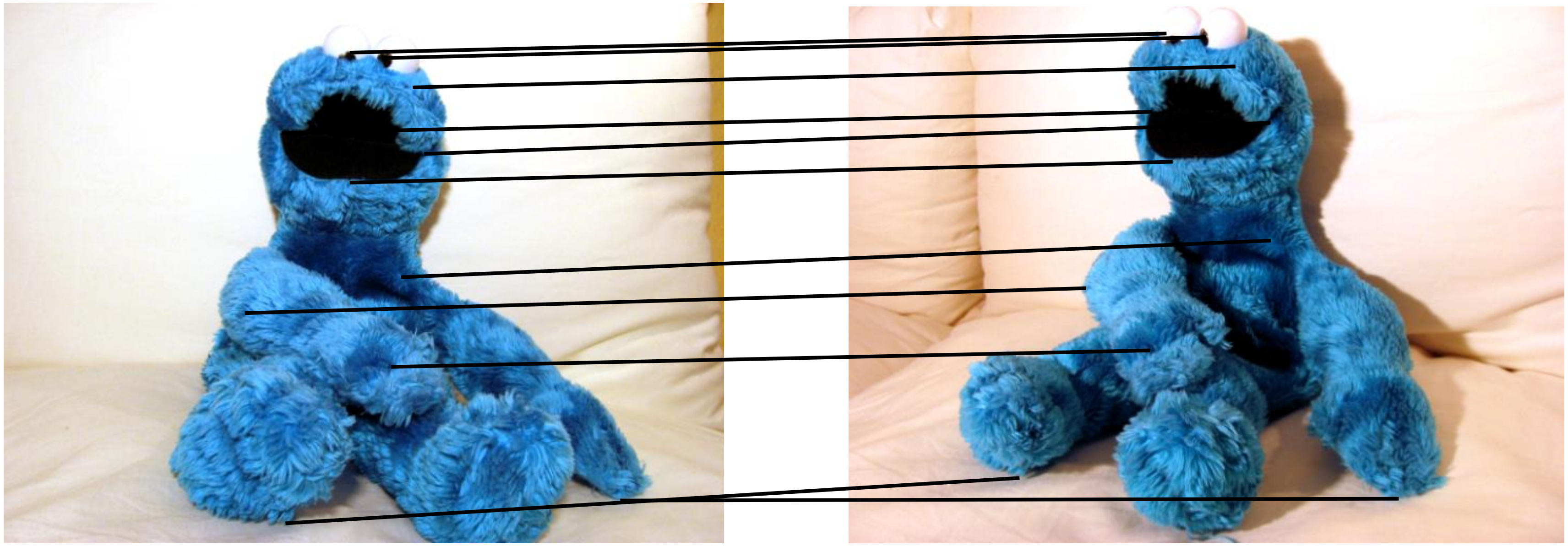
Outline

- Introduction
- Features and Feature Matching
- Geometry of Image Formation
- Calibration
- **Structure from Motion**
- Dense Stereo
- Conclusion

Structure from Motion

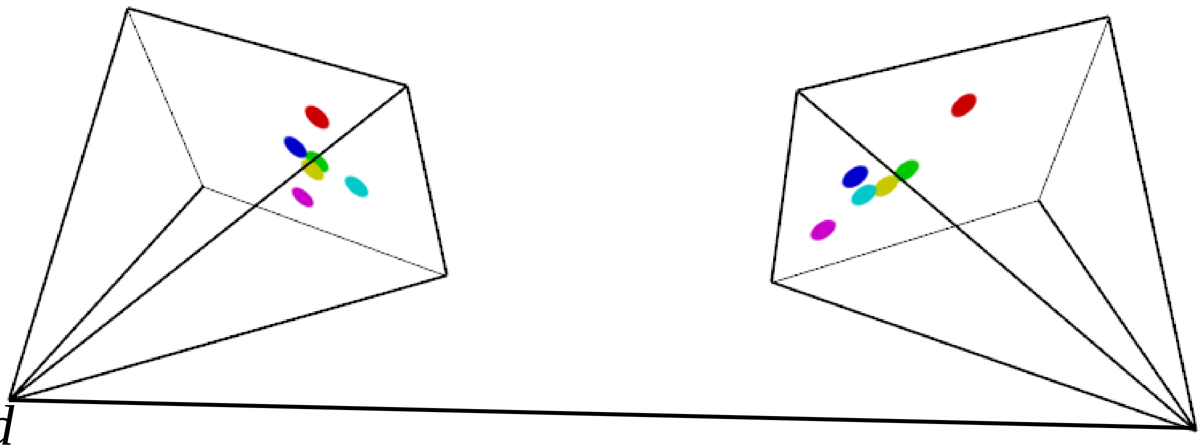
Structure from Motion



Determine corresponding feature points between consecutive images, using SIFT features

