ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
ONLINE 10 | 15 January 2021

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

Antonio Agudo

Institut de Robòtica i Informàtica industrial, Barcelona, Spain

Institut de Robòtica i Informàtica Industrial
CSIC
Consejo Superior de Investigaciones Científicas
UPC
Universitat Politècnica de Catalunya
BarcelonaTech

Shape and Motion from RGB Video

- **Input:** A monocular video.
- **Output:** 3D shape of an object, and full camera trajectory.

Type of Shape Structure

- **Rigid:** the 3D shape geometry is unknown but remains constant along time.
- **Non-Rigid:** the 3D shape geometry is dynamic, *i.e.*, the configuration can change along time. Many possibilities can be found:
  - **Isometric:** the global shape is deformable, but some local constraints can be enforced (distance between joints, etc.).
  - **Elastic:** the shape configuration can change by producing, bending- and stretching-like deformations.
  - Many others.

Rigid vs. Non-Rigid – SfM vs. SLAM

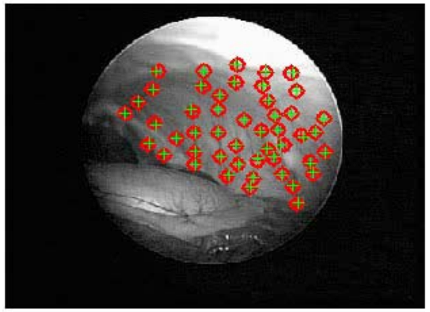
- Both problems are roughly the same.
- Rigid and non-rigid objects were previously considered in both frameworks.
- In general, SLAM and rigid SfM formulations exploit a perspective camera while NRSfM approaches use an orthographic one.
  - Calibration is known a priori and it is normally performed by using a calibration pattern.
- In this paper, *our goal is to recover the camera calibration from scratch, without assuming training data or calibration patterns.*

Our Approach for Self-Calibration

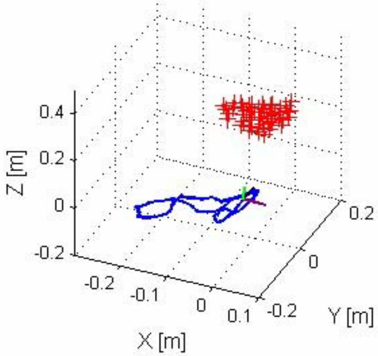
- **Input:** Just a monocular video.
- **Output:** Camera trajectory, self-calibration and the 3D shape geometry of a wide variety of objects (rigid, elastic, isometric, etc.). Our solution is sequential, as the data arrives.
- The self-calibration includes the parameters: focal length, principal point and two distortion parameters.
- Correspondences are computed automatically between consecutive images.
- No training data nor a calibration pattern are needed.

Motivation

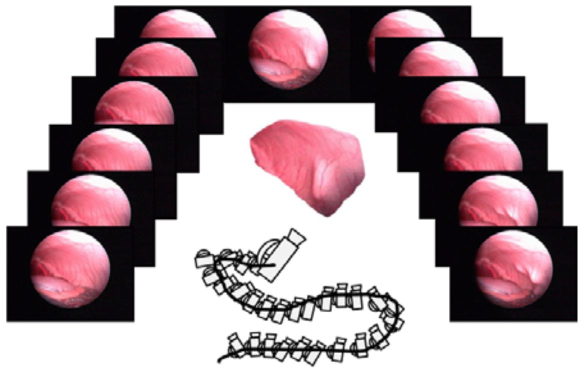
Image and tracked features



Estimated 3D scene and camera motion # 1200



Textured map



Qualitative Comparison

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

- Full calibration was considered for rigid shapes.

Qualitative Comparison

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

- Full calibration was considered for rigid shapes.
- For non-rigid ones (just isometric deformations), only the focal length was included and assuming a batch processing.

Qualitative Comparison

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

- Full calibration was considered for rigid shapes.
- For non-rigid ones (just isometric deformations), only the focal length was included and assuming a batch processing.
- *Our approach* is *the only one* that jointly retrieves the *3D reconstruction* (from rigid to non-rigid deformations), camera trajectory, and the full *self-calibration* in a *sequential* manner.

