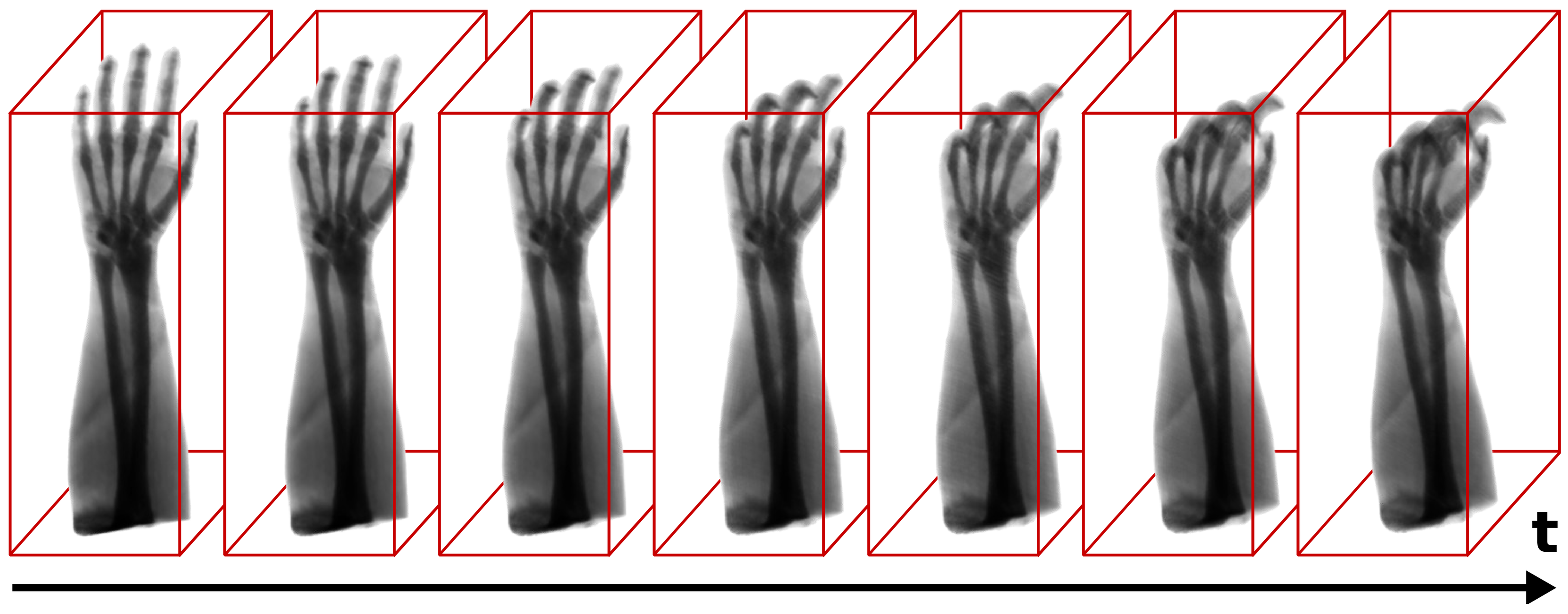
CT from Motion: Volumetric Capture of Moving Shapes with X-rays and Videos

Julien Pansiot and Edmond Boyer

Morpheo, Inria Grenoble Rhône-Alpes, France

Motivation

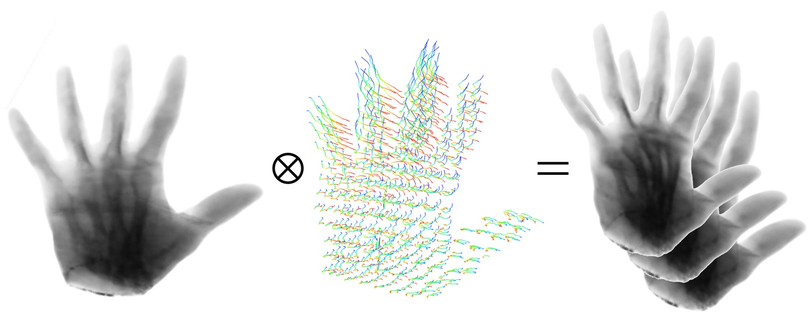
- **Capture dense 3D+t attenuation models of non-rigidly moving samples**
- **Novel volumetric in-depth motion insight**
- **Challenging complementary imaging modalities fusion**



