# CT from Motion: Volumetric Capture of Moving Shapes with X-rays and Videos

Julien Pansiot
julien.pansiot@inria.fr
Edmond Boyer
edmond.boyer@inria.fr
Inria Grenoble – Rhône-Alpes
Grenoble, France

**Abstract**

In this paper, we consider the capture of dense volumetric X-ray attenuation models of non-rigidly moving samples. Traditional 3D medical imaging apparatus, *e.g.* CT or MRI, do not easily adapt to shapes that deform significantly such as a moving hand. We propose an approach that simultaneously recovers dense volumetric shape and motion information by combining video and X-ray modalities. Multiple colour images are captured to track shape surfaces while a single X-ray device is used to infer inner attenuations. The approach does not assume prior models which makes it versatile and easy to generalise over different shapes. Results on synthetic and real-life data are presented that demonstrate the approach feasibility with a limited number of X-ray views. The resulting dense 4D attenuation data provides unprecedented insights for motion analysis.

## 1 Introduction

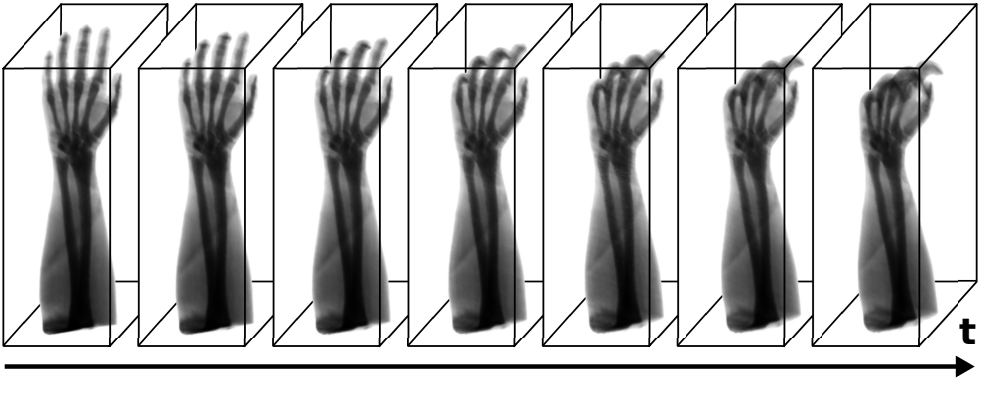


Figure 1: Results of the proposed approach for Computed Tomography from Motion (CTfM) on synthetic data (23 input frames). Dense 4D volumetric attenuation from non-rigid hand motion rendered as planar radiographic images over time from an arbitrary viewpoint. Global rigid motion has been suppressed here for clarity.

On one hand, visual motion capture systems have made significant progress over the last decade and enable fast and reliable modeling of moving shapes. Whether marker-based or markerless, relying on visible, IR, or even UV light, they can only provide observations on the shape surface and motion and it remains unclear what the internal shape structure is and how it moves. Prior models attached to the shape surface, *e.g.* a skeleton with human bodies, can help inferring information inside the shape. However such a model is usually far from being anatomically correct, especially due to the complex dynamics of soft and elastic tissues such as muscles, fat, and skin [9]. In order to estimate the real structure and motion underneath the visible surface, other observations must be considered.

On the other hand, traditional medical imaging modalities allowing for in-depth information capture can only perform on near-static subjects (*e.g.* CT, MRI), on limited volume and high ionising radiations (*e.g.* Multi-Detector CT and Electron Beam CT), have penetration limits (*e.g.* ultrasonography), or provide solely 2D images (*e.g.* planar radiography).

In this paper, we are aiming at capturing both the in-depth 3D shape and the motion of a moving sample. The resulting 4D volumetric models can find applications in motion analysis, medical diagnostic and footwear or prosthesis design, among others, by providing new insights on how movements are built. They can have a great impact in the understanding of inside motion and the design of new anatomically-correct motion models. To this purpose, we propose a novel approach that takes benefit of a visual capture system to recover a shape surface and its motion and combine it with a X-ray imaging device to model the inside moving shape structure. The proposed CTfM approach is in essence a cross-over between Computed Tomography (CT) and Structure-from-Motion (SfM). The key idea is to capture surface motion and propagate this belief in-depth whilst refining it with actual X-ray data.

