Vision-based endoscope tracking for 3D ultrasound image-guided surgical navigation

[Yang et al. 2014, *Comp Med Imaging and Graphics*]

Gustavo Sato dos Santos

IGI Journal Club

23.10.2014

Motivation

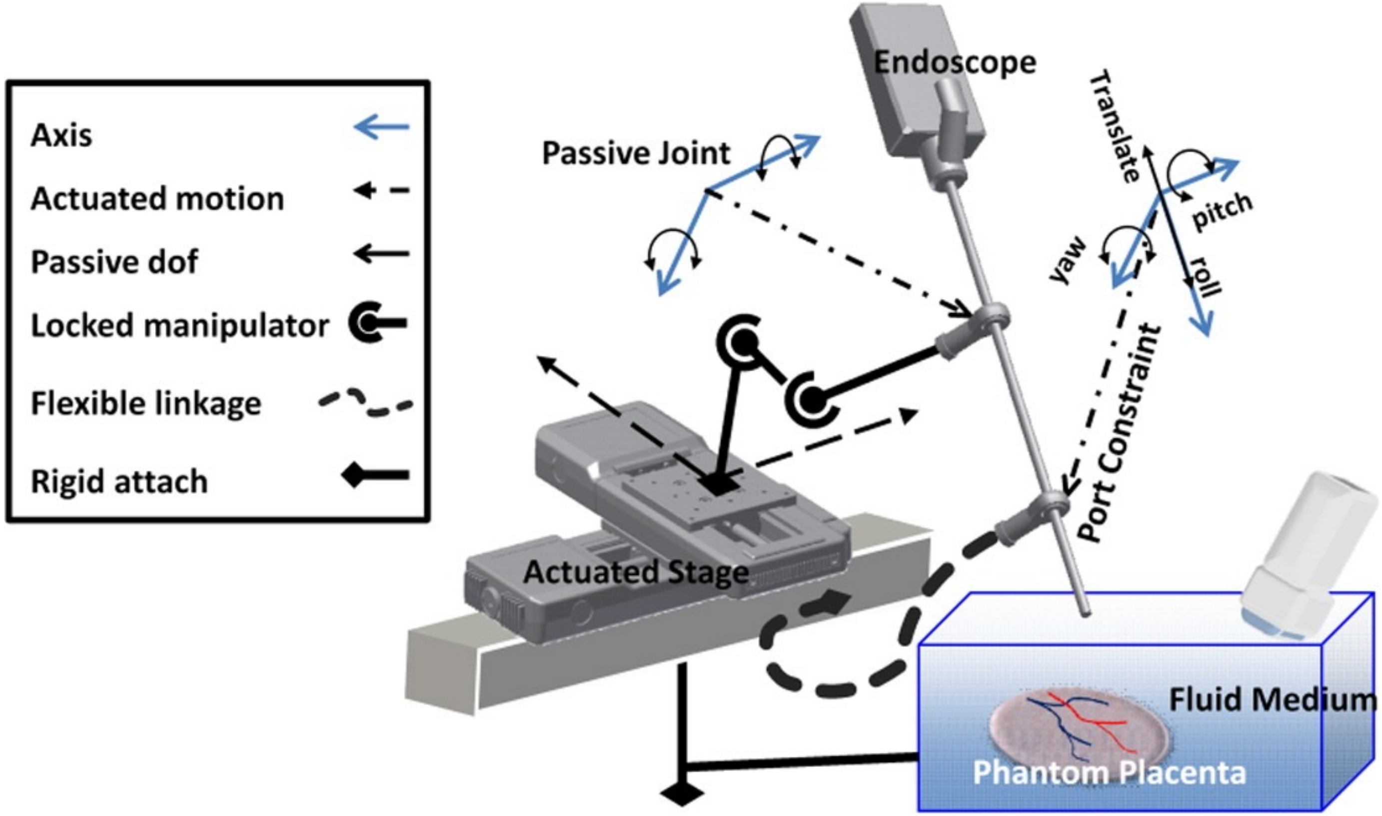
Goal: Surgical navigation for minimally-invasive fetal surgery

Disadvantages of other tracking methods

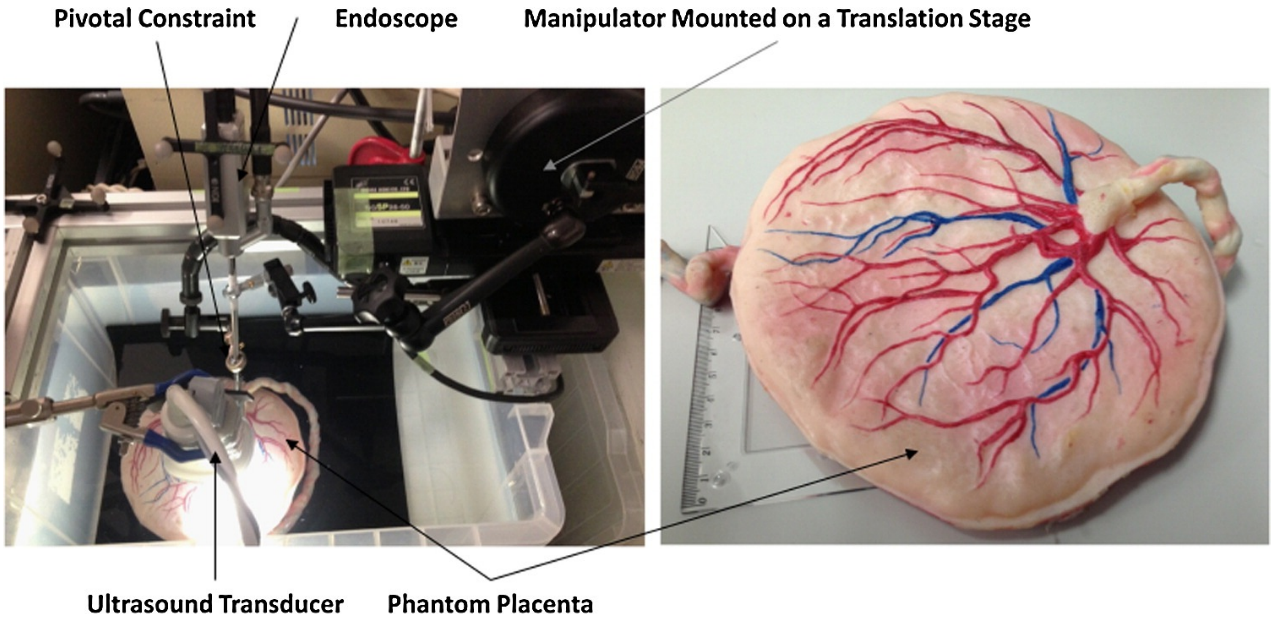
- Optical: line-of-sight between tracker and markers
- Electromagnetic (EM): prone to noise - electronic devices in OR
- EM tracker + inertia measurement unit (IMU): issues with tracking initialisation, drift errors, accuracy
- Vision-based (Structure-from-Motion): not suitable due to unpredictable amniotic fluid, need to minimise illumination