Proposed Approach

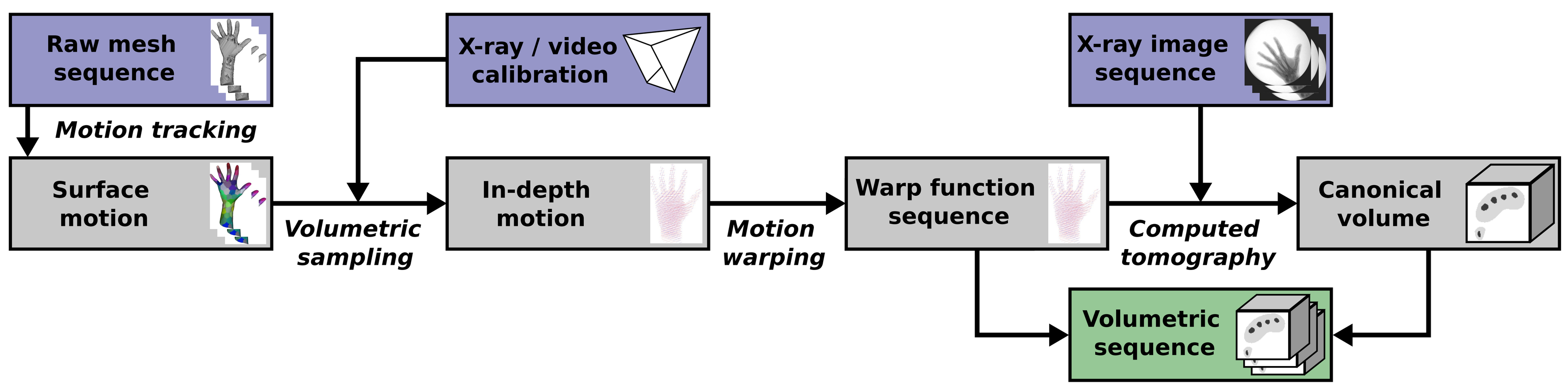
Strategy:
- Static X-ray & video sensors
- Dynamic sample: shape motion is a feature [1]
- Motion fused over time with X-ray imagery [2]

Key components:
- Shape surface motion estimated from video
- Surface motion propagated in-depth (warp)
- X-ray measures aggregated over time
- CT method to recover 3D+t attenuation



3D canonical volume | *Warp (t)* | *3D+t volume*

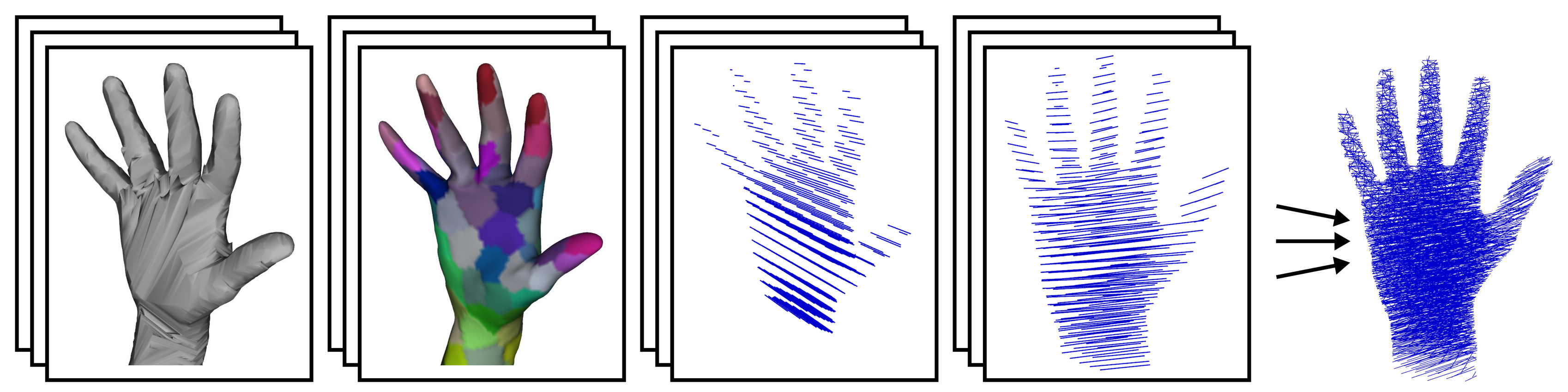
Method



Warp Function Estimation

Surface motion estimation
- Multiple video capture
- 3D mesh from video (visual hull)
- Geometric patch-based mesh tracking