The remainder of this paper is composed of the following: in Section 2 we review related works; in Section 3 we describe our shape and motion capture method; in Section 4 we present our results from synthetic and real-life data, before concluding in Section 5.

## 2 Related work

We focus below on works that consider dense volumetric reconstruction of moving objects. Little research has yet been carried out to perform tomography on moving samples. We note also that the combination of visual and X-ray imaging has likewise received little attention.

A significant amount of research in this context has focused on recovering moving features from bi-planar radiography, based on prior models [3]. While these methods produce medically relevant results, they require strong prior anatomic models, typically skeleton-based, and usually some amount of manual intervention. Without anatomical model, sparse data can be captured such as the position of purposefully implanted markers or specific, well-segmented features [16]. These methods only capture the motion of a known shape, whereas we aim at capturing both shape and motion.

Some approaches have been proposed to combine CT or SPECT imaging with vision-based motion tracking [7, 13]. These methods are however usually targeted for motion correction and hence are only able to deal with a limited range of motion. They consider motion as noise to be reduced, whereas motion is part of our acquisition process to ensure multiple viewpoints X-ray images.

One recent work by Fuerst *et al.* [6] does actually capture both shape and motion, as they coupled vision-based tracking with cone-beam radiography. In this work, video is used to register and align several 3D tomographic volumes in order to stitch them altogether. How-

ever, the tomographic reconstruction of the volumes itself was performed by conventional Cone-Beam Computed Tomography (CBCT). Moreover, only rigid motion between views was considered. In an earlier approach [11], we proposed to perform the tomographic reconstruction using solely sample motion capture from video to estimate the relative pose between the sample and the X-ray imaging device. The assumption is anyway that shapes undergo only rigid motions, which strongly limits the application scope.

The approach we propose is also related in spirit to the Kinect-based DynamicFusion [10], although with radically different data: radiographic attenuation (dense and fully transparent) in the present case against depth (sparse, localised, and fully opaque) for DynamicFusion. Hence the key sparsity assumption of the DynamicFusion approach does not hold in our case.

To the best of our knowledge, our approach is the first to simultaneously retrieve both in-depth shape and motion of a non-rigidly moving sample with very limited prior knowledge and by combining visible and X-ray images.

## 3 Shape and motion capture method

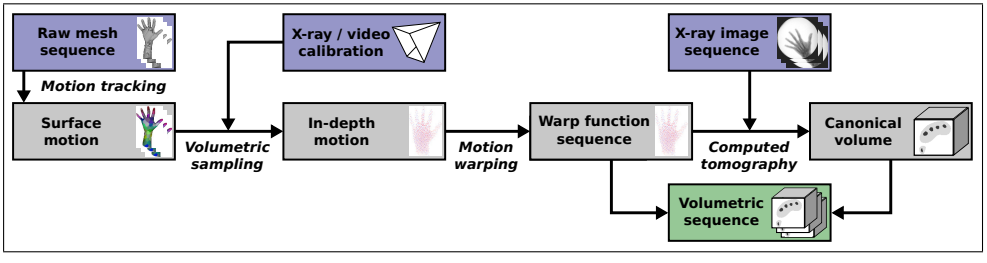


Figure 2: Proposed approach for X-ray Computed Tomography from Motion (CTfM). The captured meshes are tracked over time and the inner shape motion is derived from it. A warp function is defined that transforms any mesh into a single canonical model. This warp is combined with the X-ray images to perform Computed Tomographic (CT) reconstruction on the canonical volume. Finally, volumetric attenuation sequences can be generated.

The proposed X-ray Computed Tomography from Motion (CTfM) approach relies on the key assumption that the observed sample over time can be modelled as the composition of a static 3D canonical model and a warp function which deforms the canonical model as in [10]. This assumption is theoretically sound because the total X-ray absorbance over the sample is constant over time if we neglect the border effects (aperture). We stress that the canonical model is estimated within the proposed method, and hence in no way a prior model.

We assume that we have a canonical 3D volumetric attenuation model represented as a regular voxel grid, which is displaced non-rigidly over time using a volumetric warp function. The proposed approach consists of estimating the volumetric warp function and then performing the tomographic reconstruction on the canonical model, as illustrated in Fig. 2.

