Structure from Motion Using Rigidly Coupled Cameras without Overlapping Views

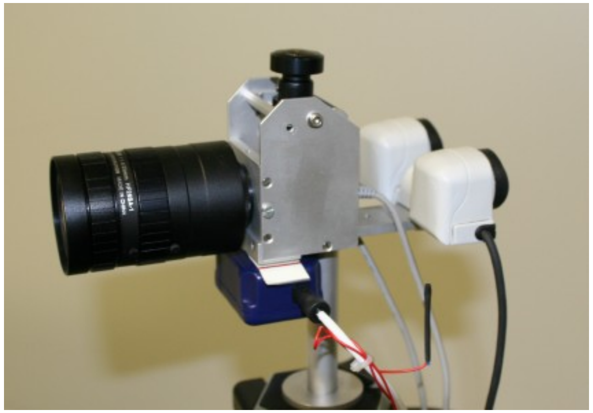
Dipl. Inf. Sandro Esquivel
Prof. Dr.-Ing. Reinhard Koch

Multimedia Information Processing
Christian-Albrechts-University, Kiel

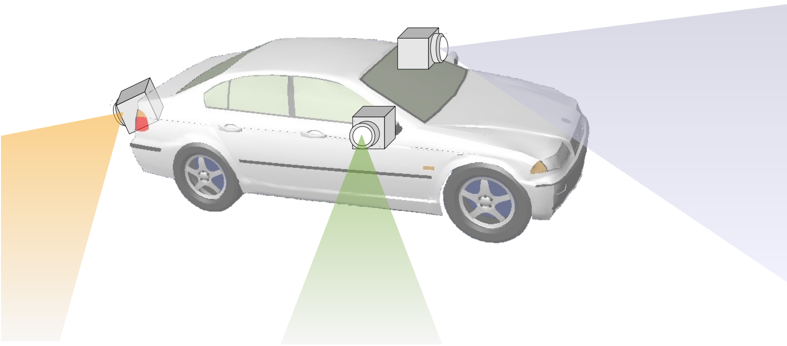
Motivation

- Various applications for camera systems with non-overlapping views
  - Omnidirectional imaging systems
  - See-through Augmented Reality devices
  - Advanced Driver Assistance Systems
- Computer vision benefits from large combined field of view
- Extrinsic calibration is complicated!

Point Grey Ladybug 5



Augmented Reality Binocular

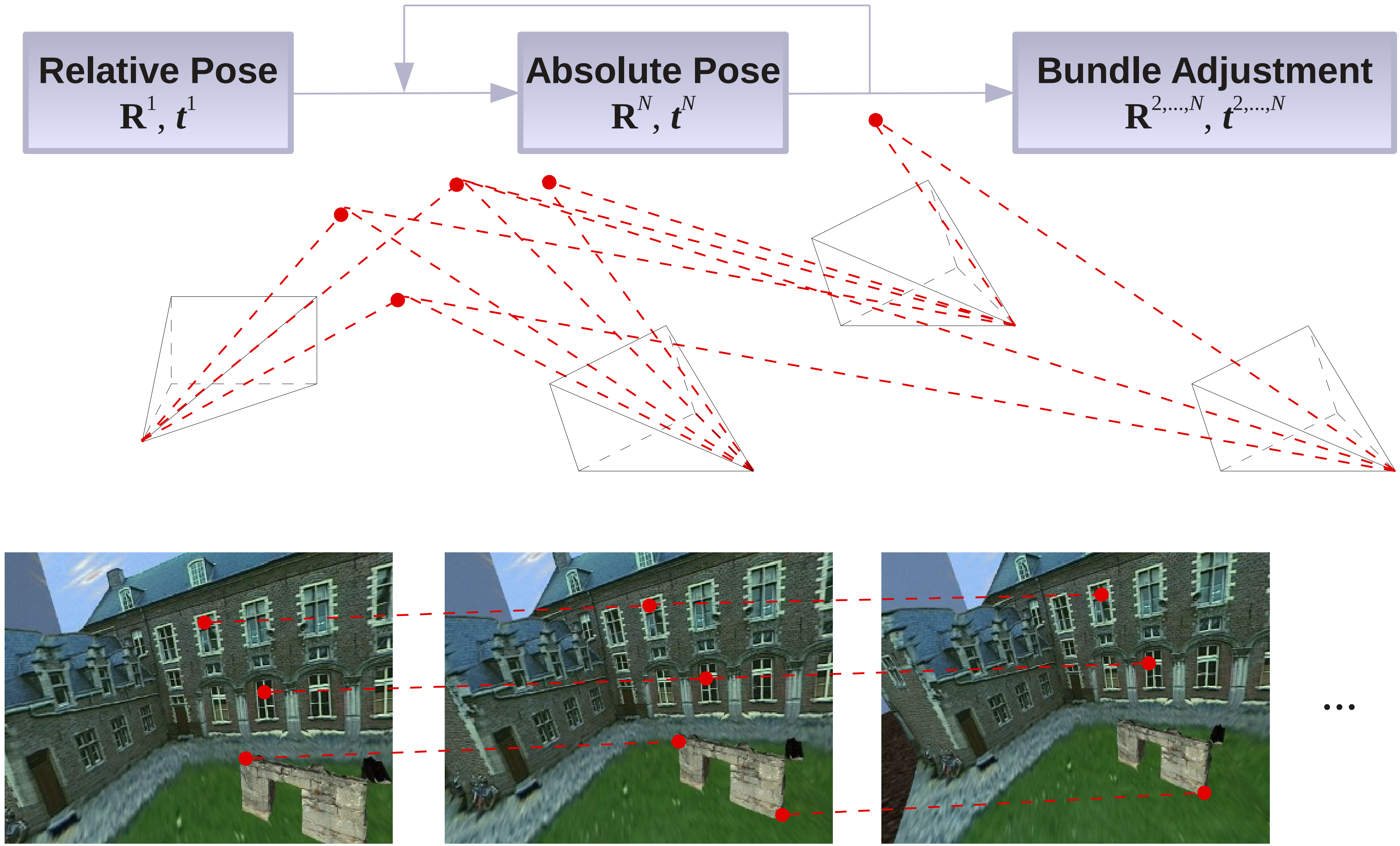


Vehicle-mounted cameras

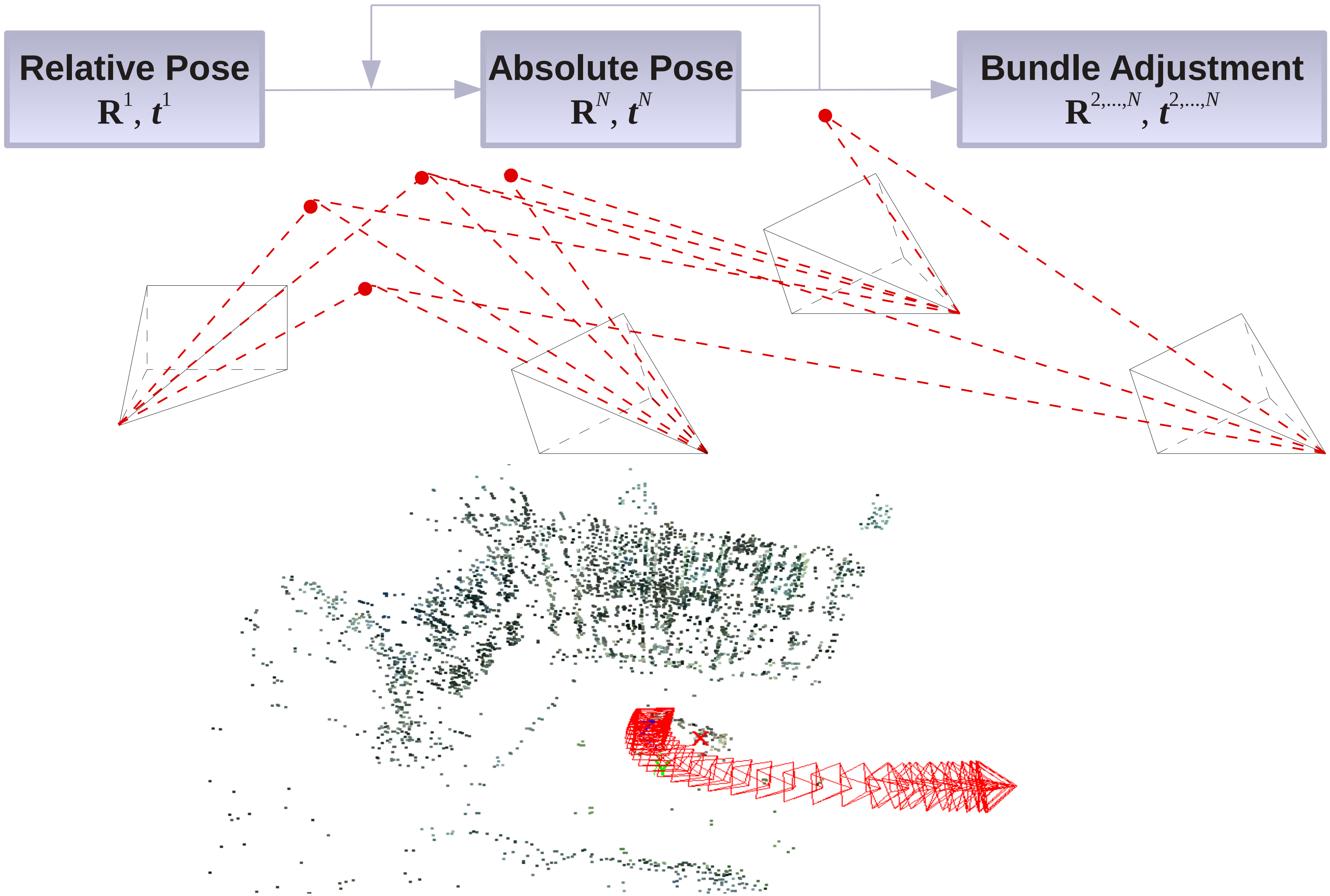
Outline

- Motivation
- Multi-Camera Structure from Motion
- Enforcing Rigid Motion Constraints
- Tests and Evaluation
- Conclusion

