman [14] recall, projective geometry establishes that 3D points and the corresponding projections are related using the following cross products:

$$\hat{\mathbf{q}}_a^j \times \mathrm{P}^j \cdot \hat{\mathsf{Q}}_a = \mathbf{0} \qquad \hat{\mathbf{q}}_a^k \times \mathrm{P}^k \cdot \hat{\mathsf{Q}}_a = \mathbf{0} \qquad (3)$$

where all points are expressed in homogeneous coordinates. Each of these relations determines two linearly independent equations that can be solved by using a Singular Value Decomposition (SVD). If this is repeated for a second point $\mathbf{Q}_b$ projected on the same pair of X-ray images j and k, it is also possible to derive the 3D length of the segment $\overleftrightarrow{\mathsf{Q}}_{a,b}$ defined by the two back-projected ends Q_a and Q_b (i.e., $\left|\overleftrightarrow{\mathsf{Q}}_{a,b}\right|$).

2.5. Experimental setup

To test the proposed methodology, we used three phantoms. The first two consisted of two anthropomorphic phantoms (head and torso, studied separately), which are shown in Fig. 4. A set of spherical lead bearings (1 mm size) was manually placed in the phantoms at well-known locations relative to a local coordinate system W′ (shown in Fig. 4) bound to each phantom. The reason for choosing W′ is that we can easily know the location of the lead bearings relative to W′ (by construction), but we do not know the exact position of each phantom relative to W.

Both phantoms were placed ∼1.40 m away from the anode, and their rotation was tracked using the techniques described in Section 2.3. Radiographs were produced at arbitrary angles relative to the X-ray imaging system, and their projection matrices were calculated with Eq. (2). Fig. 5 shows two of these X-ray snapshots and some steps of the dense surface mapping process. X-ray images are 35×43 cm with a resolution of 10^4 px/m.

We also conducted some experiments to study the accuracy of the estimation of the motion in the clinical context described in this paper. In one test, we checked whether

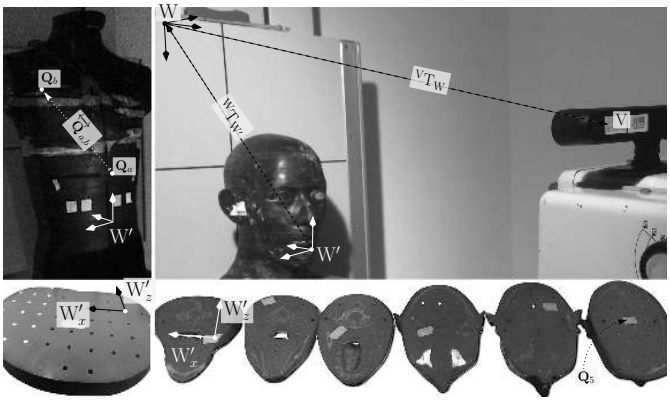

Figure 4: Tested anthropomorphic phantoms (torso and head). Several slices are shown and the local reference frame W′ is defined in two of them (one for each phantom). Spherical lead bearings are placed in each slice. As an example, two of them ($\mathbf{Q}_a$ and $\mathbf{Q}_b$) are represented, with $\overleftrightarrow{\mathbf{Q}}_{a,b}$ being the segment that connects them.