We perform the tomographic reconstruction using a specific flavour of the Simultaneous Algebraic Reconstruction Technique (SART) [1] and improve results from X-ray images directly. Finally, a volumetric model in motion can be trivially generated from the warp and the canonical model. This 4D model can be visualised in various ways over time, such

as multiplanar 2D slices stacks (*e.g.* Fig. 6 and 8) or projected and integrated into planar absorbance images (*e.g.* Fig. 1, 9, 10).

In the following, we first describe the surface tracking using visual information, then the derived warp function over the attenuation volume, and the tomographic reconstruction performed using registered information in the canonical model.

## 3.1 Surface motion estimation

Under the proposed paradigm, a sample in motion is captured simultaneously by multiple video cameras and a single X-ray imaging device. Firstly, a temporally coherent mesh is reconstructed exclusively from the video input, based on the segmented sample silhouettes, using the patch-based method proposed by [5], as illustrated in Fig. 9 (third row). This tracking method splits arbitrarily the mesh into patches and associates a rigid transformation (translation and rotation) to each of them. Local rigidity priors are then enforced to estimate the most likely deformations in a probabilistic fashion. This method is purely geometric, hence not sensitive to texture and illumination issues, but has limited capabilities in capturing motion tangential to the surface.

The captured surface motion can be interpreted as a function $\Gamma(v_{t,i}) = \overline{v_{t,i}}$ transforming any mesh $M_t = \{v_{t,i}\}$ composed of vertices $v_{t,i}$ at time $t$ into the canonical mesh $\overline{M} = \{\overline{v_{t,i}}\}$. Whilst $\Gamma$ is estimated sparsely on the mesh vertices $\{\overline{v_{t,i}}\}$, it is defined continuously on the mesh surface by linear interpolation.

## 3.2 Warp function estimation

The temporally coherent mesh sequence samples the surface motion over time. More specifically this motion is discretised spatially over a set of vertices. One such mesh in the sequence is arbitrarily selected as the canonical mesh. Since the warp is linear, this choice will theoretically have limited impact on the result, as long as the whole canonical sample surface is visible (as required by many vision-based reconstruction techniques). However the numerical stability could be affected by this choice, in particular due to sampling density variations.

In order to perform tomographic reconstruction (CT), it is necessary to encode the motion in-depth throughout the entire volume. For this purpose we cast light rays from the X-ray image and compute their intersections with the corresponding mesh. Note that depending on the mesh geometry, each of these ray may have multiple entry and exit points (multiple segments). The intersecting ray segments are then warped into the canonical mesh using the surface correspondences, as illustrated in Fig. 3. This provides a coarse discretisation of the warp function throughout the volume.

The volumetric warp function $\Phi$ is therefore the extension of $\Gamma$'s domain from the mesh surface to its entire inner volume by piece-wise linear interpolation.

## 3.3 Tomographic reconstruction

Once the warp function $\Phi$ is known for every time step we can accumulate information from all the warped rays into the canonical model, *i.e.* over space and time, and solve for the dense volumetric attenuation by computed tomography. To this aim, we iterate between actual attenuation estimation, regularisation, and image-based refinement. This method was devised to cater specifically for the challenges posed by the sample motion, such as surface motion tracking inaccuracies. The complete CT pipeline is illustrated in Fig. 4.

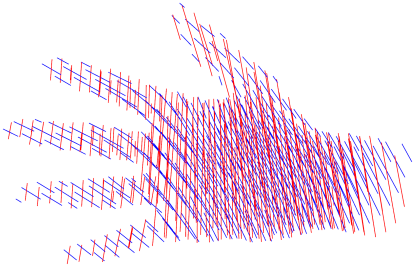

Figure 3: Volumetric warp function sampling (real-life dataset). The rays represent the X-ray light paths warped into the canonical model for two different poses, rendered in two different colours. Only a fraction of the actual rays are represented for clarity.