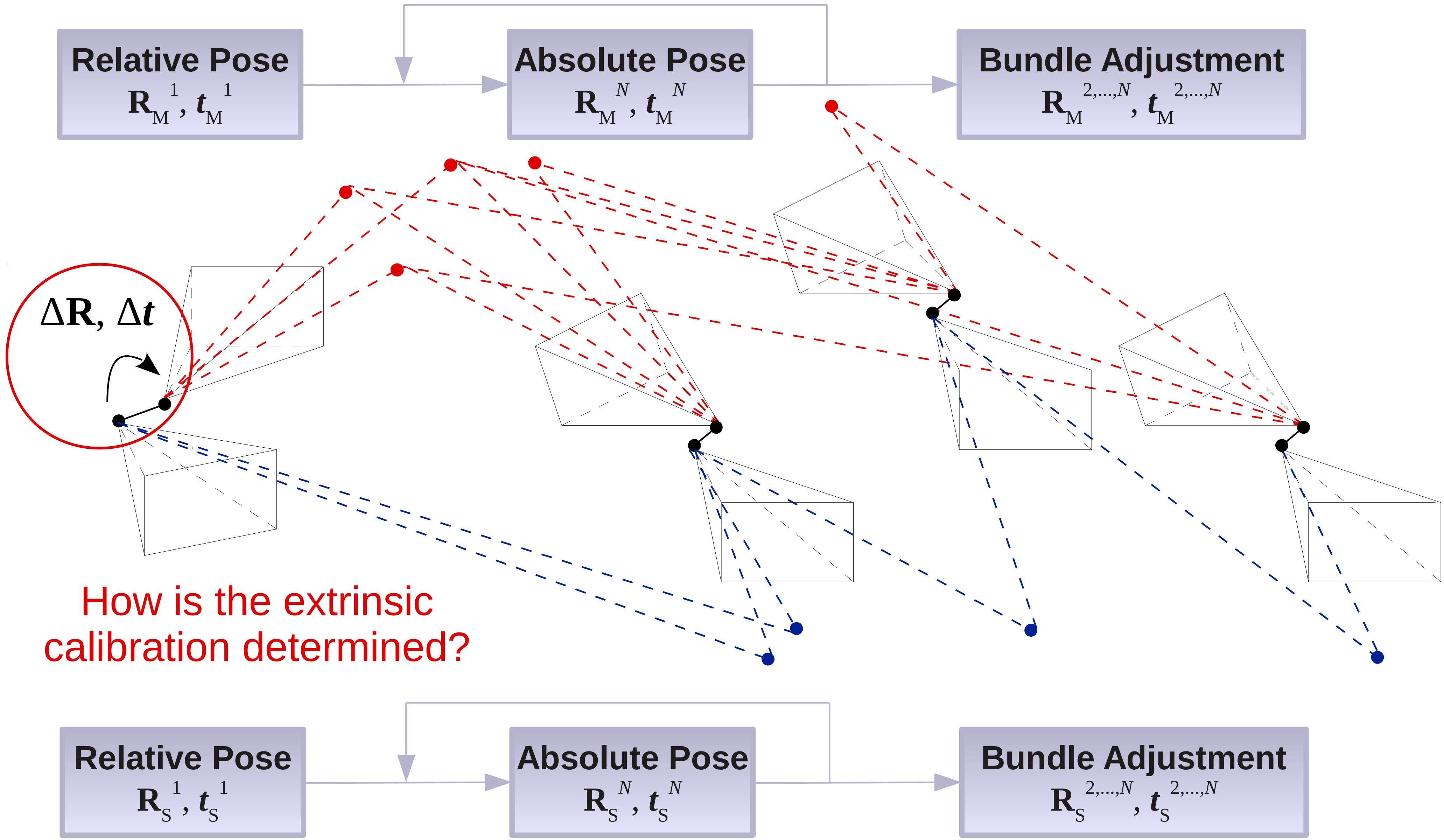
Structure from Motion



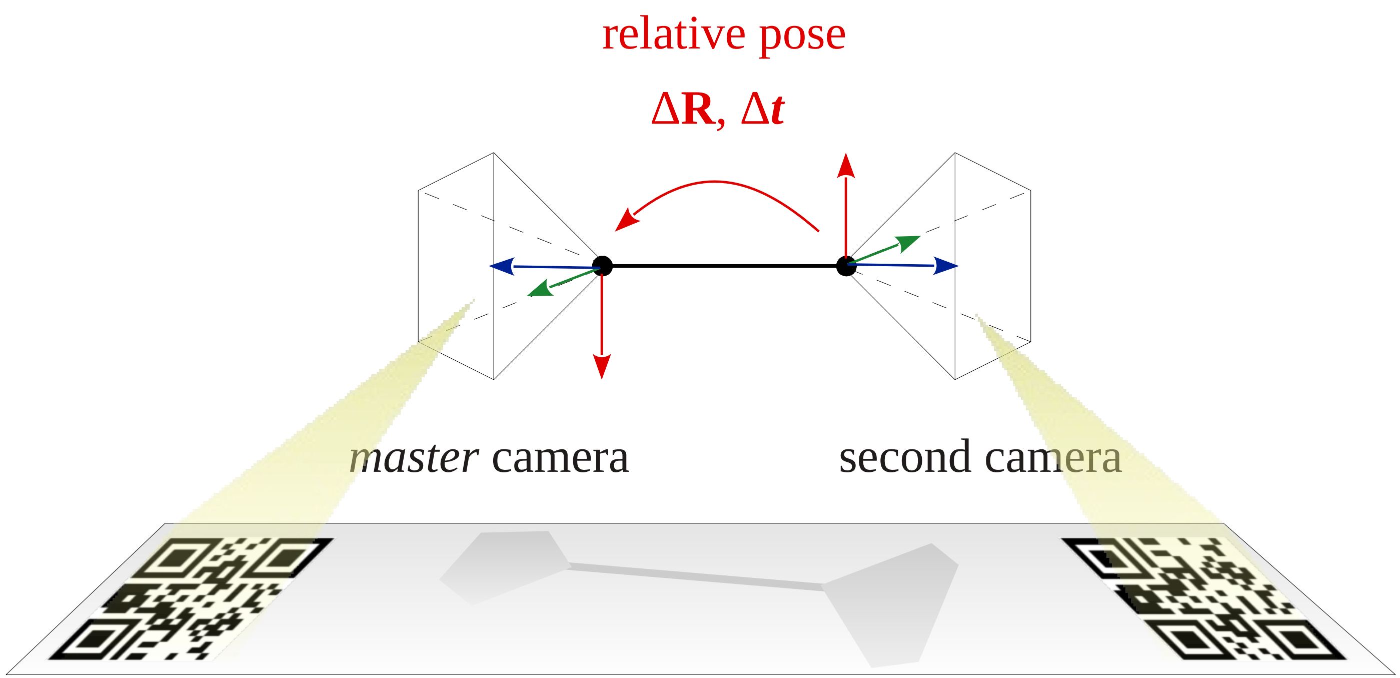
Structure from Motion



Structure from Motion

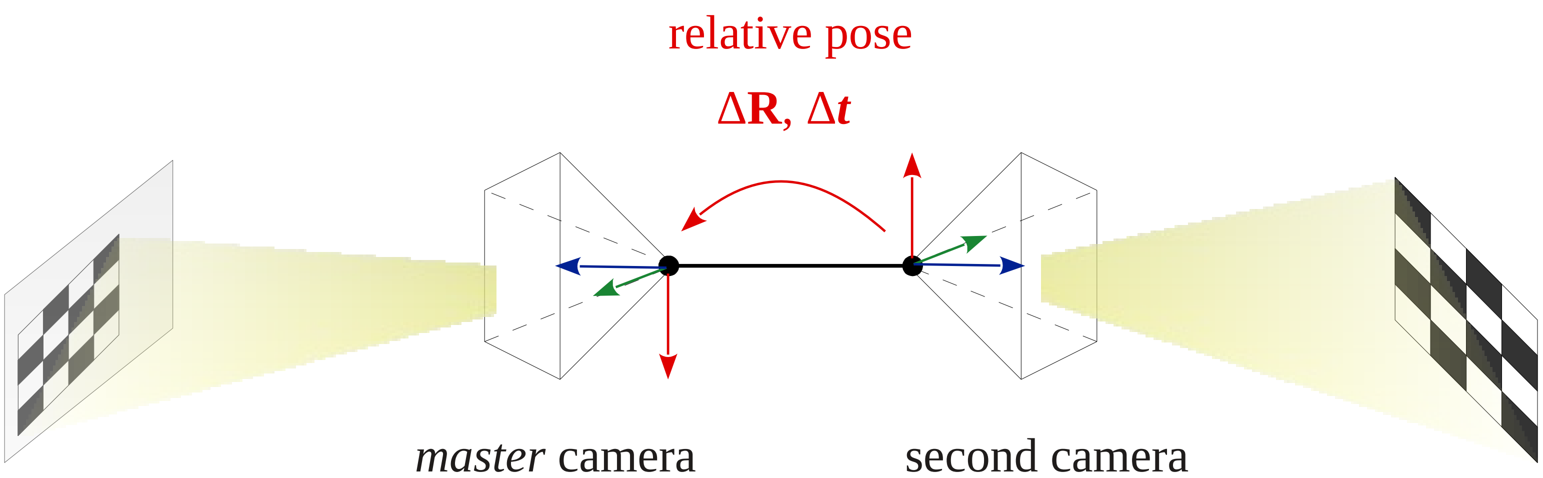


Calibration of Non-overlapping Cameras



- Use large calibration object to find $\Delta\mathbf{R}$, Δt

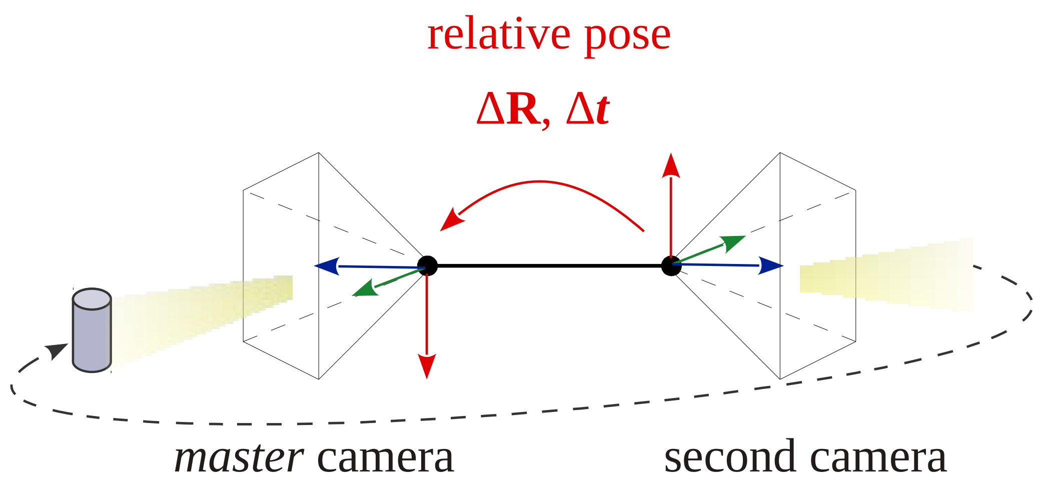
Calibration of Non-overlapping Cameras



- Use large calibration object to find **ΔR**, ***Δt***
- Use local calibration object and mirrors

[Kumar, Ilie, Frahm & Pollefeys, 2008], [Hesch, Mourikis & Roumeliotis, 2009]