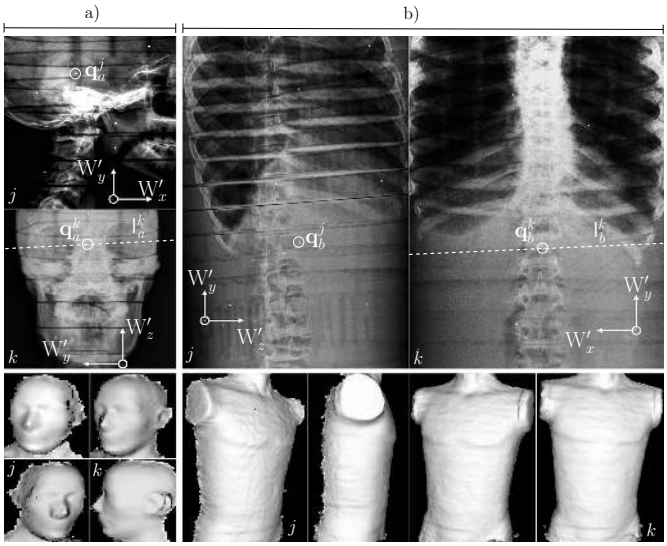

Figure 5: a) and b) Pairs of X-ray images generated from the anthropomorphic phantoms (head and torso) at two positions j and k with an angular amplitude of 85° (around the craniocaudal axis) between them. The projection of the chosen local reference frame W′ is also represented. Notice the white dots corresponding to the projections of the lead bearings. Two of these (belonging to the head ($\mathbf{q}_a$) and torso ($\mathbf{q}_b$)) are highlighted in each stereo pair of images. The dashed lines tagged with l_a^k and l_b^k are the corresponding epipolar lines in image k associated to their stereo counterparts (projections $\mathbf{q}_a^j$ and $\mathbf{q}_b^j$, respectively). Some stages of the DSM algorithm representing the continuous tracking of the patient are shown below the X-ray images.

normal breathing would interfere with the motion results derived by KinectFusion (Fig. 7-a). A patient was asked to breathe normally for 20 seconds while remaining still at some angles relative to the vertical axis. The same procedure was repeated but this time with the patient holding his breath for the same amount of time. The measured angle in both scenarios was then compared over time. In the second experiment (Fig. 7-b), we studied the accuracy of some estimated angular spans when compared against ground-truth for a relatively small tracked object (the ∼25 cm tall polystyrene head-shaped phantom presented below). In this case, the *principal rotation* (ϕ), which is discussed in Schaub et al. [20], was obtained 10 times for each ${}^{V^k}\mathrm{T}_{V^j}$ and then compared against the one read in an angle meter located beneath the rotating platform. In both experiments, the distance object-camera was ∼0.8 m.

These anthropomorphic phantoms (together with the radiograph pairs generated from them) will be used to estimate the 3D locations of the lead spherical bearings. These derived locations will then be compared against the ground-truth positions expressed in the W′ coordinate system. It is then mandatory to also find the optimal transformation (i.e., ${}^{W}\mathrm{T}_{W'}$) that best aligns the two sets: Q_i, which is derived in coordinates of W; and $\mathbf{Q}_i$, which is known relative to W′. This entails performing a Euclidean transformation, which preserves shape and size, as tackled by Besl and McKay [21]. In this work, ${}^{W}\mathrm{T}_{W'}$ is found with a Procrustes analysis described by Dryden and Mardia [22] but with scale invariance. We will apply this transformation later in Section 3 when comparing the mean Euclidean distance between the 3D coordinates of both sets.

The third phantom is made of polystyrene (Fig. 7-b). On this occasion, 15 medical needles of known and precise lengths 36.1, 18.4, and 9.2 mm (5 of each type) were introduced at random locations, in both vertical and horizontal orientations. The phantom was placed ∼1.57 m away from the X-ray anode, and its motion was tracked using the same techniques described in Section 2.3. The angular span between t_j and t_k was 70 degrees. In this case, we compared the estimated needle lengths $\left|\overleftrightarrow{\mathsf{Q}}_i\right|$ against the ground-truth. To derive these lengths, we followed the procedure described at the end of Section 2.4. The only role of this second phantom was to house the aforementioned needles at different and varied positions. This phantom could have had any shape; however, we thought the experiment would be more realistic if the phantom resembled a human body part (a head, in this case).

3. Results

Using Eq. (3), we obtained the 3D location Q_i of each lead fiducial from combinations of radiograph pairs of the anthropomorphic phantoms at different orientations as well as the 2D pixel coordinates of each spherule projection in each image. The alignment operation ${}^{W}\mathrm{T}_{W'}$ was applied to express their components relative to the W reference frame. Table 1 shows the mean Euclidean difference ($\triangle$) between all Q_i and $\mathbf{Q}_i$. Each row also specifies the angular amplitude ($\measuredangle$) around the vertical axis for each orientation/radiograph combination. The experiment was carried out for the head and torso parts separately in order to test the tracking algorithm with a relatively small body part (head) and a relatively large body part (chest).