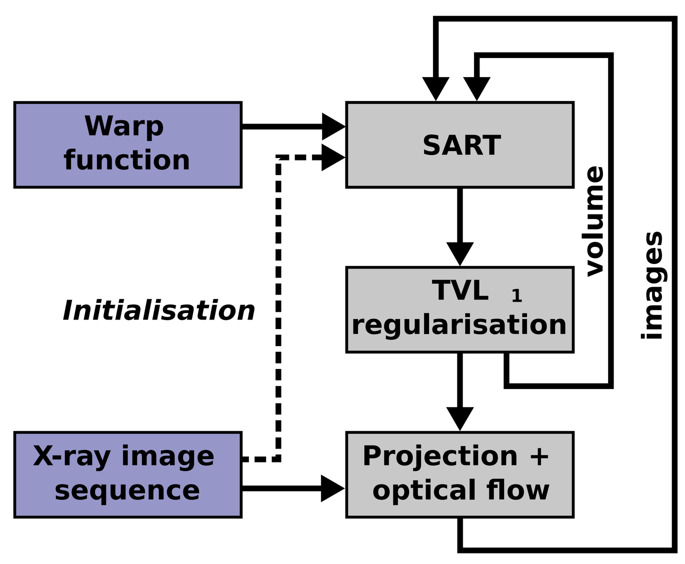
In-depth motion estimation
- In-depth volume sampling (X-ray lines-of-sight)
- Sampling warp onto canonical volume
- Line-of-sight accumulation in canonical volume



Raw mesh | *Tracked mesh* | *Sampled volume* | *Warped rays* | *Accumulated rays*

Tomographic Reconstruction

Iterative estimation of:
- Canonical model given warp function and images
  - Simultaneous Algebraic Reconstruction Technique
  - Specialised for non-homogeneous sampling
- Canonical model regularisation with TV L_1
  - Favours homogeneous volumes
  - Allows for sparse large gradients
- Image-based refinement based on residual optical flow
  - Flow between input image and projected model
  - Corrects motion estimation inaccuracies



Video link

Acknowledgements
This research was partly funded by the KINOVIS and CaMoPi projects.

References
[1] R. A. Newcombe, D. Fox, and S. M. Seitz. **DynamicFusion: Reconstruction and tracking of non-rigid scenes in real-time.** CVPR 2015, pp. 343–352.
[2] J. Pansiot and E. Boyer. **3D Imaging from Video and Planar Radiography.** MICCAI 2016, pp. 450-457.

KINOVIS X-ray Capture Platform

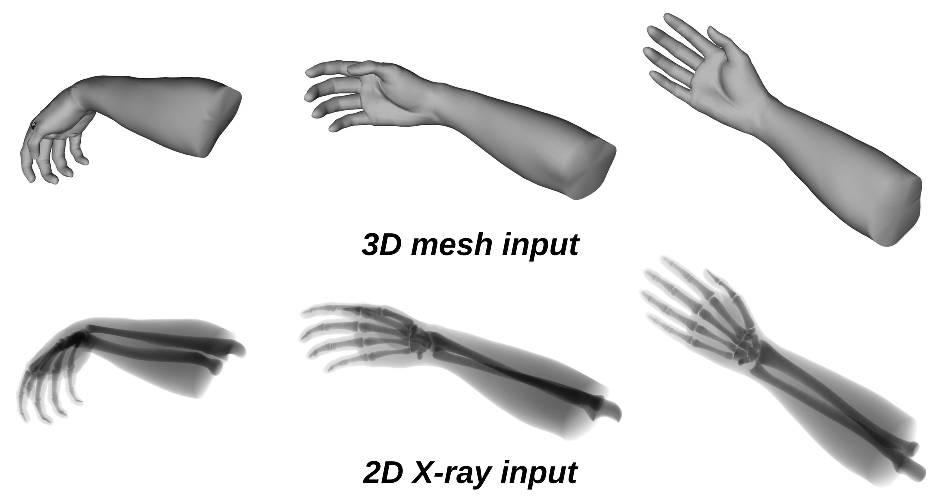
KINOVIS capture platform
- 10 video cameras
  - 3D surface motion
  - 4Mpix @ 100fps
- 2 X-ray C-arms (only 1 used)
  - In-depth motion
  - 1Mpix @ 30 fps

Synthetic Forearm

Synthetic forearm composed of:
- Articulated skeleton mesh
- Deformable skin mesh

Forearm rendered as:
- 3D mesh (linear blend skinning)
- X-ray image (raycasting)