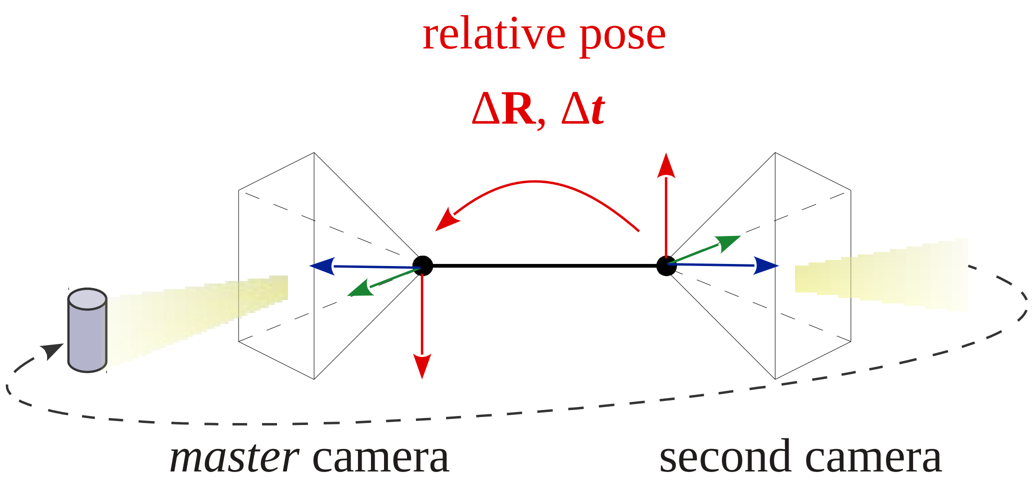
Calibration of Non-overlapping Cameras



- Use large calibration object to find **ΔR**, ***Δt***
- Use local calibration object and mirrors
- Track objects between visual gaps

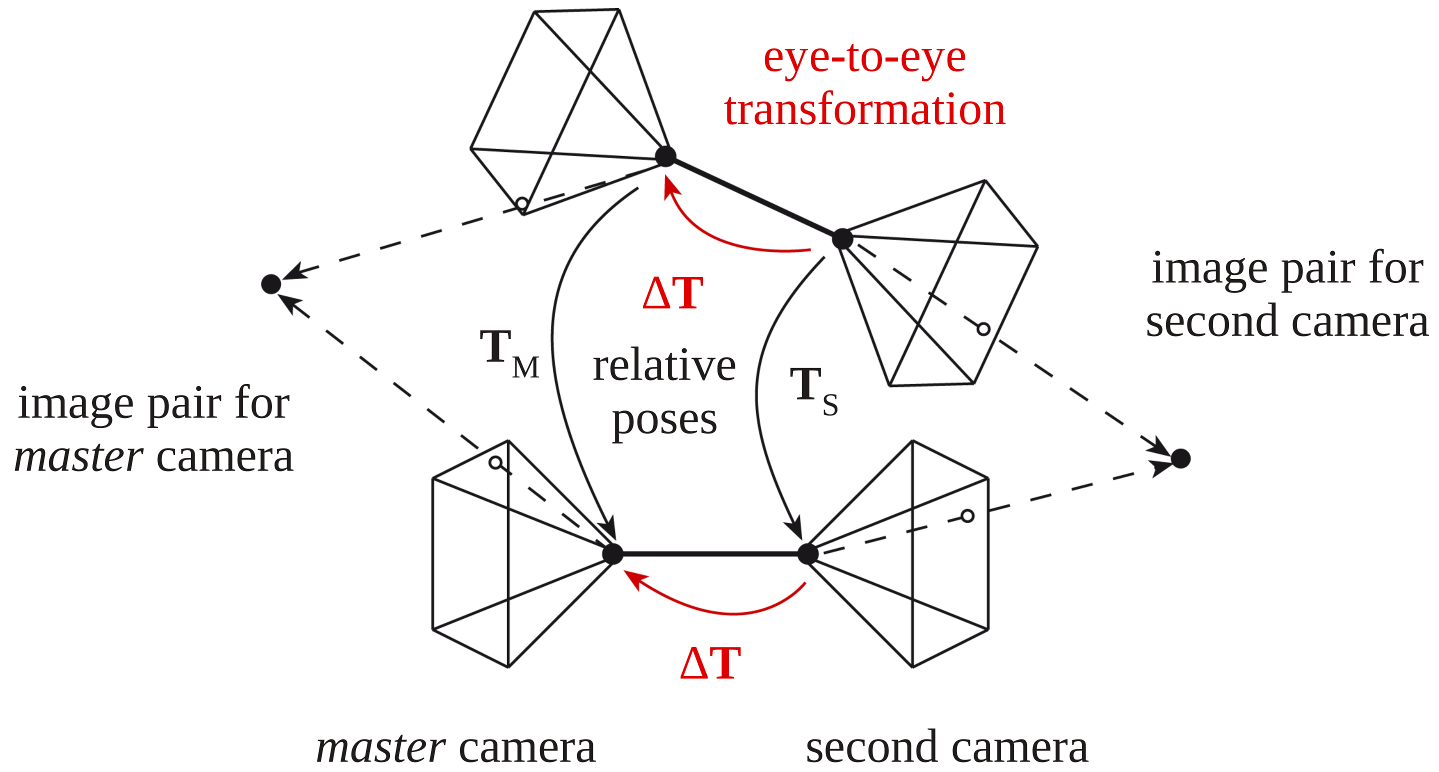
[Makris, Ellis & Black, 2004], [Jaynes, 2004], [Rahimi & Darrell, 2006]

Calibration of Non-overlapping Cameras



- Use large calibration object to find $\Delta\mathbf{R}$, Δt
- Use local calibration object and mirrors
- Track objects between visual gaps
- Knowledge about scene and motion needed!