	Head phantom		Torso phantom	
$\measuredangle$ (°)	$\triangle$ (mm)	σ (mm)	$\triangle$ (mm)	σ (mm)
25	1.4	0.7	0.8	0.5
45	0.9	0.5	0.9	0.5
60	1.3	0.6	1.0	0.6
85	1.6	0.9	1.2	0.7

Table 1: Mean Euclidean span ($\triangle$) between ground-truth positions of spherical lead bearings inside the anthropomorphic phantom (head and torso) and the derived ones for some angular extents ($\measuredangle$).

We also compared the differences in the distances between some spherical bearings in the anthropomorphic phantom (head and torso). Given that we know where each spherule is located (in the W′ system), we can also determine the ground-truth Euclidean distance $\left|\overleftrightarrow{\mathbf{Q}}_{a,b}\right|$ between any $\mathbf{Q}_a$, $\mathbf{Q}_b$ pair. This ground-truth distance is then compared against the estimated one $\left|\overleftrightarrow{\mathbf{Q}}_{a,b}\right|$ using the previously derived locations $\mathbf{Q}_i$ of each spherical lead bearing. Fig. 6 shows these differences graphically for three angles and for several known distances (ground-truth). These distances range from the closest (24.6 mm) to the widest gap (355.1 mm) between two given spherical lead bearings. Again, the experiment was executed separately for the head and torso anthropomorphic phantoms; however, for the sake of simplicity, the results are shown for both phantoms in Fig. 6. The same plot also distinguishes between which distances were measured in the head and torso phantoms.

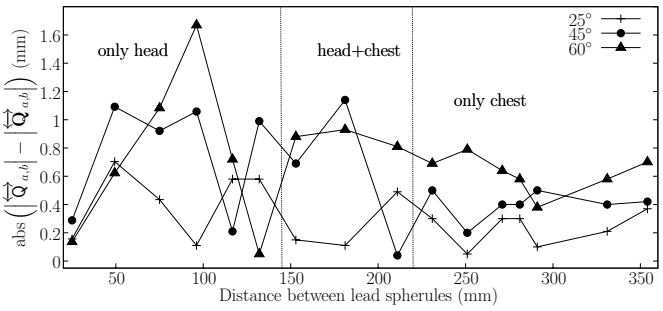

Figure 6: Difference (in absolute value) between computed distances between pairs of spherical lead bearings and ground-truth. The results for three angular amplitudes (25°, 45°, and 60°) are shown.

In the case of the polystyrene phantom, we obtain the mean needle lengths $\left|\overleftrightarrow{\mathbf{Q}}_i\right|$ shown in Table 2, and we compare them against the ground-truth $\left|\overleftrightarrow{\mathbf{Q}}_i\right|$.

$\left\|\overleftrightarrow{\mathbf{Q}}_i\right\|$	$\left\|\overleftrightarrow{\mathbf{Q}}_i\right\|$	$\left\|\overleftrightarrow{\mathbf{Q}}_i\right\|$	$\left\|\overleftrightarrow{\mathbf{Q}}_i\right\|$	$\left\|\overleftrightarrow{\mathbf{Q}}_i\right\|$	$\left\|\overleftrightarrow{\mathbf{Q}}_i\right\|$
36.1	36.1±0.3	18.4	18.5±0.1	9.2	9.1±0.2

Table 2: Derived lengths $\left|\overleftrightarrow{\mathbf{Q}}_i\right|$ and ground-truth $\left|\overleftrightarrow{\mathbf{Q}}_i\right|$ (in mm).

measured	25.0	45.0	60.0	85.0
derived	25.0±0.1	45.1±0.1	60.1±0.2	85.2±0.3

Table 3: Principal rotations (ϕ) measured with an angle meter and then obtained with KinectFusion (in °).

For the evaluation of the accuracy of the KinectFusion algorithm, after conducting the experiments described in Section 2.5, we obtain the following results. On one hand, normal breathing (Fig. 7-a and -c) does not significantly affect the estimated angle around the vertical axis of rotation when compared against the one obtained when the patient held his breath. On the other hand, when rotating the head-like polystyrene phantom (Fig. 7-b), the estimated ϕ and ground-truth (angle meter) are in good agreement, as shown in Table 3.

