Enhancing 3D Visual Odometry with Single-Camera Stereo Omnidirectional Systems

Thesis Defense

Carlos Jaramillo

Spring 2018

INTRODUCTION

Enhancing 3D Visual Odometry with Single-Camera Systems

Goal: Enhancing Visual Odometry

<Intro> <SOS> <GUMS> <VO>

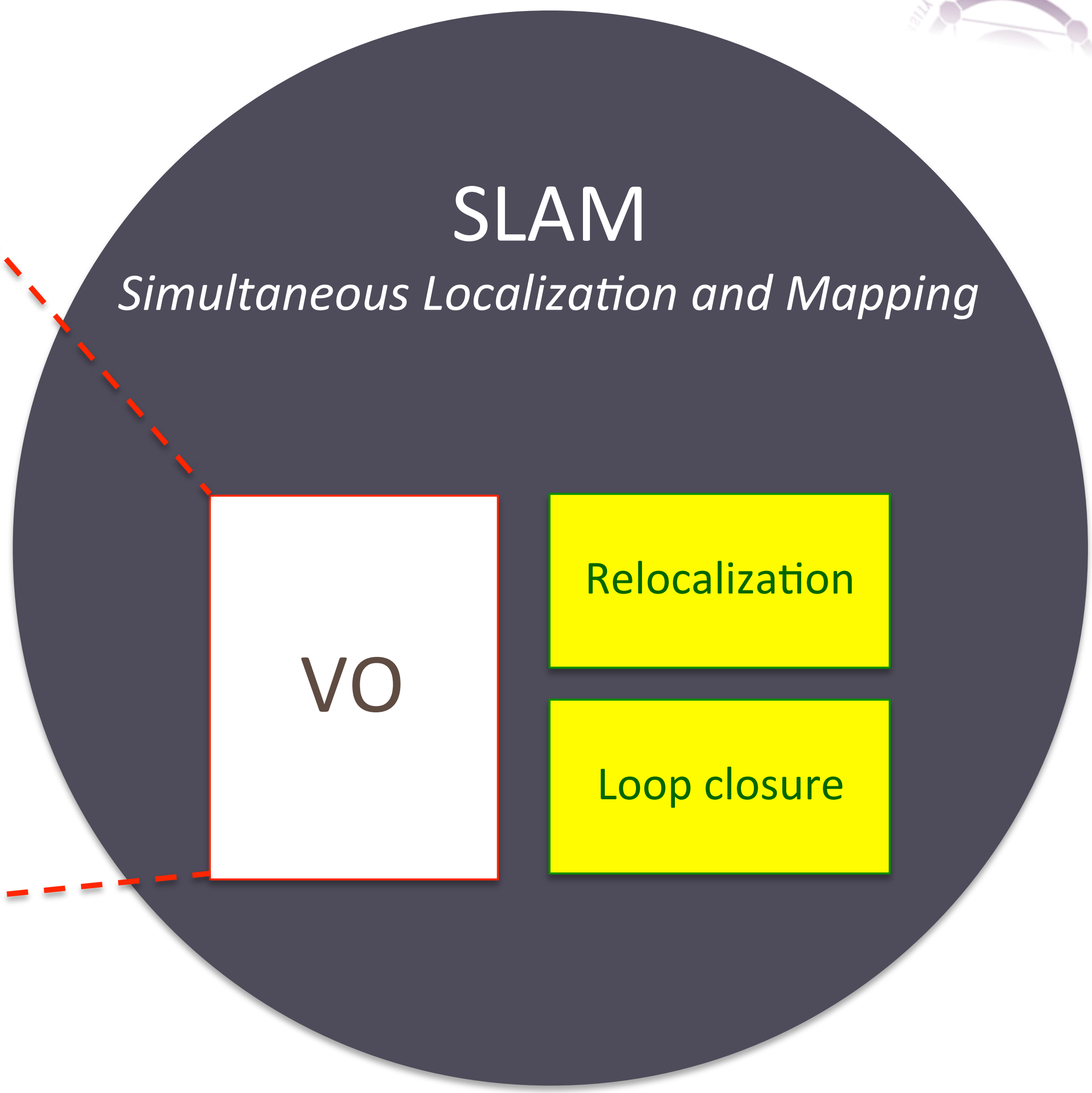
visual odometry
(VO)

=

camera pose
estimation

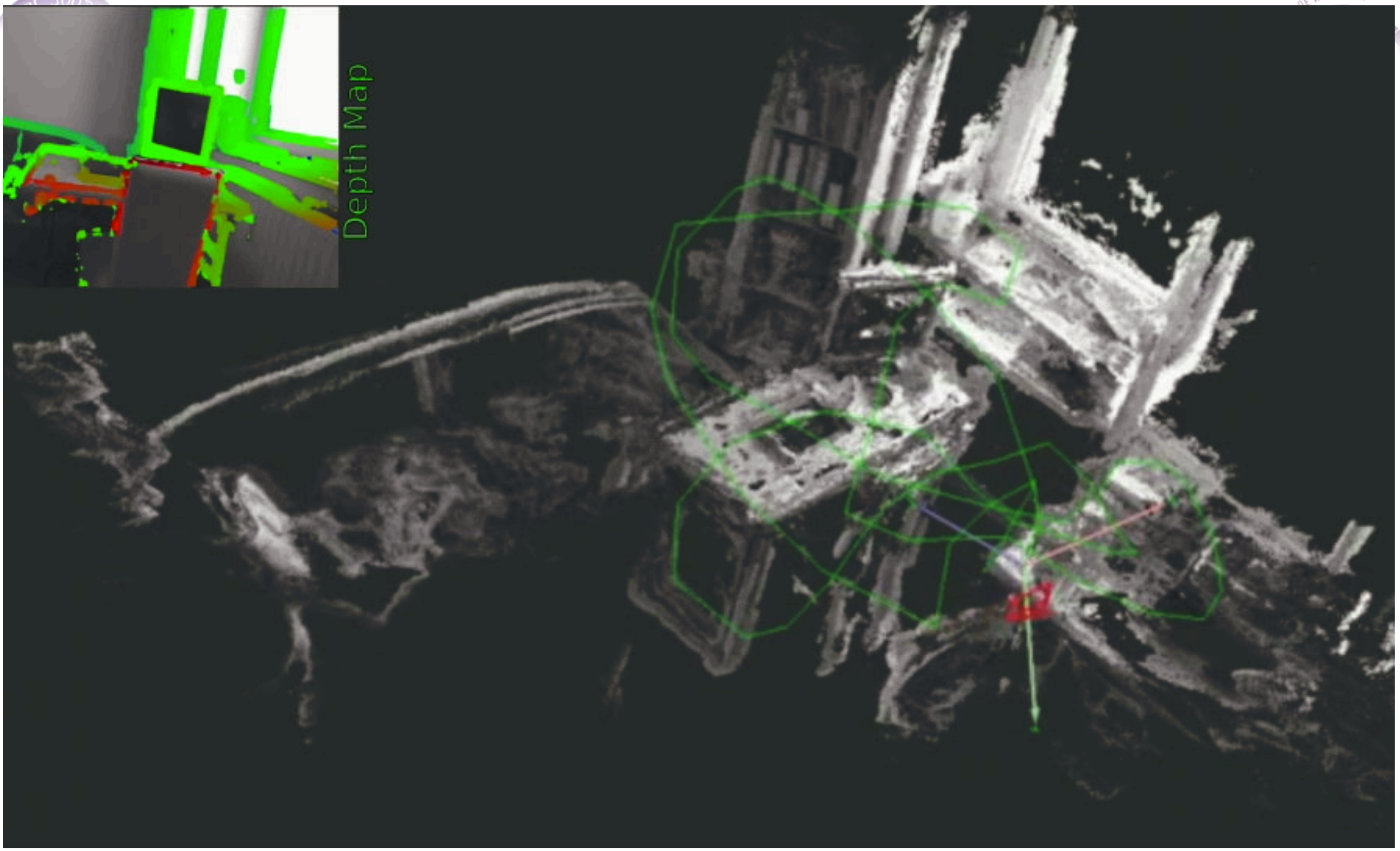
=

camera tracking



Visual Odometry Example

<Intro> <SOS> <GUMS> <VO>



Goal Oriented Contents

<Intro> <SOS> <GUMS> <VO>

1. **Problem Motivation: Omnidirectional Vision**

2. **Design of the Single-Camera Stereo Omnidirectional System (SOS)**

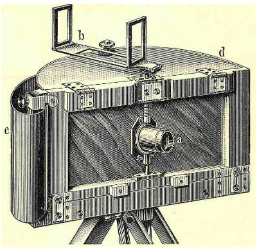
3. **Projection model (GUMS) and calibration**

4. **Visual odometry (VO) with Single-Camera SOS**

Motivation: Omnidirectional Vision

- 3-D world ←→ Surround Vision

- **3-D world ←→ Surround Vision**
- **Eye geometry of insects**
  - 360° *azimuthal field of view.*
- **Technology: Panoramic Vision via**
  - Slit photography (1843)



Puchberger's
first panoramic-camera

View from the top of Lookout Mountain, TN, Albumen prints, February, 1864, by *George N. Barnard*

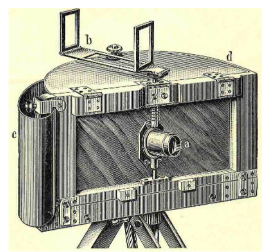
Public Domain, https://commons.wikimedia.org/w/index.php?curid=85275

Motivation: Omnidirectional Vision

<Intro> <SOS> <GUMS> <VO>

- **3-D world ←→ Surround Vision**
- **Eye geometry of insects**
  - 360° *azimuthal field of view.*

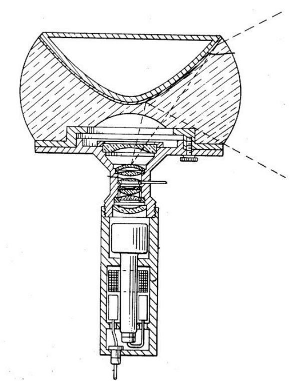
- **Technology: Panoramic Vision via**
  - Slit photography (1843)



Puchberger's first panoramic-camera

  - Catadioptric Sensors (1911)

    Cata (Mirror) + Dioptric (Lenses)



Rees hyperbolic ODVS (1970)

  - Wide-angle Lenses

Metric Range Perception Devices

<Intro> <SOS> <GUMS> <VO>

Active

Time-of-Flight

- Sonars (Ultrasonic)

- Laser Range Finders (LIDAR: Light Detection And Ranging)

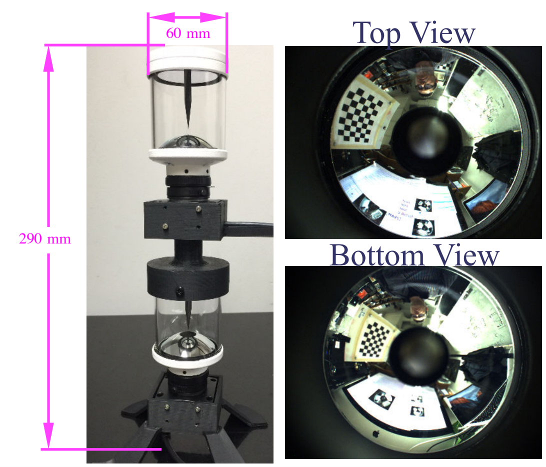
  - SwissRanger 3D camera

- RGB-D Sensors

Passive

- Stereo Cameras

- **Stereo Catadioptrics** (Mirrors + lenses)



SINGLE-CAMERA STEREO OMNIDIRECTIONAL SYSTEM (SOS)

Peer-reviewed articles:

Spherical SOS:

Igor Labutov, **Carlos Jaramillo**, and Jizhong Xiao. "*Generating near-spherical range panoramas by fusing optical flow and stereo from a single-camera folded catadioptric rig.*" Machine Vision and Applications, 24(1):1–12, 9 2011.

Hyperbolic SOS:

Carlos Jaramillo, Roberto G Valenti, Ling Guo, and Jizhong Xiao. "*Design and Analysis of a Single-Camera Omnistereo Sensor for Quadrotor Micro Aerial Vehicles (MAVs).*" *Sensors*, 16(2):217, 1 2016. ISSN 1424-8220

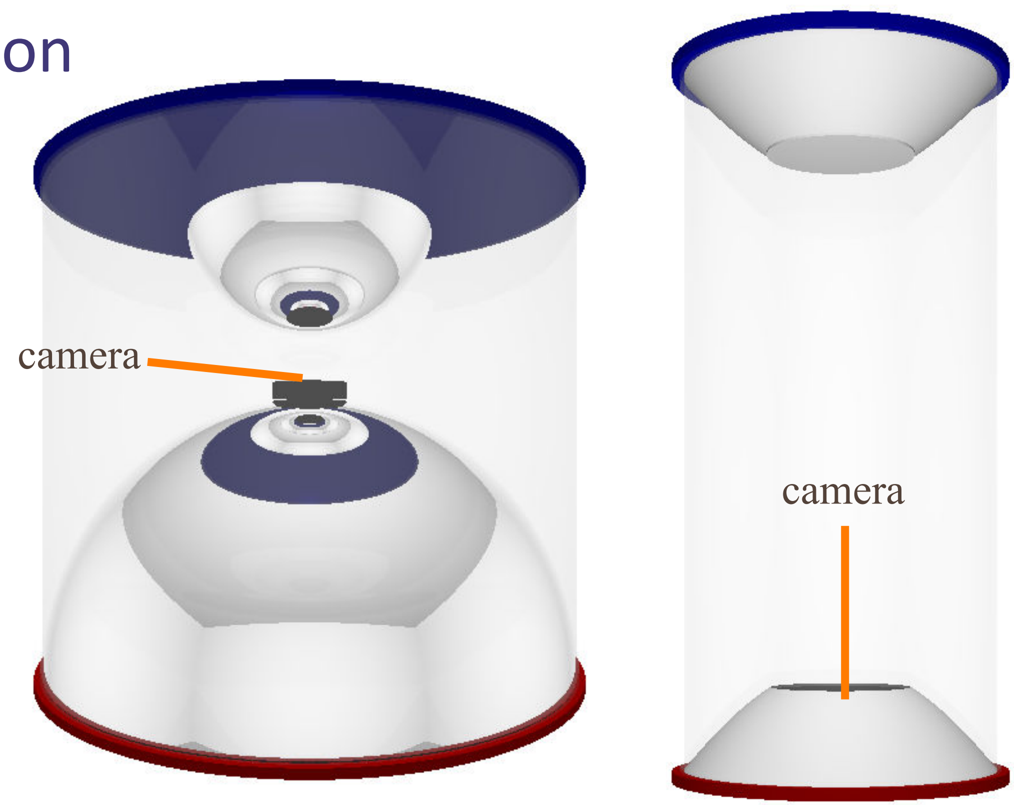
Single-Camera SOS

<Intro> **<SOS>** <GUMS> <VO>

- **Design goal:**
  – Omnidirectional 3D Vision with a single **camera** + two curved **mirrors**