Relative Pose Estimation

Given 2D-2D point correspondences: find relative camera translation between images



$|c_2| = 1$

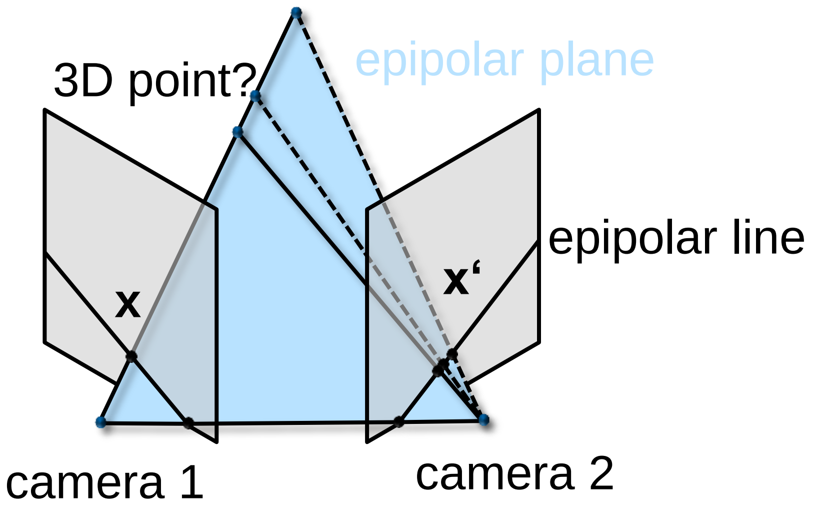
$R_1 = Id$

$c_1 = (0,0,0)^T$

Relative Pose Estimation

- epipolar geometry describes two-view relationship between images
- unknown transformation and 3D points, but:
- correspondence for point x lies on epipolar line in second image (and vice versa)

perspective two-view geometry



Hartley 2004

Perspective Relative Pose Estimation

- a correspondence is normalized by the calibrated camera matrices:

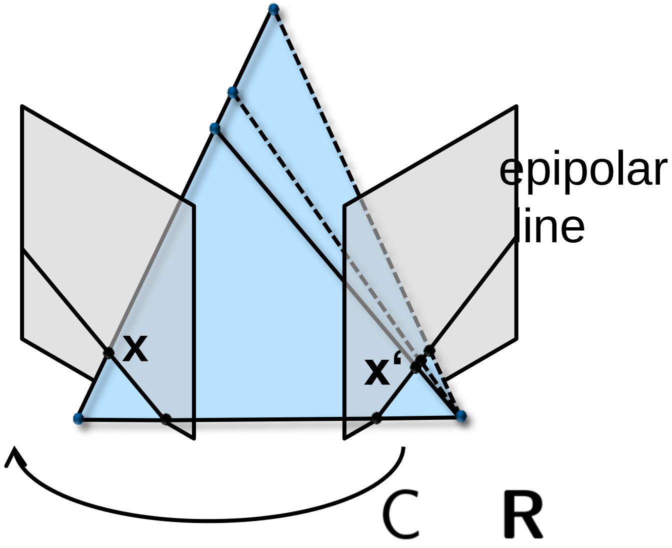
$$\mathbf{x}_\mathrm{n} = \mathbf{K}^{-1}\mathbf{x} \qquad \mathbf{x'}_\mathrm{n} = \mathbf{K'}^{-1}\mathbf{x'}$$