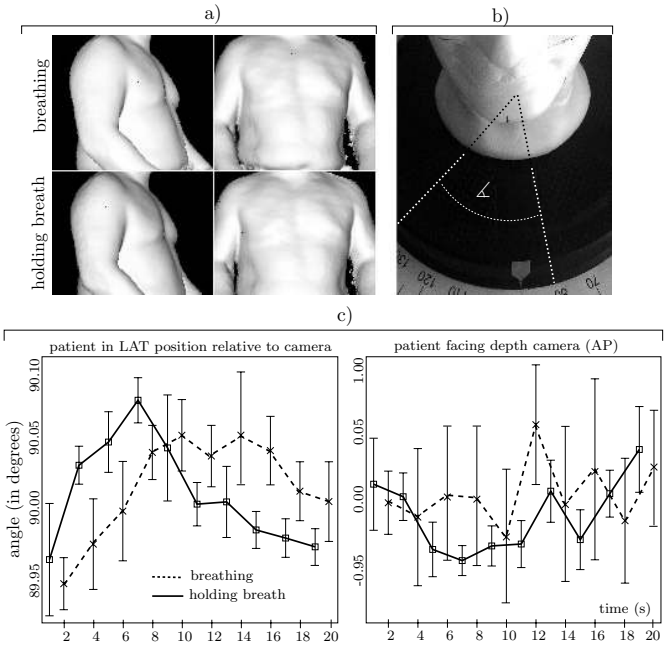

Figure 7: Evaluation of the accuracy of KinectFusion. a) LAT and AP patient surface model (changes produced by breathing are unnoticeable). b) Head phantom used to measure the precision of the derived angular span. c) Evolution of the estimated angle (around vertical axis) for the two aforementioned patient orientations while breathing (dashed lines) and holding breath (solid lines) for ∼20s. Points represent the mean value for each 2s. window (∼44 depth frames), and error bars show the dispersion for that time segment.

4. Discussion

Epipolar geometry allows the 3D reconstruction from two geometrically calibrated images. In the case of stereo X-ray images, this reconstruction can be achieved from observed *paired* projections in them. If a patient has to undergo a conventional X-ray examination (which may consist of the generation of two radiographs), 3D locations and lengths can then be derived at no additional cost.

In this paper, we have summarized the different ways in which this 3D information can be derived from pairs of X-ray images. We have presented our genuine approach based on the extraction of geometrical information with the help of an interplaying depth sensor. The gathered depth data is then analyzed by the KinectFusion algorithm to estimate the patient's rigid movements in real-time.

To test our 3D reconstruction approach, we have run and described several experiences with phantoms that resemble human body parts, such as the head and torso. In these experiments, we have successfully derived 3D data from some embedded lead bearings and metal needles. The results have demonstrated that we can achieve a millimetric level of precision. As Fig. 6 and Table 2 highlight, millimetric precision is achievable when comparing the derived lengths against ground-truth. We also obtain a nice agreement between the derived 3D locations and their ground-truth position, as shown in Table 1.

These results leave the door open for the application of this technique as a complementary tool for other radiological or clinical exams whose main goal is measuring distances or geometrically relating 3D spots whose projections are observable in two radiographs. In a real scenario, these spots can be splinters, cysts, tumors, benign corpuscles, etc., and basically anything worth measuring or locating inside the human body. Our method leaves the identification of these common points to the health professional in both radiographs. In addition, given that KinectFusion also obtains the surface of the patient, it is immediate to relate the 3D location of inner points to this surface. This enables, for instance, the determination of their depth relative to the person's skin. It should be mentioned that other clinical examinations, such as ultrasound, are also capable of outputting distances/locations. For this reason, we suggest the application of our approach only when the generation of X-ray images from the patient is, on its own, the appropriate path to achieve a diagnosis.

