Evaluation of pairwise calibration techniques for range cameras and their ability to detect a misalignment

Antoine Lejeune
<antoine.lejeune@ulg.ac.be>
David Grogna
Marc Vandroogenbroeck
Jacques Verly

Montefiore Institute, University of Liège

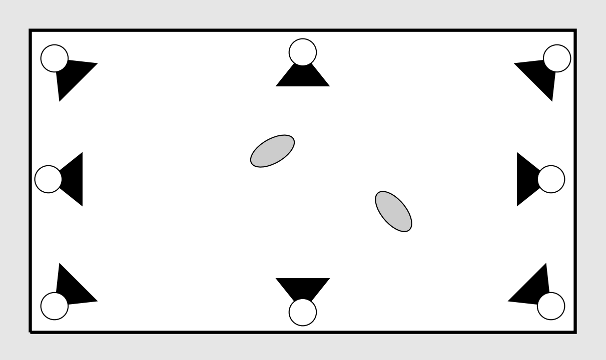
2014 International Conference on 3D Imaging (IC3D 2014)
9 December 2014

Multicamera systems

Pair-wise calibration is a building block of multicamera systems. They provide

- better coverage of large volume;
- multiple point of view of the scene; and
- can increase precision and robustness.

Examples of application: immersive virtual environment, gait analysis of humans, ...

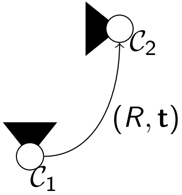


Pairwise calibration

- A multicamera system has to be calibrated, i.e. we need to estimate the relative position between couples of camera.
- For color images, we estimate the intrinsic parameters of each camera and the fundamental matrix between pairs of camera.
- In 3D, we wish to find the rigid body transformation (R, t) that brings points of one camera to the reference coordinate frame of the second camera

$$\mathbf{P}^{(2)} = R\mathbf{P}^{(1)} + t$$

where R is a rotation, t a translation vector and $\mathbf{P}^{(i)}$ denotes the coordinates of the 3D point $\mathbf{P}$ as seen from camera i.



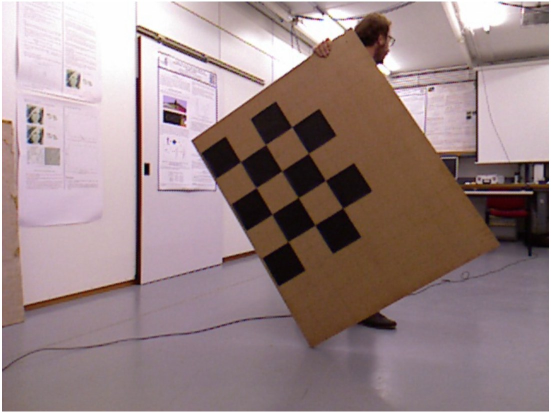
Range cameras

- Directly measure a geometric information
- Different technologies:
  - Structured light: Microsoft Kinect (version 1)
  - Time-of-flight: PMD CamCube 2.0, Microsoft Kinect (version 2)
- Nonlinear noise that can vary across the pixels of the image
  - no data at all for some parts of the image
  - there are models of the noise (depending on the technology of the camera)
- There can be problems when naively combining several range cameras with overlapping field of views

Microsoft Kinect (version 1) PMD CamCube 2.0

Classical technique

- Using a two-sided chessboard to perform a color calibration
- Minimize the reprojection error
- OpenCV implementation

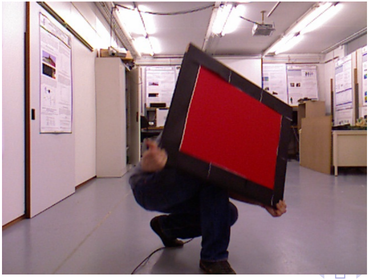


Plane pattern

Depth-based calibration using a plane:

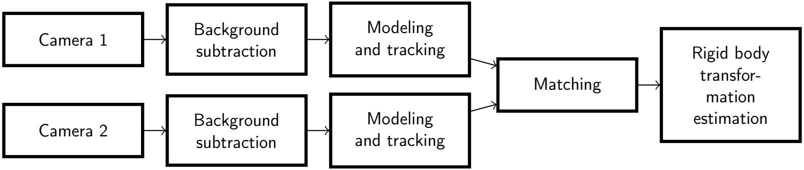
- Plane segmentation can be done in the RGB space or in the depth space
- Point correspondences between the camera are established using the center of the plane
- Rigid body transformation estimated by the least-square minimization of

$$\sum_i w_i \left\| \mathbf{P}^{(2)} - R\mathbf{P}^{(1)} - t \right\|^2 .$$



Movement based calibration

- Pairwise calibration using the movement in the scene
- No crafted calibration object
- Permit to detect a misalignment and recalibrate the system when one occurs
- Processing pipeline:



Background subtraction

- Simple background model learned over the first N_{BG} frames:

$$B_Z(\mathbf{p}) = \max_j (Z_j(\mathbf{p})), \quad j < N_{BG}.$$

- Foreground segmentation based on the estimated noise $\sigma(\mathbf{p})$ at each pixel:

$$F(\mathbf{p}) = \begin{cases} \text{true} & \text{if } Z(\mathbf{p}) \text{ is valid and} \\ & |B_Z(\mathbf{p}) - Z(\mathbf{p})| > \lambda\sigma(\mathbf{p}) \\ \text{false} & \text{otherwise} \end{cases}$$

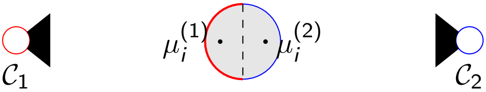
  with $\sigma_{kinect}(\mathbf{p}) = (Z(\mathbf{p}))^2$ and $\sigma_{tof}(\mathbf{p}) = (A(\mathbf{p}))^{-1}$.