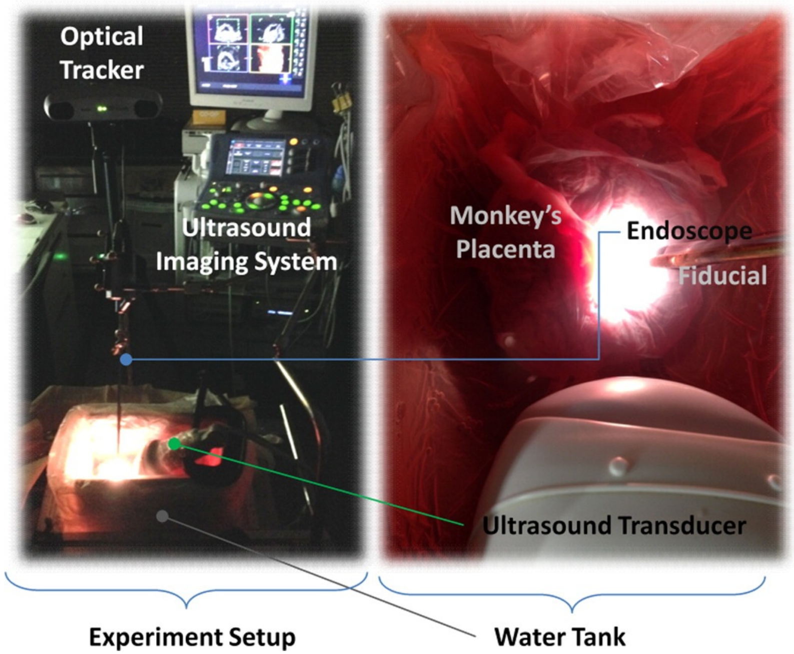
Approach

- Initial camera position by ultrasound image-based localisation;
- Vision-based tracking

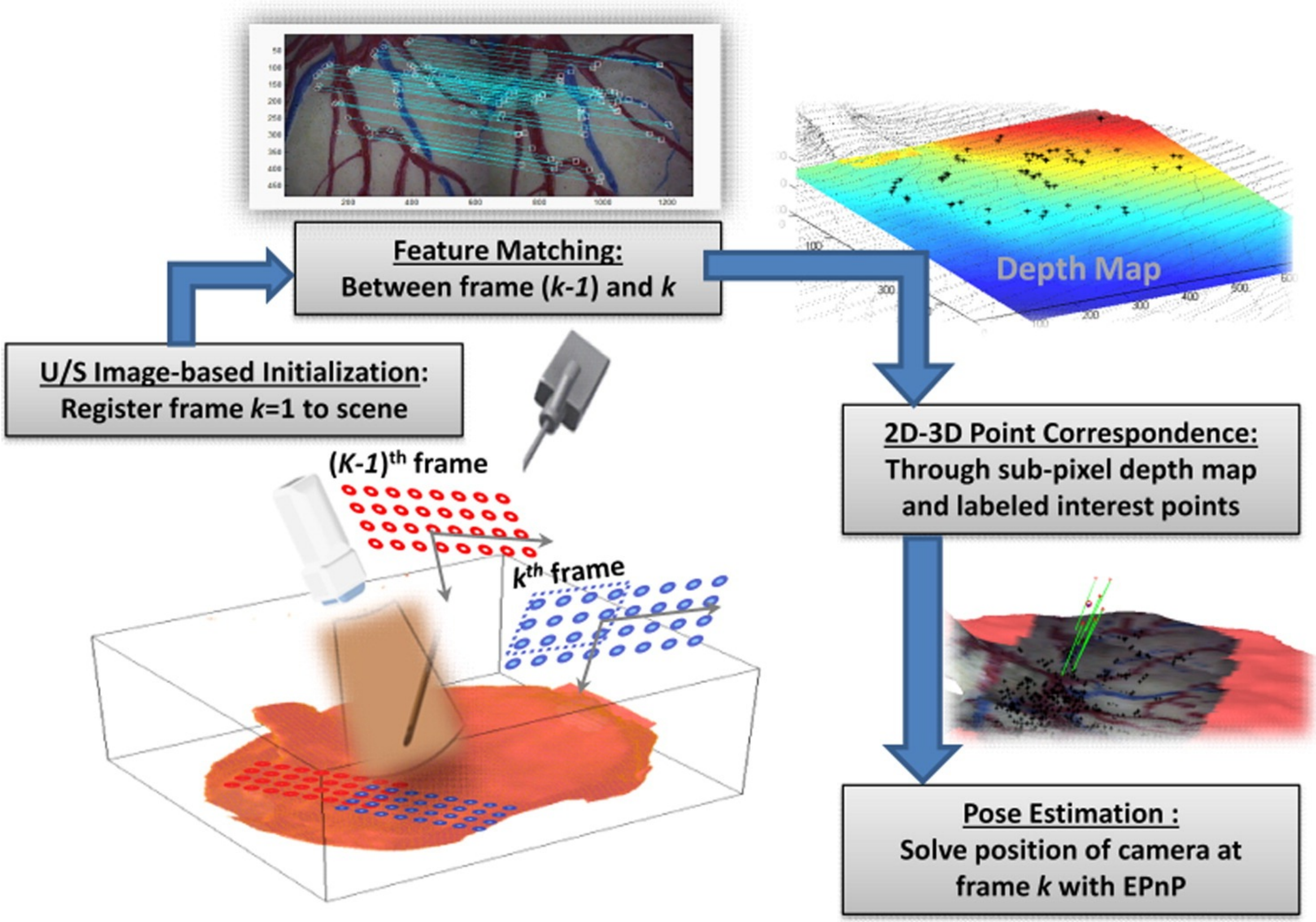


- Ultrasound: Hitachi ProSound α10 w/ 3D tilt-scanning convex sector transducer
  - mounted on rigid bracket (to minimise motion artefacts)
- Endoscope: Shinko Optical, 5.4mm diam. rigid endoscope, Xenon light source
- Translation stage: Sigma Koki, 4 μm/pulse resolution, 1 μm precision

