Combined Visible and X-Ray 3D Imaging

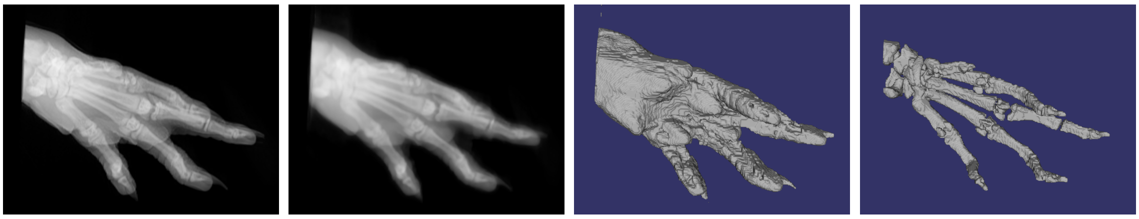
Julien Pansiot
julien.pansiot@inria.fr
Lionel Reveret
lionel.reveret@inria.fr
Edmond Boyer
edmond.boyer@inria.fr

INRIA Grenoble Rhône-Alpes
655 Avenue de l'Europe,
38334 Saint Ismier cedex
France

Abstract

This paper considers 3D imaging of moving objects and introduces a technique that exploits visible and x-ray images to recover dense 3D models. While recent methods such as tomography from cone-beam x-ray can advantageously replace more expensive and higher-dose CT scanners, they still require specific equipment and immobilised patients. We investigate an alternative strategy that combines a single x-ray source and a set of colour cameras to capture rigidly moving samples. The colour cameras allow for coarse marklerless motion tracking, which is further refined with the x-ray information. Once the sample poses are correctly estimated, a dense 3D attenuation model is reconstructed from the set of x-ray frames. Preliminary results on simulated data compared to ground-truth as well as actual in-vivo experiments are presented.

Keywords: motion capture; tomography; x-ray; colour video.



Ground truth (256 slices) | Reconstruction (from 32 views) | Skin isosurface | Skeleton isosurface

Figure 1: Rat paw volumetric reconstruction from a sparse set of simulated visible and x-ray images (32 frames from a rotational motion). Left-to-right: ground truth CT volume, volumetric reconstruction from 32 views rendered as planar x-ray, isosurface for soft tissues interface, and skeleton interface.

1 Introduction and related work

The capture of movements to study functional analysis has gained great attention in the recent years with the improvement of acquisition systems. In that respect, the ability to capture simultaneously the motion of both internal structures such as the skeleton and external surfaces is a challenge with promising applications. Whilst a range of solutions are available for motion capture and analysis in 3 dimensions using visual cues, most are solely limited to the recovery of surface information. Conversely, radiography allows for the capture of inner structures in motion, but is mostly restricted to 2 dimensions. In this paper we proceed one step towards full volumetric 3D motion capture by investigating the case of rigidly moving samples, relying on a combination of visible and x-ray cameras, and assuming no prior knowledge on the captured sample.

Our aim is to recover dense 3D models of moving objects. The existing two main strategies for radiography are the regular planar x-ray and the Computed Tomography (CT) scanner. While the latter produces high-resolution dense 3D models, it suffers three main drawbacks: higher radiation dose for the patient [10], higher cost and, more importantly here, immobility of the patient during the procedure. For these reasons, specific procedures have been developed to generate volumetric attenuation models from a limited number of regular cone-beam x-ray sources, such as full 3D models from isocentric/orbital C-arm [10], and increasingly well resolved models for breast tomosynthesis [11]. Model-based methods using bi-planar x-ray beams [1] even allow motion capture [2], however such methods require strong prior models and usually manual intervention.

In order to compute a dense 3D attenuation model while limiting the patient dose, allowing for motion capture and incidentally reducing cost, we propose a novel approach that combines regular colour cameras with standard x-ray imagery to acquire a rigidly but freely moving markerless sample. Our framework keeps all equipment static, eliminating the need for specific isocentric C-arms or other tightly controlled systems [10]. The approach does not use prior anatomic models either, hence this is neither a registration problem [11], nor a model fitting issue [1] and it allows for unknown shapes. The paradigm we follow here does not consider motion as noise but, on the contrary, as one key to 3D reconstruction.

The closest approach to ours has been proposed by Sidky *et al.* [9] where a reconstruction method for any type of calibrated x-ray imaging device under the assumption of limited angles and/or number of views is presented. Schumacher *et al.* [8] also proposed a method that combines reconstruction and motion correction in SPECT imaging, without requiring colour cameras. This method is however tailored for motion correction and hence appears to cater only for a limited range of motion. The use of colour video for motion detection during e.g. SPECT has been covered for example by [6], but requires markers. A markerless method has been proposed by [4], but is again targeted to small motion correction.