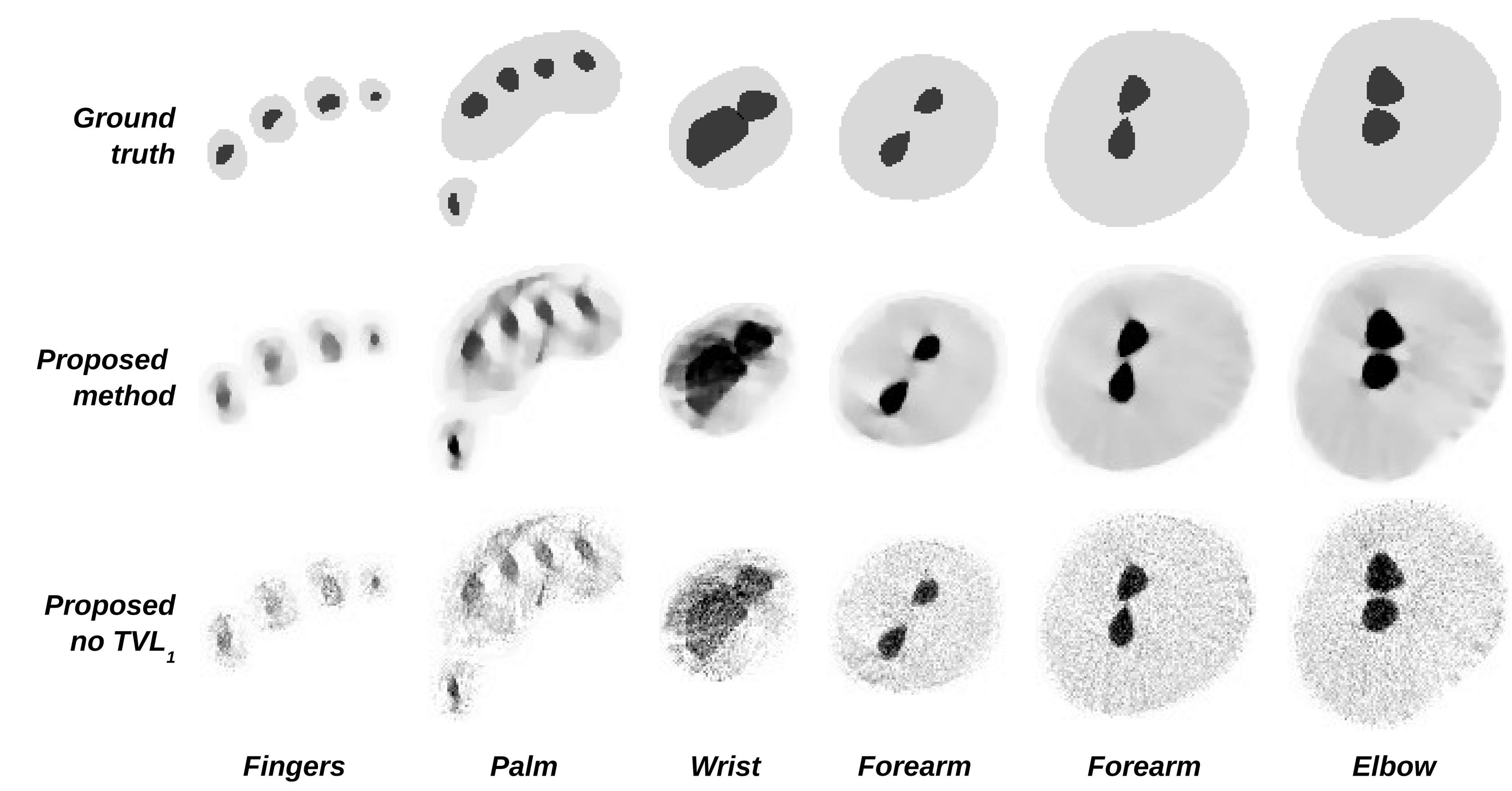
Simulated motion (23 frames)
- Global rotation + palm & finger flexion



3D mesh input

2D X-ray input

Results: 3D Canonical Volume



Real-life Hand

Real-life human hand data
- 10 static video cameras
- 1 static X-ray C-arm
- Volumetric resolution: 1mm/voxel

Actual motion
- Moved over 30 frames at 10 fps
- Free wrist rotation over ~180°
- Free finger flexion + dorsiflexion

1 X-ray input | *10 video inputs*

Results: 3D+t Dense Volume