Eye-to-Eye Calibration



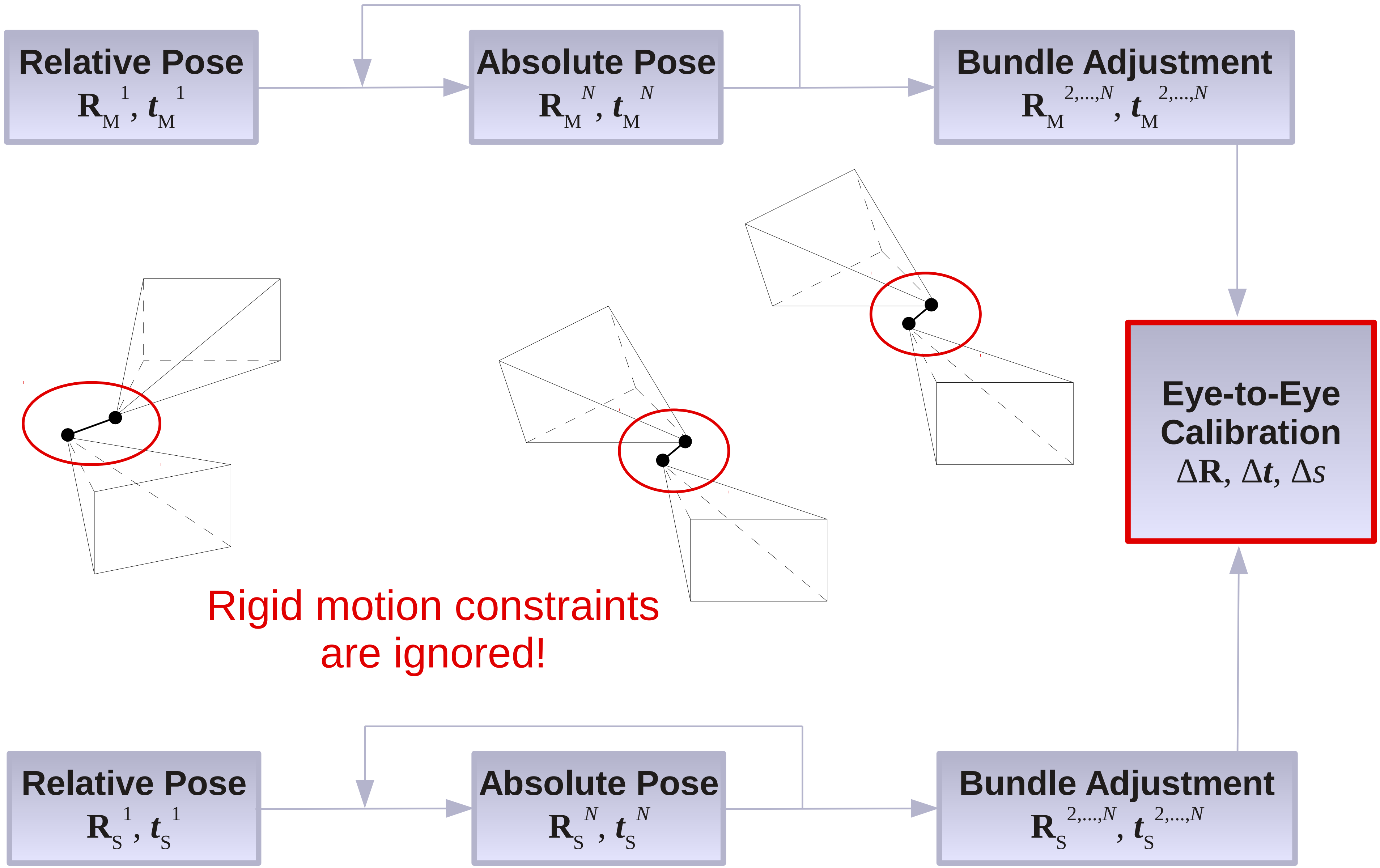
Rigid coupling equation: $\mathbf{T}_M \, \Delta\mathbf{T} = \Delta\mathbf{T} \, \mathbf{T}_S$

$\Delta\mathbf{T}$ contains $\Delta\mathbf{R} \in SO(3)$, $\Delta\boldsymbol{t} \in \mathbb{R}^3$ and $\Delta s \in \mathbb{R}$

Rotation part: $\mathbf{R}_M \, \Delta\mathbf{R} = \Delta\mathbf{R} \, \mathbf{R}_S$

Translation part: $\mathbf{R}_M \, \Delta\boldsymbol{t} + \boldsymbol{t}_M = \Delta s \Delta\mathbf{R} \, \boldsymbol{t}_S + \Delta\boldsymbol{t}$

Structure from Motion

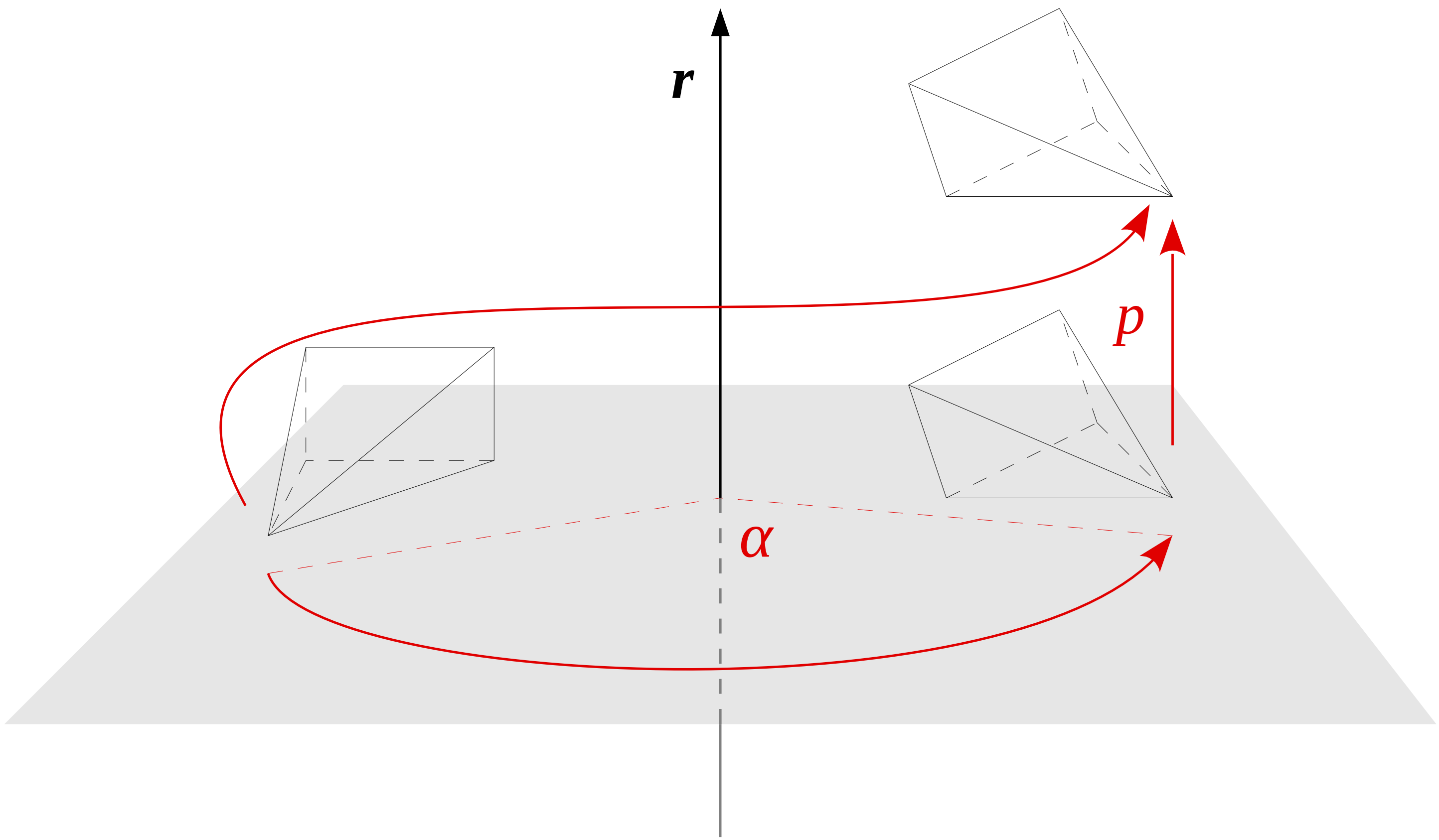


Outline

- Motivation
- Multi-Camera Structure from Motion
- Enforcing Rigid Motion Constraints (“RMCE”)
- Tests and Evaluation
- Conclusion

Rigid Motion Constraints

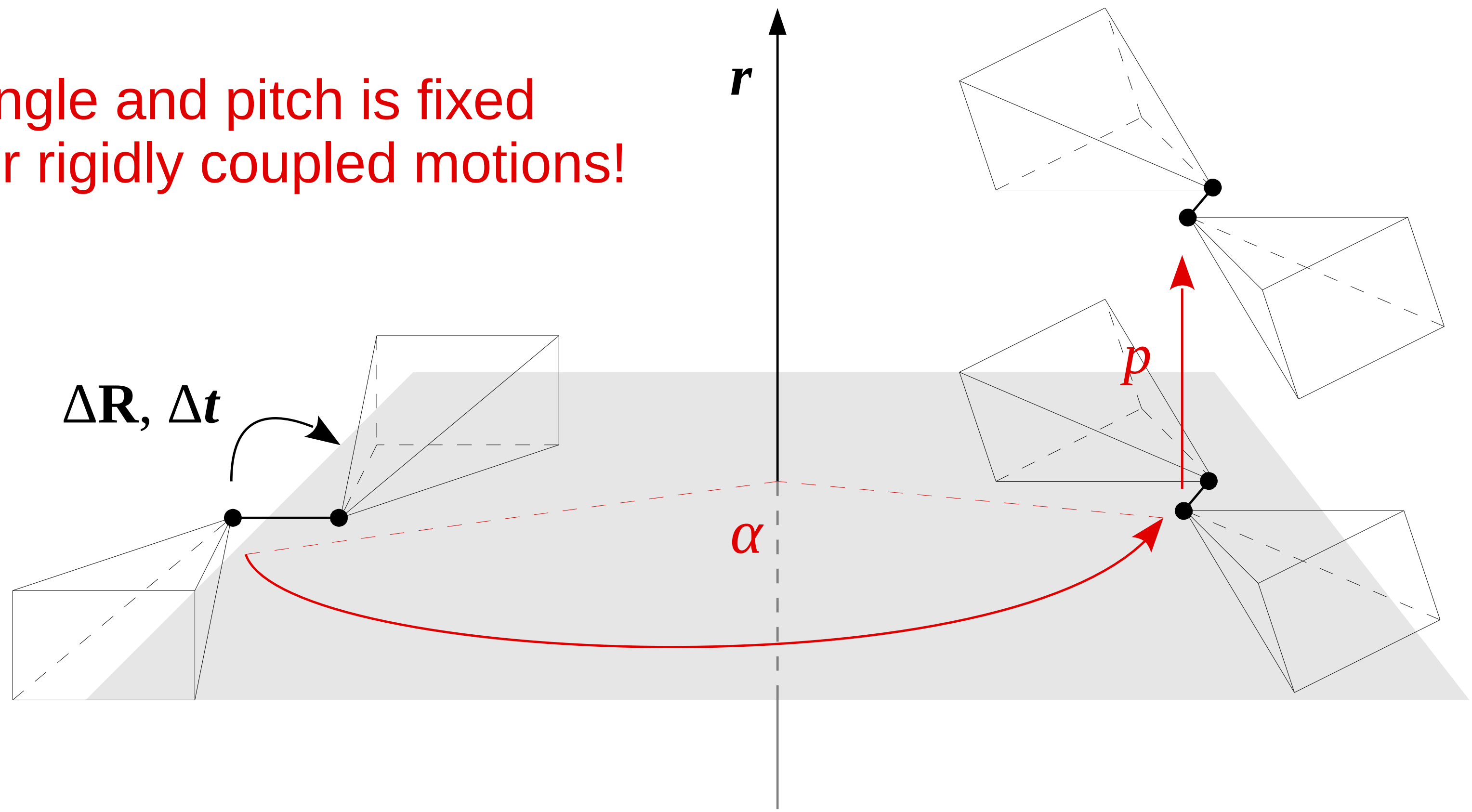
- Chasles' theorem for rigid motion representation
  - Rotation by angle α around 3d line with direction $\boldsymbol{r}$ (*"screw axis"*)
  - Translation of length p along $\boldsymbol{r}$ (*"pitch"*)



