ETH zürich

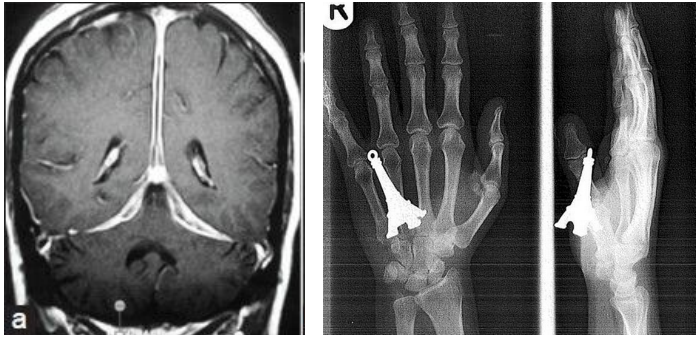
CVL Computer Vision Lab

What Correspondences Reveal about Unknown Camera and Motion Models?

Thomas Probst[1], Ajad Chhatkuli[1], Danda Pani Paudel[1], Luc Van Gool[1,2]

[1]Computer Vision Laboratory, ETH Zurich, Switzerland
[2]VISICS, ESAT/PSI, KU Leuven, Belgium

Which is the correct camera model?



Orthographic

Perspective

Affine

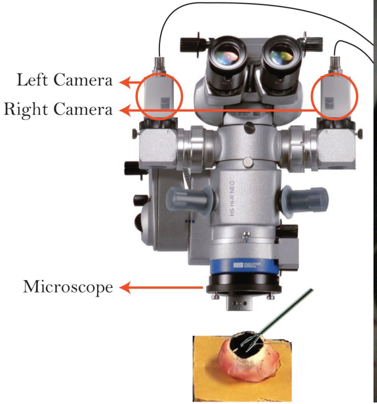
Fisheye

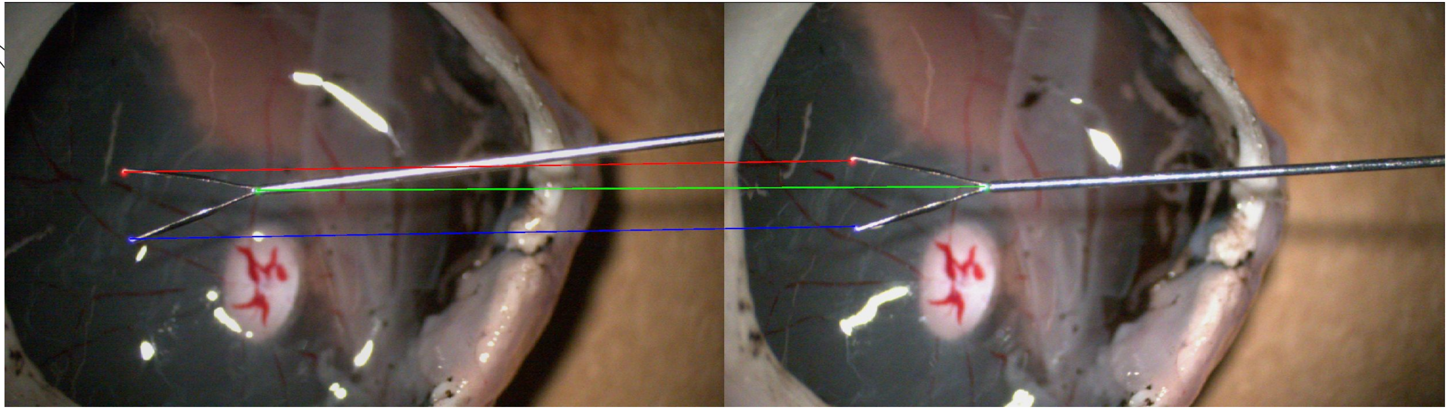
Camera-specific methods for geometry tasks.

- Registration
- Calibration
- Structure-from-Motion

Which is the correct motion type?

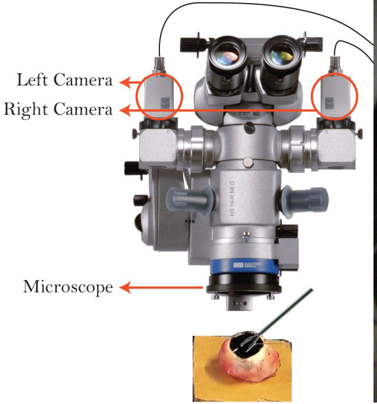
- Motion between two views is typically constrained
- Which degrees of freedom?