Self-Calibration Non-Rigid SoG

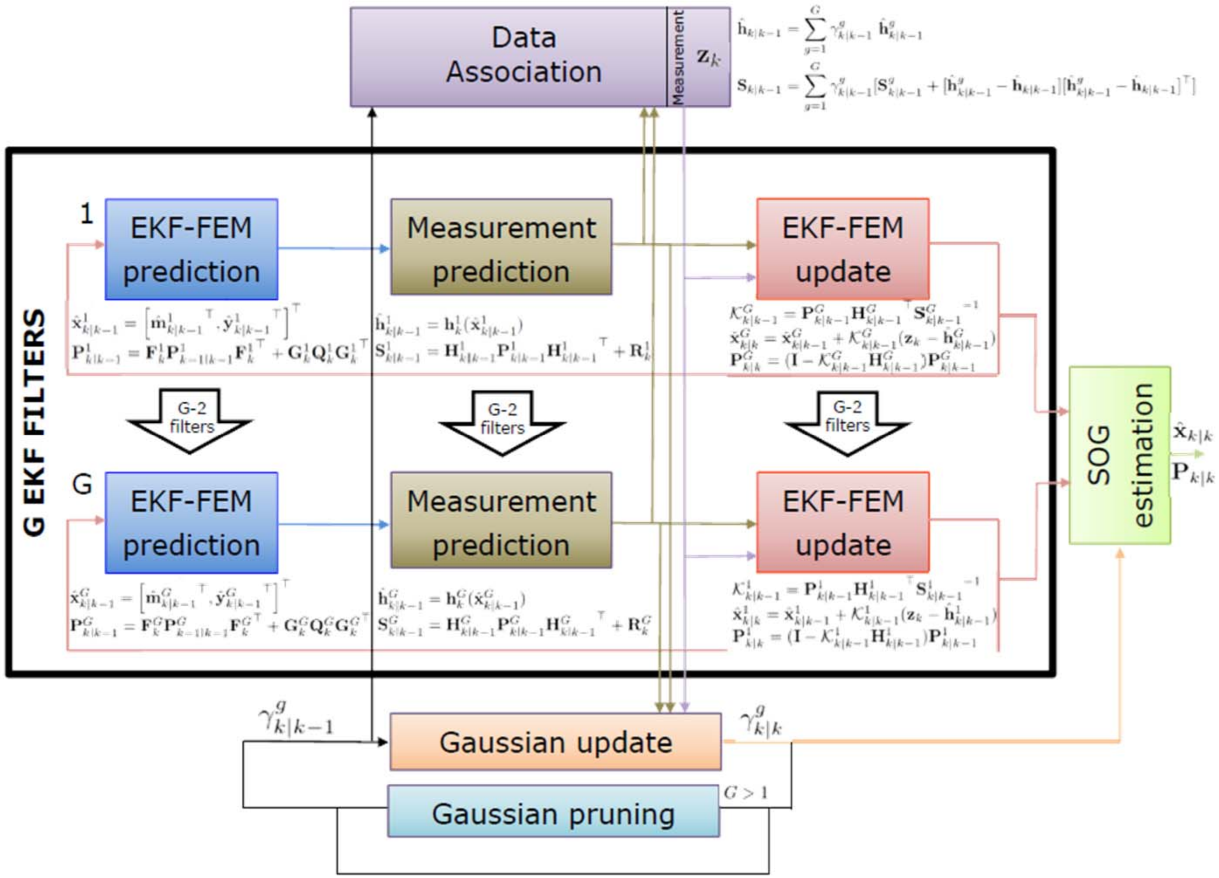
Self-Calibration Non-Rigid SoG

- Our approach consists in a Sum-of-Gaussians (SoG) filter to Gaussianly combine G multivariate distributions as:

$$p(\mathbf{x}) = \sum_{g=1}^{G} \underbrace{\gamma^g}_{\text{weight factor}} \mathcal{N}\left(\overbrace{\mathbf{x}^g}^{\text{mean}}; \underbrace{\mathbf{P}^g}_{\text{covariance}}\right)$$

- Every Gaussian distribution can come from an Extended Kalman Filter (EKF), where both mean and covariance are estimated by a prediction-update strategy.
- The SoG filter combines several EKF solutions (a filter bank), running in parallel.
- Shape deformations are encoded by a finite-element model.

Non-Rigid SoG Filter



Bank of EKF Filters

Problem Formulation

- The *state of the camera* is represented by a 18-dimensional vector, considering both intrinsic and extrinsic parameters:

$$\mathbf{m} = [\mathbf{c}^\top, \mathbf{r}^\top, \mathbf{q}^\top, \mathbf{v}^\top, \boldsymbol{\omega}^{\mathcal{C}^\top}]^\top$$

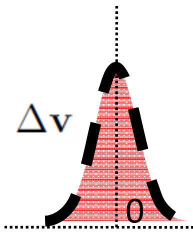
c includes the calibration, **r** and **q** denote the camera position and orientation, **v** and are the linear a$\boldsymbol{\omega}$ angular velocities.