To the best of our knowledge, our approach is the first to use motion, through the combination of visible and x-ray images, to perform 3D radiography. Potential medical applications include affordable dental 3D imaging, 3D imaging in confined environments, and, when non-rigidities will be handled, joints inner dynamics.

The remainder of this paper is composed of the following: in Section 2 we describe our reconstruction method, in Section 3 we present our results from both simulated and in-vivo data, before concluding in Section 4.

2 Proposed method

The proposed method relies on a single x-ray source combined with a set of regular colour cameras. The colour cameras are used to build a 3D surface model that is tracked over time. The motion estimation of this model is refined using the x-ray images. Once the motion is known, a dense 3D attenuation model can be reconstructed from the x-ray data. The overall pipeline is illustrated in Figure 2.

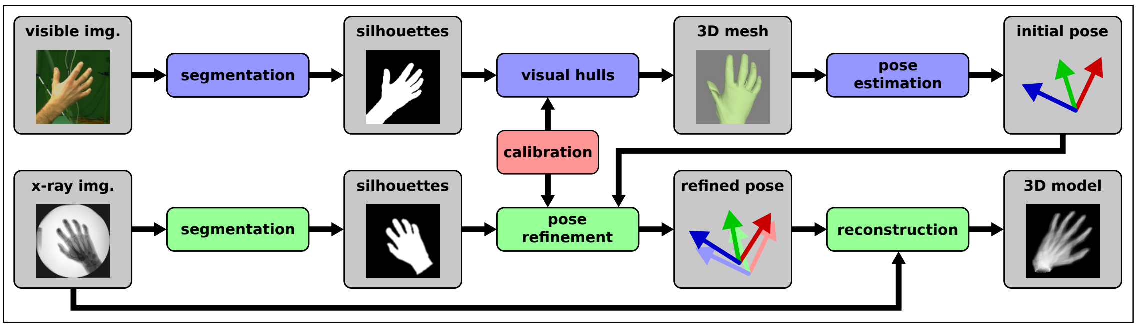


Figure 2: Overview of the proposed pipeline for combined visible and x-ray 3D imaging.

2.1 Calibration

Both visible and x-ray cameras are fully calibrated in world coordinate system prior to acquisition. This operation is performed in two stages to minimise the x-ray usage. Firstly, the eight video cameras are fully calibrated (intrinsic and pose) on their own, using a moving LED wand within a single bundle adjustment framework. Secondly, a lightly coloured lead ball is displaced within the entire capture volume, triangulated by the visible cameras, before the x-ray camera is finally fully calibrated knowing the 3D from calibrated video cameras.

2.2 Initial pose estimation

Since the sample can move freely within the capture volume, it is necessary to estimate its pose prior to reconstruction. Using the colour cameras only, the silhouettes of the sample segmented from the background are processed with a polyhedral visual hull algorithm [3] to generate a 3D mesh of the sample per frame (see fig. 2). The remaining meshes of the sequence are then registered to a reference pose model, e.g. the first frame, with a robust Iterative Closest Point (ICP) implementation with outlier detection. This provides an initial pose, i.e. rotation and translation, estimation for each frame and with respect to the reference frame.

2.3 Pose refinement

Since the visual hull meshes exhibit artefacts due to the limited number of colour cameras, the initial pose estimation is corrupted. To improve the pose, we exploit directly the x-ray silhouettes, segmented from the x-ray background light.

To this aim, we rely on the assumption that if the poses were perfectly registered, the volume reprojection on the x-ray plane should equate perfectly to the x-ray silhouettes. For simplicity, we consider the model to be static and captured by moving cameras, with intrinsic parameters from the initial calibration, and extrinsic from the initial pose estimation. A cost function penalising differences between the original x-ray silhouettes and the reprojected model is used within a gradient descent framework to refine the poses iteratively. If C is the set of N camera parameters, I the set of N segmented images, M a mesh, 'proj' the camera projection operator, and 'vh' the visual hull operator, we try to estimate:

$$\arg\min_{C} \left(\sum_{i=1}^{N} |\mathrm{proj}_{C_i}(M) - I_i| \right) \quad st. \quad M = \mathrm{vh}_C(I). \tag{1}$$

This extra stage also reduces slight spatial and temporal misalignment between the visible and x-ray systems, providing more robust results.

2.4 Reconstruction

Once the pose is known, the Kaczmarz method [5], also referred to as Algebraic Reconstruction Technique (ART) is used to reconstruct iteratively the volumetric attenuation. For enhanced performance, the views are ordered such that two successive points of view are most orthogonal, and the reconstruction takes places only within the x-ray visual hull.