- **Design advantages:**
  – Low cost ($ and ⚡)
  – Light weight (portable)
  – Wide field-of-view
  – Passive sensing

"Folded" SOS Configurations



Spherical mirrors   Hyperbolic mirrors

Single-Camera SOS

Omnistereo Intuition

Top mirror mostly captures the ground

Single-Camera SOS

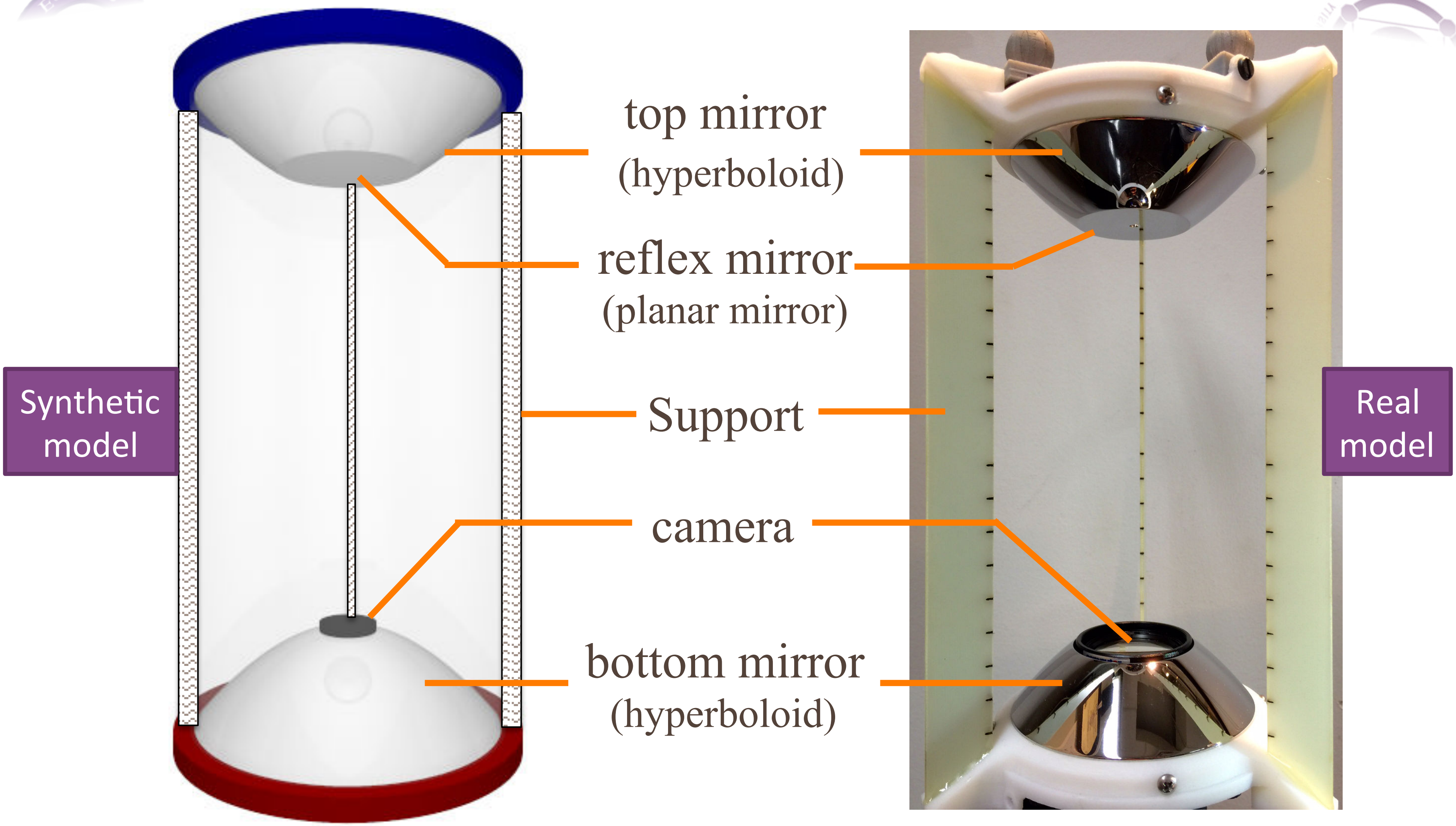
Omnistereo Intuition

Bottom mirror mostly captures the ceiling

Single-Camera SOS

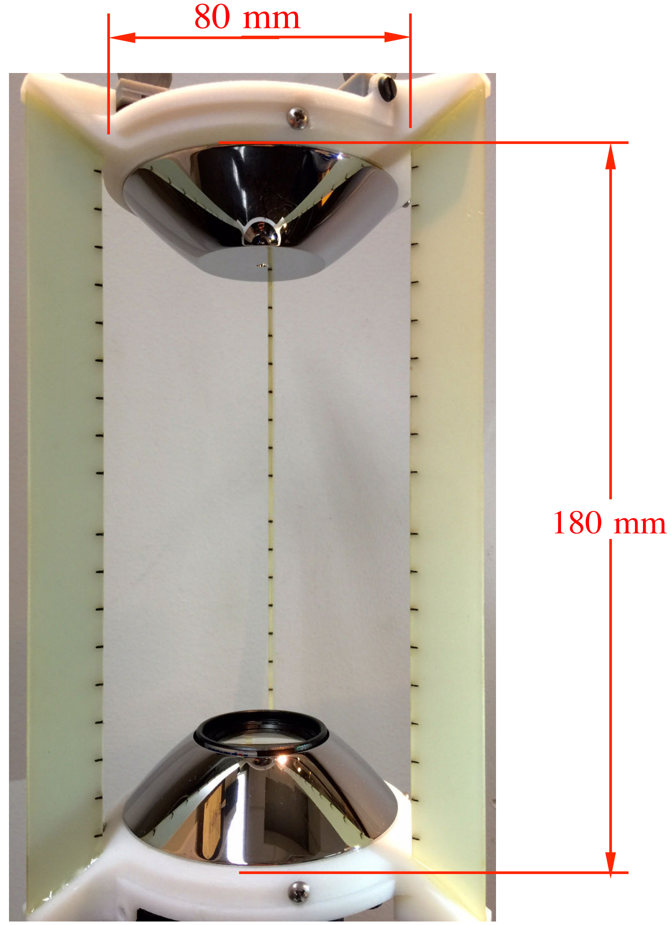
Omnistereo Intuition

Hyperbolic Single-Camera SOS



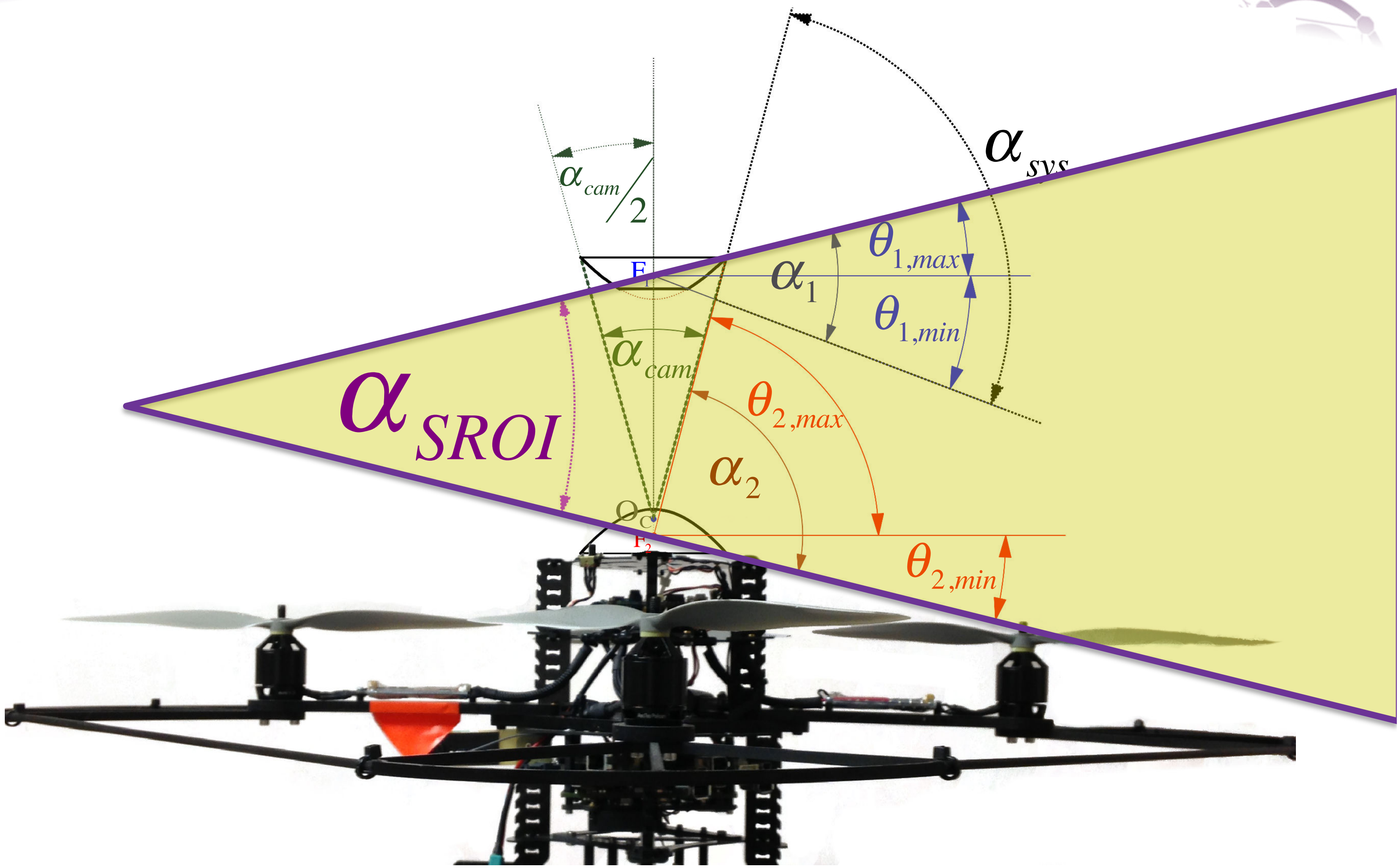
Hyperbolic Single-Camera SOS

- **Design goal:**
  - Omnidirectional stereo vision with **configurable** stereo region (SROI)
- **Design pros:**
  - Low radial distortion
  - Central (SVP), or slighty-central (real)
  - Optimal FOV (SROI) + baseline
- **Design cons**
  - Custom hyperbolic mirrors
  - Hard to assemble
    - cannot satisfy the theoretical SVP
    - misalignment issues