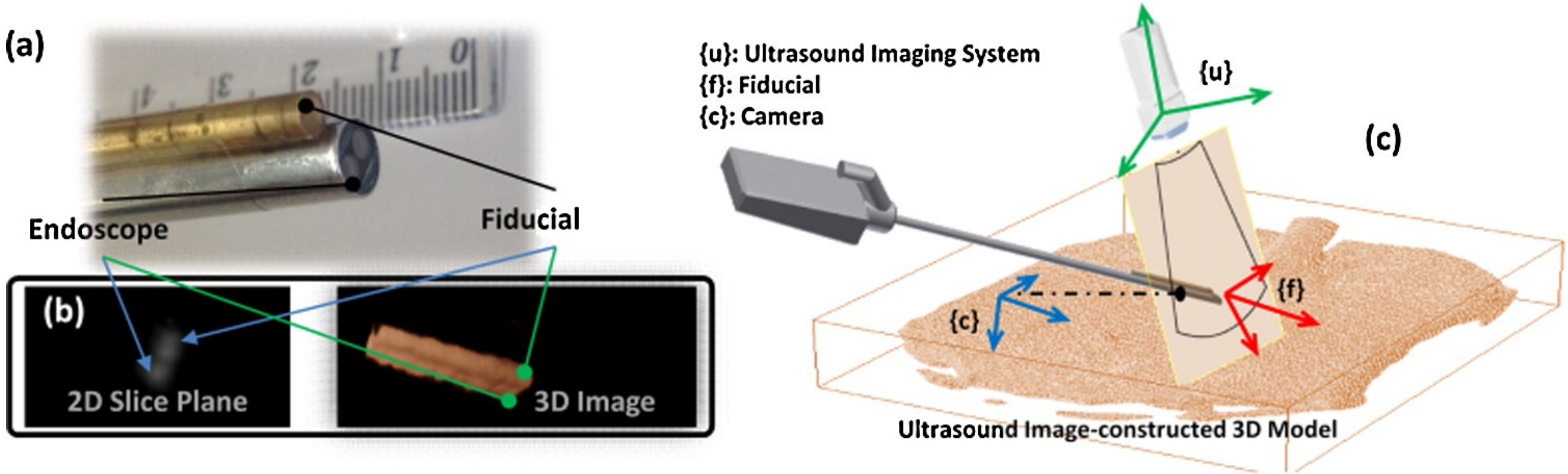

Pivotal Constraint
Endoscope
Manipulator Mounted on a Translation Stage
Ultrasound Transducer
Phantom Placenta


Optical Tracker
Ultrasound Imaging System
Monkey's Placenta
Endoscope
Fiducial
Ultrasound Transducer
Experiment Setup
Water Tank

Workflow


Feature Matching:
Between frame (k-1) and k
Depth Map
U/S Image-based Initialization:
Register frame k=1 to scene
(K-1)th frame
kth frame
2D-3D Point Correspondence:
Through sub-pixel depth map
and labeled interest points
Pose Estimation :
Solve position of camera at
frame k with EPnP

U/S image-based initialisation



- Scene geometry acquired by 3D ultrasound imaging
  - Manual selection of placenta ROI;
  - Thresholding with isovalue → meshed surface model (50,000 vertices)
- Camera position acquired by localising fiducial (8 cm length, 0.3 cm diam.)
  - Prior fiducial-camera calibration (f → c transformation)
- Localisation error ≈ 1.32 mm
- Low acquisition rate, multiple sampling required for robustness

Underwater camera calibration

Optical properties of medium → intrinsic parameters of camera

- Camera pre-calibrated in saline solution used for experiments
  - Camera calibration toolbox for Matlab (Bouguet JY, 2004)

- Images corrected for radial and tangential lens distortions
  - Brown-Conrady model (Brown DC, *Photon Eng* 1971)

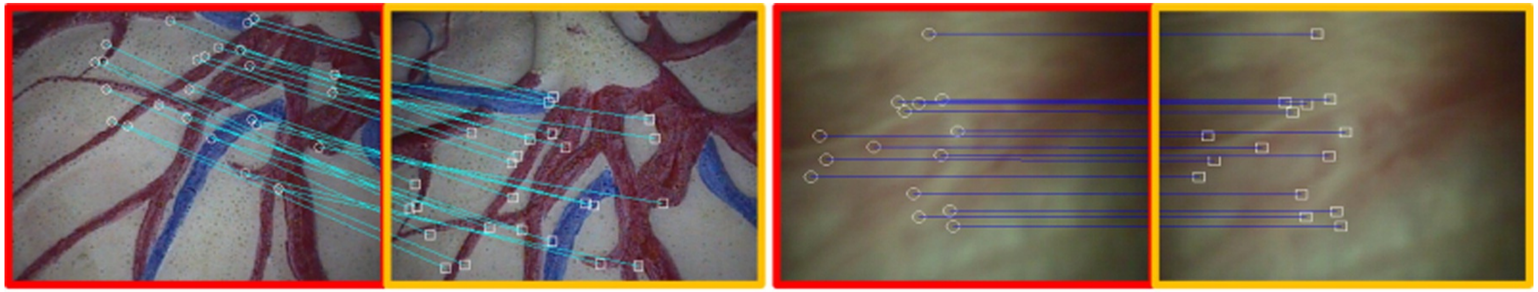
Inter-frame feature matching

Speed-Up-Robust-Features (SURF) algorithm [Bay et al., *Comput Vis Image Und* 2008]

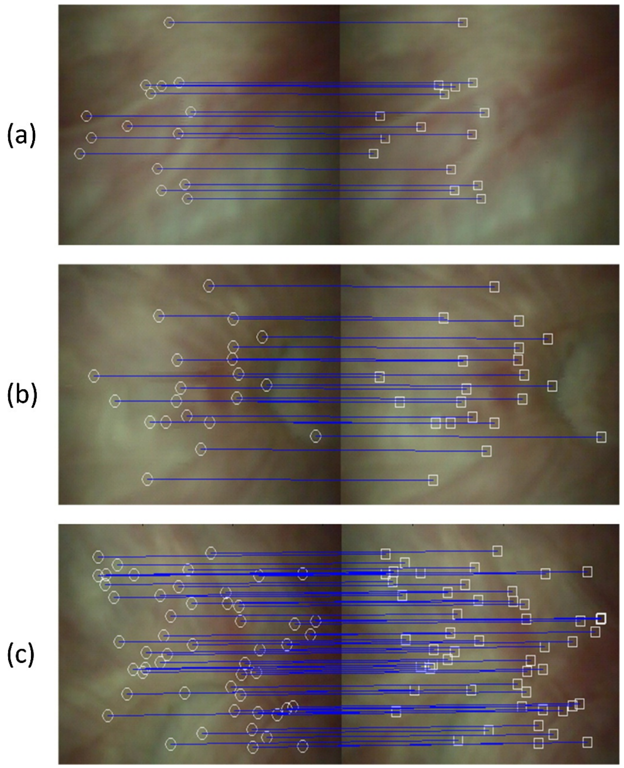
- Scale and rotation invariant features
- FAST-Hessian feature detection, 64-element descriptor representing distribution of Haar-wavelet responses of feature neighbourhood
- Robust even in scenes with poor texture (important for tissue imaging)
- Outlier removal: RANSAC algorithm
- Result: 10-30 reliable feature matches (20 required for subsequent processing)