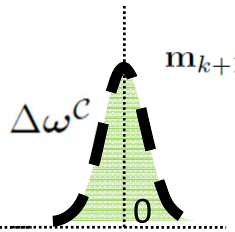
- The *state of the shape* is represented by a 3N-dimensional vector, considering N 3D points as $\mathbf{y} = [\mathbf{g}_1, \ldots, \mathbf{g}_N]^\top$ with $\mathbf{g}_n = [x_i, y_i, z_i]^\top$

- Given an image $\mathcal{I}_k$, the goal is to estimate the EKF vector $\mathbf{x} = [\mathbf{m}^\top, \mathbf{y}^\top]^\top$

Camera and Surface Motion Models

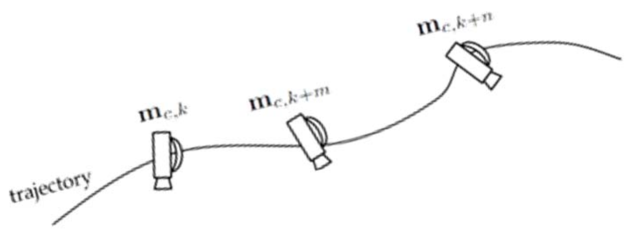
- The *camera state function* is represented by:





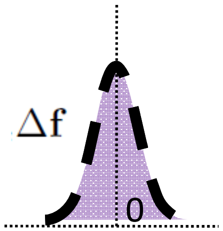
$$\mathbf{m}_{k+1} = \begin{bmatrix} \alpha_{k+1} \\ \beta_{x_{k+1}} \\ \beta_{y_{k+1}} \\ k_{1_{k+1}} \\ k_{2_{k+1}} \\ \mathbf{r}_{k+1} \\ \mathbf{q}_{k+1} \\ \mathbf{v}_{k+1} \\ \omega^{\mathcal{C}}_{k+1} \end{bmatrix} = \begin{bmatrix} \alpha_k \\ \beta_{x_k} \\ \beta_{y_k} \\ k_{1_k} \\ k_{2_k}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

Constant calibration model



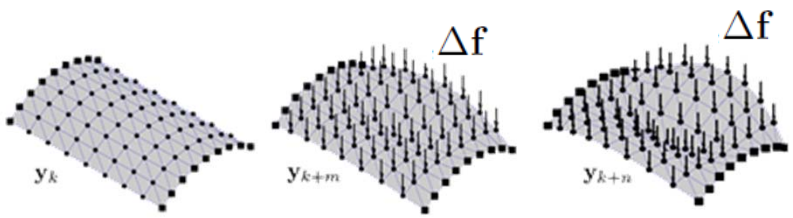
Constant velocity model

- The *shape state function* is represented by:



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

FEM model with
Gaussian forces



Gaussian Update and Pruning

- The contribution of every EKF filter is Gaussianly combined by means of a weight coefficient, that is updated as:

$$\gamma^g_{k|k} = \frac{\gamma^g_{k|k-1}\,\mathcal{N}(\mathbf{i}^g_{k|k-1};\mathbf{S}^g_{k|k-1})}{\sum_{g=1}^{G}\gamma^g_{k|k-1}\,\mathcal{N}(\mathbf{i}^g_{k|k-1};\mathbf{S}^g_{k|k-1})}$$

with $\mathbf{i}^g_{k|k-1} = \mathbf{z}_k - \mathbf{h}^g_k(\hat{\mathbf{x}}^g_{k|k-1})$ the innovation vector.

- The number of Gaussians is reduced by means of sequential probability ratio test removing those with a low weight factor.
- An overall solution can be considered (only for visualization):

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

Experimental Evaluation

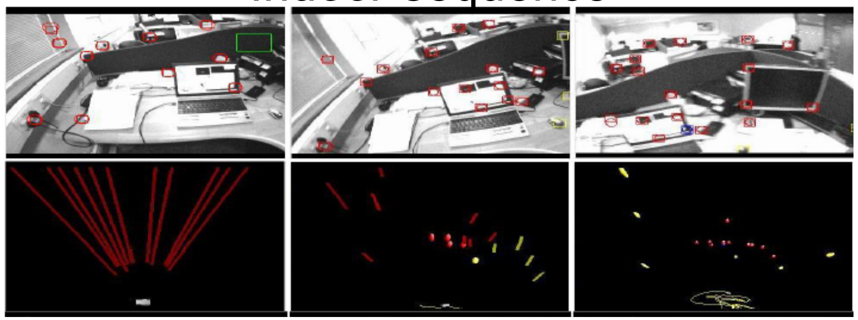
Experimental Results

- Two blocks of real experiments:
  - Non-Critical motion sequences.
    - *Rigid*. Indoor general scenarios with loop closing.
    - *Non-Rigid*. Rigid, elastic and quasi-isometric deformations are included.
  - Critical motion sequences.
    - *Rigid*. Shape and/or calibration are ill-posed.
- In total, seven hand-held camera videos are used.
- An offline calibration in employed to validate our estimation.