Rigid Motion Constraints

- Local motion of rigidly coupled camera
  - Local rotation by angle α around axis with direction $\Delta\mathbf{R}^\mathsf{T}\boldsymbol{r}$
  - Local translation of length p along $\Delta\mathbf{R}^\mathsf{T}\boldsymbol{r}$

Angle and pitch is fixed
for rigidly coupled motions!



Dual Quaternions

- Formal description of screw motions by dual quaternions:

$$\check{\mathbf{q}} = \mathbf{q} + \varepsilon\,\mathbf{q}' = \mathbf{q} + \varepsilon\,\tfrac{1}{2}\,\boldsymbol{t}\cdot\mathbf{q}$$

$$\mathbf{q} = (\boldsymbol{q}, q) = \left(\sin\left(\tfrac{\alpha}{2}\right)\boldsymbol{r}, \boxed{\cos\left(\tfrac{\alpha}{2}\right)}\right)$$

fixed for rigidly coupled motions

$$\mathbf{q}' = (\boldsymbol{q}', q') = \tfrac{1}{2}\left(\cos\left(\tfrac{\alpha}{2}\right)\boldsymbol{t} + \sin\left(\tfrac{\alpha}{2}\right)\boldsymbol{t}\times\boldsymbol{r}, \boxed{-\sin\left(\tfrac{\alpha}{2}\right)\underbrace{\boldsymbol{t}^{\top}\boldsymbol{r}}_{p}}\right)$$

Quadratic constraints $\|\mathbf{q}\| = 1,\ \mathbf{q}^{\top}\mathbf{q}' = 0$

- 3D transformation is quadratic in parameters:

$$\mathbf{R}\,\boldsymbol{X} + \boldsymbol{t} = \check{\mathbf{q}}\cdot\boldsymbol{X}\cdot\check{\mathbf{q}}^{*} = \mathbf{q}\cdot\boldsymbol{X}\cdot\mathbf{q}^{*} + 2\,\mathbf{q}'\cdot\mathbf{q}^{*}$$

$$\mathbf{R}(\mathbf{q}) = \mathbf{I} + 2q[\boldsymbol{q}]_{\times} + 2[\boldsymbol{q}]_{\times}^{2}$$

$$\boldsymbol{t}(\mathbf{q}, \mathbf{q}') = 2\left(q\,\boldsymbol{q}' - q'\boldsymbol{q} + \boldsymbol{q}\times\boldsymbol{q}'\right)$$

Parallel Structure from Motion

initialization for unknown scenes

Relative Pose $\mathbf{R}_M^{\;1}, \boldsymbol{t}_M^{\;1}$

Relative Pose $\mathbf{R}_S^{\;1}, \boldsymbol{t}_S^{\;1}$

2d-2d correspondences

RMCE

local 3d points (up to different scales)

repeat for $n = 2, \ldots, N$

Absolute Pose $\mathbf{R}_M^{\;n}, \boldsymbol{t}_M^{\;n}$

Absolute Pose $\mathbf{R}_S^{\;n}, \boldsymbol{t}_S^{\;n}$

2d-3d correspondences

if poses are suitable for eye-to-eye calibration

Eye-to-Eye Calibration $\Delta\mathbf{R}, \Delta\boldsymbol{t}, \Delta s$

global 3d points

Joint Bundle Adjustment $\mathbf{R}_M^{\;2,\ldots,N}, \boldsymbol{t}_M^{\;2,\ldots,N}, \Delta\mathbf{R}, \Delta\boldsymbol{t}$

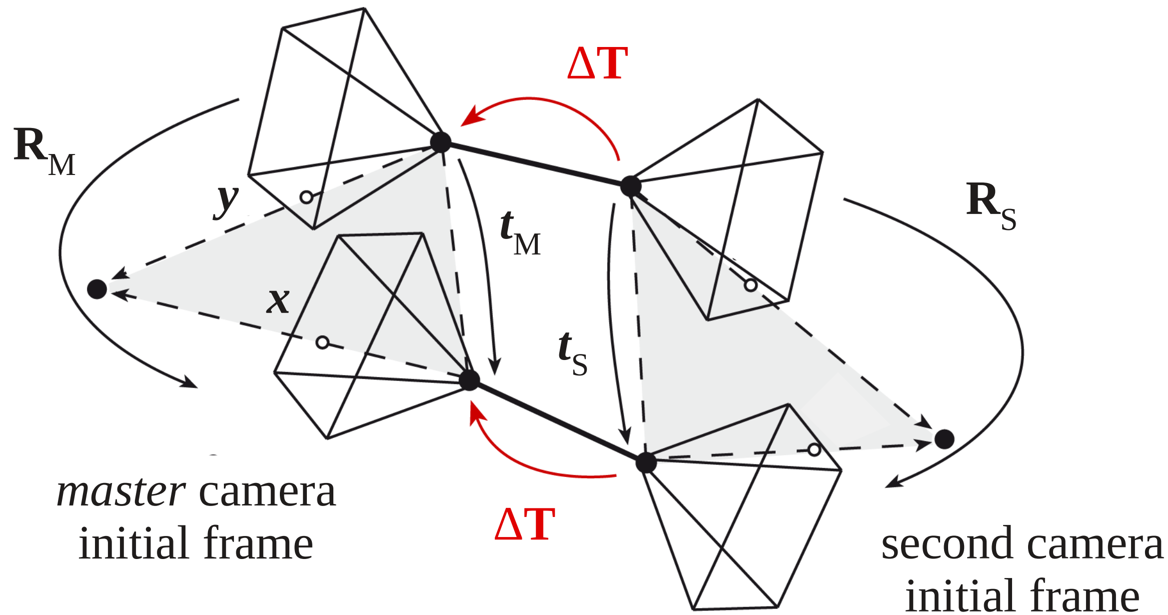
refined global 3d points

Relative Pose Estimation

$$\boldsymbol{y}^{\mathsf{T}}\mathbf{E}\,\boldsymbol{x} = 0$$

$$\mathbf{E} \sim \mathbf{R}^{\mathsf{T}}[\boldsymbol{t}]_{\times}$$

$$\mathbf{E}(\mathbf{q},\boldsymbol{t}) = \mathbf{R}(\mathbf{q})^{\mathsf{T}}[\boldsymbol{t}]_{\times}$$



$$f(\mathbf{q},\boldsymbol{t};\boldsymbol{x},\boldsymbol{y}) = d(\boldsymbol{y},\mathbf{E}(\mathbf{q},\boldsymbol{t})\boldsymbol{x})$$

$$\min \sum_{(\boldsymbol{x},\boldsymbol{y})\in \boldsymbol{C}_\mathrm{M}} f((\boldsymbol{q}_\mathrm{M},q),\boldsymbol{t}_\mathrm{M};\boldsymbol{x},\boldsymbol{y})^2 + \sum_{(\boldsymbol{x},\boldsymbol{y})\in \boldsymbol{C}_\mathrm{S}} f((\boldsymbol{q}_\mathrm{S},q),\boldsymbol{t}_\mathrm{S};\boldsymbol{x},\boldsymbol{y})^2$$