Inter-frame feature matching

Phantom | Ex-vivo monkey placenta



(a) (b)



Texture conditions:

(a) Desirable

(b) Moderate

(c) Poor

2D-3D point correspondence

Mapping image coordinates *(i_p,j_p)* to 3D coordinates *(x_p,y_p,z_p)*

- Project 3D vertices of ultrasound image model to the camera plane to obtain their image coordinates:

$$^{k-1}\mathbf{z}^{\bullet}\left({}^{k-1}\begin{pmatrix} i & j & 1\end{pmatrix}^{\mathrm{T}}\right) = \mathbf{K}\left({}^{k-1}\mathbf{R}_u \;\middle|\; {}^{k-1}\mathbf{t}_u\right){}^{u}\begin{pmatrix} x & y & z & 1\end{pmatrix}^{\mathrm{T}}$$

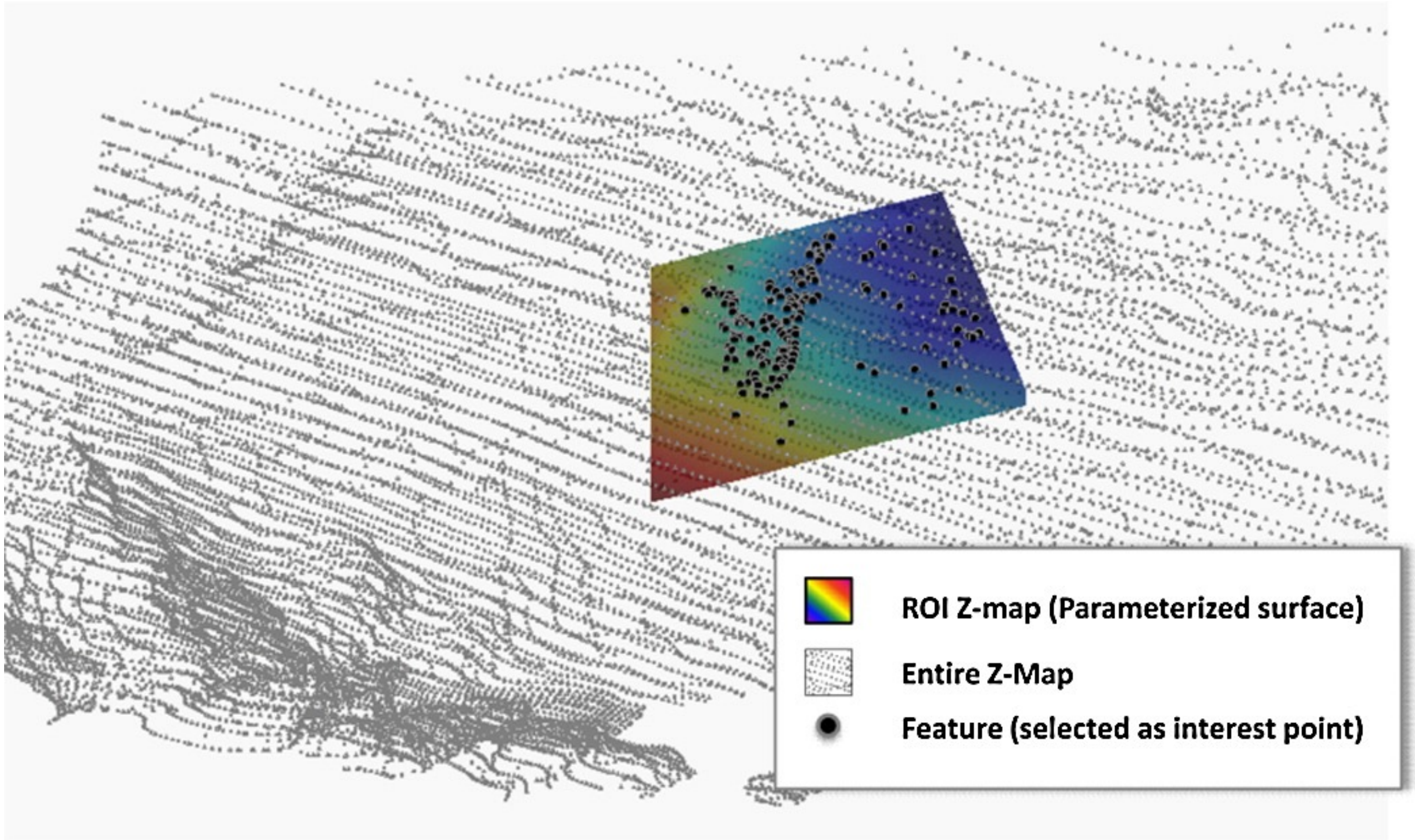
K: intrinsic camera parameter matrix; $^{k-1}\mathbf{R}_u$,$^{k-1}\mathbf{t}_u$: rotation matrix and translation vector from camera's viewpoint at frame *k-1*

- Delaunay triangulation of points $(i,j,{}^{k-1}z)$ → dense depth map $\boldsymbol{Z}(i,j)$
- 3D camera-centric coordinates of interest points:

$$^{k-1}\mathbf{p}_l = \left({}^{k-1}\left(i_p - i_o\right)\cdot\frac{{}^{k-1}Z\left(i_p,j_p\right)}{f_x},\quad {}^{k-1}\left(j_p - j_o\right)\cdot\frac{{}^{k-1}Z\left(i_p,j_p\right)}{f_y},\quad {}^{k-1}Z\left(i_p,j_p\right)\right)$$

(i_0,j_0) and (f_x,f_y): principal point and focal length from **K**

2D-3D point correspondence



3D interest points are updated every frame, according to matching features across two adjacent images

Pose estimation

Pose estimation as Perspective-n-Point (PnP) problem

- Better accuracy and stability than Direct Linear Transformation (DLT)
- **EPnP algorithm** (Lepetit V et al., *Int J Comput Vision* 2009):
  - non-iterative
  - solves coordinates of M=4 virtual control points $\boldsymbol{q}=\{q_1,\ldots,q_M\}$