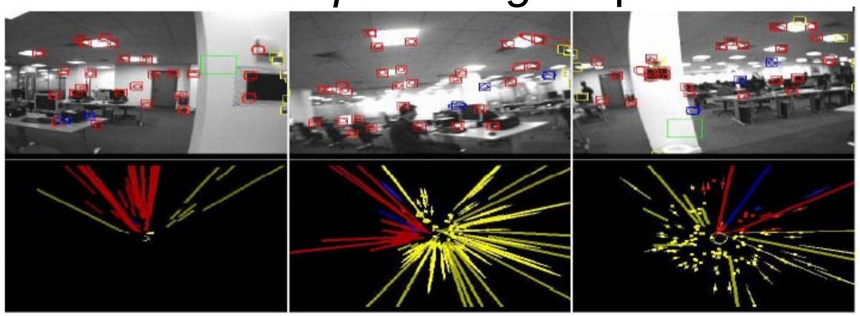
Non-Critical Motion Sequences

- The rigid case (general indoor conditions):
  - *Indoor* sequence: a RGB camera with smooth motion is observing a rigid scene.
  - *Loop closing* sequence: in this case, the camera follows a loop trajectory.

Indoor sequence



Loop Closing sequence

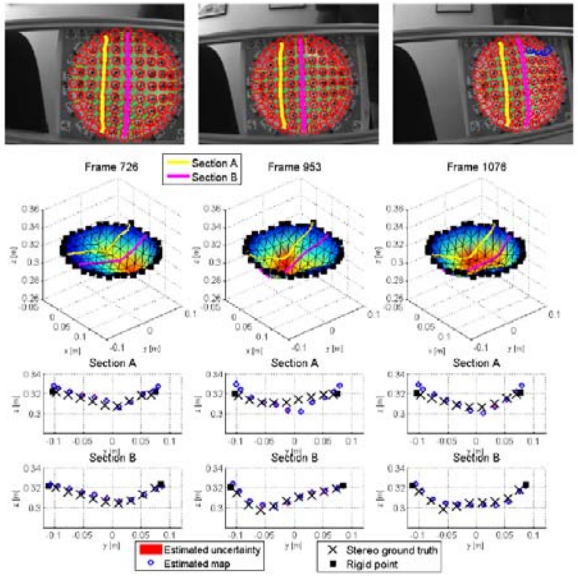


Non-Critical Motion Sequences

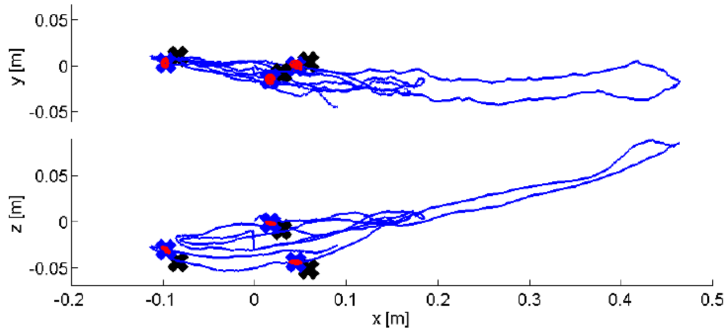
- The non-rigid case (synthetic and in-vivo materials):
  - *Silicone cloth* sequence: a RGB camera with smooth motion is observing a deformable silicone cloth.
  - *Laparoscopy* sequence: an endoscope is observing a rabbit abdominal cavity in a medical exploration.

Non-Critical Motion Sequences

- The non-rigid case (synthetic and in-vivo materials):
  - *Silicone cloth* sequence: a RGB camera with smooth motion is observing a deformable silicone cloth.