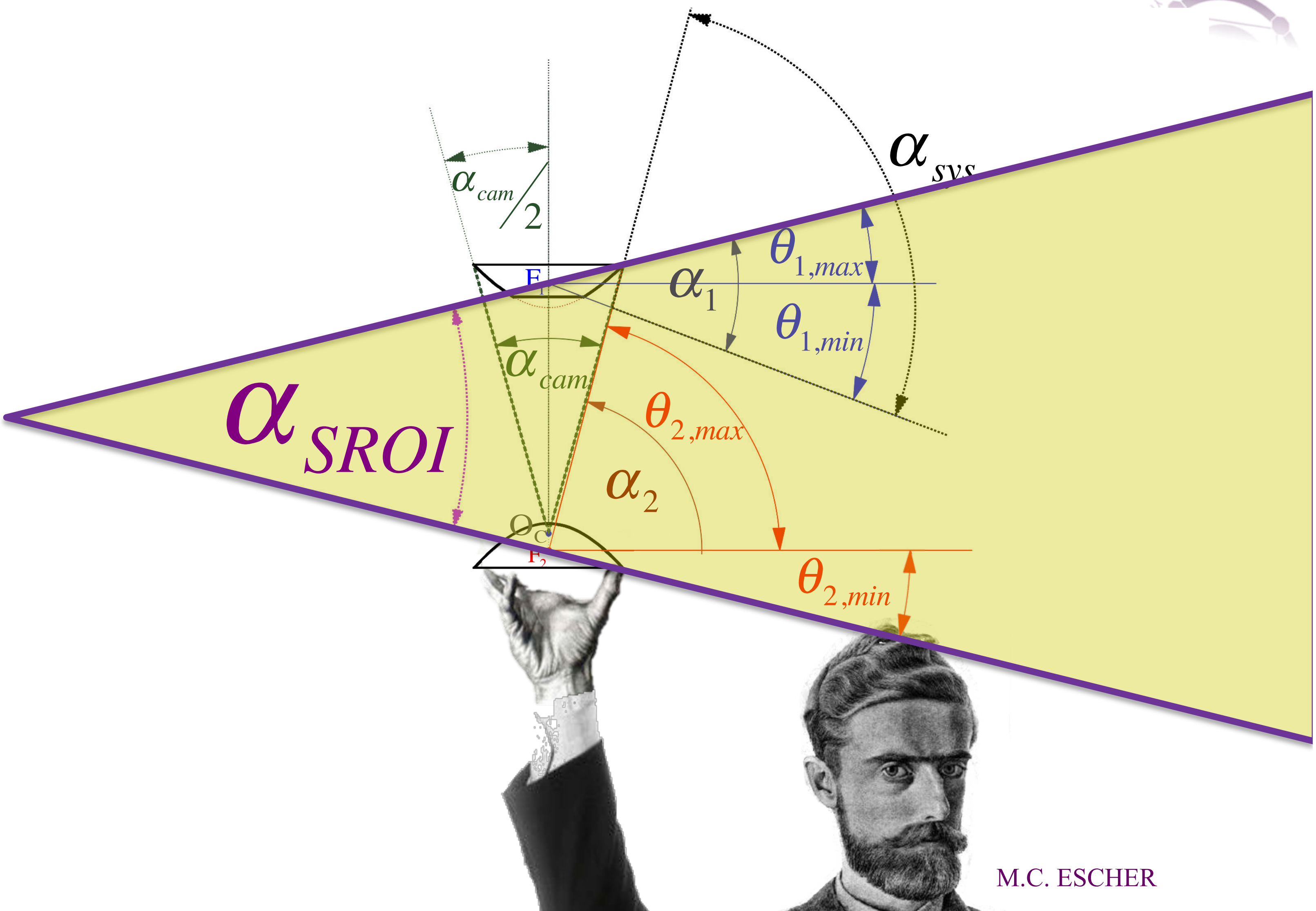


baseline ≈ 150 mm

Hyperbolic SOS: Field-of-View

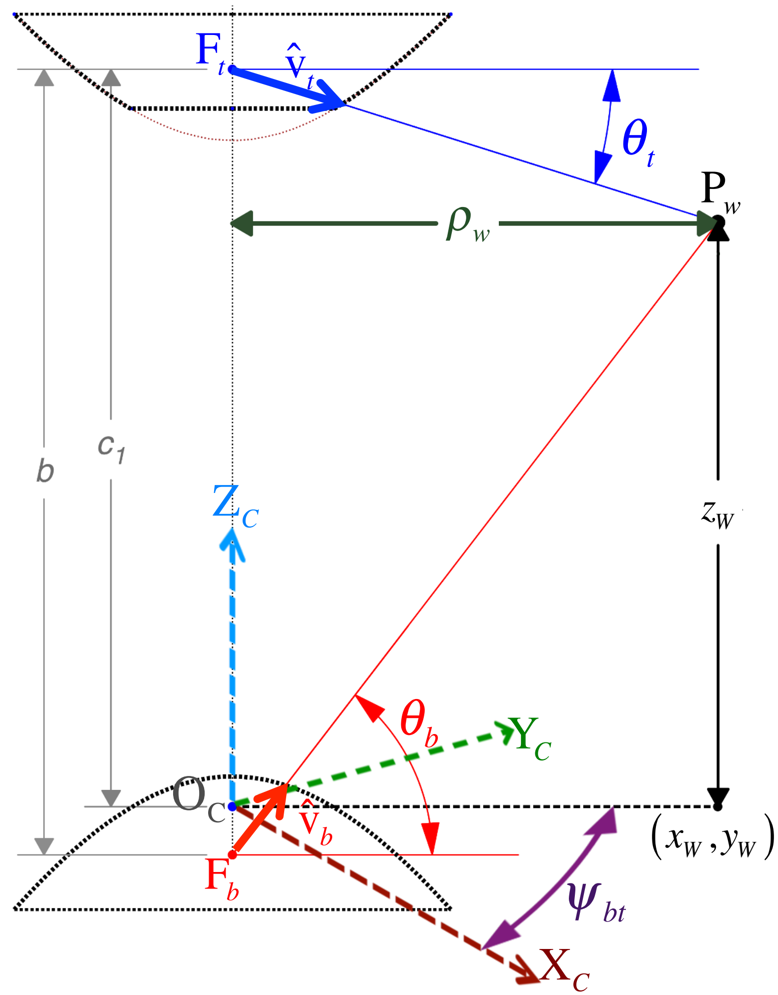


Hyperbolic SOS: Field-of-View



Hyperbolic SOS: Triangulation

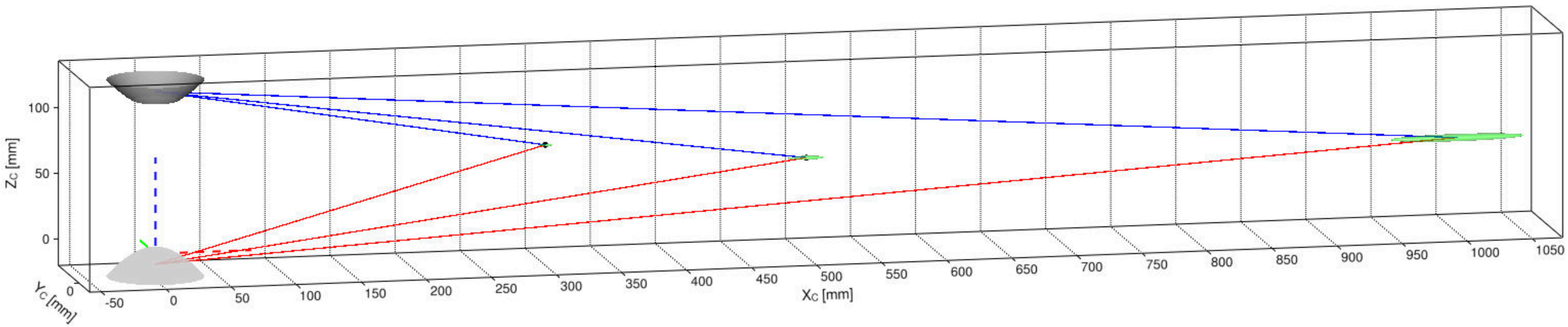
Goal: Back-projected direction vectors $\hat{\mathrm{v}}_t$ and $\hat{\mathrm{v}}_b$ of a point's correspondences in the image are triangulated as P_w



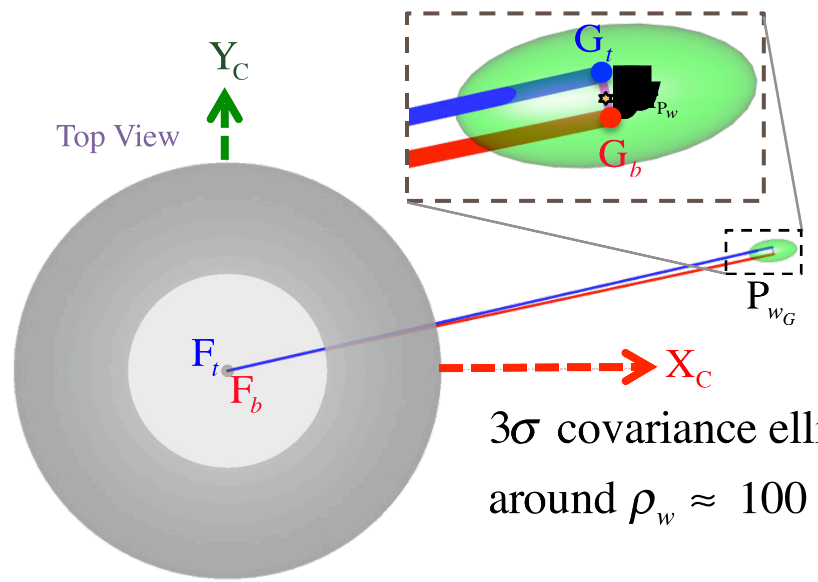
Triangulation & Depth Uncertainty

<Intro> **<SOS>** <GUMS> <VO>

Uncertainty ellipsoids (viz at 1-sigma) for triangulated points at ranges $\rho_w \approx \{0.3, 0.5, 1.0\}$ meters



For a std. dev. of 1 pixel
in the image correspondences



3σ covariance ellipsoid
around $\rho_w \approx$ 100 mm

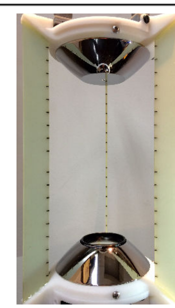
Hyperbolic SOS: Parameters

Rig Geometric Parameters (Theoretical Values)

Parameter	Dimension on Rig Type	
	Near-Sighted	Far-Sighted
r_{sys}[mm]	37.0	40.0
r_{ref}[mm]	17.2	19.0
r_{cam}[mm]	7.0	18.0
b[mm]	131.6	150.0
h_{sys}[mm]	150.0	176.6
α_{sys}[°]	66.8	40.9
α_{SROI}[°]	25.0	33.5
$\rho_{\min}$[mm]	65.0	250.0

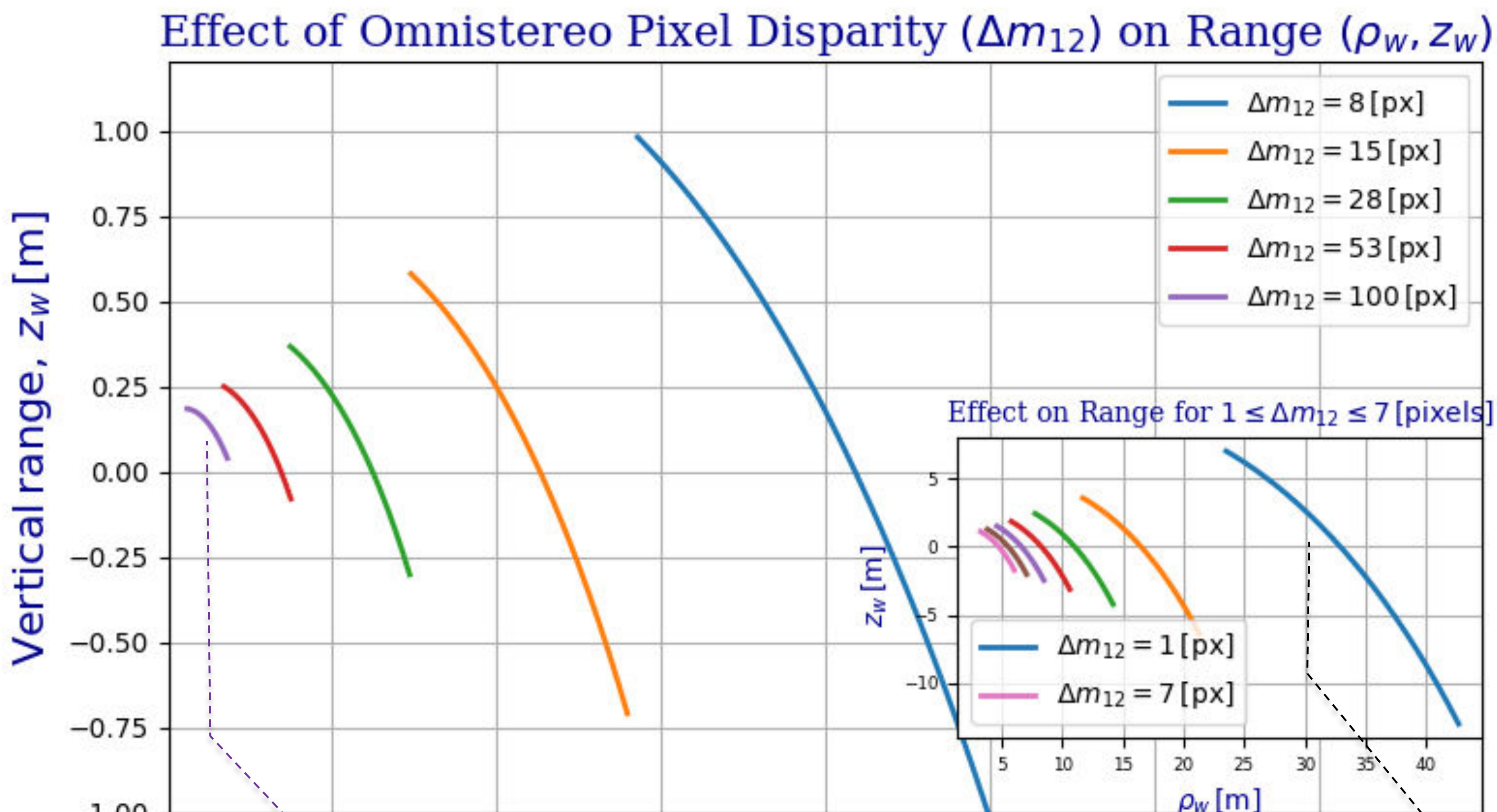
Min Range [mm]

based on SROI near vertices



Hyperbolic SOS: Range Limits

<Intro> **<SOS>** <GUMS> <VO>



min range: ~0.25 m
for <100 px disparity

MAX range: ~30 m
for >1 px disparity

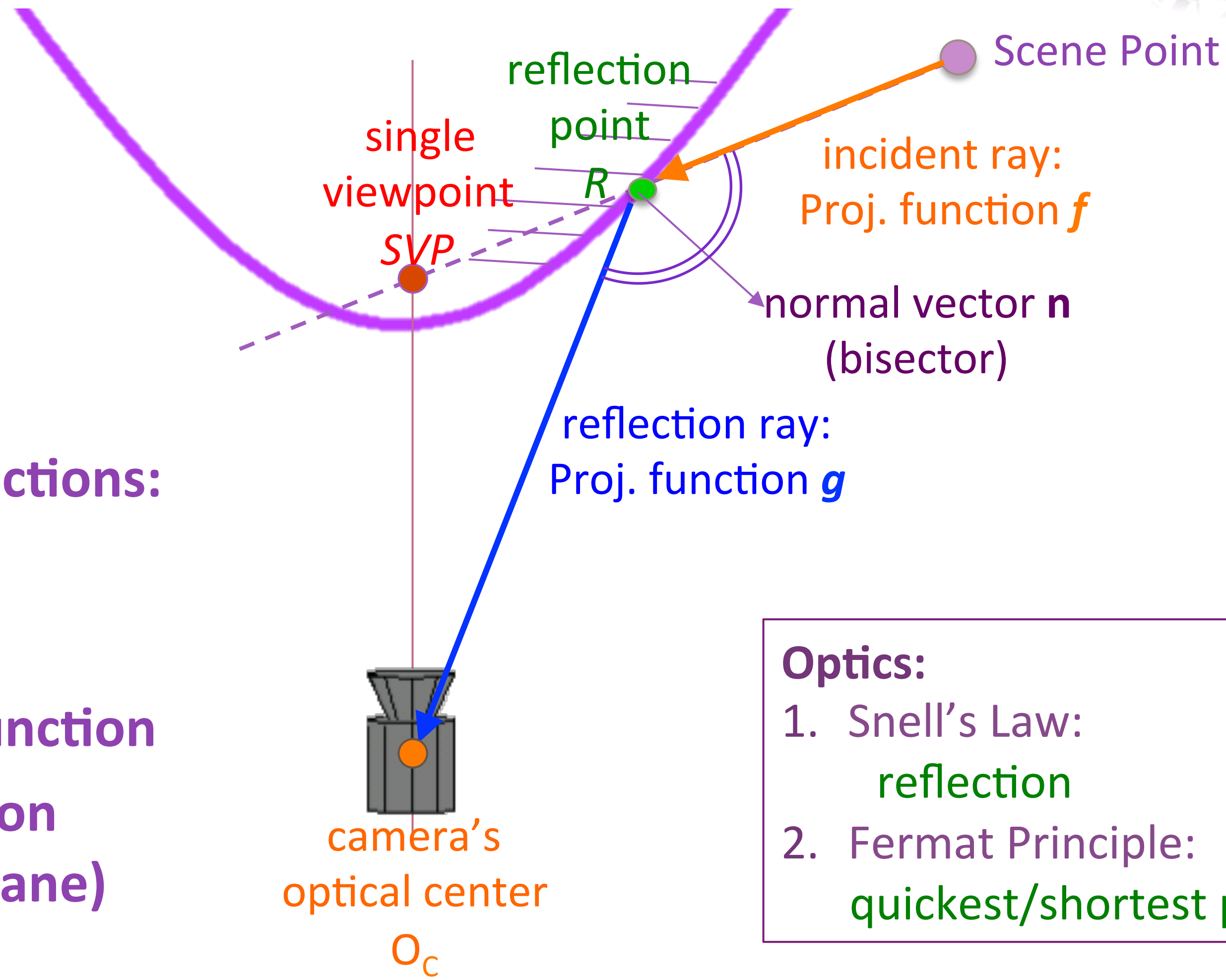
NOTE: using "far-sighted" rig

GUMS:
A GENERALIZED UNIFIED MODEL FOR SOS

Peer-reviewed publication:

Carlos Jaramillo, Roberto G Valenti, and Jizhong Xiao.
"*GUMS: A Generalized Unified Model for Stereo Omnidirectional Vision (Demonstrated Via a Folded Catadioptric System)*."
In IEEE International Conference on Intelligent Robots and Systems (IROS*),*
2016, 2528–33. Korea

Single Viewpoint (SVP) Constraint



Composition of 2 functions:

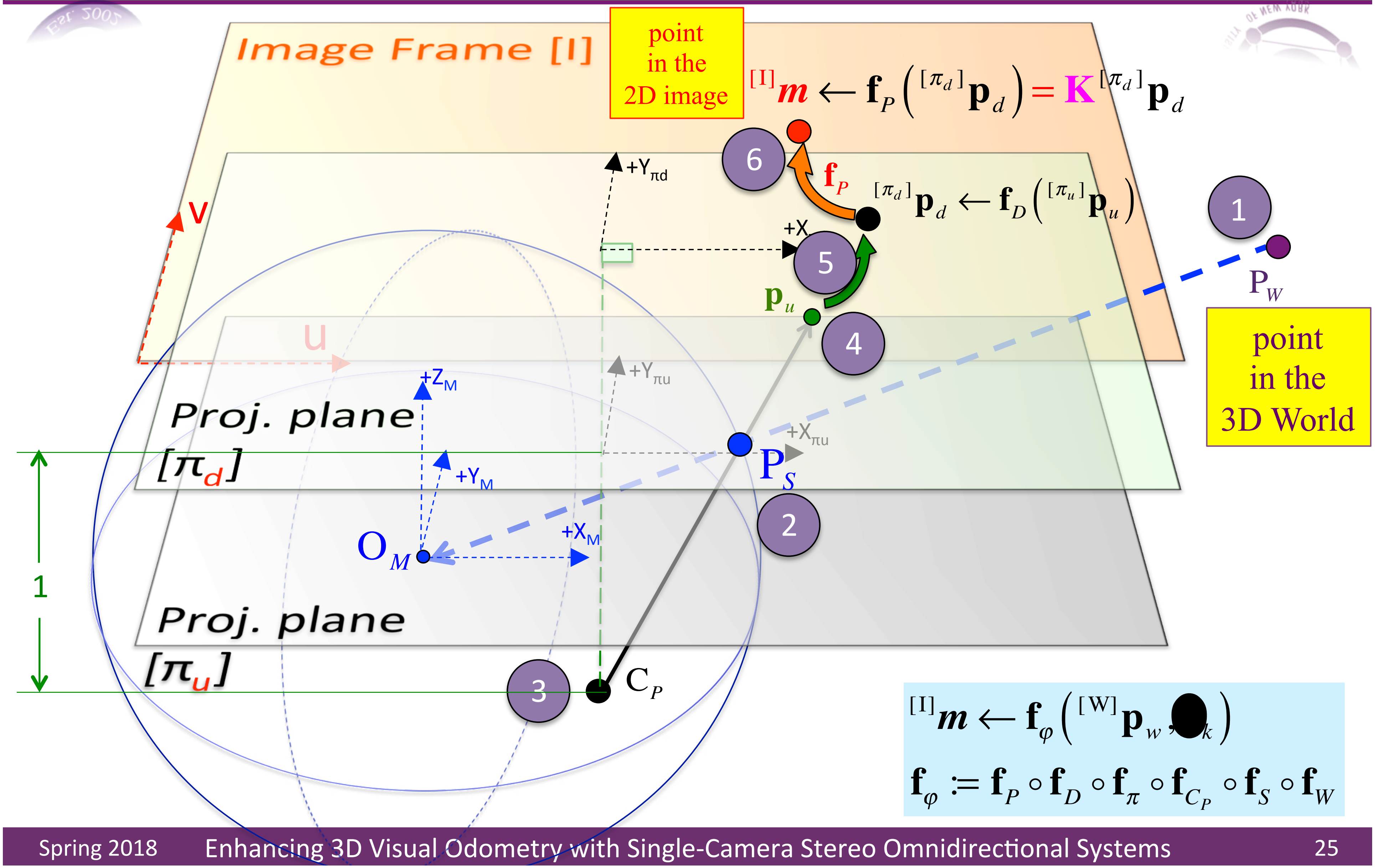
$$g \circ f$$

- ***f*** **is a non-linear function**
- ***g*** **is a linear function (projection to a plane)**

Optics:

1. Snell's Law: reflection
2. Fermat Principle: quickest/shortest path

Generalized Unified Model (GUM)

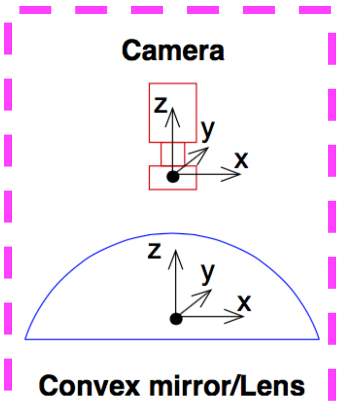


$$^{[I]}m \leftarrow \mathbf{f}_\varphi\left({}^{[W]}\mathbf{p}_w, \text{[redacted]}_k\right)$$

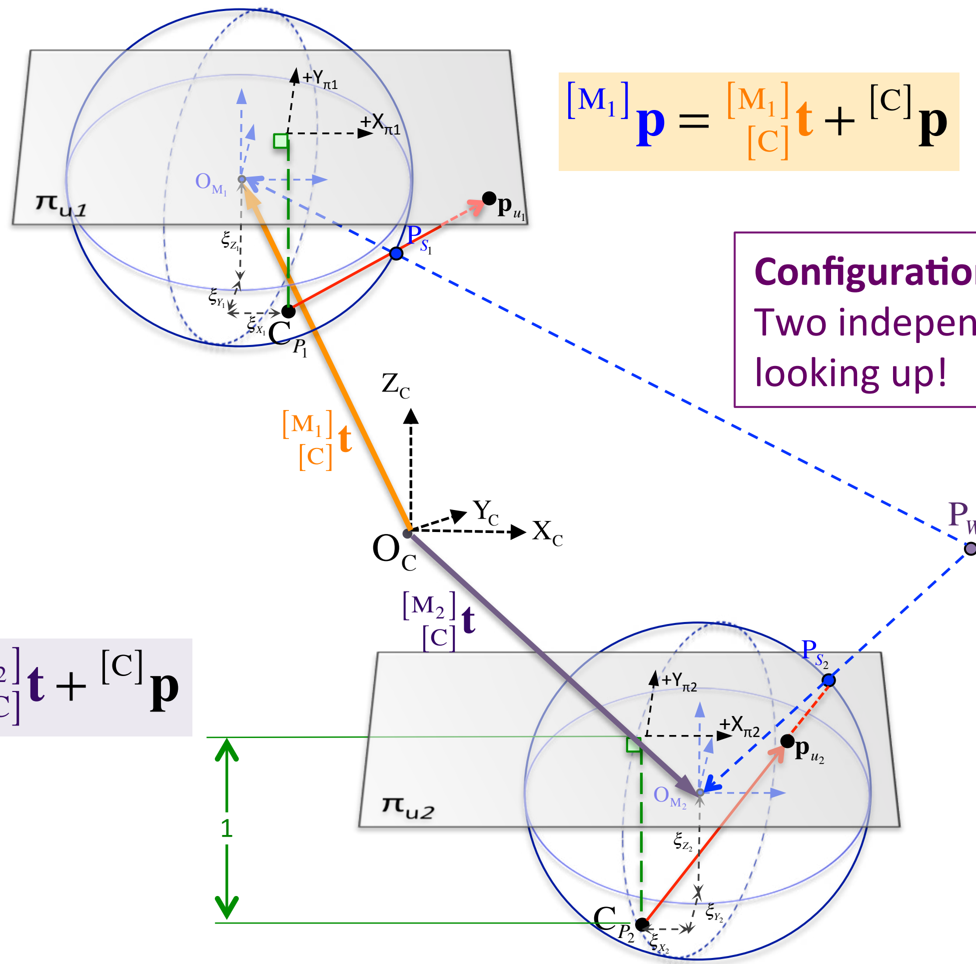
$$\mathbf{f}_\varphi \coloneqq \mathbf{f}_P \circ \mathbf{f}_D \circ \mathbf{f}_\pi \circ \mathbf{f}_{C_P} \circ \mathbf{f}_S \circ \mathbf{f}_W$$

GUMS in a general omnistereo configuration

View 1



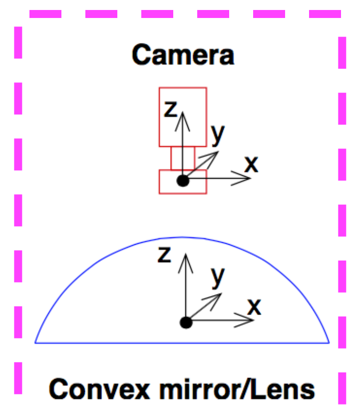
$$^{[M_1]}\mathbf{p} = {}^{[M_1]}_{[C]}\mathbf{t} + {}^{[C]}\mathbf{p}$$



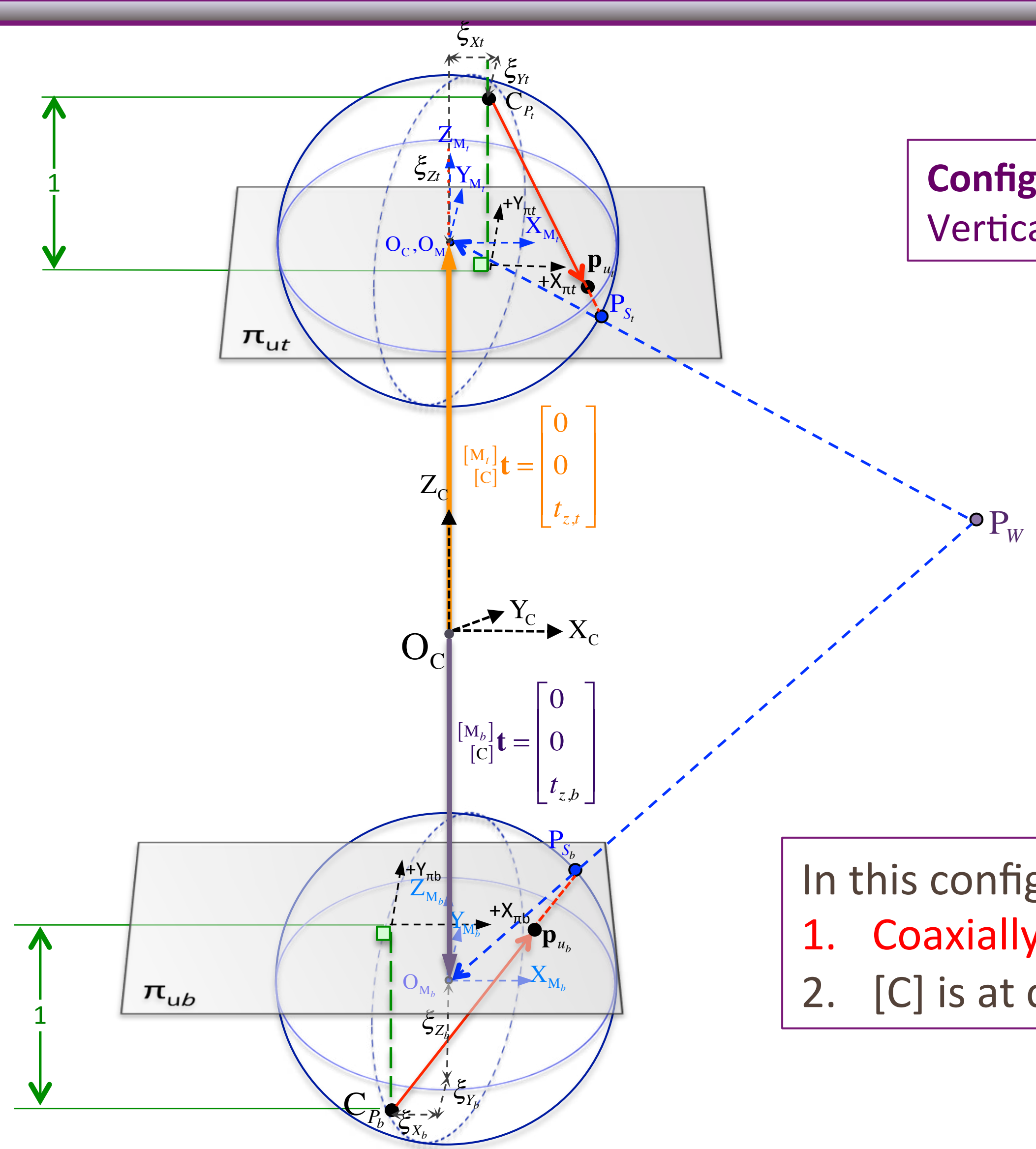
Configuration:
Two independent catadioptric ODVS looking up!

$$^{[M_2]}\mathbf{p} = {}^{[M_2]}_{[C]}\mathbf{t} + {}^{[C]}\mathbf{p}$$

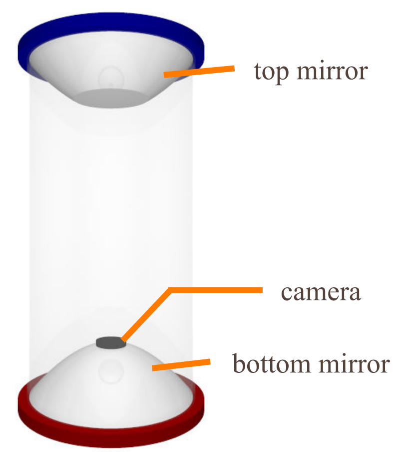
View 2



Generalized Unified Model for SOS



Configuration:
Vertical and *"folded"* omnistereo



In this configuration, **GUMS** is:

1. Coaxially constrained
2. [C] is at coincident with [M_t]

GUMS Parameters

For the coaxial-alignment constraint:

Extrinsic parameters: $^{[\mathrm{M}_k]}_{[\mathrm{C}]}\mathbf{t} = \begin{bmatrix} 0 \\ 0 \\ t_{z,k} \end{bmatrix}$ $\quad$ $\text{for } k = \{t, b\}$

Intrinsic parameters:

[redacted] $_k = [\xi_X, \xi_Y, \xi_Z]_k;$ $\quad$ $\mathbf{d}_k = [d_1, d_2, d_3]_k;$ $\quad$ $\mathbf{c}_k = [\alpha, \gamma_1, \gamma_2, u_c, v_c]_k$

generalized camera's intrinsic parameters

We get: $\mathbf{x}_k = \begin{bmatrix} \xi_k, & \mathbf{d}_k, & \mathbf{c}_k \end{bmatrix}_{(1\times 11)}, \quad \text{for } k = \{t, b\}$

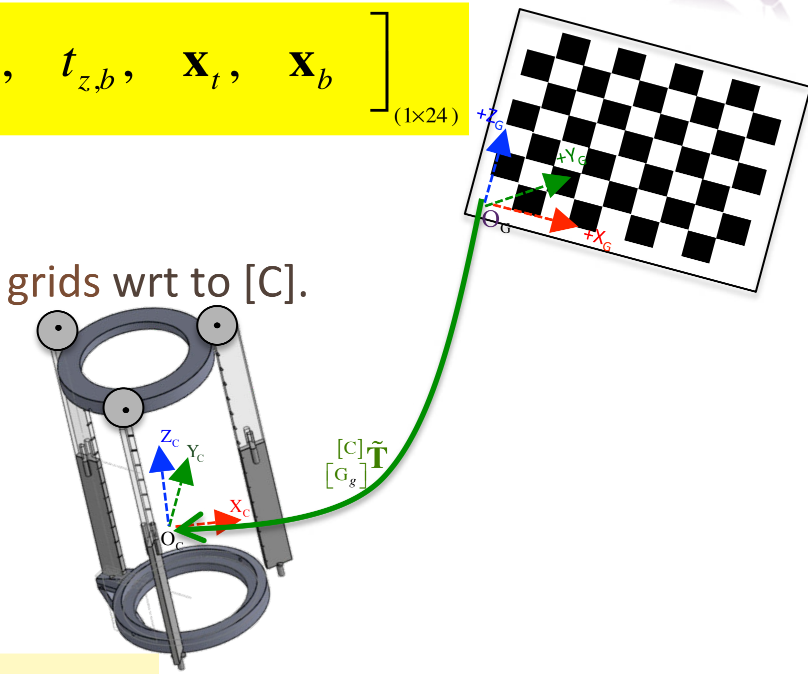
This coaxial **GUMS** has 24 parameters: [redacted] $_{GUMS} = \begin{bmatrix} t_{z,t}, & t_{z,b}, & \mathbf{x}_t, & \mathbf{x}_b \end{bmatrix}_{(1\times 24)}$

GUMS Parameters (Calibration)

This coaxial **GUMS** has 24 parameters: [redacted]$_{GUMS} = \begin{bmatrix} t_{z,t}, & t_{z,b}, & \mathbf{x}_t, & \mathbf{x}_b \end{bmatrix}_{(1\times 24)}$

We add the "extrinsic" transformations of the *L* planar grids wrt to [C].

$${}^{[C]}_{[G_g]}\mathbf{T} = \begin{bmatrix} {}^{[C]}_{[G_g]}\mathbf{R}, & {}^{[C]}_{[G_g]}\mathbf{t} \end{bmatrix}, \forall g \in \{1, \ldots, L\}$$



Using rotation quaternions:

$$\mathbf{g}_g = \begin{bmatrix} {}^{[C]}_{[G_g]}\mathbf{q}, & {}^{[C]}_{[G_g]}\mathbf{t} \end{bmatrix} = {}^{[C]}_{[G_g]}\begin{bmatrix} \hat{q}_0, \hat{q}_1, \hat{q}_2, \hat{q}_3, t_x, t_y, t_z \end{bmatrix}$$

quaternions of form: $\mathbf{q} := \begin{bmatrix} q_0, & (q_1, q_2, q_3) \end{bmatrix}$

also represented by: $\mathbf{q} := q_0 + q_1\hat{\mathbf{i}} + q_2\hat{\mathbf{j}} + q_3\hat{\mathbf{k}}$

This **GUMS** has 24+7*L* parameters for **calibration**:

$$\mathbf{v}_{tb} = \begin{bmatrix} \{\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_L\}, & t_{z,t}, & t_{z,b}, & \mathbf{x}_t, & \mathbf{x}_b \end{bmatrix}_{(1\times(24+7L))}$$

Calibration: Explained for GUMS

Goal: To estimate a vector of parameters $\mathbf{v}_{tb} = \left[\{\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_L\}, \quad t_{z,t}, \quad t_{z,b}, \quad \mathbf{x}_t, \quad \mathbf{x}_b \right]_{(1\times(24+7L))}$ such that it minimizes an objective function *J* that accumulates the *squared reprojection error*

NOTE: For L grid patterns of N points, **now** we have **4**xNxL points on the image.

$$\mathbf{v}_{tb}^{*} = \underset{\mathbf{v}_{tb}}{\arg\min}(J), \text{ where } J(\mathbf{v}) := \frac{1}{2} \sum_{k=[t,b]} \sum_{g=1}^{L} \sum_{i=1}^{N} f_m\left(\tilde{\boldsymbol{m}}_{ikg}, \boldsymbol{m}_{ikg}\right)^2$$

$\tilde{\boldsymbol{m}}_{ikg} \leftarrow \mathbf{f}_{\varphi_k}\left({}^{[C]}\tilde{\mathbf{p}}_{ig}, \text{[redacted]}_k\right)$, where ${}^{[C]}\tilde{\mathbf{p}}_{ig} \leftarrow \mathbf{f}_g\left({}^{[G_g]}\mathbf{p}_i, \tilde{\mathbf{g}}_g\right)$

$$\text{where } f_m\left(\tilde{\boldsymbol{m}}_{ikg}, \boldsymbol{m}_{ikg}\right) := \sqrt{\left(\tilde{u}_{ikg} - u_{ikg}\right)^2 + \left(\tilde{v}_{ikg} - v_{ikg}\right)^2}$$

$\boldsymbol{m}_{ikg}$ is the **true** image position of corner point *i* in its pattern view *g* (from corner detection)

$\mathbf{f}_{\varphi_k}$ is the projection function to estimate $\tilde{\boldsymbol{m}}_{ikg}$ image coordinates via GUM *k* corresponding to point ${}^{[W]}\mathbf{p}_{ig}$ from grid pattern *g* using [redacted] $_k = \left[\tilde{t}_{z,k}, \quad \tilde{\mathbf{x}}_k \right]_{(1\times12)}$ parameters

Recall the projection function:

$${}^{[I_k]}\boldsymbol{m}_{ig} \leftarrow \mathbf{f}_{\varphi_k}\left({}^{[W]}\mathbf{p}_i, \text{[redacted]}_k\right) := \mathbf{f}_P \circ \mathbf{f}_D \circ \underbrace{\mathbf{f}_\pi \circ \mathbf{f}_{C_P} \circ \mathbf{f}_S}_{\mathbf{f}_H} \circ \overbrace{\mathbf{f}_W}^{\mathbf{f}_M \circ \mathbf{f}_G}$$

Results from a Misaligned SOS

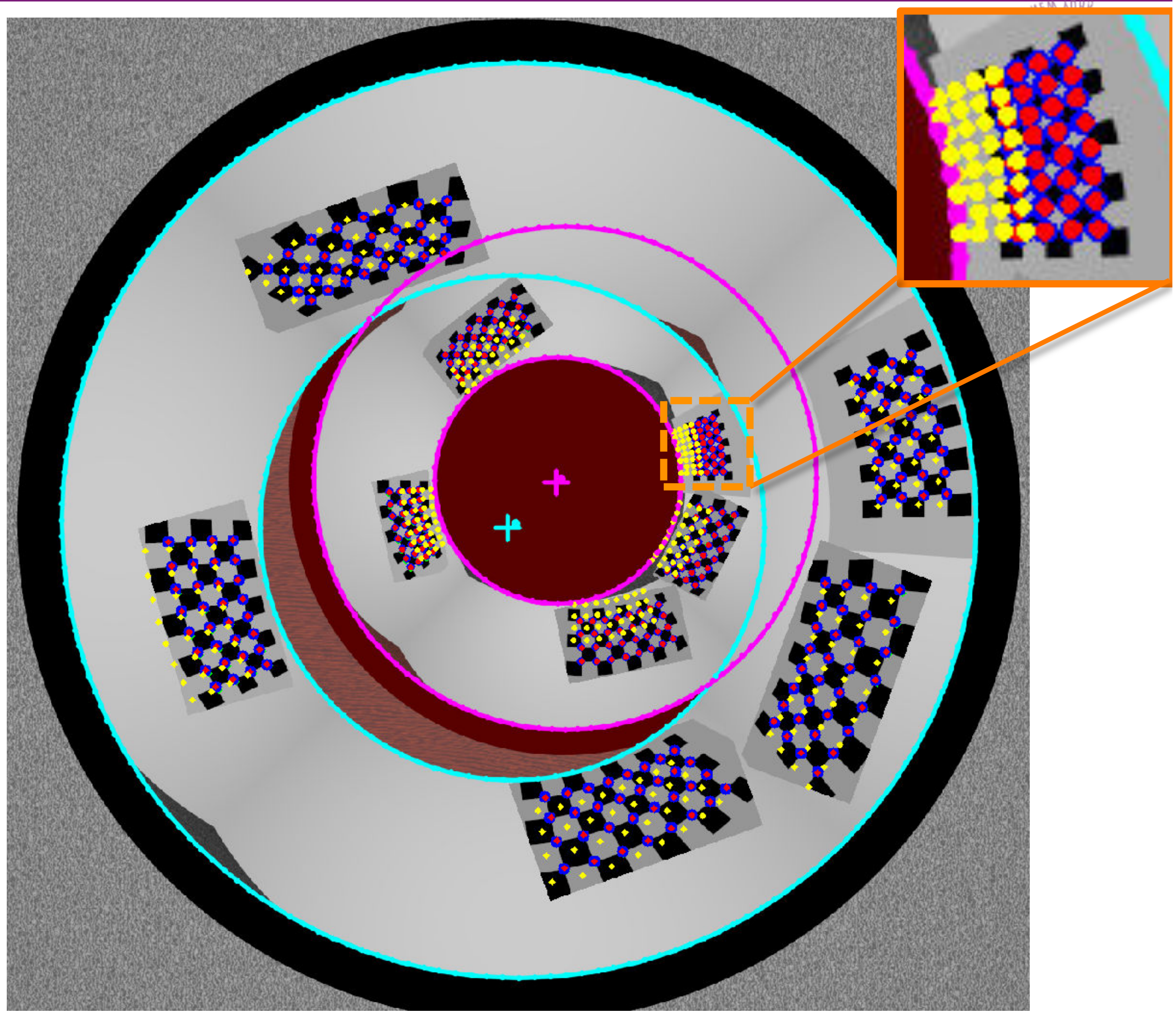
misaligned synthetic case

Projection Error

Cyan: top mirror bounds
Mag: bottom mirror bounds
Centers:
- • initial
- ✚ calibrated

Blue: **true** detected points ($\boldsymbol{m}$)
Green: ground-truth pose
RMSE = 2.24 [pixels]
Red: estimated pose
RMSE = 0.39 [pixels]
Yellow: initial pose
RMSE = 21.42 [pixels]

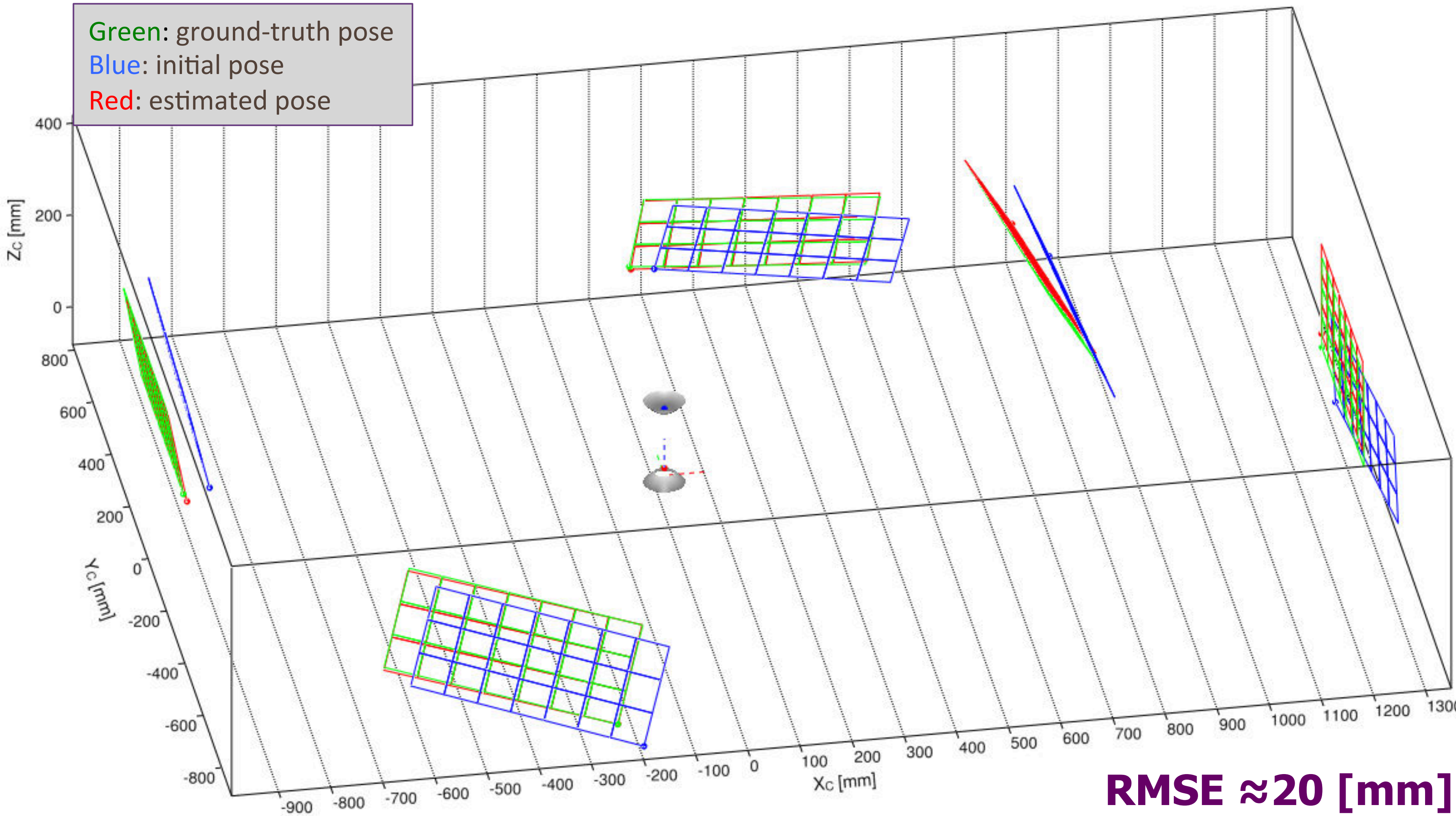
(Green, Red, Yellow: $\tilde{m}$)



