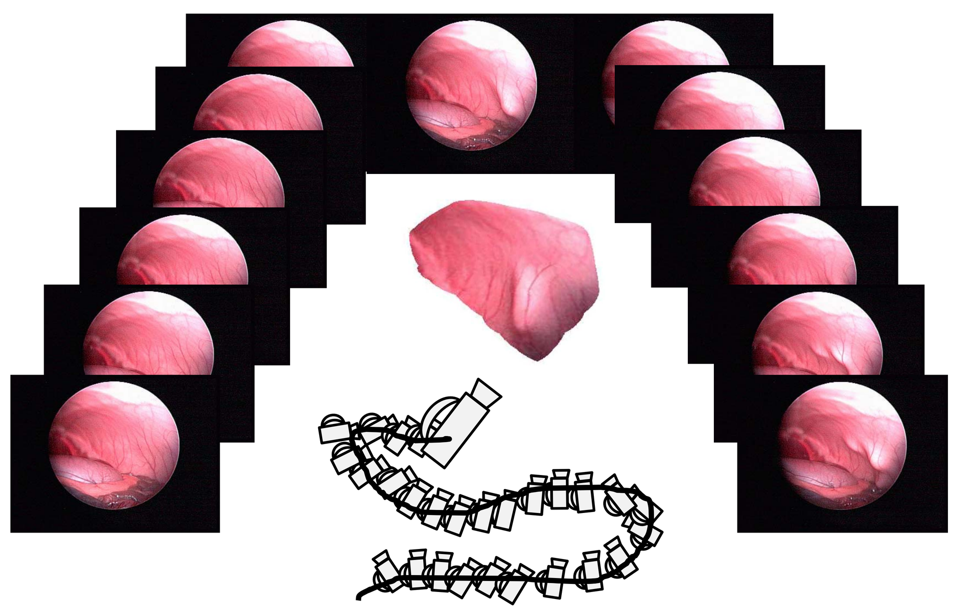
Total Estimation from RGB Video: On-line Camera Self-calibration, Non-Rigid Shape and Motion

ICPR 2020 25th International Conference on Pattern Recognition
ONLINE 10 | 15 January 2021

Antonio Aguado
Institut de Robòtica i Informàtica Industrial, CSIC-UPC, Barcelona, Spain
Institut de Robòtica i Informàtica Industrial
CSIC Consejo Superior de Investigaciones Científicas
UPC

Problem Statement



- Given an uncalibrated *monocular video* where both rigid and non-rigid scenes can be observed.
- We want to jointly and sequentially retrieve *the self-calibration* of the camera, the *3D non-rigid shape model* and the *full camera trajectory*.
- We propose a Bayesian filtering approach based on a sum-of-Gaussians filter composed of a bank of non-rigid extended Kalman filters. Neither training data nor a calibration pattern are needed.