- Essential-matrix transforms point from one image onto epipolar line in second image:

$$\mathbf{E} = [\mathsf{C}]_\times \mathbf{R}$$

$$\mathbf{l} = \mathbf{E}\mathbf{x'}_\mathrm{n}$$

$$\mathbf{l'} = \mathbf{E}^\mathrm{T}\mathbf{x}_\mathrm{n}$$



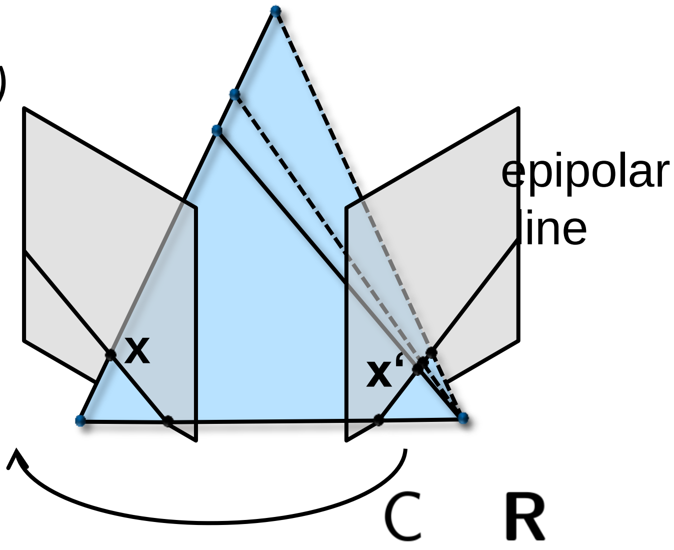
Hartley 2004

Perspective Relative Pose Estimation

- epipolar constraint derived from epipolar geometry:

$$\mathbf{x}_{\mathrm{n}}^{\mathrm{T}} \mathbf{E} \mathbf{x}'_{\mathrm{n}} = 0$$

- is used to stack linear system of equations
- solved by either 8-point algorithm *(e.g. Hartley '97)* or five-point algorithm *(Nistér, 2004)*



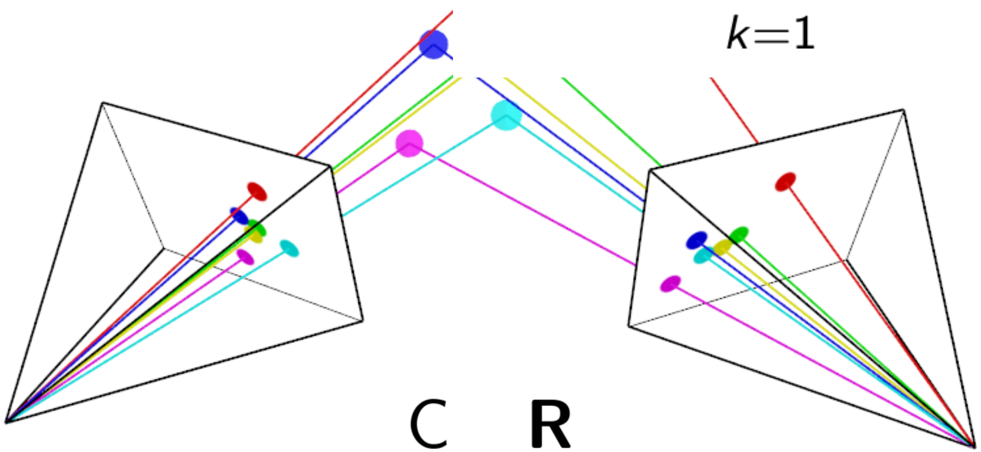
Hartley 2004

GEOMAR

Triangulation

For each correspondence:
- intersect 3D rays in space
- such that they meet in common 3D point

$$\mathrm{argmin}_{\mathrm{X}} \sum_{k=1}^{K} \| \mathbf{R}_k \mathrm{X}_{\mathrm{s}_k} + \mathrm{C}_k + \kappa_k \mathbf{R}_k \tilde{\mathrm{X}}_{\mathrm{a}_k} - \mathrm{X} \|_2^2$$