- Connected component analysis to filter out small components

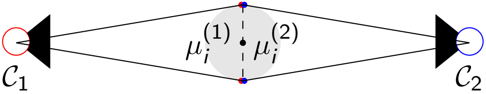
Modeling

- We model the segmented objects $\mathcal{O}_i$ as gaussian blobs with a center of mass μ_i and a covariance matrix Σ_i.
  - Center of mass computed using only the border pixels

With all pixels:

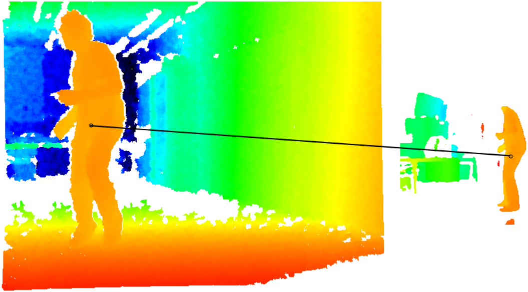


With border pixels:



Tracking and matching

- Tracking within a single camera is based on the center of mass μ_i and the covariance Σ_i
- Matching between cameras is performed by only using the covariance Σ_i
- We use the Kullback-Leibler divergence as a similarity measure



Misalignment detection

- A "ground truth" pairwise calibration is previously obtained: (R_{GT}, t_{GT})
- We estimate regularly the current transformation using the method based on movement: (R_t, t_t)
- A misalignment is detected when

$$t_{err} = \|t_{GT} - t_t\| > \tau$$

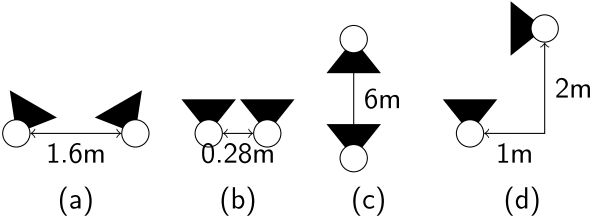
or

$$R_{err} = \left\|\log\left(R_{GT}^T R_t\right)\right\|_F > \theta,$$

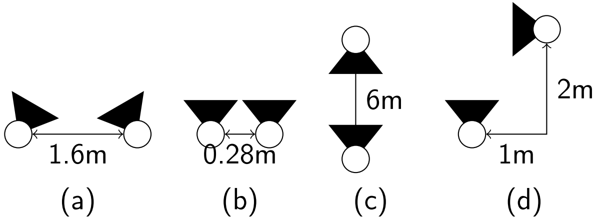
i.e. when the translational error or the angular error are above some threshold.

Evaluation

- Groundtruth (R_{GT}, t_{GT}) is computed using the chessboard-based method.
- Evaluation metrics
  - Translational error:
$$t_{err} = \|t_{GT} - t\|$$
  - Angular error:
$$R_{err} = \left\|\log\left(R_{GT}^T R\right)\right\|_F \in [0; 90°]$$
- 4 spatial configurations and 2 sets of cameras (Kinect-Kinect and Kinect-CamCube) tested

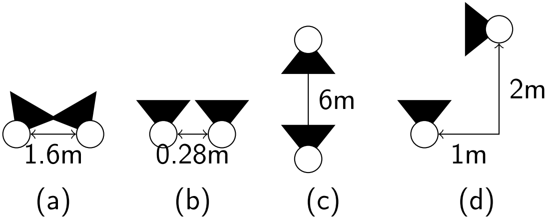


Comparison (Kinect-Kinect)



Configuration	(a)	(b)	(c)	(d)
	Translation error (in meter)			
Chessboard (GT)	0	0	0	0
Plane	0.094	0.057	0.047	0.155
Movement	0.076	0.069	0.128	0.189
	Angular error (in degree)			
Chessboard (GT)	0	0	0	0
Plane	2.59	1.34	0.45	5.53
Movement	2.49	1.37	1.96	3.58

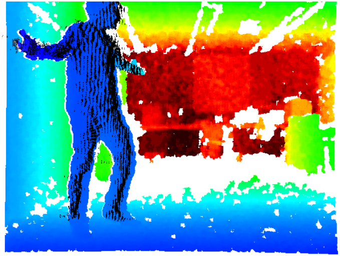
Comparison (Kinect-Camcube)



Configuration	(a)	(b)	(c)	(d)
	Translation error (in meter)			
Chessboard (GT)	0	0	0	0
Plane	0.176	0.042	0.044	0.064
Movement	0.10	0.223	0.169	0.339
	Angular error (in degree)			
Chessboard (GT)	0	0	0	0
Plane	4.42	4.94	0.54	1.61
Movement	3.46	2.88	1.95	11.28

Conclusion

- Techniques based on range values don't reach the same level of precision as state of the art pairwise calibration technique for color images
- However,
  - they can provide a good approximation in some cases
  - thay are easier to set-up
  - movement based calibration permits to detect a misalignment and can offer a temporary calibration when it happens



Questions

Thank you for listening,

any questions?