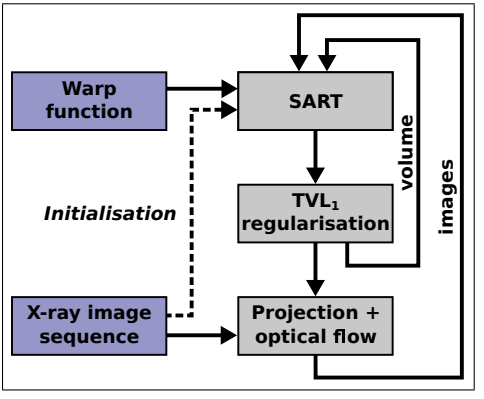


Figure 4: Iterative Computed Tomography (CT) reconstruction scheme: the canonical volume is estimated from the warp function and the X-ray images, regularised using a $TVL_1$ norm, and residual noise is handled in 2D using optical flow comparison with the reprojected volume.

The proposed method converges within a few iterations in practical cases. This can be explained empirically because the problem is ill-posed, and hence at least a local minima will be found which satisfies all constraints (observations, $TVL_1$ regularisation, flow correction).

### 3.3.1 X-ray imaging

X-ray imaging relies on a light source which is attenuated by photoelectric absorption as it travels through a partly transparent sample. The attenuation coefficient generally varies with the actual material and hence spatially in the sample.

Given partly transparent material with a linear attenuation coefficient $\mu(x)$ over a line-of-sight (*e.g.* a ray $\omega_i$), illuminated with a intensity $L_0$, the intensity $L(\omega_i)$ transmitted to the receiver is governed by the Beer-Lambert Law [2] in integral form:

$$L(\omega_i) = L_0 e^{-\int_{\omega_i} \mu(x)dx}. \tag{1}$$

We discretise the problem in 3D and reformulate it as a weighted sum over the voxels $v_j$ along the given ray $\omega_i$, $d_j$ being the distance covered within the voxel $v_j$ and $\mu_j$ the attenuation assumed uniform within $v_j$, defining the absorbance $I(\omega_i)$:

$$I(\omega_i) = \log \frac{L(\omega_i)}{L_0} = -\sum_{j \in \omega_i} d_j \mu_j. \tag{2}$$

In the reminder of this paper, we assume that $L_0$ is either known, or the reconstruction will be performed up to scale. Therefore we work only on absorbance images $A = \{I(\omega_i)\}$, in which each pixel value is equal to the absorbance $I(\omega_i)$ over its corresponding ray $\omega_i$ given the X-ray camera projection model.

### 3.3.2 Attenuation estimation

Computed Tomography (CT) consists in integrating information from multiple intersecting rays $\omega_i$ in order to solve for the original $\mu_j$ absorbance values.

In this paper, we tackle the dense volumetric tomographic reconstruction with a variation of the Simultaneous Algebraic Reconstruction Technique (SART) [1] from the X-ray images and the warp function. The modified SART method relies on the warp function sampling rather than on regular sampling of the volume.

To this end, we discretise the rays $\omega_i$ with pixel value $I(\omega_i)$, as the list of the $N$ intersected voxels $v$ in the reconstructed volume $X$ and their associated projection weights $w$: $\omega_i = \{(v_{i,1}, w_{i,1}), ..., (v_{i,N}, w_{i,N})\}$ and define its norm: $\|\omega_i\|^2 = \sum_{l=1}^{N} w_{i,l}^2$.

We then iteratively apply equations (3) and (4) to reconstruct the attenuation volume $X = \{\mu_j\}$ from the observed images with a relaxation parameter $\alpha$:

$$\tilde{I}(\omega_i)^k = \sum_{l=1}^{N} w_{i,l} X^k \left[v_{i,l}\right] \tag{3}$$

$$X_j^{k+1} = X_j^k + \alpha \frac{\sum_i w_{i,j} \frac{I(\omega_i) - \tilde{I}(\omega_i)^k}{\|\omega_i\|^2}}{\sum_i \|\omega_i\|^2} \tag{4}$$

### 3.3.3 Total variation regularisation

Furthermore, in order to regularise an otherwise ill-posed problem when the number of X-ray views is limited (with respect to the required volumetric resolution), a Total Variation norm ($TVL_1$) [12] smoothing pass is applied between SART iterations, as originally proposed by [14] (albeit combined with ART [8]).