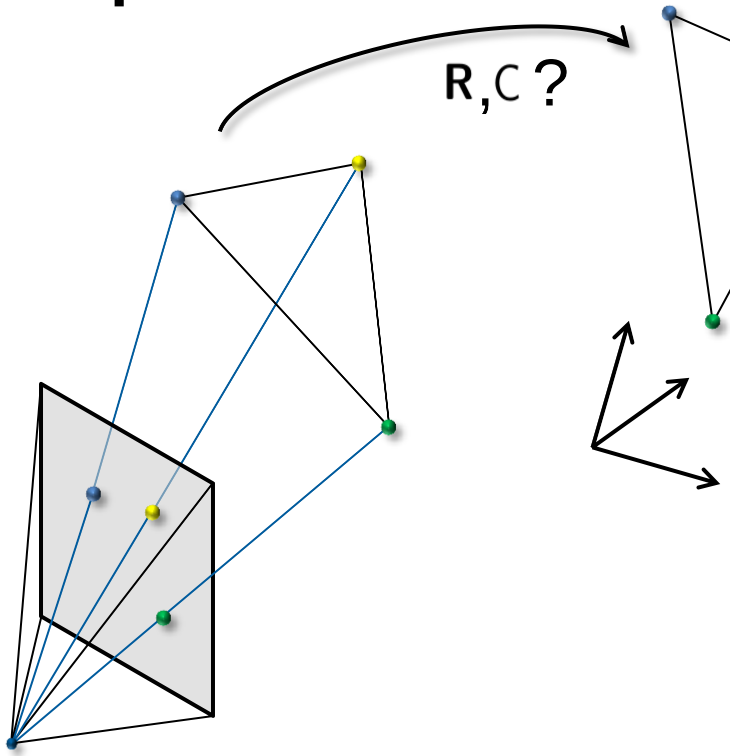
Result:
- ➢ second camera pose relative to first
- ➢ 3D points for all 2D-2D correspondences

Hartley and Sturm: Triangulation 1997

HELMHOLTZ ASSOCIATION

GEOMAR

Perspective Absolute Pose Estimation

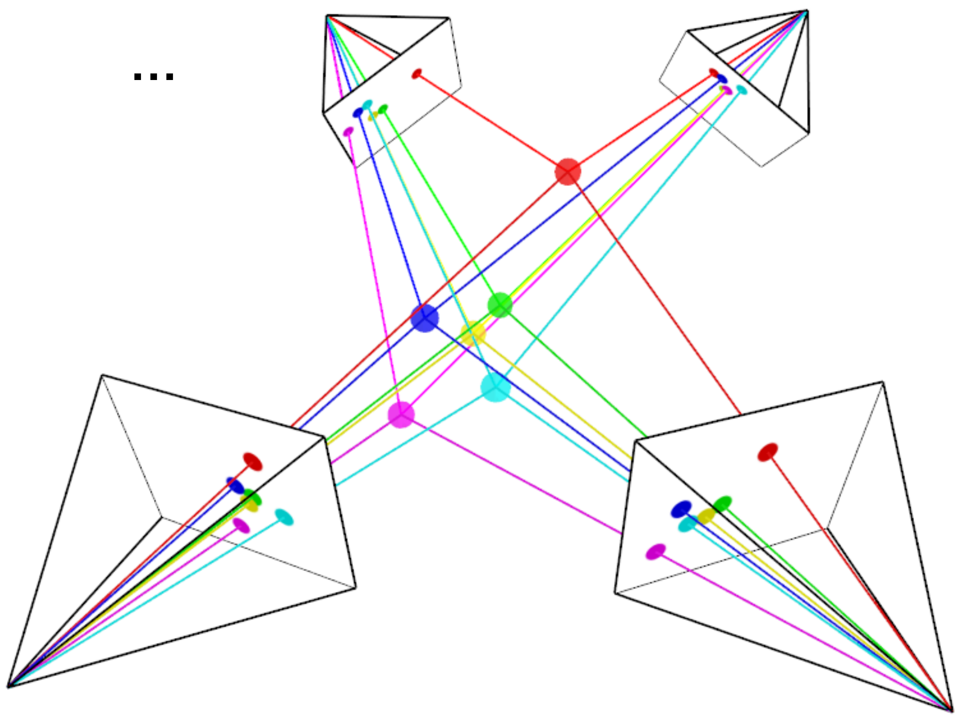


- numerous methods e.g. POSIT, DLT
- P3P pose assuming that distances in camera and world coordinate system are equal *(Haralick 1994)*
- yields equations that are solved for R and C