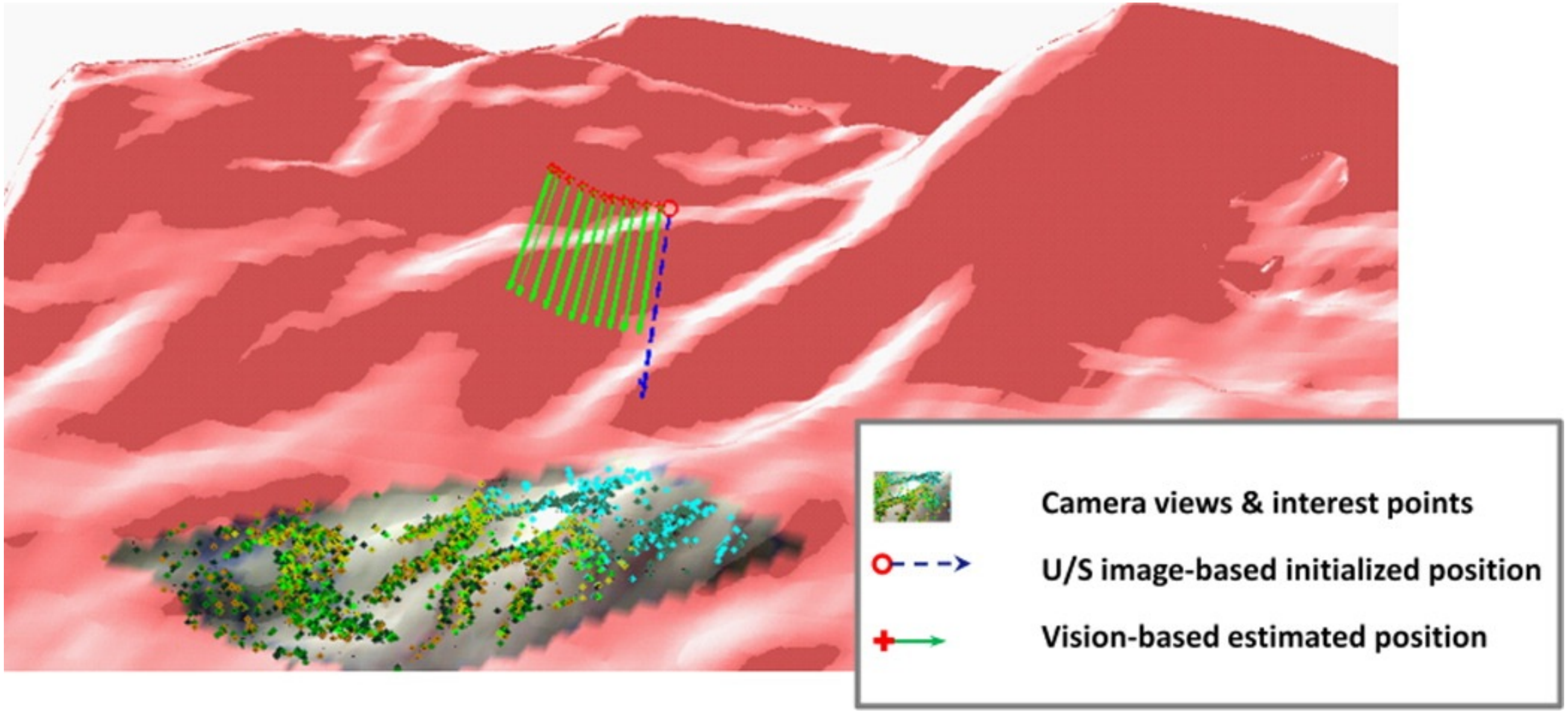
$$ {}^{k}\mathbf{z}_l^{\bullet}\left({}^{k}\begin{pmatrix} i_l & j_l & 1 \end{pmatrix}^{\mathrm{T}}\right) = \mathbf{K}\sum_{m}^{M}\lambda_{lm}{}^{c}q_m $$

λ_{lm}: homogeneous barycentric coordinates summing to one

  - Control points **q** consist of the centroid of interest points **p** and another 3 points that align closely to the principal direction of **p**
  - Computational time: O(n)
  - Performs well even with noisy non-fixed interest points