Sum of Gaussian (SoG) Filter

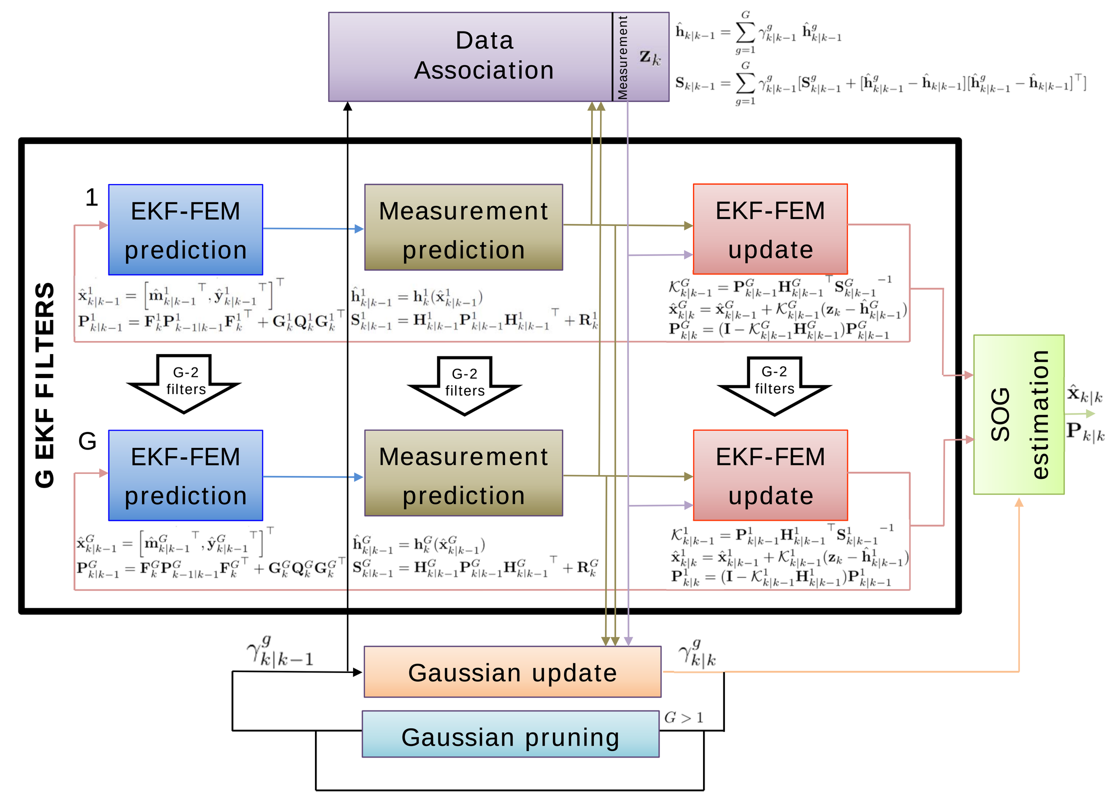
- In our SoG filter, we approximate a probability density function $p(\mathbf{x})$ as a combination of G weighted multivariate Gaussians as:

$$p(\mathbf{x}) = \sum_{g=1}^{G} \gamma^g \, \mathcal{N}(\mathbf{x}^g; \mathbf{P}^g)$$

- The SoG filter exploits a bank of Extended Kalman Filters (EKF). Every of them, it is estimated by means of a prediction-update strategy.
- The weight coefficients are updated every frame, removing those with low factor $\gamma^g_{k|k}$.
- An overall mean $\hat{\mathbf{x}}_{k|k}$ and covariance $\mathbf{P}_{k|k}$ for the SoG filter can be considered for visualization:

$$\hat{\mathbf{x}}_{k|k} = \sum_{g=1}^{G} \gamma^g_{k|k} \, \hat{\mathbf{x}}^g_{k|k}$$

$$\mathbf{P}_{k|k} = \sum_{g=1}^{G} \gamma^g_{k|k} \, [\mathbf{P}^g_{k|k} + [\hat{\mathbf{x}}^g_{k|k} - \hat{\mathbf{x}}_{k|k}][\hat{\mathbf{x}}^g_{k|k} - \hat{\mathbf{x}}_{k|k}]^\top]$$



Meth. \ Feat.	Tracking	Self-Calibration: Focal	Self-Calibration: Full	Process: Batch	Process: Sequential	Shape: Rigid	Shape: Non-Rigid Isometric	Shape: Non-Rigid Elastic
[12, 34]	✓	✓			✓	✓		
[16]	✓		✓		✓	✓		
[19, 25, 38]				✓		✓	✓	✓
[2, 32]					✓	✓	✓	✓
[8]	✓				✓	✓	✓	✓
[14, 33]	✓	✓		✓		✓	✓	
Ours	✓		✓		✓	✓	✓	✓

Self-Calibration Non-Rigid SOG

- The state of the camera is represented by a 18-dimensional vector, including calibration $(\alpha, \beta_x, \beta_y, k_1, k_2)$, camera pose $\mathbf{r}$ and orientation $\mathbf{q}$, and linear $\mathbf{v}$ and angular $\omega^{\mathcal{C}}$ velocities:

$$\mathbf{m}_{k+1} = \begin{bmatrix} \alpha_{k+1} \\ \beta_{x_{k+1}} \\ \beta_{y_{k+1}} \\ k_{1_{k+1}} \\ k_{2_{k+1}} \\ \mathbf{r}_{k+1} \\ \mathbf{q}_{k+1} \\ \mathbf{v}_{k+1} \\ \omega^{\mathcal{C}}_{k+1} \end{bmatrix} = \begin{bmatrix} \alpha_k \\ \beta_{x_k} \\ \beta_{y_k} \\ k_{1_k} \\ k_{2_k} \\ \mathbf{r}_k + (\mathbf{v}_k + \Delta\mathbf{v})\Delta t \\ \mathbf{q}_k \times \mathbf{q}((\omega^{\mathcal{C}}_k + \Delta\omega^{\mathcal{C}})\Delta t) \\ \mathbf{v}_k + \Delta\mathbf{v} \\ \omega^{\mathcal{C}}_k + \Delta\omega^{\mathcal{C}} \end{bmatrix}$$

- The state of the scene is represented by a 3n-dimensional vector, with n the number of points.
- We use an elastic model based on finite elements, defining a compliance matrix $\mathbf{C}_k$ to relate the deformation of all points, and a vector of Gaussian acting forces $\Delta\mathbf{f}$. For every frame, the state function is:

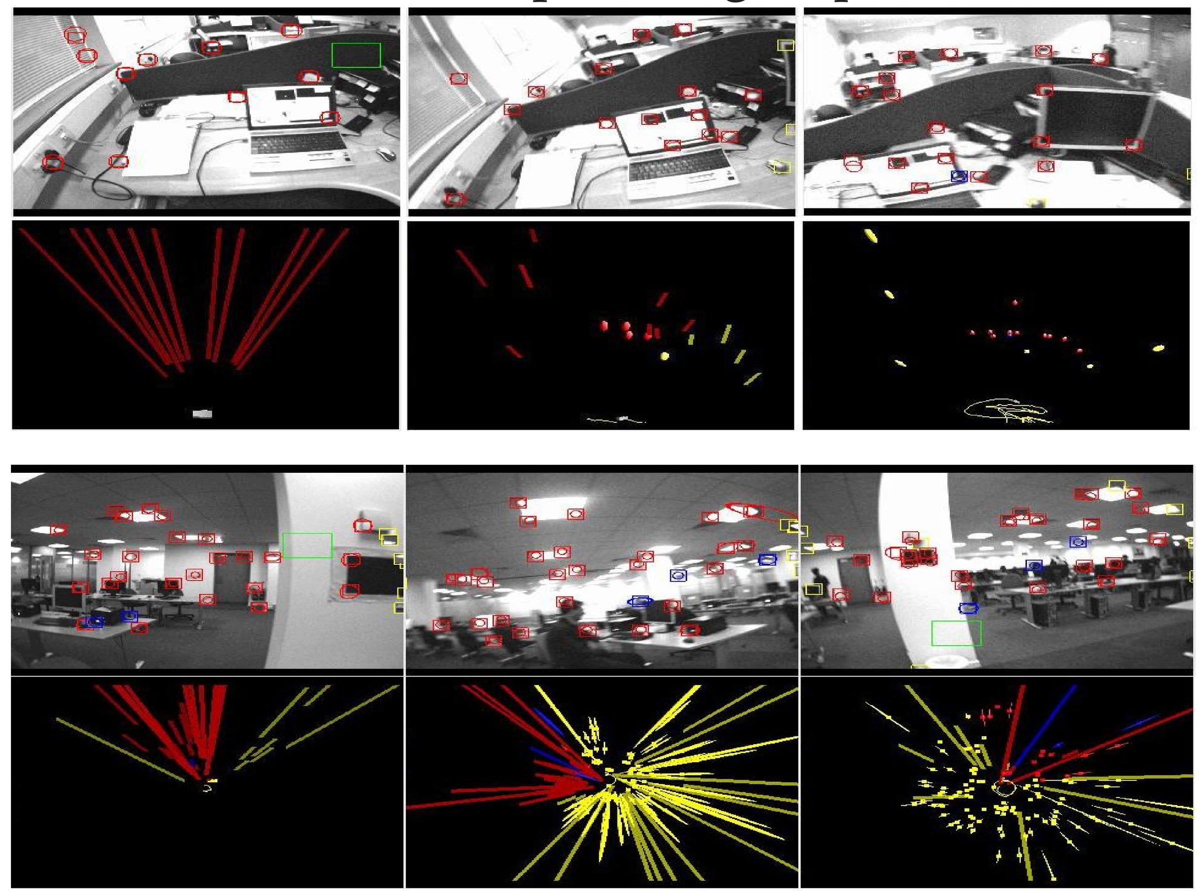
$$\mathbf{y}_{k+1} \equiv \mathbf{y}_{k+1}(\mathbf{y}_k, \Delta\mathbf{f}) = \mathbf{y}_k + \mathbf{C}_k \Delta\mathbf{f}$$

- Our model can handle both inelastic and elastic materials, and it is computed every frame.
- A full perspective camera model is assumed.