HELMHOLTZ ASSOCIATION

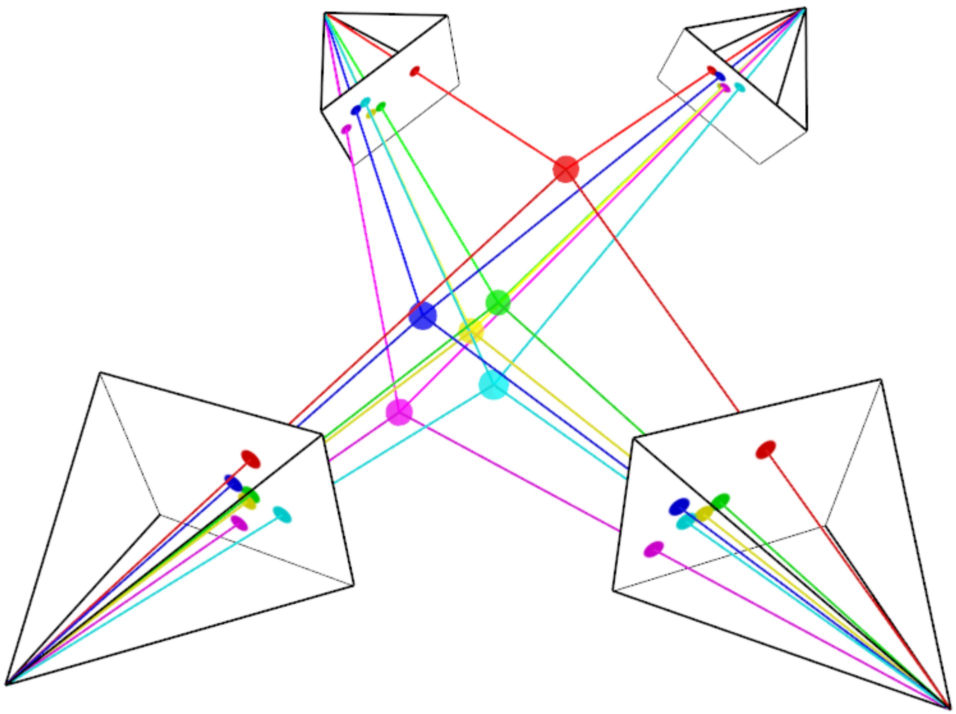
GEOMAR

Perspective Absolute Pose Estimation



HELMHOLTZ ASSOCIATION

Bundle Adjustment



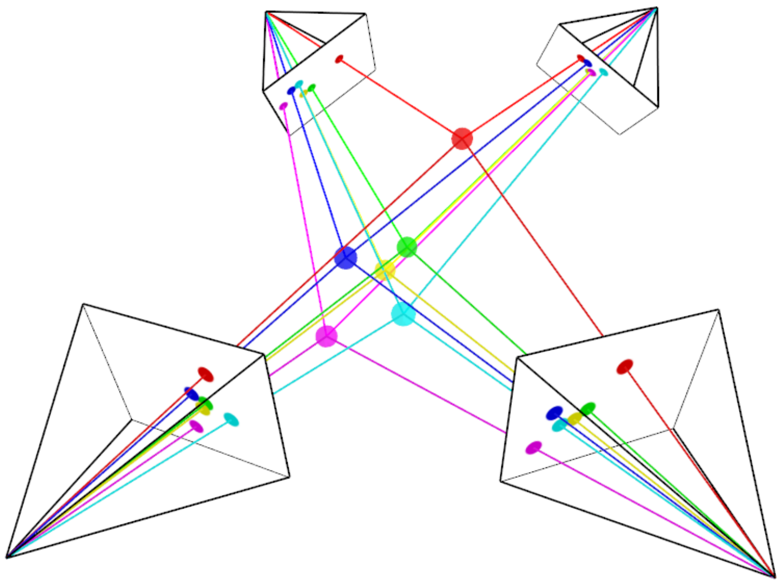
Optimization of non-linear error function over reconstruction with **un-kowns** (parameters):
- rotation matrices
- translation vectors
- 3D points