Results from a Misaligned SOS

<Intro> <SOS> **<GUMS>** <VO>

Estimated grid poses: misaligned synthetic SOS



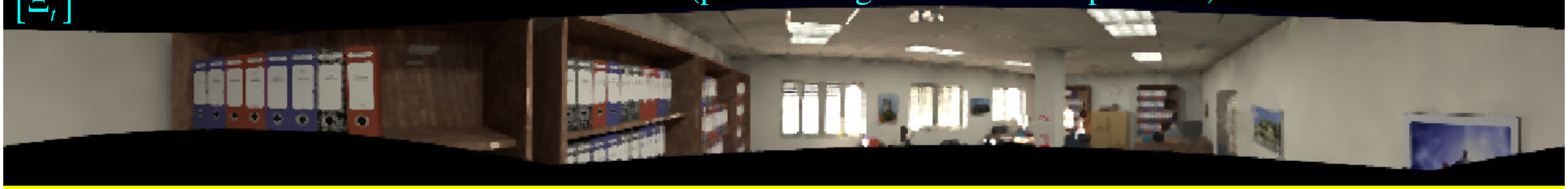
RMSE ≈20 [mm]

Calibrated Omnistereo Panoramas

<Intro> <SOS> **<GUMS>** <VO>

$\left[\Xi_{\Delta m_{tb}}\right]$ Depth Map Image (using Semi-Global Block Matching)

$\left[\Xi_{t}\right]$ Reference Panorama (provides higher view via Top Mirror)



$\left[\Xi_{b}\right]$ Target Panorama (provides lower view via Bottom Mirror)

circular
bottom view
(masked)

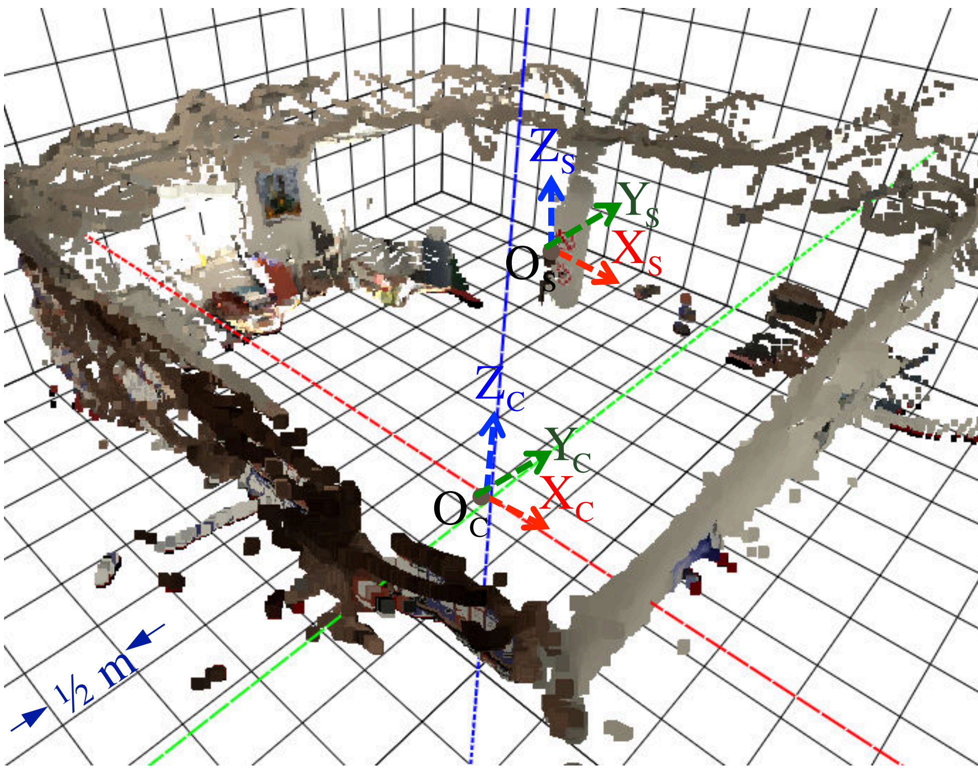
$$^{[\Xi_b]}m_b = \begin{bmatrix} u_t \\ v_t + \Xi_{\Delta m_{tb}} \end{bmatrix} \Bigg|_{^{[\Xi_t]}m_t}$$

from misaligned synthetic experiment

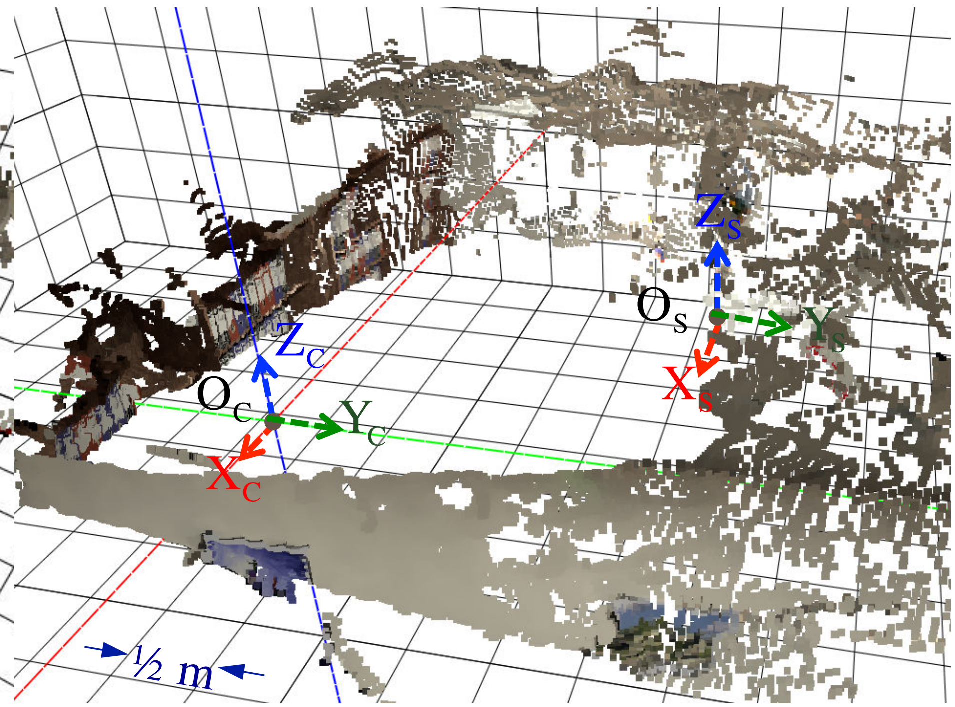
Calibrated Omnistereo 3D Points

3D Views

with Perfect Alignment



with Misalignments



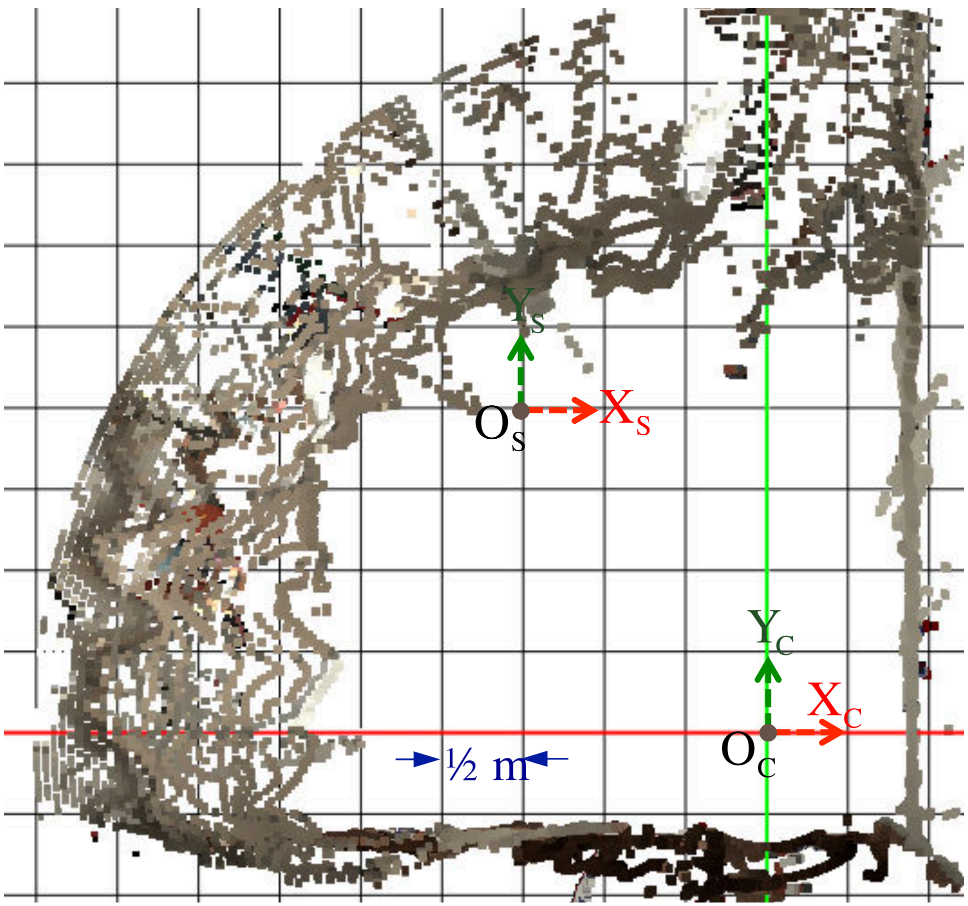
NOTE: The screenshot is taken from different viewpoints