Given the volume X being reconstructed, the observed images A with a saturated attenuation A_{sat}, and a relaxation parameter λ_i, we exploit the projection matrix sparseness by pre-computing each camera ray defined as the list of the N intersected voxels in the reconstructed volume p and their associated weights w. For each ray $R_i = \{(p_{i,1}, w_{i,1}), \ldots, (p_{i,N}, w_{i,N})\}$ associated to the image pixel value A_i, we then compute the norm, the current integrated attenuation on the ray x_i, and the weighted difference with the observed pixel value, d_i:

$$\|R_i\|^2 = \sum_{j=1}^{N} w_{i,j}^2, \qquad x_i = \sum_{j=1}^{N} w_{i,j} X\left[p_{i,j}\right], \qquad d_i = \lambda_i \frac{A_i - x_i}{\|R_i\|^2}. \tag{2}$$

The voxel attenuations are then updated as follows:

$$X^{n+1}\left[p_{i,j}\right] = \begin{cases} 0 & \text{if } X^n\left[p_{i,j}\right] + d_i w_{i,j} < 0 \text{ (positivity constraint)}, \\ X^n\left[p_{i,j}\right] & \text{if } d_i < 0 \text{ and } A_i = A_{sat} \text{ (prevent saturation artefacts)}, \\ X^n\left[p_{i,j}\right] + d_i w_{i,j} & \text{in other cases.} \end{cases} \tag{3}$$

The relaxation parameter λ_i is weighted for each pose based on the leave-one-out score calculated as in eq. (1) in order to favour poses which appear more reliably estimated.

Since the amount and nature of the observed data may render the problem locally ill-posed, high-frequency noise is observed in the results. Using the assumption that living organisms can be modeled by a set of relatively homogeneous tissues, a 3D adaptation of Rudin *et al.* seminal non-linear Total Variation (TV) [7] is enforced on the model in between iterations for regularisation, as proposed by [9]. This improves dramatically the results, as illustrated in Figure 3.

3 Experiments and results

3.1 Simulated data

The combined x-ray and visible images can be simulated by rendering real volumetric CT data using raycasting. In this experiment, we rendered the CT model of a rat paw (256^3 voxels), simulating a system of eight colour cameras and one x-ray device. The combined set of 2D images was then processed through the whole pipeline. The effective projection resolution for both systems was 600×600 pixels. The model was artificially moved continuously for 32 frames, following roughly a 180-degree rotation. This allowed for comparison with ground-truth data. Figures 1 and 3 show the original model as well as the reconstructed volume.

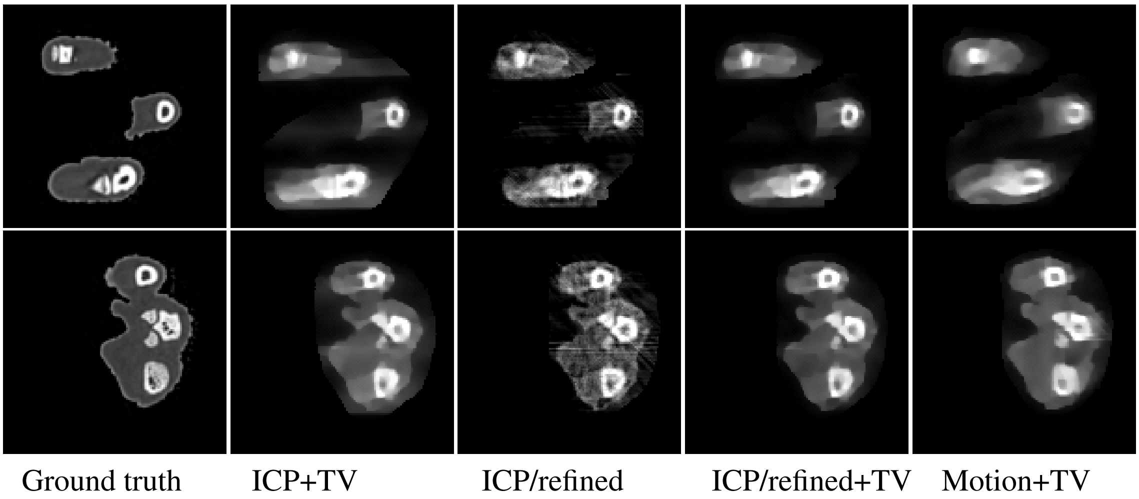


Figure 3: Two rat paw reconstructed slices from simulated visible and x-ray images (32 frames). Left-to-right: ground-truth CT data (256 slices), motion estimated with ICP and TV regularisation, motion refined but without TV, motion refined and TV, true motion and TV. Slight motion estimation errors are clearly corrected by the refinement stage, but not as well as with true motion. TV regularisation reduces significantly the high-frequency noise.