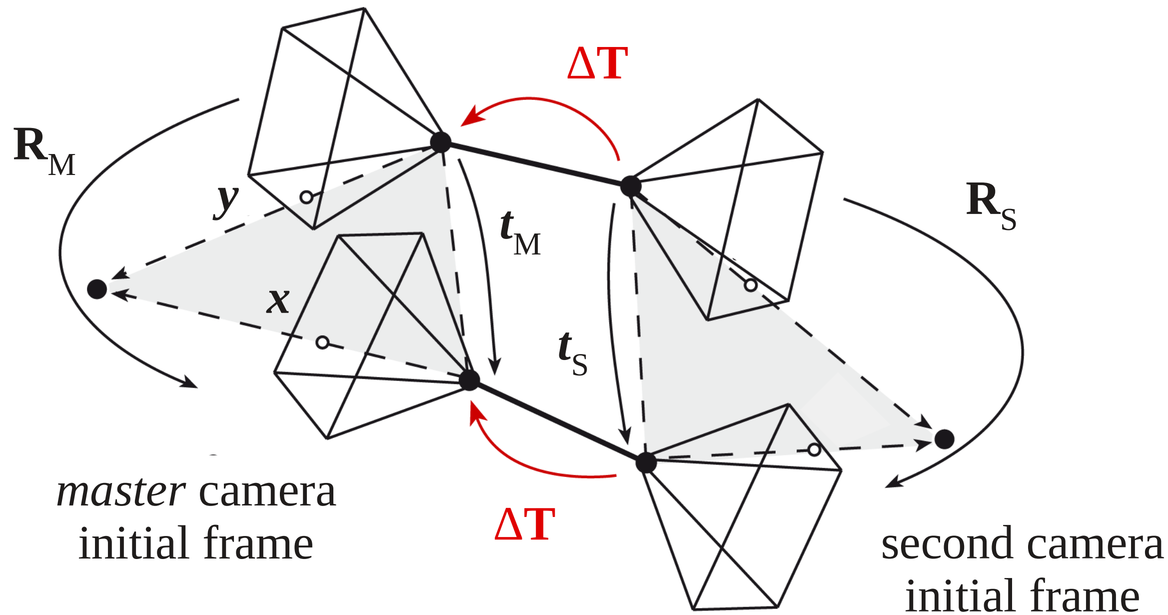
$$\text{s.t.}\quad \boldsymbol{q}_\mathrm{M}^{\mathsf{T}}\boldsymbol{q}_\mathrm{M} = \boldsymbol{q}_\mathrm{S}^{\mathsf{T}}\boldsymbol{q}_\mathrm{S} = 1-q^2 \quad\text{and}\quad \boldsymbol{t}_\mathrm{M}^{\mathsf{T}}\boldsymbol{t}_\mathrm{M} = \boldsymbol{t}_\mathrm{S}^{\mathsf{T}}\boldsymbol{t}_\mathrm{S} = 1$$

Relative Scale Estimation

$$\boldsymbol{y}^{\mathsf{T}} \mathbf{E}\, \boldsymbol{x} = 0$$

$$\mathbf{E} \sim \mathbf{R}^{\mathsf{T}} [\boldsymbol{t}]_\times$$

$$\mathbf{E}(\mathbf{q}, \boldsymbol{t}) = \mathbf{R}(\mathbf{q})^{\mathsf{T}} [\boldsymbol{t}]_\times$$



- Translations $\boldsymbol{t}_\mathrm{M}$, $\boldsymbol{t}_\mathrm{S}$ are known up to (different!) scale
- Rescale $\boldsymbol{t}_\mathrm{S}$ with respect to the equal pitch constraint:

  $\boldsymbol{t}_\mathrm{S} \leftarrow \Delta s\ \boldsymbol{t}_\mathrm{S}$ with $\Delta s = (\boldsymbol{t}_\mathrm{M}{}^{\mathsf{T}} \boldsymbol{r}_\mathrm{M}) \,/\, (\boldsymbol{t}_\mathrm{S}{}^{\mathsf{T}} \boldsymbol{t}_\mathrm{S})$

- For absolute scale, measurement in master camera is needed!

Rigidly Coupled Structure from Motion

initialization for unknown scenes

Relative Pose
$\mathbf{R}_{\mathrm{M}}^{1}, \boldsymbol{t}_{\mathrm{M}}^{1}$

Relative Pose
$\mathbf{R}_{\mathrm{S}}^{1}, \boldsymbol{t}_{\mathrm{S}}^{1}$

2d-2d correspondences

Joint Relative Pose and Scale
$\mathbf{R}_{\mathrm{M}}^{1}, \boldsymbol{t}_{\mathrm{M}}^{1}, \mathbf{R}_{\mathrm{S}}^{1}, \boldsymbol{t}_{\mathrm{S}}^{1}, \Delta s$

local 3d points (up to single scale)

---

repeat for $n = 2, \ldots, N$

Absolute Pose
$\mathbf{R}_{\mathrm{M}}^{n}, \boldsymbol{t}_{\mathrm{M}}^{n}$

Absolute Pose
$\mathbf{R}_{\mathrm{S}}^{n}, \boldsymbol{t}_{\mathrm{S}}^{n}$

2d-3d correspondences

RMCE

---

if poses are suitable for eye-to-eye calibration

Eye-to-Eye Calibration
$\Delta\mathbf{R}, \Delta\boldsymbol{t}$

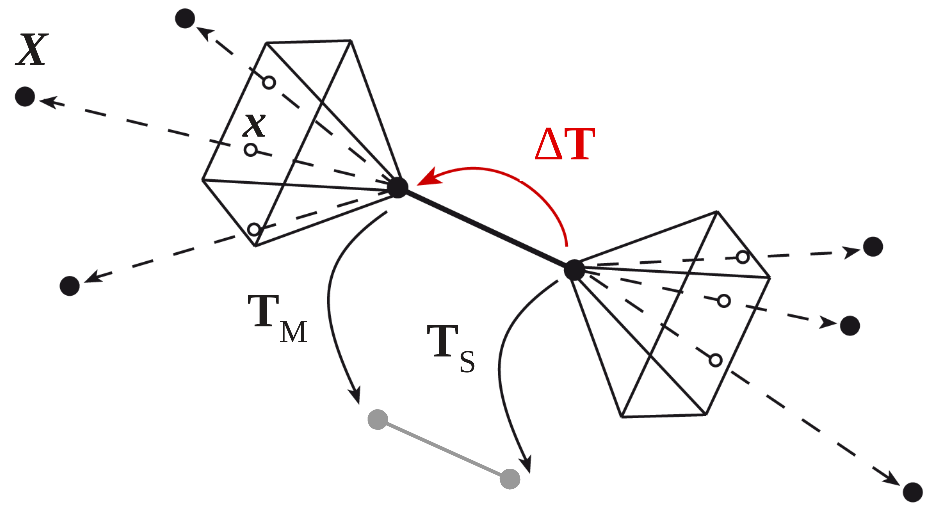
global 3d points

Joint Bundle Adjustment
$\mathbf{R}_{\mathrm{M}}^{2,\ldots,N}, \boldsymbol{t}_{\mathrm{M}}^{2,\ldots,N}, \Delta\mathbf{R}, \Delta\boldsymbol{t}$

refined global 3d points

Absolute Pose Estimation

$$\boxed{x \sim \mathbf{R}^{\mathsf{T}}(X - t)}$$



$$f(\mathbf{q},\mathbf{q}';x,X) = d\left(x,\mathbf{R}(\mathbf{q})^{\mathsf{T}}\left(X - t(\mathbf{q},\mathbf{q}')\right)\right)$$

$$\min \sum_{(x,X)\in C_{\mathrm{M}}} f\left((\mathbf{q}_{\mathrm{M}},q),(\mathbf{q}'_{\mathrm{M}},q');x,X\right)^2 + \sum_{(x,X)\in C_{\mathrm{S}}} f\left((\mathbf{q}_{\mathrm{S}},q),(\mathbf{q}'_{\mathrm{S}},q');x,X\right)^2$$

$$\text{s.t.}\quad \mathbf{q}_{\mathrm{M}}^{\mathsf{T}}\mathbf{q}_{\mathrm{M}} = \mathbf{q}_{\mathrm{S}}^{\mathsf{T}}\mathbf{q}_{\mathrm{S}} = 1 - q^2 \quad \text{and} \quad \mathbf{q}_{\mathrm{M}}^{\mathsf{T}}\mathbf{q}'_{\mathrm{M}} = \mathbf{q}_{\mathrm{S}}^{\mathsf{T}}\mathbf{q}'_{\mathrm{S}} = -qq'$$

Rigidly Coupled Structure from Motion

initialization for unknown scenes

Relative Pose $\mathbf{R}_M^1, \boldsymbol{t}_M^1$

Relative Pose $\mathbf{R}_S^1, \boldsymbol{t}_S^1$ ← 2d-2d correspondences

Joint Relative Pose and Scale $\mathbf{R}_M^1, \boldsymbol{t}_M^1, \mathbf{R}_S^1, \boldsymbol{t}_S^1, \Delta s$ → local 3d points (up to single scale)

---

repeat for $n = 2, \ldots, N$

Absolute Pose $\mathbf{R}_M^n, \boldsymbol{t}_M^n$

Absolute Pose $\mathbf{R}_S^n, \boldsymbol{t}_S^n$ ← 2d-3d correspondences

Joint Absolute Pose $\mathbf{R}_M^n, \boldsymbol{t}_M^n, \mathbf{R}_S^n, \boldsymbol{t}_S^n$

---

if poses are suitable for eye-to-eye calibration

Eye-to-Eye Calibration $\Delta\mathbf{R}, \Delta\boldsymbol{t}$ → global 3d points