The choice of this norm is guided by the typical nature human tissues, usually a set of compact elements with homogeneous attenuation within them. Indeed, unlike a least-square minimisation which would lead to excessive smoothing, the $TVL_1$ norm allows for sharp, albeit localised, gradients. This assumption also holds for a wide range of objects. We would like to stress here that this rather generic regularisation is the one and only prior required by our approach to solve for the attenuation volume. This improves significantly the quality of the results, as illustrated in Fig. 6 (bottom).

The TV minimisation process is a non-linear iterative process detailed in equations (5), (6), and (7). In these equations $\Delta t$ is a step size, $h$ the voxel size, $\sigma$ is the expected standard deviation of the noise, $X^0$ is the original volume, and $\Delta^{\{x,y,z\}}$ is the derivative operator along the $x$, $y$, or $z$ axis.

$$\|\nabla X_{ijk}^n\| = \sqrt{(\Delta^x X_{ijk}^n)^2 + (\Delta^y X_{ijk}^n)^2 + (\Delta^z X_{ijk}^n)^2} \tag{5}$$

$$\lambda^n = -\frac{h}{2\sigma^2} \sum_{i,j,k} \left( \|\nabla X_{ijk}^n\| - \frac{\Delta^x X_{ijk}^0 \Delta^x X_{ijk}^n + \Delta^y X_{ijk}^0 \Delta^y X_{ijk}^n + \Delta^z X_{ijk}^0 \Delta^z X_{ijk}^n}{\|\nabla X_{ijk}^n\|} \right) \tag{6}$$

$$X_{ijk}^{n+1} = X_{ijk}^n - \Delta t \lambda^n (X_{ijk}^n - X_{ijk}^0) + \frac{\Delta t}{h} \left[ \Delta^x \left( \frac{\Delta^x X_{ijk}^n}{\|\nabla X_{ijk}^n\|} \right) + \Delta^y \left( \frac{\Delta^y X_{ijk}^n}{\|\nabla X_{ijk}^n\|} \right) + \Delta^z \left( \frac{\Delta^z X_{ijk}^n}{\|\nabla X_{ijk}^n\|} \right) \right] \tag{7}$$

The parameters $\Delta t$ and $\sigma$ must be pre-defined. To this aim, the noise level is measured in a non-regularised reconstruction.

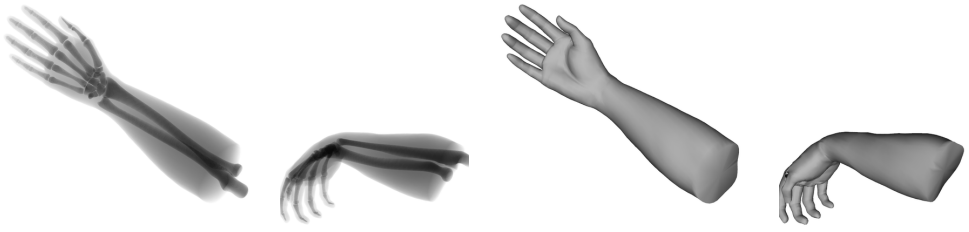

Figure 5: Synthetic input data (2 selected frames over time). Left: X-ray absorbance images rendered by raycasting. Right: skin mesh used for surface tracking.

#### 3.3.4 X-ray based refinement

The volumetric warp function is solely estimated from the surface motion which is not necessarily correct due to the complexity of the inner motion within the samples, as well as the motion tracking limitations. Hence this initial solution must be corrected using the X-ray images themselves.

To this end, we compare and correct the projected estimated volume with the actual image as proposed by [11]. A coarse-to-fine approach to optical flow is used [4]. This is equivalent to displace the warp function orthogonally to the sampling rays. Whilst this alone does not suffice to handle large motion estimation errors, it does eliminate streak artefacts due to slightly misaligned images and surfaces. We then warp the original image accordingly, and iterate.

## 4 Experimental results

The proposed method was validated using synthetic data, allowing for quantitative assessment, and with a real-life moving hand, demonstrating its capabilities for actual applications with elaborate motions and in presence of multiple noise sources.

### 4.1 Synthetic results

In order to validate quantitatively the proposed approach, a synthetic dataset was generated. To this aim, a mesh model of a human forearm was considered. The skeletal mesh components were "rigged" with standard skeletal animation techniques *i.e.* bones linked together with forward kinematics actuators, and the skin mesh was deformed using linear blend skinning. The motion was mainly composed of an overall translation and rotation (approx. 180°), a palm flexion (approx. 90°), and a finger flexion (over 180° along the phalanges), as illustrated in Fig. 5.