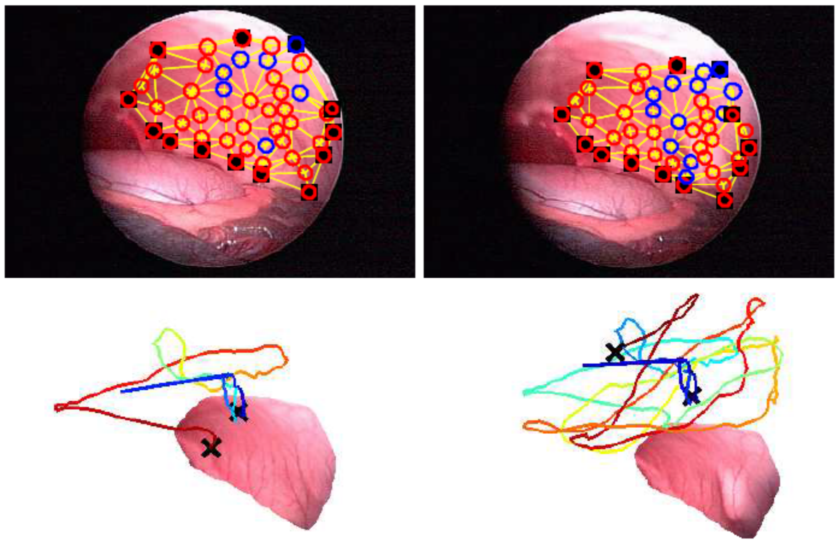


Mean 3D reconstruction error of 3.96 mm



Non-Critical Motion Sequences

- The non-rigid case (synthetic and in-vivo materials):
  - *Laparoscopy* sequence: an endoscope is observing a rabbit abdominal cavity in a medical exploration.



Critical Motion Sequences

- Shape and/or calibration is ill-posed:
  - *Pure rotation* sequence: a RGB camera with smooth motion is observing a rigid scene.
  - *Pure translation* sequence: in this case, the camera follows a loop trajectory.
  - *Parallel optical axis* sequence.

Calibration Results

Param. \ Data	Non-Critical Motion Sequences							
	Indoor		Loop Closing		Silicone Cloth		Laparoscopy	
	Off-line	On-line (Ours)	Off-line	On-line (Ours)	Off-line	On-line (Ours)	Off-line	On-line (Ours)
α[pixels]	194.10	195.24$\pm$1.27	196.90	196.97$\pm$0.53	312.89	309.30$\pm$0.30	280.91	274.36$\pm$0.32
β_x[pixels]	160.20	158.94$\pm$0.92	153.50	159.14$\pm$1.41	157.66	158.60$\pm$0.11	184.48	166.00$\pm$0.18
β_y[pixels]	128.90	128.85$\pm$0.99	130.80	131.22$\pm$1.19	121.32	119.21$\pm$0.11	133.48	136.06$\pm$0.17
k_1[mm^{-2}]	.0623	.0661$\pm$.0023	.0693	.0721$\pm$.0028	.0094	.0056$\pm$.0002	.0054	.0078$\pm$.0004
k_2[mm^{-4}]	.0139	.0122$\pm$.0008	.0109	.0107$\pm$.0007	.00011	.00036$\pm$.00003	.00026	.0004$\pm$.00004

- Our approach can sequentially estimate the auto-calibration in a wide variety of scenarios together with shape and motion:
  - The principal point is better estimated for cycle-torsion motions, as it is observed in the *Indoor* experiment.
  - In general, the two distortion parameter estimations are correlated, being difficult to recover everyone of them.
  - Competitive solutions with respect to ground truth.

Calibration Results

Data / Param.	Non-Critical Motion Sequences								Critical Motion Sequences					
	Indoor		Loop Closing		Silicone Cloth		Laparoscopy		Pure Rotation		Pure Translation		Parallel Optical Axis	
	Off-line	On-line (Ours)	Off-line	On-line (Ours)	Off-line	On-line (Ours)	Off-line	On-line (Ours)	Off-line	On-line (Ours)	Off-line	On-line (Ours)	Off-line	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
$k_1[\text{mm}^{-2}]$	.0623	.0661±.0023	.0693	.0721±.0028	.0094	.0056±.0002	.0054	.0078±.0004	.0623	.0626±.0073	.0623	.0676±.0048	.0623	.0679±.0109
$k_2[\text{mm}^{-4}]$	.0139	.0122±.0008	.0109	.0107±.0007	.00011	.00036±.00003	.00026	.0004±.00004	.0139	.0098±.0024	.0139	.0088±.0015	.0139	.0121±.0032

- As expected, for some motions we cannot recover a good solution. As no prior is assumed about the type of motion:
  - A general estimation is obtained.
  - When the shape and/or motion is incorrect, the calibration is anecdotic. The joint solution is bad (see *Pure Rotation*).
  - Some scenarios allow the estimation of shape and motion, but not the calibration (see *Parallel Optical Axis*).

Thank you so much!

Antonio Agudo

Institut de Robòtica i Informàtica industrial, Barcelona, Spain