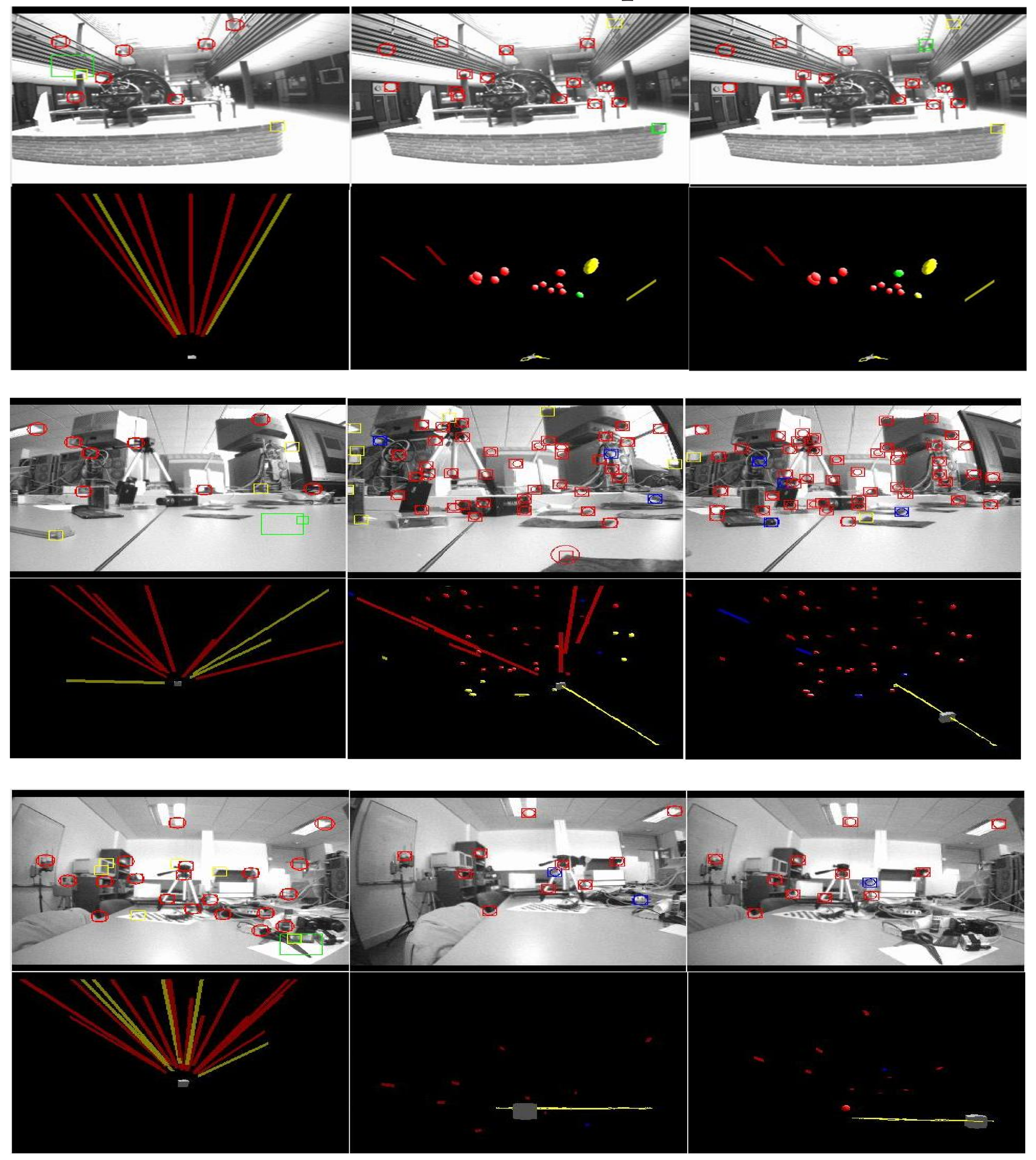
Calibration, Motion and Non-Rigid Shape Reconstruction