The simulated forearm in motion was then rendered as X-ray images over time using a specific raycasting method: each (closed) mesh was associated with a constant inner attenuation, and the total X-ray absorbance was integrated based on the entry and exit points. As for the surface information, we used directly the raw skin mesh (albeit not tracked). The complete recorded sequence consisted of 104 frames, but was down-sampled using 1 every 4 frames (without specific selection), and only 23 were use for reconstruction in order to demonstrate the value of the proposed method with a very limited number of views.

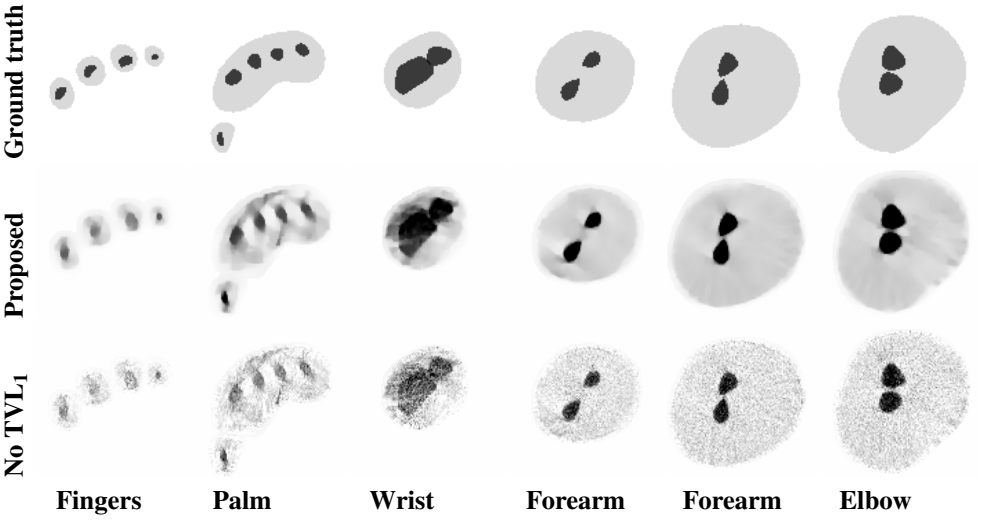


Figure 6: Synthetic data results based on 23 input frames only (6 selected slices, fingers to elbow). Top: ground truth sampling, middle: canonical model reconstruction, bottom: reconstruction without $TVL_1$ regulariser. Skeletal features are well reconstructed, even on the finger areas (left), where structures are thinner and motion significantly non-rigid. The regulariser is key to ensure convergence and reduce noise.

We then applied the full proposed pipeline, *i.e.* temporal mesh tracking, warp function estimation, and tomographic reconstruction on this data.

In order to evaluate quantitatively the resulting canonical model, we discretised the original mesh into a voxel grid to obtain volumetric ground truth as illustrated in Fig. 6 (top) and computed the difference with the resulting volume. The average voxel value difference was relatively low, at $7.9 \times 10^{-4}$, and the RMS was $4.0 \times 10^{-3}$.

The raw results illustrated in Fig. 6 demonstrate that the proposed method is able to infer accurately the volumetric attenuation canonical model from a non-rigidly moving sample. In particular, thin features such as the finger tips (only a few voxels wide) are recovered by the proposed method, despite a strongly non-rigid deformation. The main remaining artefacts lie between the phalanges in the palm (second from left in Fig. 6) where a resonance effect between the model and the sampling directions is observed.

The reprojected results illustrated in Fig. 1 show the practical value of the proposed method, as the combination of the estimated canonical model and warp function allows to visualise the dense 4D model at any instant in time and under any chosen viewpoint, whereas a single one was captured originally. Better still, the complete sequence (104 frames) can be visualised using the whole skin mesh sequence for motion estimation, whereas the canonical volumetric model is estimated from 23 frames only. In other words, the proposed method is able to render images in configurations where neither the requested viewpoint nor the sample pose were actually captured by X-ray imagery.

We did not attempt to evaluate the motion estimation on its own, since good performance on the canonical model reconstruction is intrinsically linked to correct motion estimation.