We also extract two procedural outcomes: large tracked areas produce better results, as shown in Fig. 6. The second conclusion is that normal breathing does not especially affect the rotation tracking process as shown in Fig. 7-c. Breathing slightly increases the inaccuracy of the measured angle around the vertical axis when it is compared against the *holding breath* scenario. This small drift is caused by tiny involuntary movements that are easily trackable by the depth sensor and KinectFusion. However, we recall that X-ray examinations usually require gowns, which help to mask ribcage variations. Also, as indicated in Section 2.3, radiograph pairs should be generated in the same breathing state (required by the specific X-ray protocol) to maximize the accuracy of the proposed techniques.

Similar research studies like the one carried out by Laporte et al. [23] achieve a little higher accuracy by using a novel *Non Stereo Corresponding Contour* Method (NSCC), whose principle is the elastic 3D model deformation relative to 2D contours that are available in two calibrated X-rays films. Their authors have tested the performance of NSCC in the case of the femur, achieving a *mean point to surface distance* between the NSCC-reconstructed models and the CT-obtained ones of $\sim$1 mm.

One of the key advantages of the presented techniques is that they do not require a dedicated fiducial system during the examination, which contributes to the patient's calmness and reduces the sense of invasiveness. The patient just feels he/she is undergoing an ordinary standing erect X-ray examination that produces two radiographs, as many clinical protocols and trauma evaluations already require. The only difference is that he/she has to smoothly rotate between X-ray snapshots. This movement can be further eased with the help of a simple rotating platform.

In the work by Albiol et al. [13] (mentioned in Section 1), the authors make use of visible fiducials of the same Aruco type introduced in Section 2.2. These binary patterns remain present even during the examination phase. However, in the present research, visual fiducials are no longer necessary after calibration because the patient's body now behaves as a reference mark on its own. The main drawback when not using these visual markers is that our present setup is compelled to remain fixed to avoid recalibration. However, it is easy to devise an extension of this work to integrate the techniques discussed in [13], allowing spatial modifications of the X-ray system while continuously tracking the motion of the patient.

5. Conclusions

We have presented a method for deriving 3D locations from plain radiographs in fiducial-less ordinary diagnostic settings. With this methodology, which is based on depth cameras and the tracking of the patient's motion, projection matrices can be derived in ordinary X-ray settings and 3D inner locations and lengths can be determined through epipolar geometry. Tests with X-ray phantoms show that a millimetric level of precision can be achieved. The proposed techniques can be used in healthcare scenarios where 3D measurements are relevant and as an alternative to other modalities which may require higher doses or which may be felt more invasive by the patient.

Funding

This research has the support of Information Storage S.L., University of Valencia (grant CPI-15-170), CSD-2007-00042 Consolider Ingenio CPAN (grant CPAN-13TR01), IFIC (Severo Ochoa Centre of Excellence SEV20140398) as well as the support of the Spanish Ministry of Industry, Energy, and Tourism (grant TSI1001012013019).

Ethical approval

The patient data used in this research was handled according to the Spanish government's policies and procedures (specifically, Organic Law 15/1999 on *Protection*

of Personal Data, Law 14/2007 on *Biomedical Research* and Law 41/2002 on *Patient Autonomy and Health Documentation and Information-Related Rights and Obligations*) and the regulations recommended by the Spanish Research Council (`http://www.csic.es/normativa`).

Acknowledgments

The authors would like to thank the Radiation Oncology Department of the Physics Section at "La Fe" Hospital for the anthropomorphic phantom used in this work.

Competing interests

The authors have no conflict of interest.

References

[1] 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 and Physics*, vol. 36, pp. 968–974, 2014.

[2] L. Caponetti and A. Fanelli, "3d bone reconstruction from two x-ray views," in *Engineering in Medicine and Biology Society, 1990., Proceedings of the Twelfth Annual International Conference of the IEEE.* IEEE, 1990, pp. 208–210.

[3] K. Khoshelham and S. O. Elberink, "Accuracy and resolution of Kinect depth data for indoor mapping applications," *Sensors*, vol. 12, no. 2, pp. 1437–1454, 2012.

[4] K. Khoshelham, "Accuracy Analysis of Kinect Depth Data," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 38, pp. 133–138, 2012.