from synthetic experiments

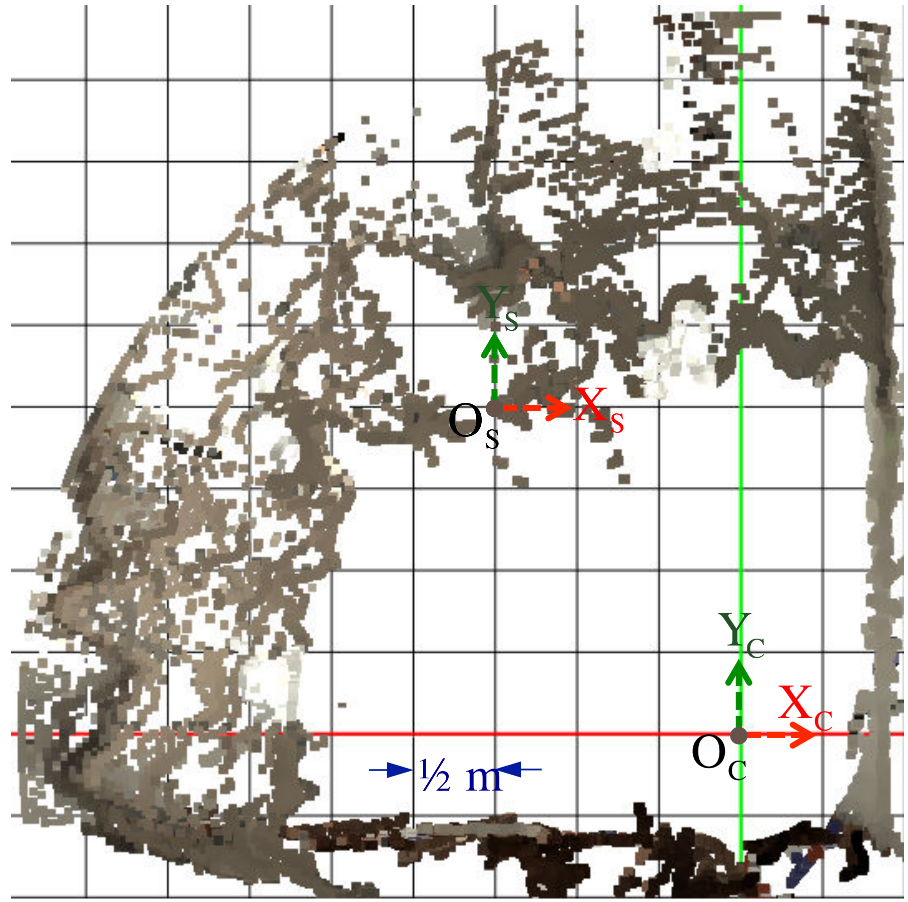
Calibrated Omnistereo 3D Points

Orthographic Views

with **Perfect Alignment**

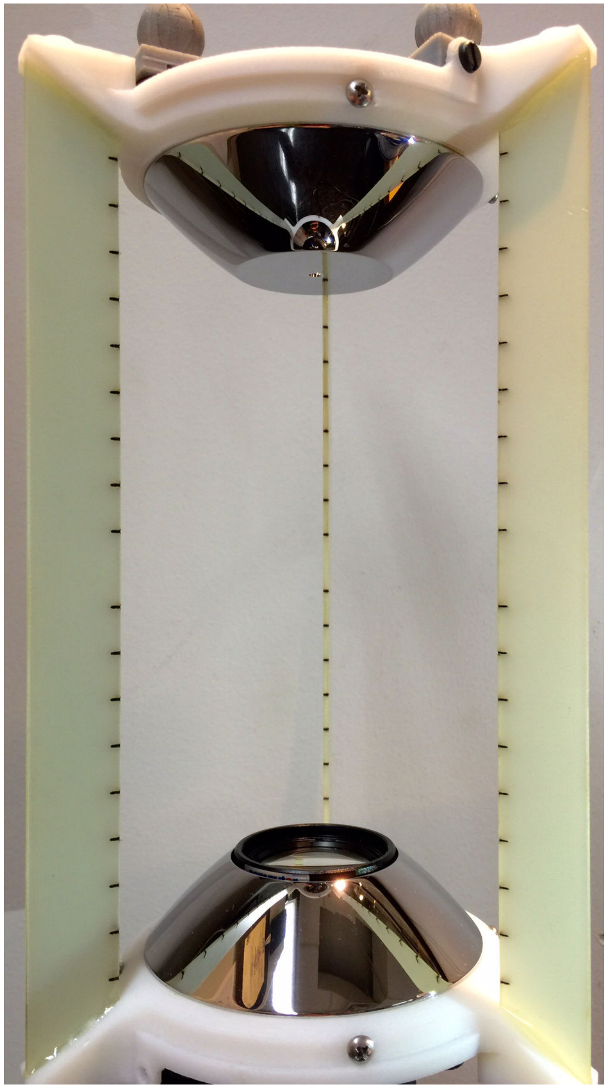


with **Misalignments**



from **synthetic** experiments

Real-Life Calibration Experiments



Calibration Approach Comparison

<Intro> <SOS> **<GUMS>** <VO>

RMSE for Real-Life Calibration Experiments

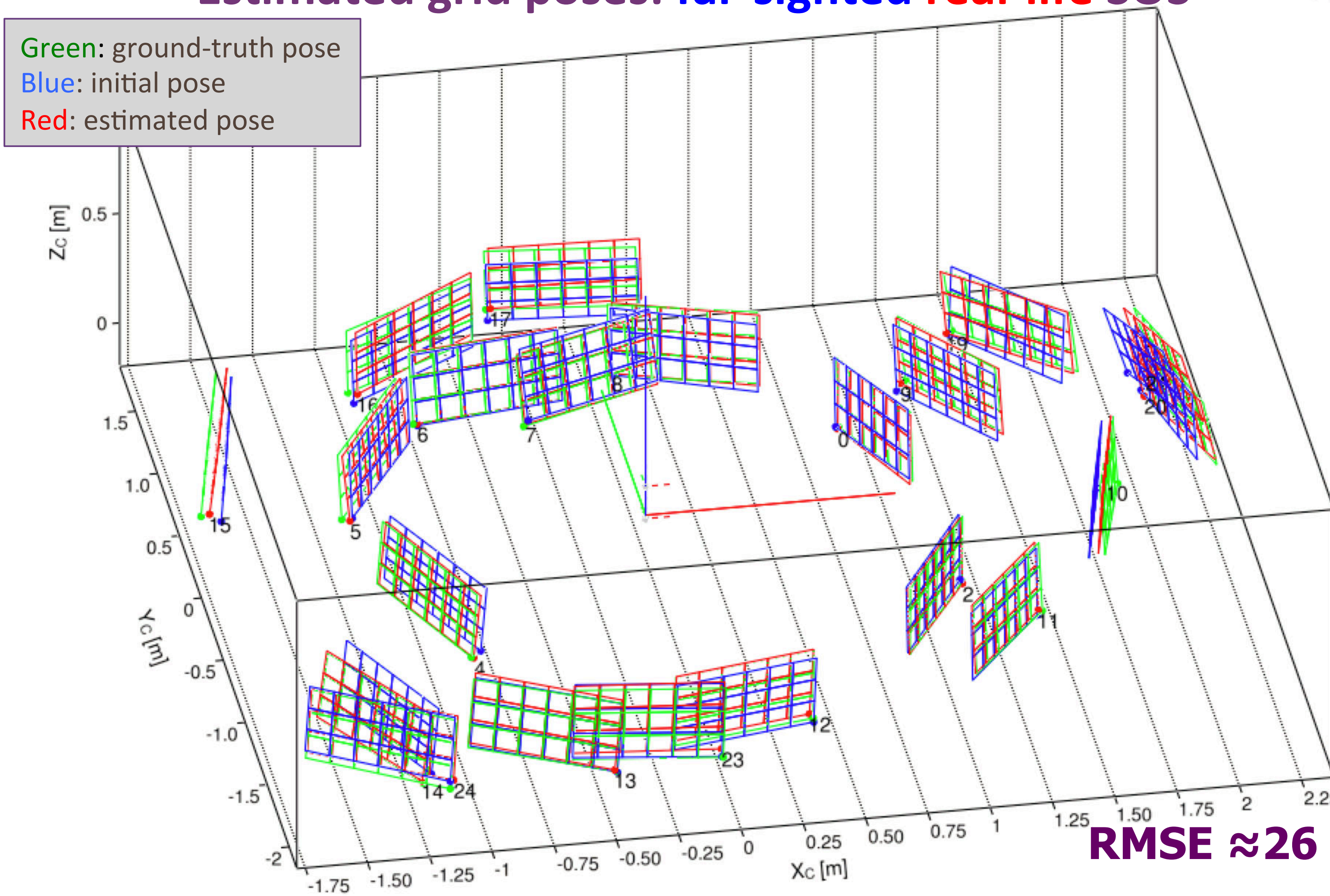
Sight	Coupled	2D Error [px]		3D Error from GT [mm]	
		GT	no GT	Triang.	[G] Pose
Near	Yes	**4.86**	**2.56**	**6.06**	**3.66**
	No	11.02	10.94	22.34	4.64
Far	Yes	**3.71**	**1.64**	**126.81**	**26.38**
	No	5.65	3.27	267.69	45.38

GUMS **coupled** approach **reduces** the overall error

Newest hyperbolic single-camera SOS

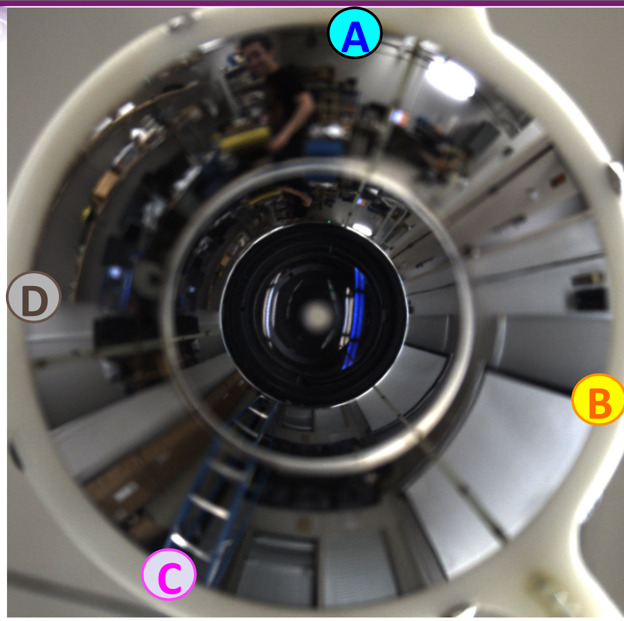
3D Calibration Results for Real SOS

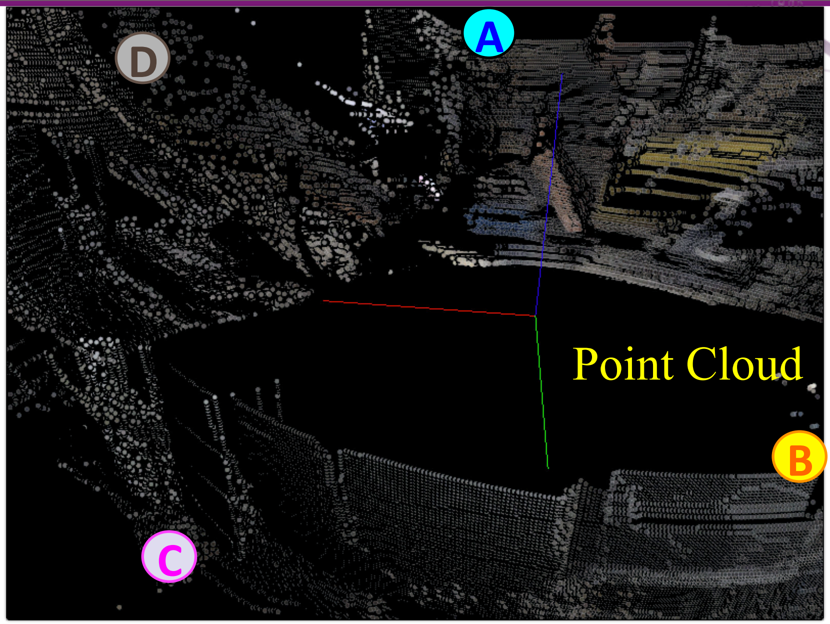
Estimated grid poses: far-sighted real-life SOS

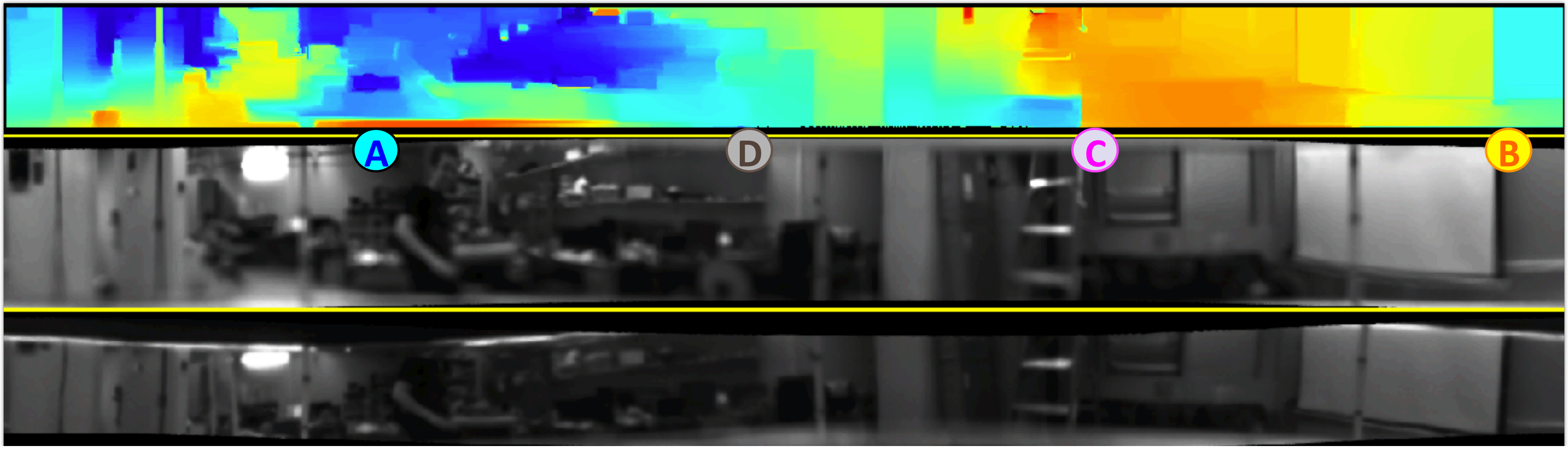


RMSE ≈26 [mm]

Real-Life Calibrated Example







VISUAL ODOMETRY WITH A SINGLE-CAMERA SOS

Publication under review:

Carlos Jaramillo, Liang Yang, J. Pablo Muñoz, Yuichi Taguchi, and Jizhong Xiao.
"*Visual Odometry with a Single-Camera Stereo Omnidirectional System*."
In IEEE Robotics and Automation Letters. Received in May 2018

Visual Odometry in General

VO Definition: Estimation of the 3D pose of a vision sensor's frame wrt a reference frame

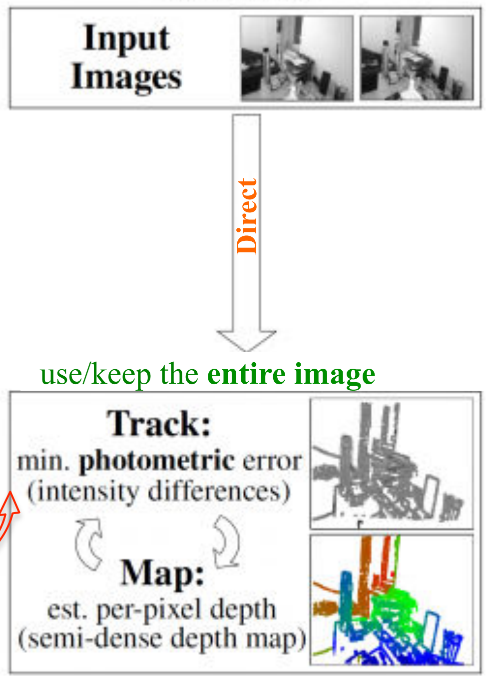
- **The reference frame:**
  - A current tracking keyframe chosen heuristically based on some criteria
  - The world or map frame (absolute pose)
- **Tracking methods:**

Lucas-Kanade algorithm:

brightness constraint

$$\mathbf{x}' = \mathbf{w}(\mathbf{x}; \xi)$$

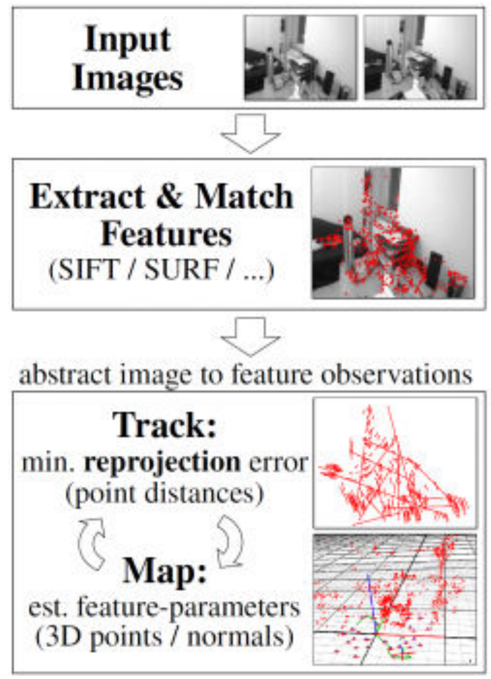
Direct



As done for: LSD-SLAM, DSO, & **DMT**

VS.

Feature-based

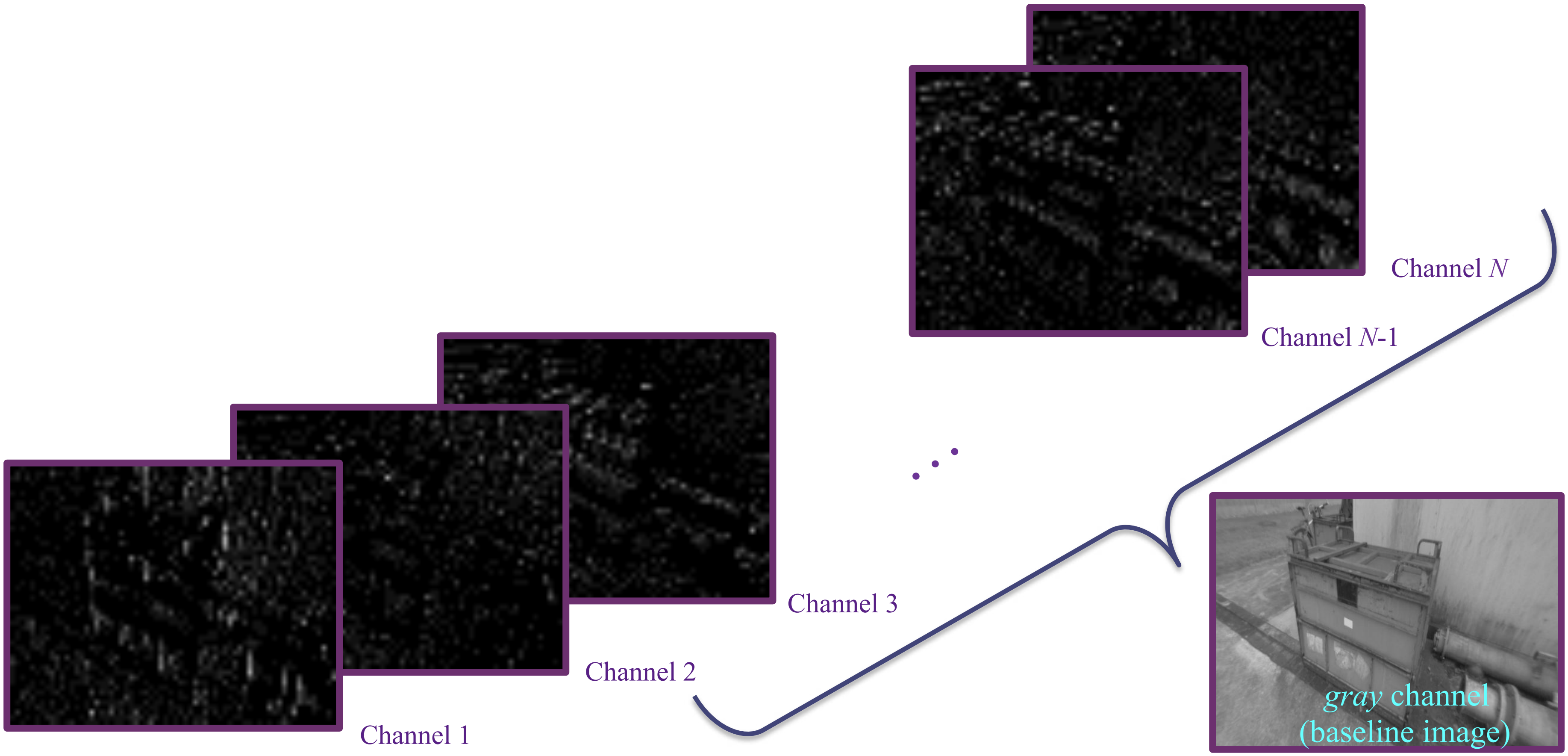


As done for: PTAM, ORB SLAM, etc.