### 4.2 Real-life results

In order to evaluate our algorithm in a real-life scenario, a subject was asked to freely move a hand in a non-rigid fashion.

The scene was recorded by 10 video cameras and a single static X-ray amplifier at 30 fps, as illustrated in Fig. 7. Video cameras and X-ray imaging device were calibrated in a common coordinate system, using a method similar to [15]. The video segmentation required for 3D reconstruction was performed half-automatically (*i.e.* a limited number of parameters to be adjusted manually). The number of video cameras is sufficient to deal with self-occlusions exhibited by relatively simple geometry such as that of a human hand.

Figure 7: Proposed capture platform actual implementation: X-ray amplifiers and 10 video cameras surrounding the acquisition volume. Note that only a single amplifier is actually used in the proposed approach.

The recorded motion was composed of an overall rotation and translation as well as a finger flexion and dorsiflexion. Only a subset of these frames was actually used in order to demonstrate the performance of the proposed method with sparse input. Indeed, we achieve a tomographic reconstruction with as little as 30 views, whereas traditional (CB)CT systems typically require 100 to 1000 views.

To the best of our knowledge, there is currently no mean to acquire 4D volumetric data of this kind, and hence no ground-truth data was available. Qualitative results illustrated in Fig. 8, 9, and 10 demonstrate the performance of the proposed approach. The real-life moving hand sequence can be played back over time as an X-ray image under an arbitrarily chosen viewpoint, independently from the captured one.

Figure 8: Real-life results (3 selected slices). Main inner structures are clearly visible, including bone cavities only a few voxels wide, although some streak artefacts remain due to motion estimation innacuracies.

These results clearly demonstrate that the proposed approach is able to reconstruct a dense volumetric attenuation model from real-life non-rigidly moving sample. Indeed, the results show that even thin structures are reconstructed, the inner structure of the bones being visible, in particular in the palm.

## 5 Conclusion and future work

In this paper we proposed CTfM, a novel approach to X-ray Computed Tomography purely based on the sample motion. The proposed approach models a non-rigidly moving sample

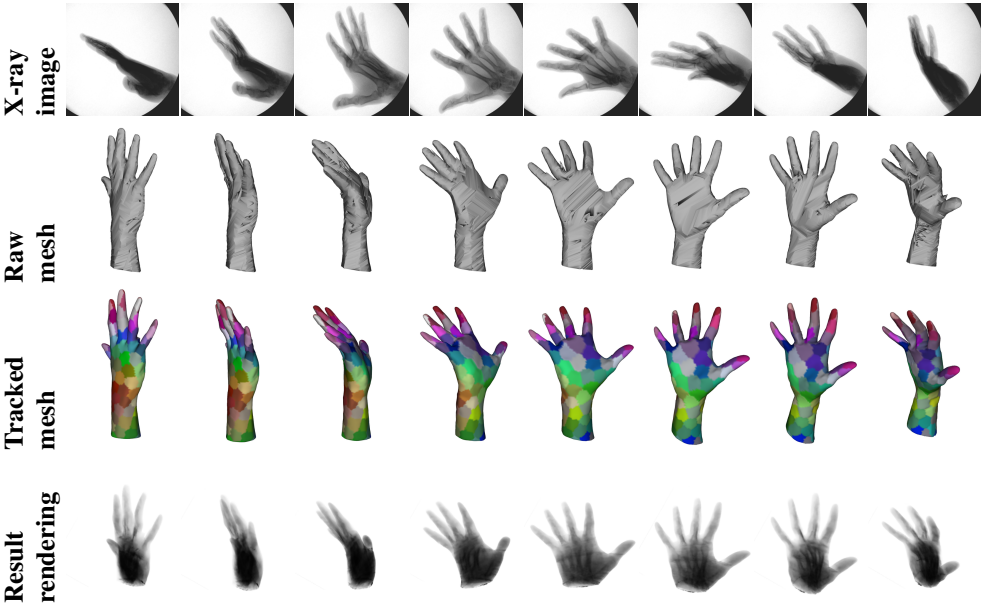


Figure 9: Real-life data and results (30 frames used as input). Left-to-right: motion over time (8 selected frames). Top-to-bottom: X-ray input image, raw mesh generated from 10 videos, tracked mesh (patches), resulting 3D volume rendered as planar X-ray images from a novel view-point. High level of details are visible in the novel rendering – see Fig. 10 for zoom.