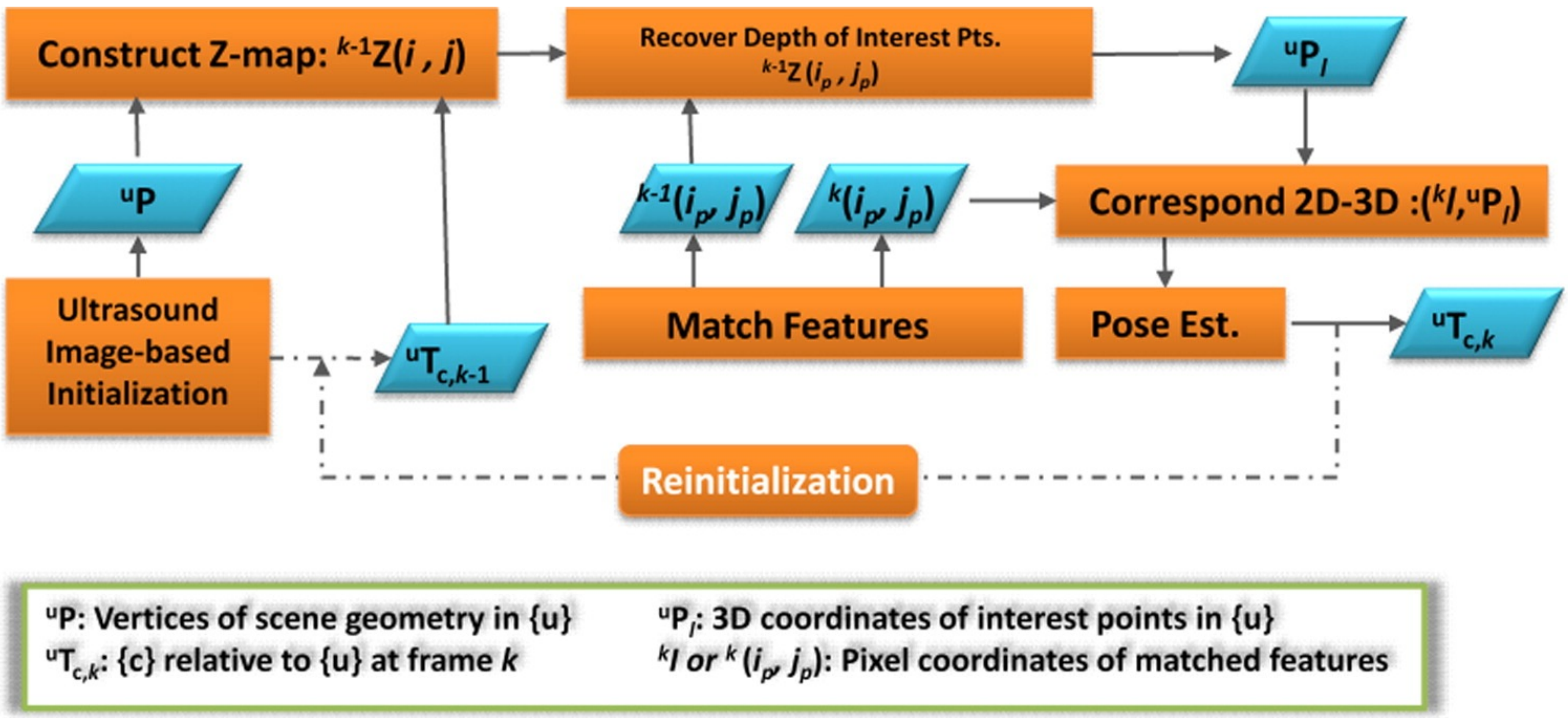
GIFT Surg UCL

Pose estimation

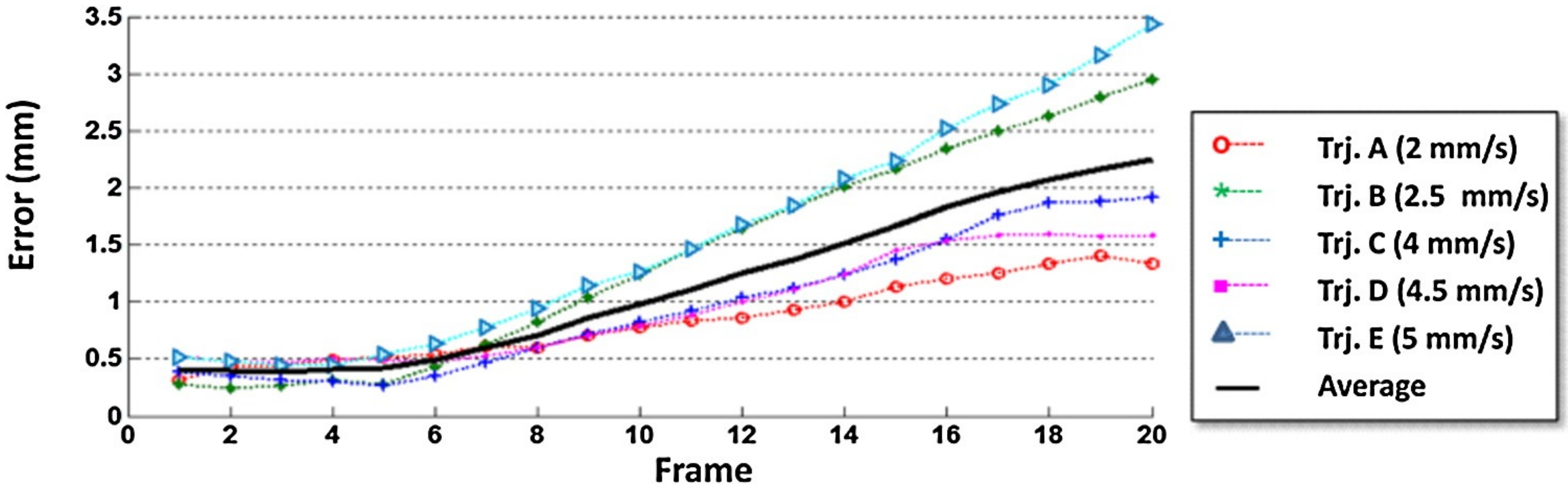


EPnP implementation. Pink surface = placental scene geometry;
textured patch = camera views projected onto constructed surface model.