knowns (observations):
- 2D points

Triggs et al. 2000, Mc Glone 2004

GEOMAR

Bundle Adjustment

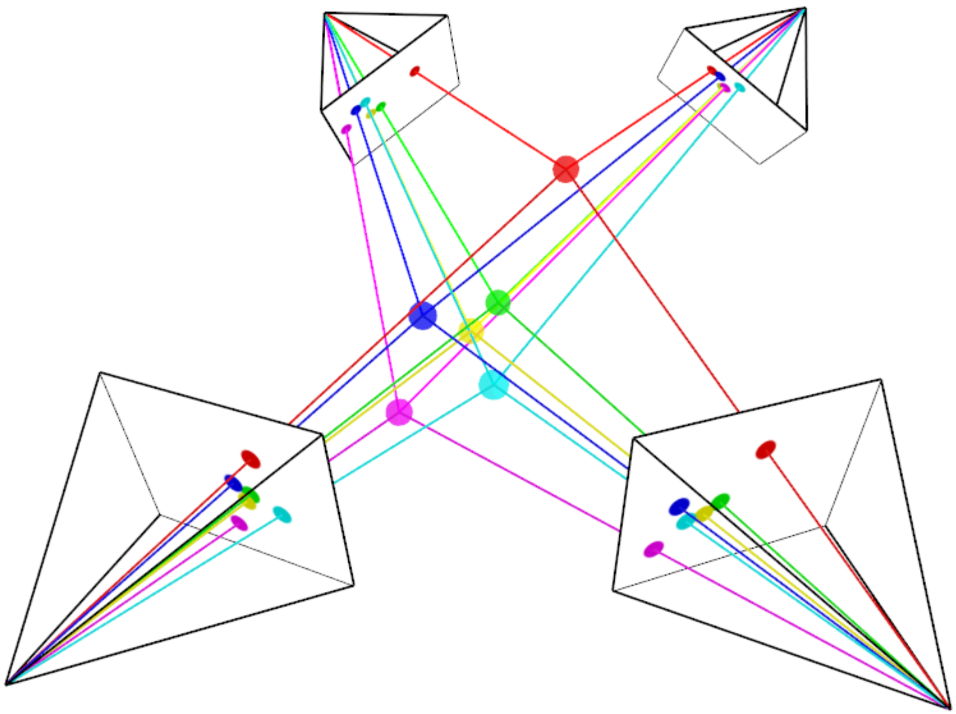


- N images (index i)
- M 3D points (index j)
- note: not every 3D point is visible in every image

$$\mathrm{argmin}_{\forall \mathbf{R}_i, \mathbf{C}_i, i \in \{1,\ldots,N\}, \forall \mathbf{X}_j, j \in \{1,\ldots,M\}} \sum_{i=1}^{N} \sum_{j=1}^{M} \| \mathrm{proj}(\mathbf{X}_j, \mathbf{R}_i, \mathrm{C}_i) - \mathbf{x}_{ij} \|_2^2$$

Triggs et al. 2000, Mc Glone 2004

Bundle Adjustment



Challenging:

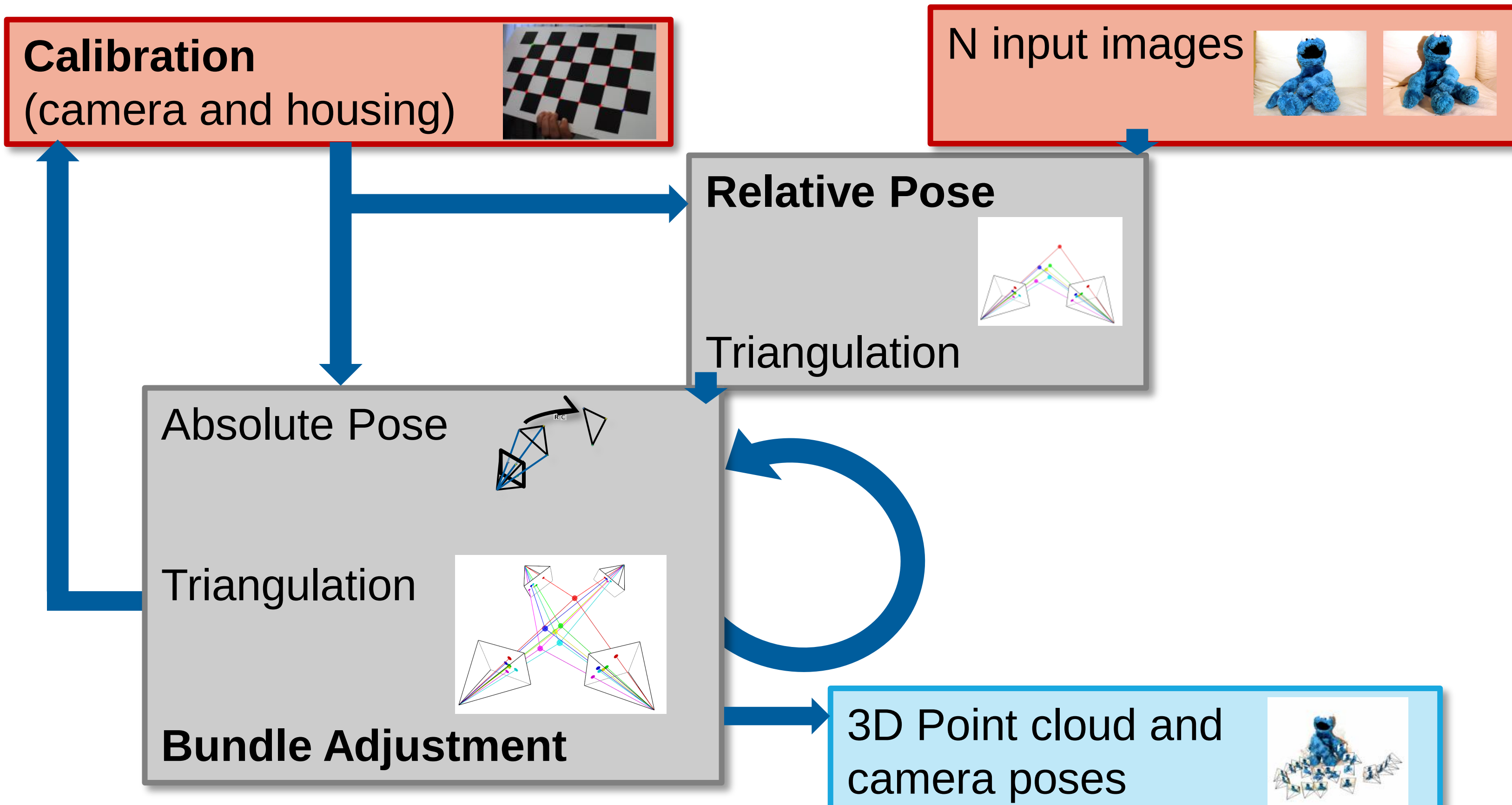
- many parameters (imaging thousands of images with many more 3D points
- potential outliers
- gauge freedom (scale, coordinate system origin)