Figure 10: Real-life results: raycasting rendering (zoom from Fig. 9). The resulting images exhibit high level of details, including the bones inner structure, particularly visible in the palm area, even in presence of non-rigid motion, *i.e.* between the metacarpals and the proximal phalanges (green arrow).

by the composition of a static 3D canonical model and a warp function encoding the motion throughout the volume over time. A very limited number of X-ray views is required, the motion field being mostly estimated on the surface and propagated in-depth. This method provides a unique insight in moving samples inner structure.

As for the improvements, some artefacts related to large locally non-linear motion could be reduced by devising and estimating a higher order warp function.

## Acknowledgements

This research was partly funded by the ANR projects KINOVIS (ANR-11-EQPX-0024) and CaMoPi (ANR-16-CE33-0014).

# References


[1] A. H. Andersen and A. C. Kak. Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm. *Ultrasonic imaging*, 6(1): 81–94, 1984.

[2] P. Bouguer. *Essai d'optique sur la gradation de la lumière*. Chez Claude Jombert, 1729.

[3] E. L. Brainerd, D. B. Baier, S. M. Gatesy, T. L. Hedrick, K. A. Metzger, S. L. Gilbert, and J. J. Crisco. X-ray reconstruction of moving morphology (XROMM): precision, accuracy and applications in comparative biomechanics research. *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology*, 313(5):262–279, 2010.

[4] A. Bruhn, J. Weickert, and C. Schnörr. Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. *IJCV*, 61(3):211–231, 2005.

[5] C. Cagniart, E. Boyer, and S. Ilic. Probabilistic deformable surface tracking from multiple videos. In *ECCV*, pages 326–339, Berlin, Heidelberg, 2010. Springer-Verlag.

[6] B. Fuerst, J. Fotouhi, and N. Navab. Vision-based intraoperative cone-beam ct stitching for non-overlapping volumes. In *MICCAI 2015*, pages 387–395. Springer International Publishing, 2015.

[7] B. F. Hutton, A. Z. Kyme, Y. H. Lau, D. W. Skerrett, and R. R. Fulton. A hybrid 3-D reconstruction/registration algorithm for correction of head motion in emission tomography. *Nuclear Science, IEEE Transactions on*, 49(1):188–194, 2002.

[8] S. Kaczmarz. Angenäherte auflösung von systemen linearer gleichungen. *Bulletin International de l'Académie Polonaise des Sciences et des Lettres. Classe des Sciences Mathématiques et Naturelles. Série A, Sciences Mathématiques*, pages 355–357, 1937.

[9] D. L. Miranda, M. J. Rainbow, J. J. Crisco, and B. C. Fleming. Kinematic differences between optical motion capture and biplanar videoradiography during a jump-cut maneuver. *Journal of Biomechanics*, 46(3):567–573, 2013.

[10] R. A. Newcombe, D. Fox, and S. M. Seitz. Dynamicfusion: Reconstruction and tracking of non-rigid scenes in real-time. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 343–352, 2015.

[11] J. Pansiot and E. Boyer. 3D imaging from video and planar radiography. In *MICCAI*, pages 450–457, Athens, October 2016. Springer.

[12] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.

[13] H. Schumacher, J. Modersitzki, and B. Fischer. Combined reconstruction and motion correction in SPECT imaging. *Nuclear Science, IEEE Transactions on*, 56(1):73–80, 2009.

[14] E. Y. Sidky, C.-M. Kao, and X. Pan. Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT. *Journal of X-ray Science and Technology*, 14 (2):119–139, 2006.

[15] X. Wang, S. Habert, M. Ma, C.-H. Huang, P. Fallavollita, and N. Navab. Precise 3D/2D calibration between a RGB-D sensor and a C-arm fluoroscope. *International Journal of Computer Assisted Radiology and Surgery*, 11(8):1385–1395, 2016.

[16] M. Yam, M. Brady, R. Highnam, C. Behrenbruch, R. English, and Y. Kita. Three-dimensional reconstruction of microcalcification clusters from two mammographic views. *IEEE Transactions on Medical Imaging*, 20(6):479–489, June 2001.