Overview of workflow

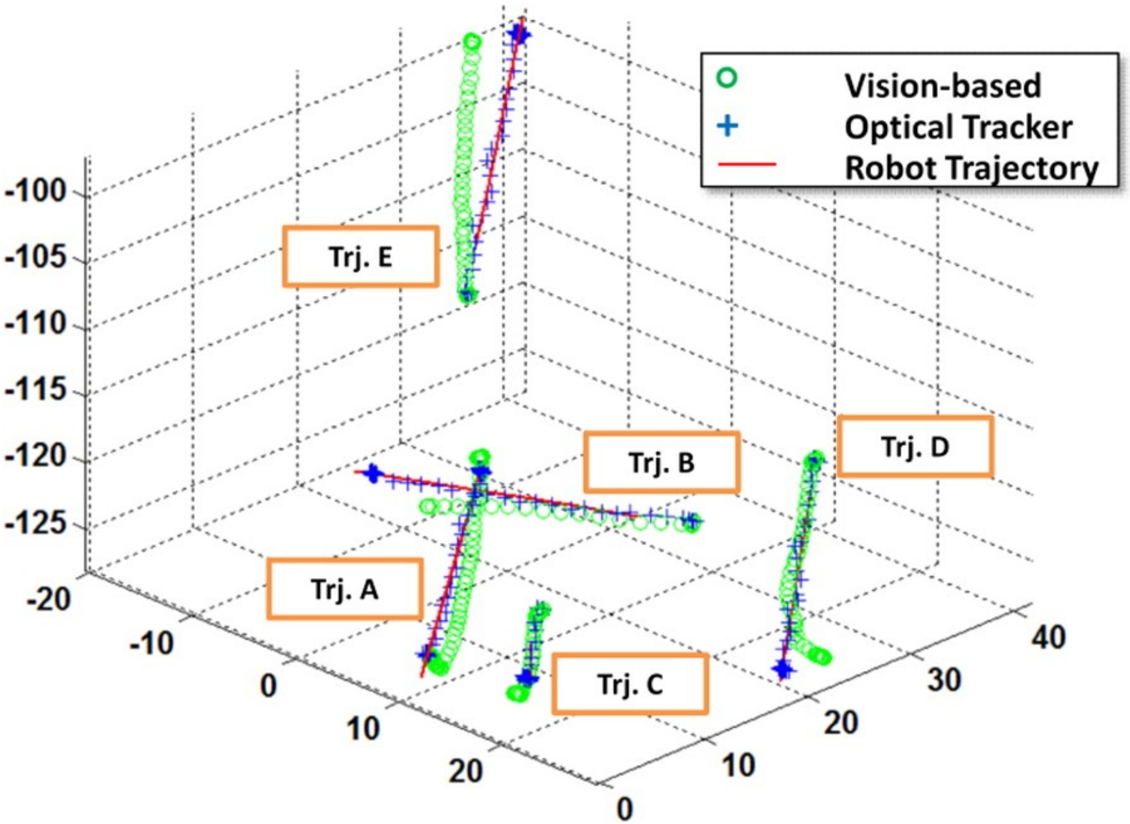


Results: phantom study, controlled trajectory



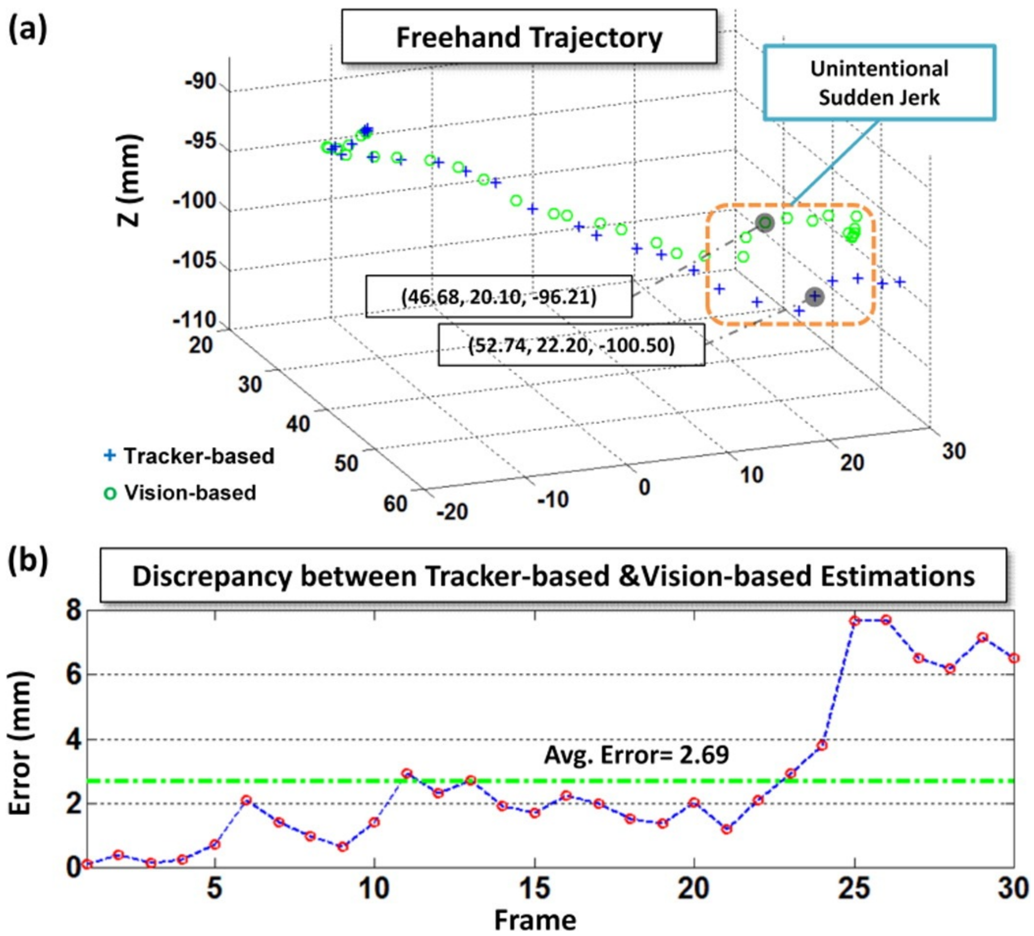
Each processing frame was at an interval of 10 acquisition frames (approx. 7s). Total displacement = 15-25 mm.

Results: phantom study, controlled trajectory



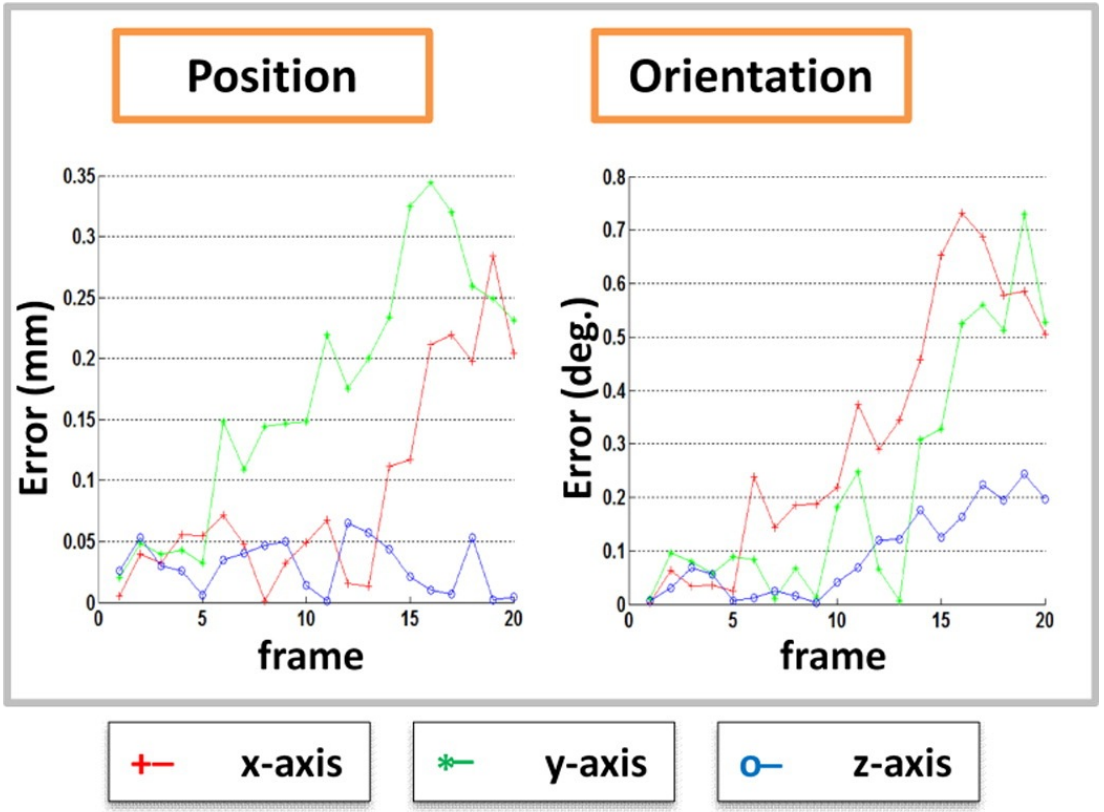
Each processing frame was at an interval of 10 acquisition frames (approx. 7s). Total displacement = 15-25 mm.

Results: phantom study, freehand trajectory



In the trajectory of approx. 30 mm, the mean absolute error was 2.69 mm in the 30 processed frames (300 acquired frames in 10 s).