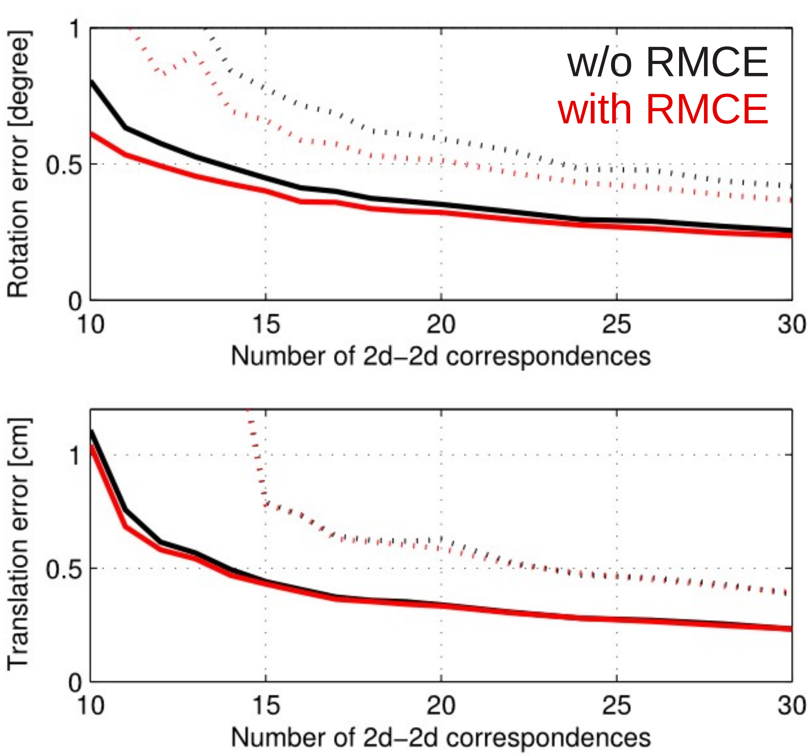
Joint Bundle Adjustment $\mathbf{R}_M^{2,\ldots,N}, \boldsymbol{t}_M^{2,\ldots,N}, \Delta\mathbf{R}, \Delta\boldsymbol{t}$ → refined global 3d points

Outline

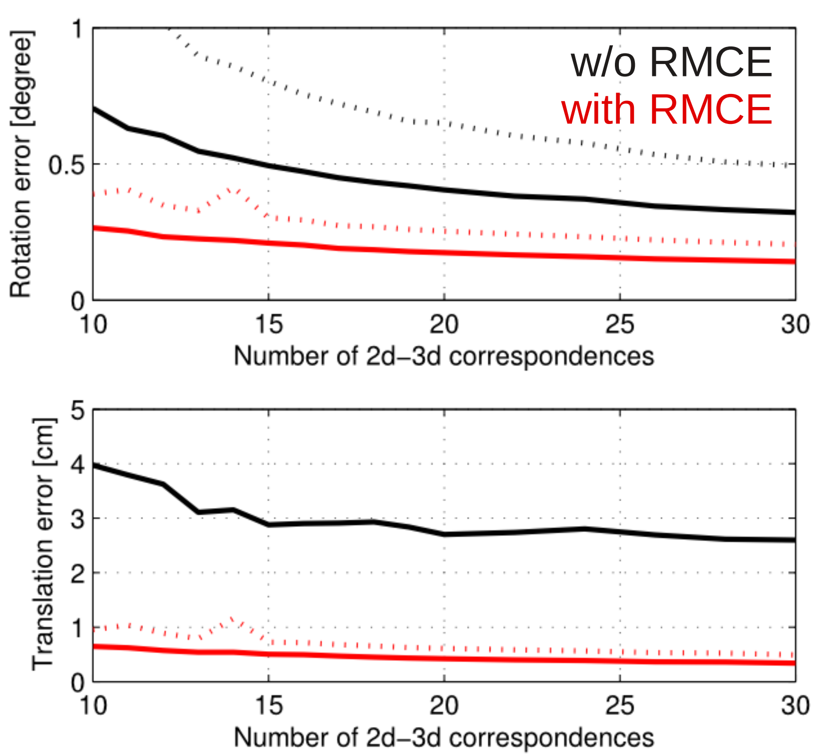
- Motivation
- Multi-Camera Structure from Motion
- Enforcing Rigid Motion Constraints
- Tests and Evaluation
- Conclusion

Evaluation with Synthetic Data

- Camera setup: 10 cm distance, rotated by 30 to 120 degree
- Motion: 25 cm translation, 30 degree rotation
- 2D error: σ_{2d} = 1 pixel (640 x 480 image), 3D error: σ_{3d} = 0.5 cm



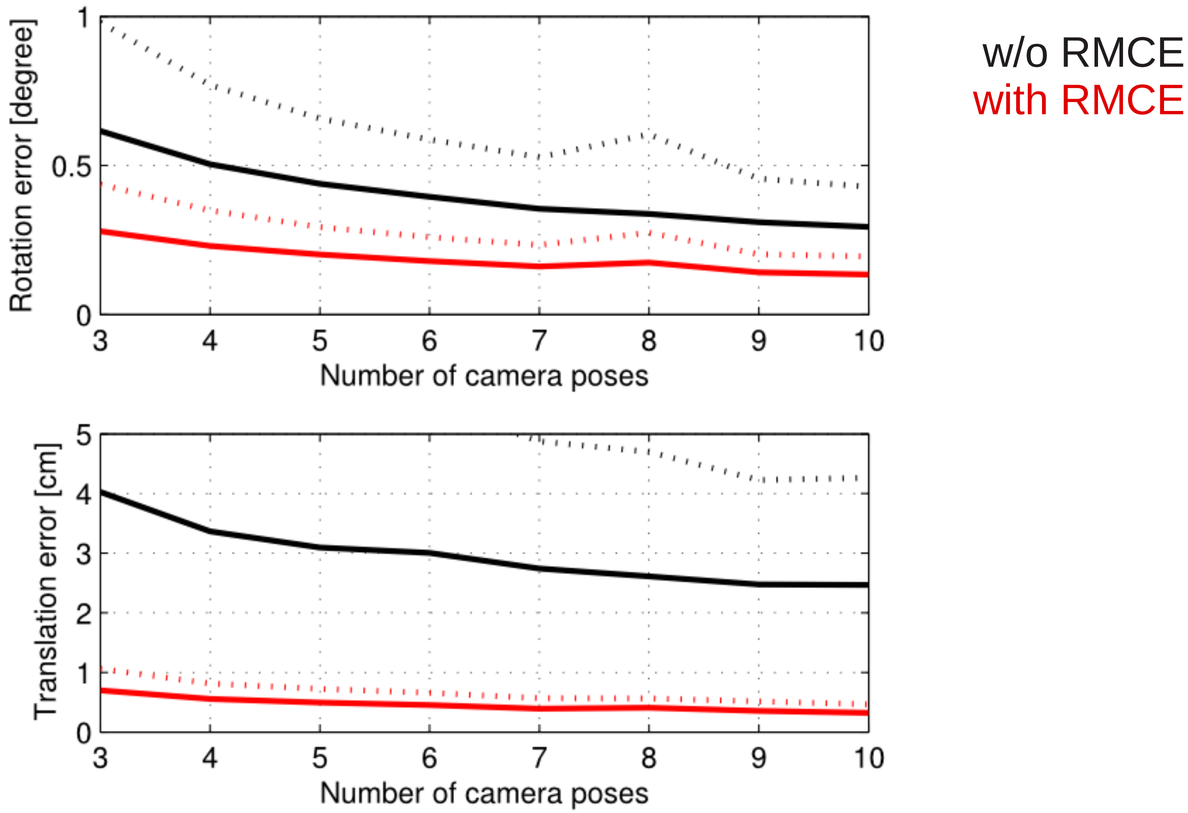
Relative pose from 2d-2d



Absolute pose from 2d-3d

Evaluation with Synthetic Data

- Compute absolute poses from 20 correspondences per camera
- 2D error: σ_{2d} = 1 pixel (640 x 480 image), 3D error: σ_{3d} = 0.5 cm
- Compute eye-to-eye calibration following [Strobl & Hirzinger, 2006]



Test with Rendered Video Sequence

- Image size is 640 x 480 pixels, 140 images in sequence
- Virtual cameras are 25 cm apart, rotated by 15 degree
- Camera motion spans 1.5 meter translation, 60 degree rotation
- View overlap is <u>not</u> used in Structure from Motion!