How it's done:

- Levenberg Marquardt optimization
- sparse matrices (not every 3D point is seen by each camera)

Triggs et al. 2000, Mc Glone 2004

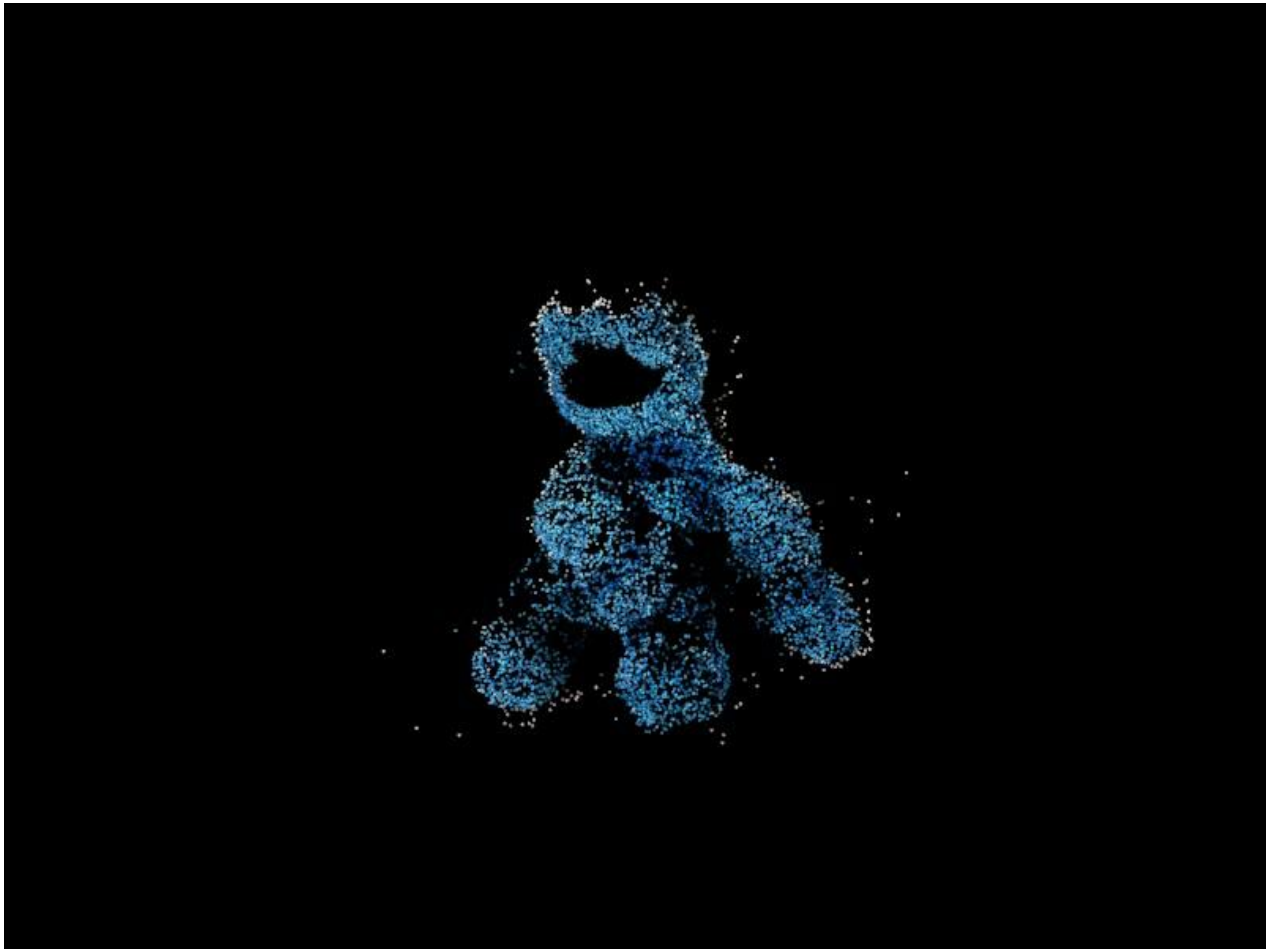
Structure from Motion - Overview



Structure from Motion - Remarks

- note that reconstruction is not metric, i.e. baseline between first and second camera is usually set to one
- bundle adjustment usually the performance bottleneck
- outlier handling required (out of scope)
- different systems exist:
  - sequential
  - hierarchical
  - find best fitting images to start with
  - stereo camera system

Structure from Motion



Structure from Motion

GEOMAR

Bundler – Example for State-of-The-Art System

Example images of the bundler system have been removed
in this version due to copy right reasons.
Please refer to the link below.

Bundler reconstructs 3D information from www.flickr.com large image sets (image from http://www.cs.cornell.edu/~snavely/bundler/).

Snavely et al. 2007

HELMHOLTZ
ASSOCIATION

References

R. Hartley and A. Zisserman. Multiple View Geometry in Computer Vision (Second Edition). Cambridge University Press, second edition, 2004.

R. Szeliski. Computer Vision: Algorithms and Applications. Springer-Verlag, 2011.

D. Nistér. An efficient solution to the five-point relative pose problem. TPAMI, 26:756–777, 2004.

R. I. Hartley and P. F. Sturm. Triangulation. Computer Vision and Image Understanding, 68(2):146–157, 1997.

D. I. DeMenthon and L. S. Davis. Model-based object pose in 25 lines of code. International Journal of Computer Vision, 15:123--141, June 1995.

GEOMAR

References

B. M. Haralick, C.-N. Lee, K. Ottenberg, and M. Nölle. Review and analysis of solutions of the three point perspective pose estimation problem. International Journal of Computer Vision, 13(3):331–356, 12 1994.

M. Pollefeys, R. Koch and L. Van Gool. Self-Calibration and Metric Reconstruction in spite of Varying and Unknown Internal Camera Parameters. International Journal of Computer Vision, 32(1), 7-25, 1999.

B Triggs, P.F. McLauchlan, R.I. Hartley, A.W. Fitzgibbon. Bundle adjustment – a modern synthesis. Vision algorithms: theory and practice, 2000

N. Snavely, S.M. Seitz, R. Szeliski. Photo tourism: exploring photo collections in 3D. ACM transactions on graphics (TOG), 2006

J. C. McGlone, editor. Manual of Photogrammetry. ASPRS, 5th edition, 2004.

HELMHOLTZ ASSOCIATION

Wrap Up

- SfM allows to reconstruct camera path and sparse 3D point cloud simultaneously
- relative pose computation utilizes epipolar geometry to determine camera movement between 2 images from 2D-2D correspondences
- absolute pose estimation computes camera pose from 2D-3D correspondences
- triangulation determines 3D point by intersecting 3D rays from 2D-2D correspondence with known camera pose
- bundle adjustment optimizes camera poses and 3D points such that reprojection error is minimized