Results: *ex vivo* study, static estimation

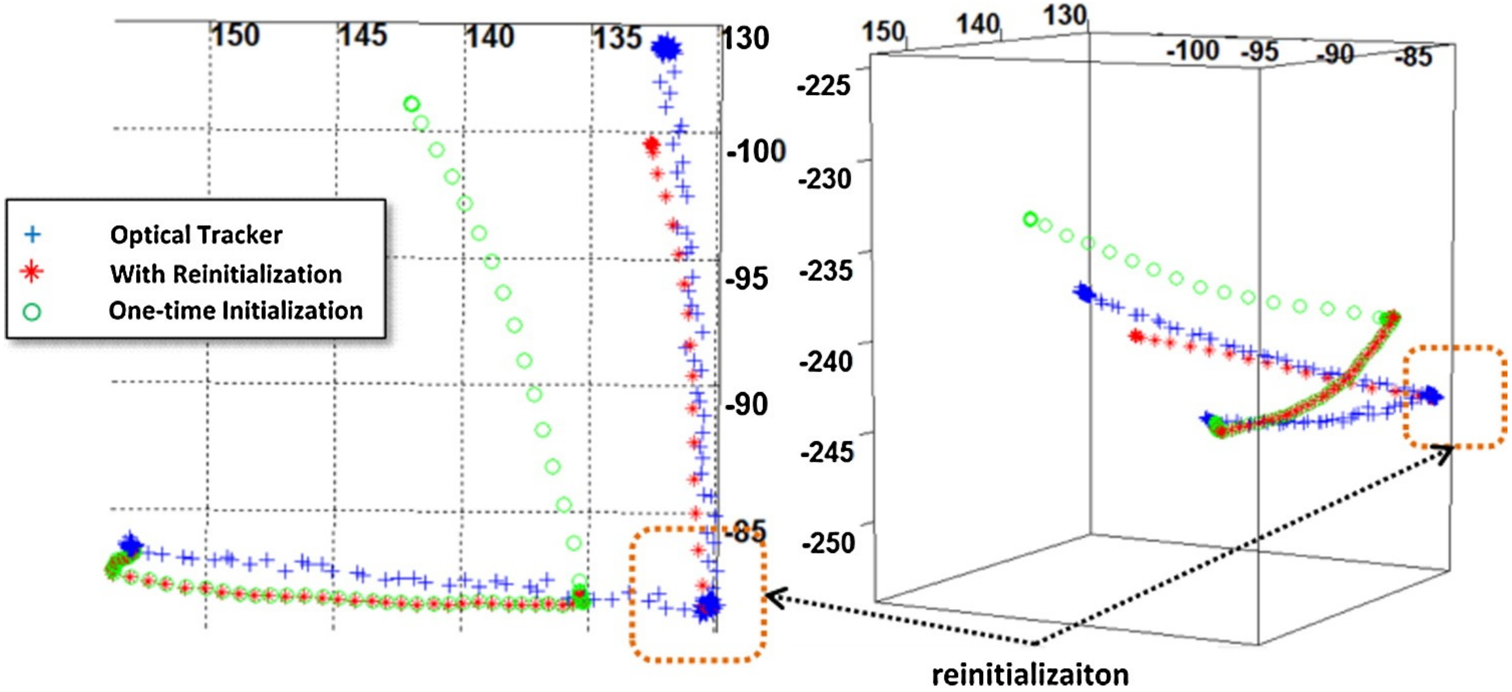


Position estimation error for ex vivo placenta imaging.

Axis	Monkey's placenta imaging			
	Orientation (Deg.)		Position (mm)	
	Mean	Standard deviation	Mean	Standard deviation
x	*0.317*	*0.089*	*0.10*	*0.11*
y	*0.225*	*0.251*	*0.18*	*0.11*
z	*0.095*	*0.246*	*0.03*	*0.03*

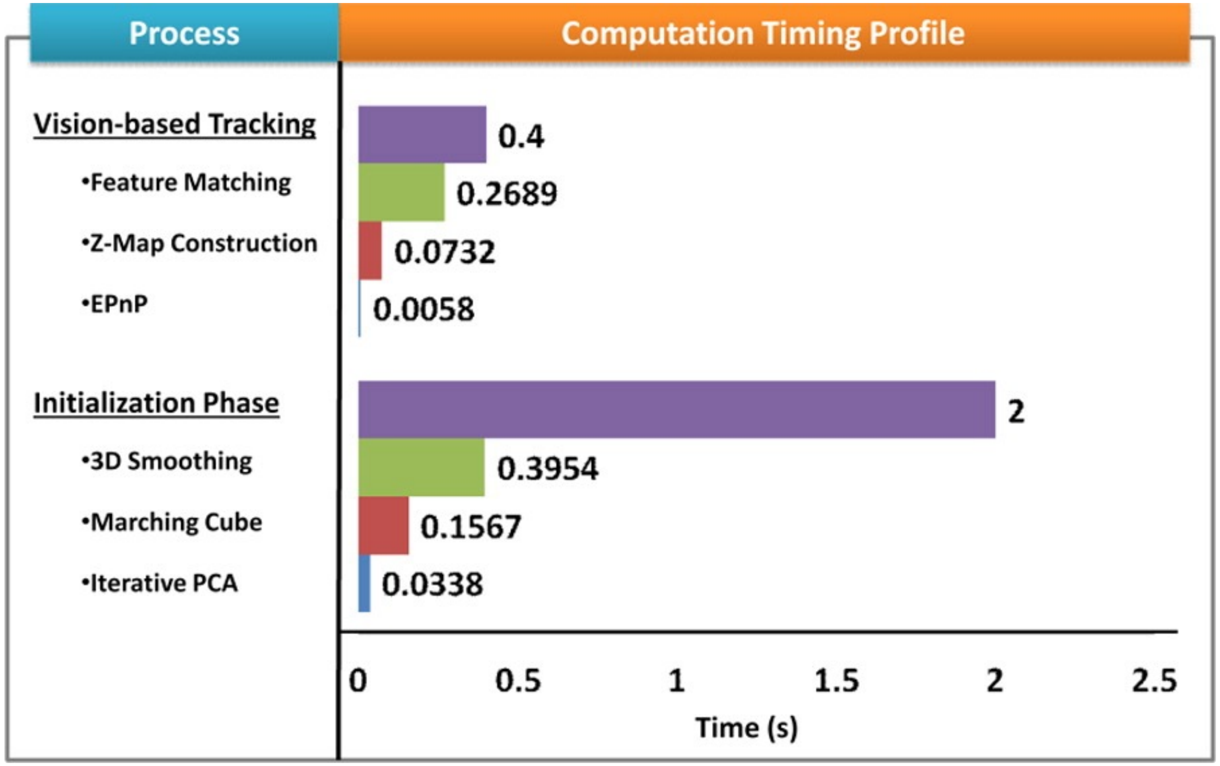
- Analysis of 100 estimations (5 positions x 20 frames)
- Validation against optical tracking, which has ~0.17 mm error
- Errors larger than phantom validation of static estimation

Results: effect of relocalisation (phantom)



- Ultrasound image-based delocalisation at 200th frame of 400-frame video
- Rectification of cumulative errors in vision-based tracking
- Final positional error reduced from 11.35 mm to 4.61 mm over total displacement of 45 mm

Results: computation time



- On a workstation with Intel Core i7-2600 3.4 GHz processor

Contributions / Future work

- Approach essentially vision-based, augmented with scene geometry information from ultrasound
- Relocalisation corrects cumulative errors or tracking failures
- Need to check performance under conditions closer to clinical setting (various kinematics, scene geometries, and illumination)
- Limitations in quality of endoscopic images can be addressed by:
  - fluorescence endoscope;
  - ultra-high sensitive endoscopic camera;
  - hyperspectral imaging of placental vasculature
- Ultrasound image artefacts lowered accuracy in the *ex vivo* study