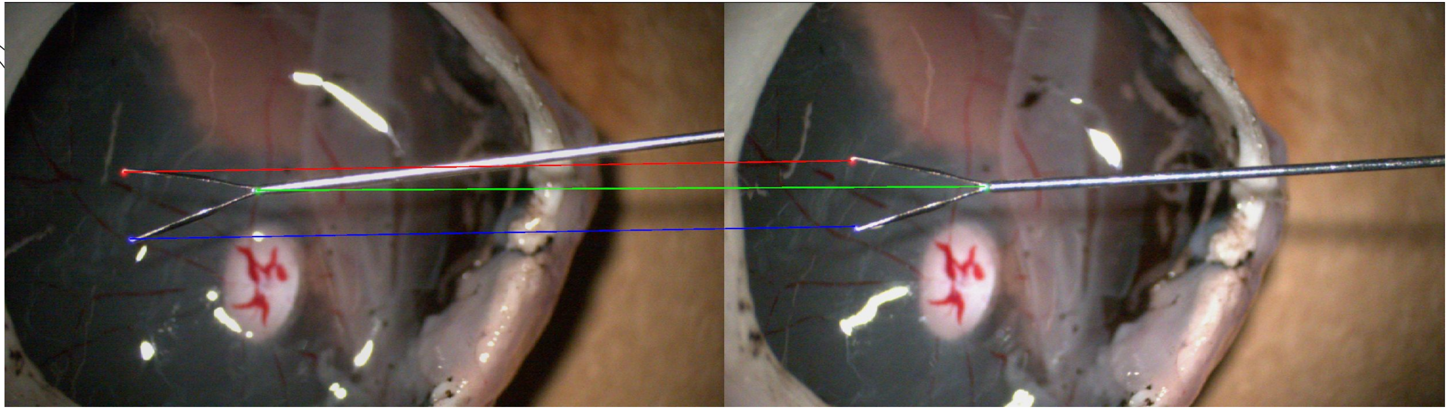


Stereo microscope
for retinal surgery

How to calibrate for 3D tasks?
Perspective camera? Pure Translation?



Stereo microscope for retinal surgery



How to calibrate for 3D tasks?
Perspective camera? Pure Translation?

How can we express the two-view relationship, when camera model and motion type are unknown?

Problem Statement

- Given only point correspondences,

  I. what is the underlying two-view relationship?
  e.g. Homography/Fundamental Matrix

  II. what is the camera model and motion type?
  e.g. perspective/affine + pure translation/rotation

I. What is the underlying two-view relationship?

- Given point correspondences $\Omega = \{\mathrm{u}_i, \mathrm{v}_i\}$
  - Noise and outliers

- Consensus Maximization not optimal
  - Model unknown
  - Less constrained model gives more inliers
  - How to select the best model?

I. Recovering the underlying two-view relationship

- **Assumption**:
  Optimal camera and motion model $(\mathcal{M}, \theta)$
  - minimizes the joint degrees of freedom of $(\mathcal{M}, \theta)$
  - explains a significantly large subset of Ω

- Camera and motion models $(\mathcal{M}, \theta)$ represented as polynomials

- *Find the sparsest set of polynomials, that agrees with more than half of the correspondences.*

II. Disambiguating camera model and motion type

Summary assuming constant camera type.

Camera model →

Motion type ↓

	Cal. Perspective	Uncal. Perspective	Orthographic	Affine
Full motion				
Rotation		$HH^T \neq I,\ H = KRK^{-1}$ ✓ ✓ ✓		
Translation				
Rotation x				
Rotation y				
Rotation z				
Translation x/y				
Translation z				

What is the two-view relationship?

Can the camera be disambiguated?
Can the motion type be disambiguated?
Can the metric motion be recovered knowing camera and motion type?