[5] M. Aoki, M. Ono, Y. Kamikawa, K. Kozono, H. Arimura, and F. Toyofuku, "Development of a real-time patient monitoring system using Microsoft Kinect," in *World Congress on Medical Physics and Biomedical Engineering*, 2013, pp. 1456–1459.

[6] A. Badal, F. Zafar, H. Dong, and A. Badano, "A real-time radiation dose monitoring system for patients and staff during interventional fluoroscopy using a GPU-accelerated Monte Carlo simulator and an automatic 3D localization system based on a depth camera," in *Medical Imaging 2013: Physics of Medical Imaging*, SPIE, Ed., vol. 8668, 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," *Journal of Digital Imaging*, vol. 26, no. 4, pp. 657–662, 2013.

[8] K. Kozono, "A study on a real-time X-ray entrance dose monitoring system in interventional radiology using Microsoft Kinect," in *Japan Radiology Congress*, 2013.

[9] F. Tahavori, M. Alnowami, J. Jones, P. Elangovan, E. Donovan, and K. Wells, "Assessment of Microsoft Kinect technology (Kinect for Xbox and Kinect for Windows) for patient monitoring during external beam radiotherapy," in *Nuclear Science Symposium and Medical Imaging Conference*, 2013, pp. 1–5.

[10] J. Wheat, S. Choppin, and A. Goyal, "Development and assessment of a Microsoft Kinect based system for imaging the breast in three dimensions," *Medical Engineering and Physics*, vol. 36, no. 6, pp. 732–738, 2014.

[11] S. Bauer, J. Wasza, S. Haase, N. Marosi, and J. Hornegger, "Multi-modal surface registration for markerless initial patient setup in radiation therapy using Microsoft's Kinect sensor," in *IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, 2011, pp. 1175–1181.

[12] P. Noonan, J. Howard, D. Tout, I. Armstrong, H. Williams, T. Cootes, W. Hallett, and R. Hinz, "Accurate markerless respiratory tracking for gated whole body PET using the Microsoft Kinect," in *IEEE Nuclear Science Symposium and Medical Imaging Conference*, 2012, pp. 3973–3974.

[13] F. Albiol, A. Corbi, and A. Albiol, "Geometrical calibration of X-ray imaging with RGB cameras for 3D reconstruction," *IEEE Transaction in Medical Imaging*, vol. 35, pp. 1952–1961, 2016.

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

[15] S. Garrido-Jurado, R. Muñoz-Salinas, F. Madrid-Cuevas, and M. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognition*, vol. 47, no. 6, pp. 2280–2292, 2014.

[16] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohli, J. Shotton, S. Hodges, and A. Fitzgibbon, "KinectFusion: Real-time dense surface mapping and tracking," in *10th IEEE international symposium on Mixed and augmented reality (ISMAR)*, 2011, pp. 127–136.

[17] J. Tong, J. Zhou, L. Liu, Z. Pan, and H. Yan, "Scanning 3D full human bodies using Kinects," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, pp. 643–650, 2012.

[18] S. Meister, S. Izadi, P. Kohli, M. Hämmerle, C. Rother, and D. Kondermann, "When can we use kinectfusion for ground truth acquisition," in *Workshop on Color-Depth Camera Fusion in Robotics, IROS*, vol. 2, 2012.

[19] K. K. Biswas and S. K. Basu, "Gesture recognition using microsoft kinect," in *International Conference on Automation, Robotics and Applications*, 2011, pp. 100–103.

[20] H. Schaub, P. Tsiotras, and J. L. Junkins, "Principal rotation representations of proper orthogonal matrices," *International Journal of Engineering Science*, vol. 33, pp. 2277–2295, 1995.

[21] P. J. Besl and N. D. McKay, "Method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.

[22] I. L. Dryden, *Statistical shape analysis.* Wiley, 1998.

[23] S. Laporte, W. Skalli, J. De Guise, F. Lavaste, and D. Mitton, "A biplanar reconstruction method based on 2d and 3d contours: application to the distal femur," *Computer Methods in Biomechanics & Biomedical Engineering*, vol. 6, pp. 1–6, 2003.