Param. \ Data	Non-Critical Motion Sequences: Indoor Off-line	Indoor On-line (Ours)	Loop Closing Off-line	Loop Closing On-line (Ours)	Silicone Cloth Off-line	Silicone Cloth On-line (Ours)	Laparoscopy Off-line	Laparoscopy On-line (Ours)	Critical Motion Sequences: Pure Rotation Off-line	Pure Rotation On-line (Ours)	Pure Translation Off-line	Pure Translation On-line (Ours)	Parallel Optical Axis Off-line	Parallel Optical Axis On-line (Ours)
α[pixels]	194.10	195.24±1.27	196.90	196.97±0.53	312.89	309.30±0.30	280.91	274.36±0.32	194.10	211.65±12.90	194.10	204.14±4.44	194.10	202.84±8.31
β_x[pixels]	160.20	158.94±0.92	153.50	159.14±1.41	157.66	158.60±0.11	184.48	166.00±0.18	160.20	158.68±6.07	160.20	156.47±3.48	160.20	158.89±7.08
β_y[pixels]	128.90	128.85±0.99	130.80	131.22±1.19	121.32	119.21±0.11	133.48	136.06±0.17	128.90	121.48±6.92	128.90	129.14±3.13	128.90	116.51±5.98
k_1[mm^{-2}]	.0623	.0661±.0023	.0693	.0721±.0028	.0094	.0056±.0002	.0054	.0078±.0004	.0623	.0626±.0073	.0623	.0676±.0048	.0623	.0679±.0109
k_2[mm^{-4}]	.0139	.0122±.0008	.0109	.0107±.0007	.00011	.00036±.00003	.00026	.0004±.00004	.0139	.0098±.0024	.0139	.0088±.0015	.0139	.0121±.0032