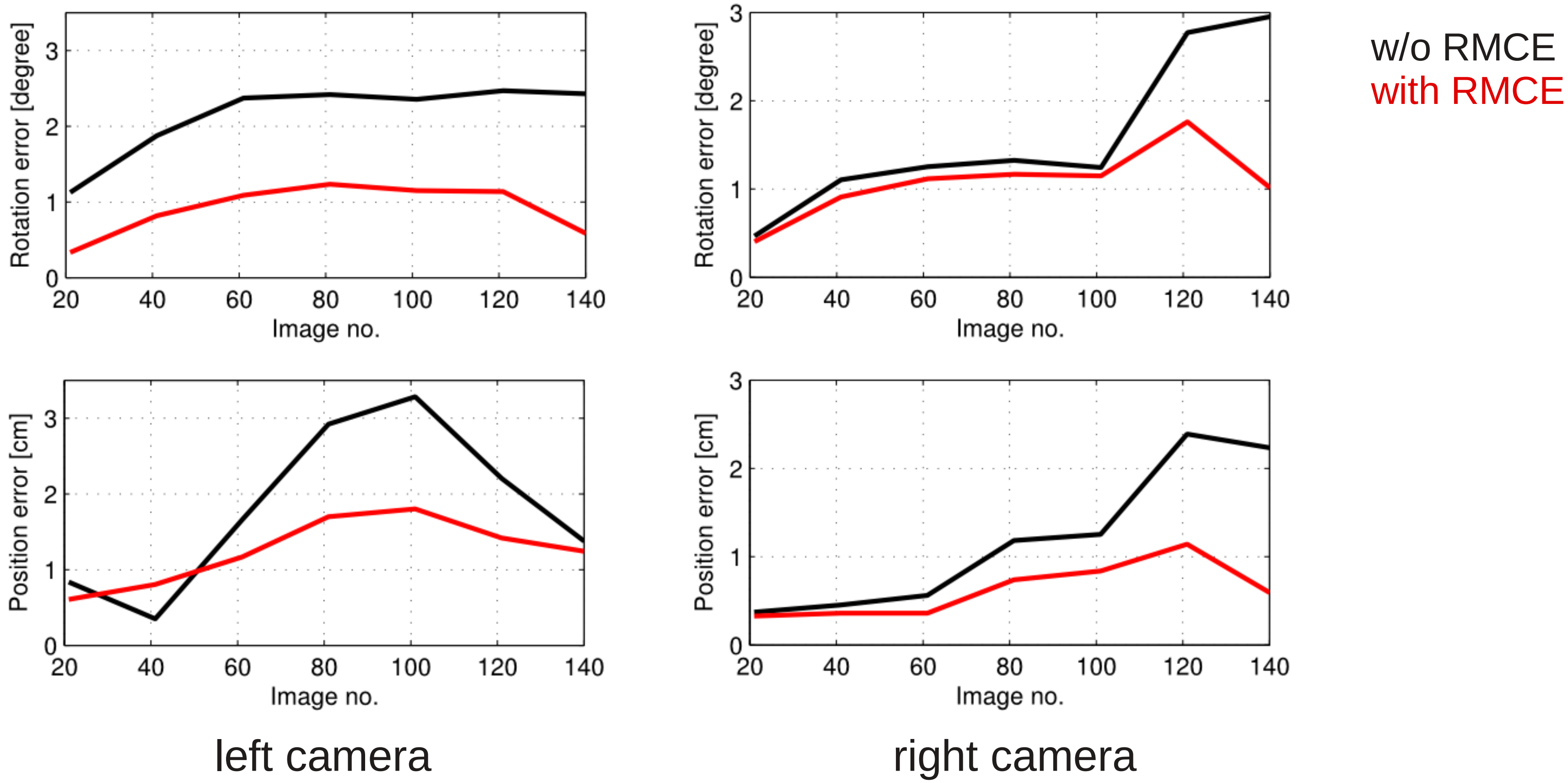
left camera

right camera

3D model

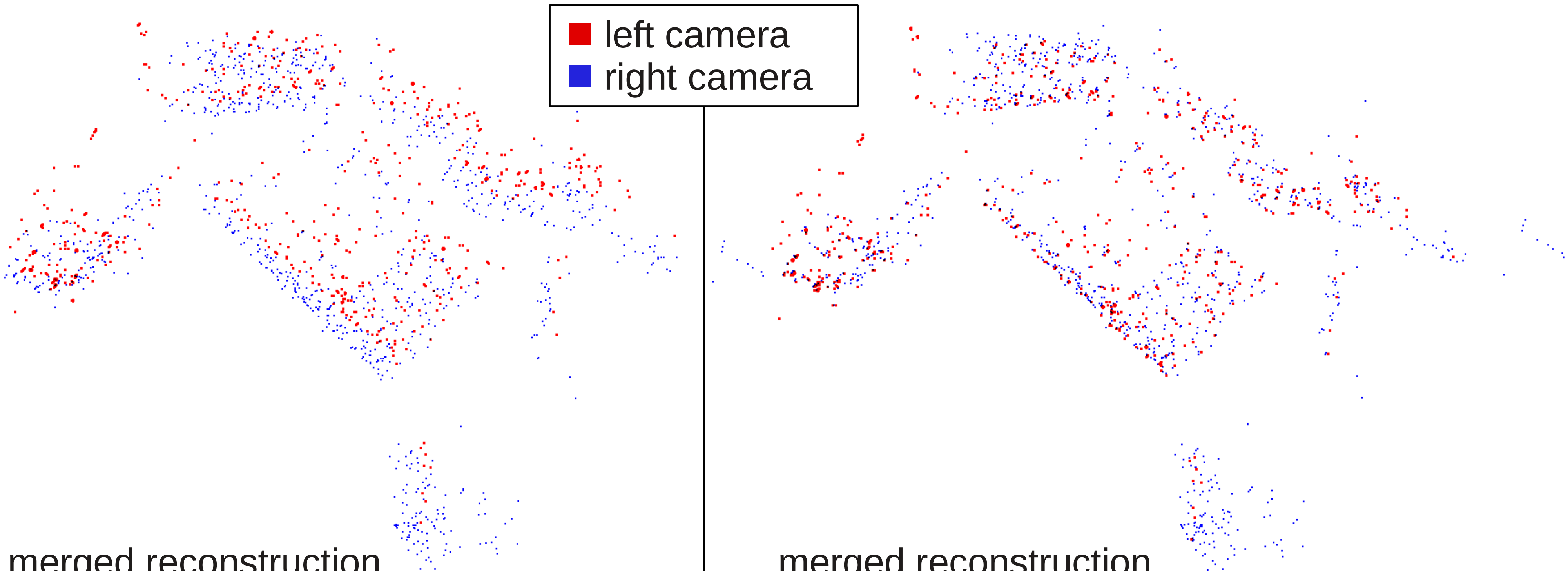
Test with Rendered Video Sequence

- Evaluate absolute pose estimation error for each camera and image
- Error typically accumulates over time (“drift”)
- RMCE reduces drift effect



Test with Rendered Video Sequence

- Error of final eye-to-eye calibration:
  - w/o RMCE: 1.21 degree orientation, 1.9 cm position
  - with RMCE: 0.49 degree orientation, 1.27 cm position



merged reconstruction
w/o rigid motion constraints

merged reconstruction
with rigid motion constraints

Test with Real Video Sequence

- Image size is 640 x 480 pixels, 120 images in sequence
- Cameras are approx. 24.5 cm apart, rotated by 17.2 degree
- Compute stereo calibration to compare with eye-to-eye calibration

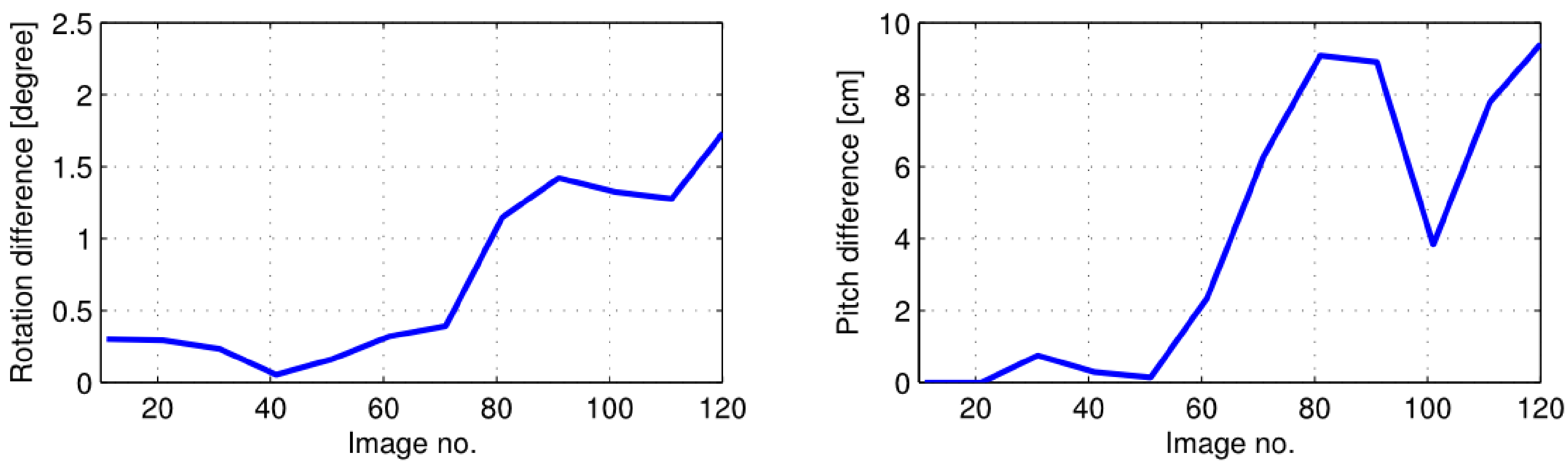
left camera

right camera

Test with Real Video Sequence

- Difference between eye-to-eye calibration and stereo calibration:
  - w/o RMCE: 0.71 degree orientation, 4.1 cm position
  - with RMCE: 0.54 degree orientation, 2.5 cm position

Violation of rigid motion constraints without RMCE



Outline

- Motivation
- Multi-Camera Structure from Motion
- Enforcing Rigid Motion Constraints
- Tests and Evaluation
- Conclusion

Conclusion

Results

- Rigid motion constraint enforcement (RMCE) can be easily integrated into the Structure from Motion pipeline
- Robustness of egomotion estimation is increased
- Eye-to-eye calibration from local egomotion is significantly improved

Future Work

- Consider degenerate motion cases (e.g. planar motion)
- Investigate analytical solution of rigidly coupled pose estimation

Thank you for your attention!

Questions?

Error Approximation for Relative Scale

- Relative scale is computed from initial motion pitchs $\Delta s = p_M / p_S$
- Approximate relative position error ε resulting from relative scale error:

  $$1 + \varepsilon < (1 + \varepsilon_M)\, p_M / (1 - \varepsilon_S)\, p_S$$

  where ε_M, ε_S are relative position errors of initial camera poses.
- For given upper bound E on ε, errors ε_M, ε_S are bounded by E':

  $$E' = ((1 + E)\, p_S - p_M) / ((1 + E)\, p_S + p_M)$$
- For $p_M \approx p_S$ the upper bound is $E' \approx E / (2 + E) \approx E / 2$

Runtime Evaluation

- Runtime is exponential in number of cameras
- Runtime is basically not effected by number of matches (up to ~200)
- Implementation can still be optimized