3.2 In-vivo experiment

A system composed of eight colour cameras (2048×2048 pixels) and one Siemens ARCADIS Avantic x-ray C-arm (1024×1024 pixels) was deployed for this preliminary experiment. The reconstruction volume (256^3 voxels) covered a cube of 20×20×20 cm, *ie.* a resolution of 0.78mm/voxel.

Three volunteers (one of them shown here) were asked to perform roughly a 180-degree wrist circumduction with fingers kept as rigid as possible within the working volume, during which 32 frames were captured. The colour cameras recorded at 30 fps, whilst the x-ray acquired at 10 fps to minimise the dose, both being only coarsely synchronised. Figure 4 illustrates the preliminary results of the proposed method.

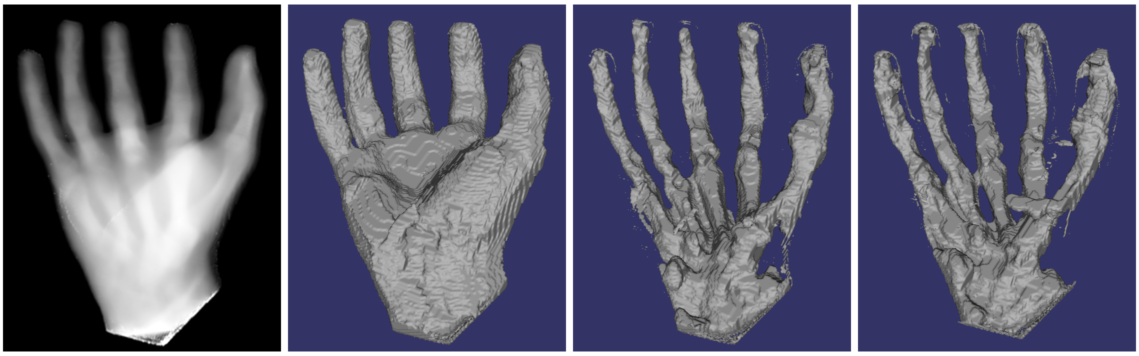


Figure 4: Human hand reconstruction with a combined visible & x-ray system (32 frames). Left-to-right: volume rendered as planar x-ray, soft tissues thresholding, skeleton thresholding, skeleton thresholding without pose refinement: some phalanges are thicker, the distal noisier. The soft tissues exhibit a concavity that cannot be modelled solely by visual hulls.

It is evident from the results that whilst the simulated data demonstrates the proposed approach's great potential, the preliminary reconstruction of in-vivo data is noisier. We attribute this difference to the fast shutter noise in x-ray images (in particular in the metacarpal area) as well as non strictly rigid motion (on the finger tips and the thumb in particular). The former should be improved by carefully planned x-ray acquisition parameters.

4 Conclusion and future work

In this paper we have shown that a multi-camera system can advantageously complement a single static x-ray imaging device and enable 3D motion imaging of rigidly moving samples. Future work includes the analysis of the achievable quality based on the recorded motion and non-rigid motion handling using rigid-by-part modelling.

Acknowledgments

This project was funded by KINOVIS (ANR-11-EQPX-0024) and MORPHO (ANR-10-BLAN-0206), with help from Amiqual4Home (ANR-11-EQPX-0002).

References

[1] S. Benameur, M. Mignotte, H. Labelle, and J. A. De Guise. A hierarchical statistical modeling approach for the unsupervised 3-D biplanar reconstruction of the scoliotic spine. *Biomedical Engineering, IEEE Transactions on*, 52(12):2041–2057, 2005.

[2] 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.

[3] J.-S. Franco and E. Boyer. Exact polyhedral visual hulls. In *BMVC*, pages 329–338, 2003.

[4] 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.

[5] 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.

[6] J. E. McNamara, P. H. Pretorius, K. Johnson, J. M. Mukherjee, J. Dey, M. A. Gennert, and M. A. King. A flexible multicamera visual-tracking system for detecting and correcting motion-induced artifacts in cardiac SPECT slices. *Medical physics*, 36(5):1913–1923, 2009.

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

[8] 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.

[9] 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.

[10] J. H. Siewerdsen, D. J. Moseley, S. Burch, S. K. Bisland, A. Bogaards, B. C. Wilson, and D. A. Jaffray. Volume CT with a flat-panel detector on a mobile, isocentric C-arm: pre-clinical investigation in guidance of minimally invasive surgery. *Medical physics*, 32(1):241–254, 2005.

[11] G. Yang, J. H. Hipwell, D. J. Hawkes, and S. R. Arridge. A nonlinear least squares method for solving the joint reconstruction and registration problem in digital breast tomosynthesis. In *MIUA*, pages 87–92, 2012.