Indoor and Loop Closing Sequences

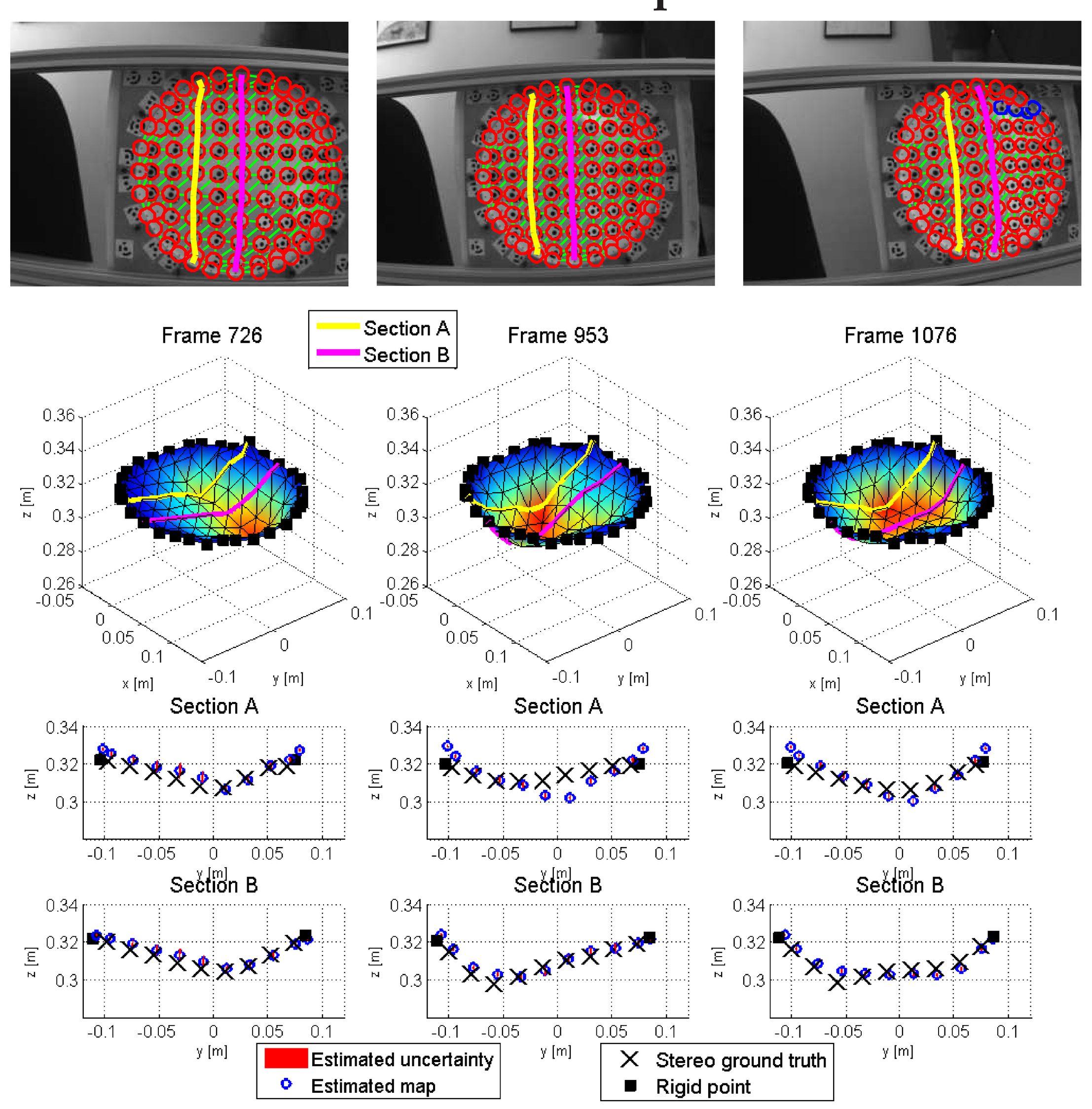


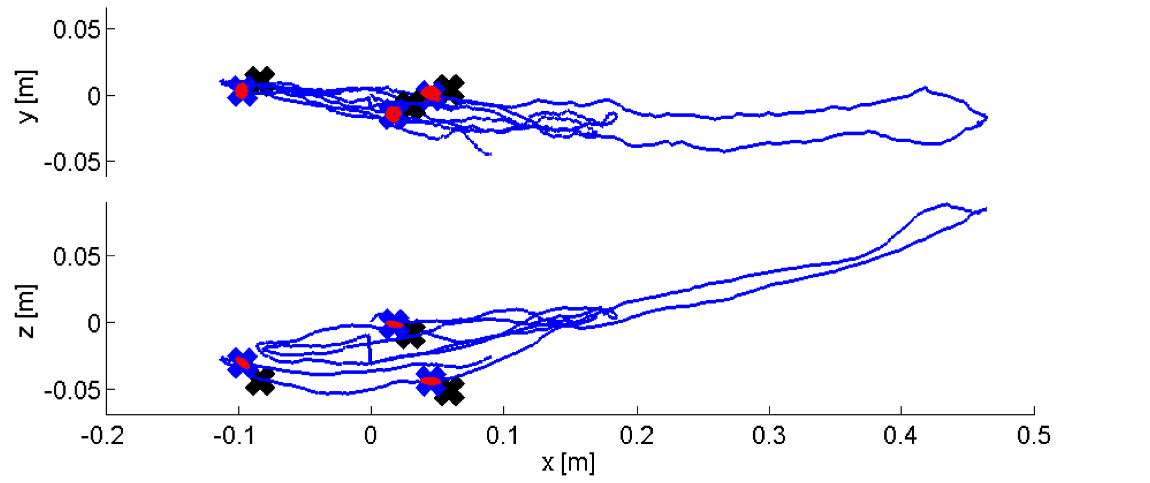
Critical Motion Sequences



- **Non-Critical Motion Scenarios**
  Hand-held 320 $\times$ 240 camera/endoscope.
  Rigid and elastic sequences.
- **Critical Motion Scenarios**
  Hand-held 320 $\times$ 240 IEEE1394 camera.
  Scene and/or calibration is not possible.

Silicone Cloth Sequence





Laparoscopy Sequence