II. Disambiguating camera model and motion type

Camera model →

Motion type ↓

	Cal. Perspective		Uncal. Perspective		Orthographic		Affine	
Full motion	$\sigma_1(\mathsf{E}) = \sigma_2(\mathsf{E})$	✓ ✓ ✓	$\mathsf{F} \in \mathcal{U}$	✓ ✓ ✗	$\mathsf{E}_O \in \mathcal{O}$	✓ ✓ ✗	$\mathsf{F}_A \in \mathcal{A}$	✓ ✓ ✗
Rotation	$\mathsf{H} = \mathsf{R}$	✓ ✓ ✓	$\mathsf{H}\mathsf{H}^\top \neq \mathsf{I},\ \mathsf{H} = \mathsf{K}\mathsf{R}\mathsf{K}^{-1}$	✓ ✓ ✓	$\mathsf{E}_O \in \mathcal{O},\ \mathsf{E}_{O3,3} = 0$	✗ ✗ ✗	$\mathsf{F}_A \in \mathcal{A},\ \mathsf{F}_{A3,3} = 0$	✓ ✓ ✗
Translation	$\mathsf{E} = [\mathsf{t}]_\times = -\mathsf{E}^\top$	✗ ✗ ✓	$\mathsf{F} = \mathsf{K}[\mathsf{t}]_\times\mathsf{K}^\top = -\mathsf{F}^\top$	✗ ✗ ✗	$\mathsf{H} \in \mathcal{T}$	✗ ✓ ✓	$\mathsf{H} \in \mathcal{T}$	✗ ✓ ✗
Rotation x	$\mathsf{H} = \mathsf{R}_x$	✓ ✓ ✓	$\mathsf{H}\mathsf{H}^\top \neq \mathsf{I},\ \mathsf{H} = \mathsf{K}\mathsf{R}_x\mathsf{K}^{-1}$ $\mathsf{H}_{2,1} = \mathsf{H}_{3,1} = 0$	✓ ✓ ✗	$\mathsf{E}_O = [\mathsf{r}_x]_\times = -\mathsf{E}_O^\top$	✗ ✗ ✗	$\mathsf{F}_A \in \mathcal{O}, \mathsf{F}_{A3,3} = 0$	✗ ✗ ✗
Rotation y	$\mathsf{H} = \mathsf{R}_y$	✓ ✓ ✓	$\mathsf{H}\mathsf{H}^\top \neq \mathsf{I},\ \mathsf{H} = \mathsf{K}\mathsf{R}_y\mathsf{K}^{-1}$	✓ ✗ ✗	$\mathsf{E}_O = [\mathsf{r}_y]_\times = -\mathsf{E}_O^\top$	✗ ✗ ✗	$\mathsf{F}_A = [\mathsf{r}_y]_\times = -\mathsf{F}_A^\top$	✗ ✗ ✗
Rotation z	$\mathsf{H} = \mathsf{R}_z$	✗ ✓ ✓	$\mathsf{H} \in \mathrm{Aff}(2, \mathbb{R})$	✗ ✓ ✗	$\mathsf{H} = \mathsf{R}_z$	✗ ✓ ✓	$\mathsf{H} \in \mathrm{Aff}(2, \mathbb{R})$	✗ ✓ ✗
Translation x/y	$\mathsf{E} = [\mathsf{t}_{x/y}]_\times = -\mathsf{E}^\top$	✗ ✗ ✓	$\mathsf{F} = \mathsf{K}[\mathsf{t}_{x/y}]_\times\mathsf{K}^\top = -\mathsf{F}^\top$	✗ ✗ ✗	$\mathsf{H} \in \mathcal{T}_{x/y}$	✓ ✓ ✓	$\mathsf{H} \in \mathcal{T}$	✗ ✗ ✗
Translation z	$\mathsf{E} = [\mathsf{t}_z]_\times = -\mathsf{E}^\top$	✓ ✓ ✓	$\mathsf{F} = \mathsf{K}[\mathsf{t}_z]_\times\mathsf{K}^\top = -\mathsf{F}^\top$	✗ ✗ ✗	$\mathsf{H} = \mathsf{I}$	✗ ✗ ✗	$\mathsf{H} = \mathsf{I}$	✗ ✗ ✗

Results - Relative Pose on Oxford Dataset

- Relative pose between two consecutive frames (short baseline)
- Multi-view SfM as GT
- Compare to RANSAC baseline
- Sparse motion assumption gives more robust poses

W. Maddern, G. Pascoe, C. Linegar and P. Newman, "1 Year, 1000km: The Oxford RobotCar Dataset", The International Journal of Robotics Research (IJRR), 2016.

Thank you for your attention.
Please visit our poster #83!