Direct Multichannel Tracking

- **Direct Approach** (*not presented here*)**:**

Direct Tracking of Multi-dimensional Features



Feature-Based VO with SOS

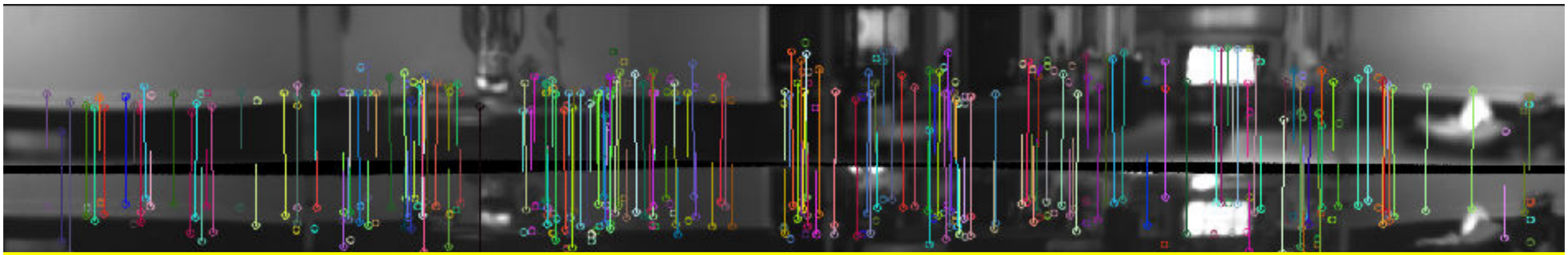
- **Frame-to-Frame Visual Odometry framework**
  - **Goal**: solve for the relative SE3 pose ${}^{[\mathrm{K}]}_{[\mathrm{C}_t]}\tilde{\mathbf{T}}$ between frames
  - **Feature-based** approach:
    - Keypoints detected via "*Good Features to Track*"
    - Keypoints described as Oriented Robust Binary (*ORB*) features
  - Frame $[\mathrm{C}_t]$ was tracked with respect to its reference keyframe $[\mathrm{K}]$
  - Keyframe creation:
    - If tracking correspondences are at least 10% of the current average:
      - Due to change in translation: $\Delta\mathbf{t}_{C_t \to K} > 1\ [\mathrm{cm}]$
      - Due to change in rotation angle: $\theta_{\Delta\mathbf{R}_{C_t \to K}} > 1^{\circ}$
  - Using Kneip's ***NP3P*** algorithm in a RANSAC fashion:
    - 3D-to-2D feature point registration approach from 3 points
    - Using a non-central model for bearing angle projection minimization
  - Final RANSAC pose model ${}^{[\mathrm{K}]}_{[\mathrm{C}_t]}\tilde{\mathbf{T}}^{(s_r)}$ is refined further via Levenberg-Marquardt

Tracking Features (Static Stereo)

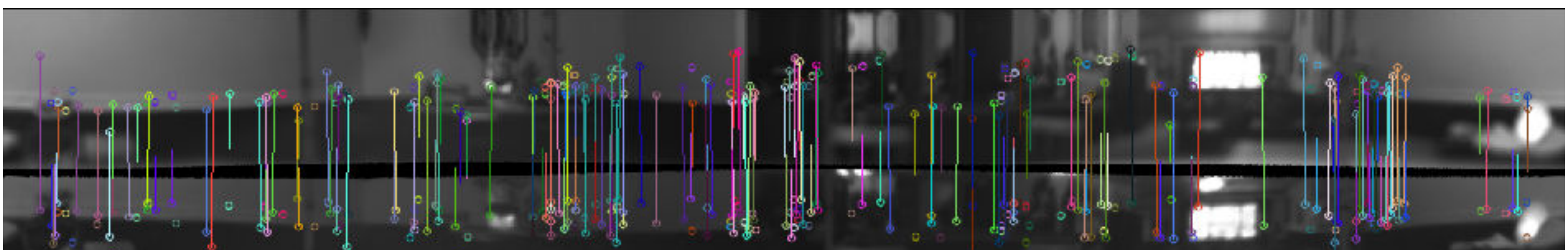
<Intro> <SOS> <GUMS> **<VO>**

KeyPoints: Detected as GFT, Described with ORB

$M_{top2bot}^{(\mathrm{K})} \coloneqq \left\{ {}^{[\Xi_{\mathrm{K}}]}(\boldsymbol{m}_{top}, \boldsymbol{m}_{bot})_i \text{ matched point-pairs from static omnistereo for the reference frame } [\mathrm{K}] \right\}$



$M_{top2bot}^{(\mathrm{C}_t)} \coloneqq \left\{ {}^{[\Xi_{\mathrm{C}_t}]}(\boldsymbol{m}_{top}, \boldsymbol{m}_{bot})_i \text{ matched point-pairs from static omnistereo for tracking frame } [\mathrm{C}_t] \right\}$

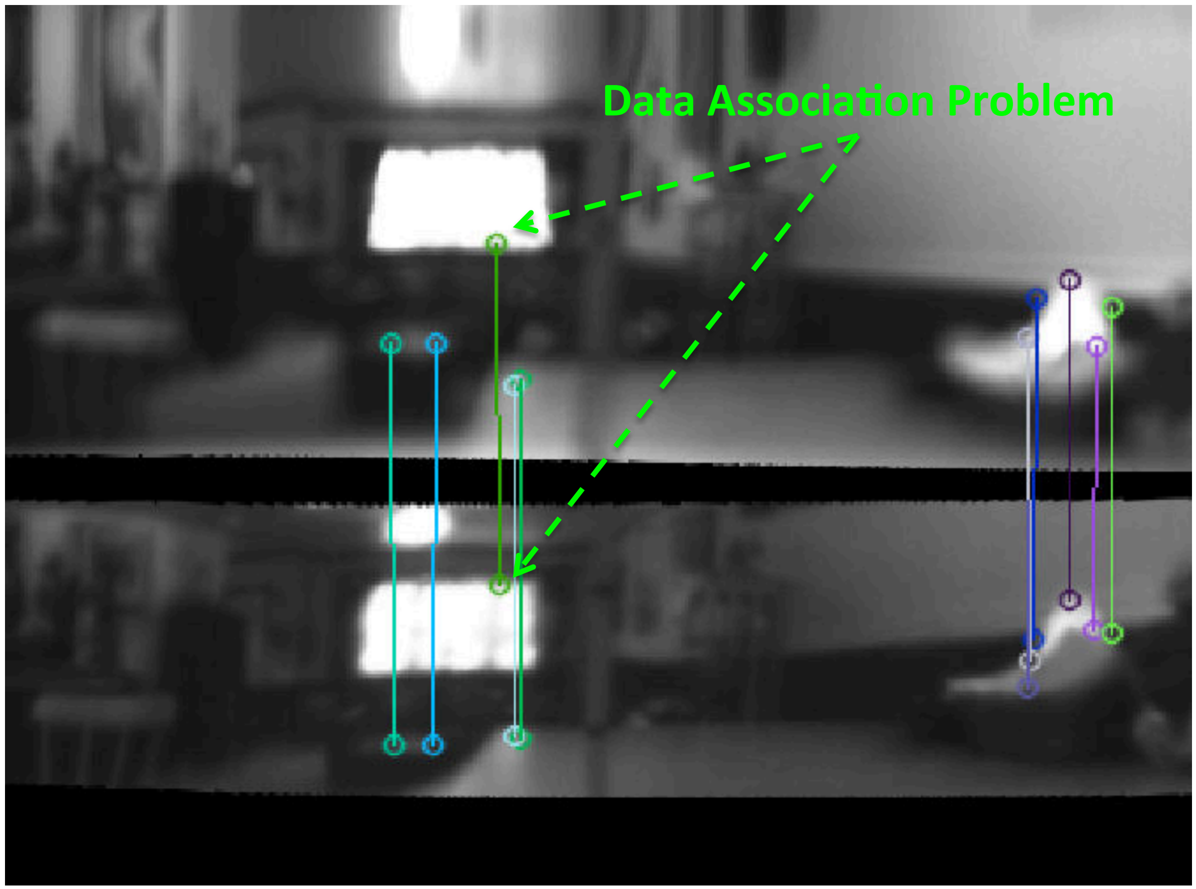


$M_{top2bot}^{(\mathrm{C}_{t+1})} \coloneqq \left\{ {}^{[\Xi_{\mathrm{C}_{t+1}}]}(\boldsymbol{m}_{top}, \boldsymbol{m}_{bot})_i \text{ matched point-pairs from static omnistereo for tracking frame } [\mathrm{C}_{t+1}] \right\}$

Tracking Features (Static Stereo)

<Intro> <SOS> <GUMS> **<VO>**

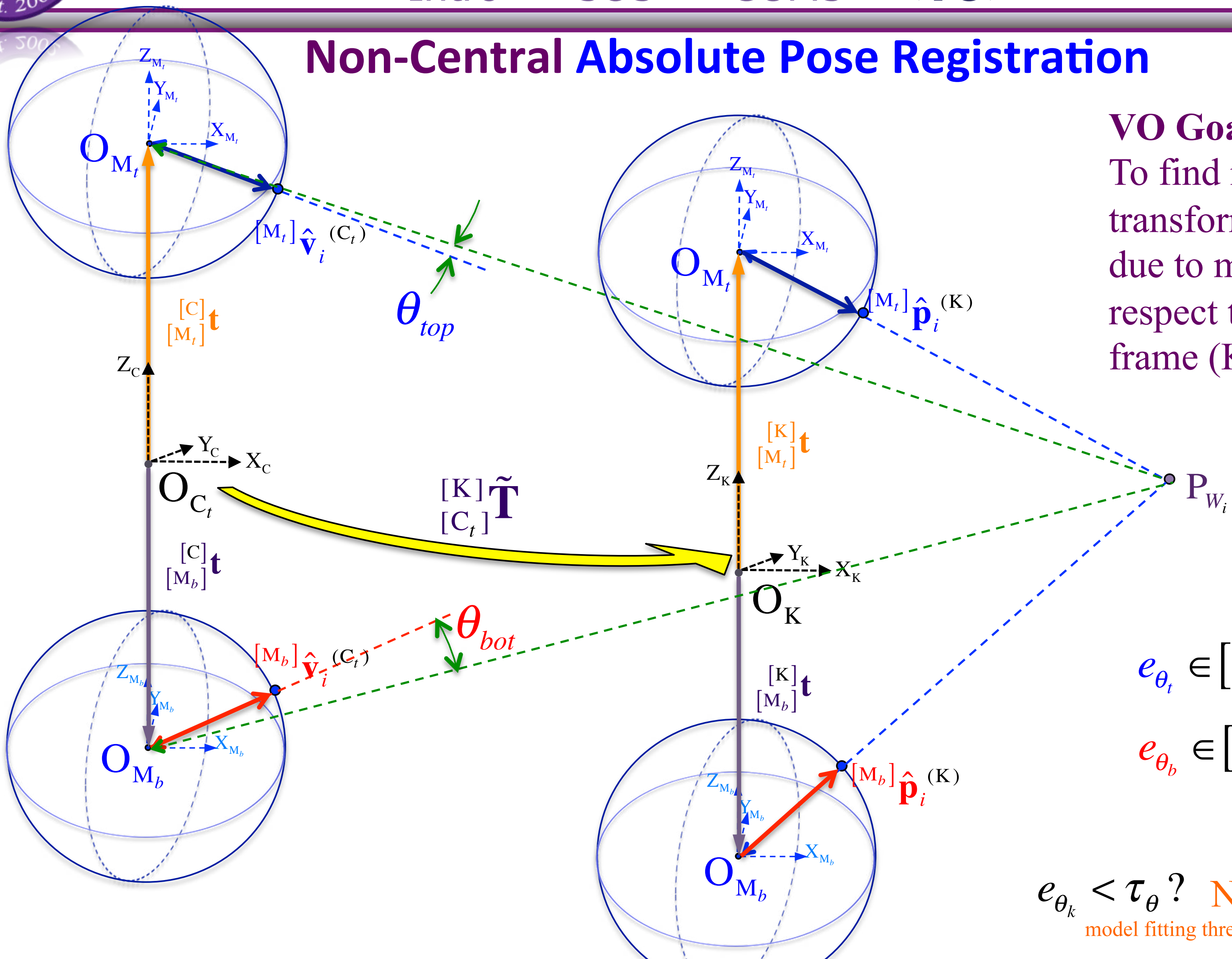
Erroneous point correspondences (outliers) must be removed



Example of some point correspondences among panoramas

Feature-Based VO: 3D-to-2D

Non-Central Absolute Pose Registration



VO Goal:
To find relative transform ${}^{[K]}_{[C_t]}\mathbf{T}$ due to motion with respect to a *ref*erence frame (Keyframe K)

$e_{\theta_t} \in [0,2]$

$e_{\theta_b} \in [0,2]$

$e_{\theta_k} < \tau_\theta$? No → outlier

model fitting threshold

Feature-Based VO: 3D-to-2D

Non-Central Absolute Pose Registration (among Top views)

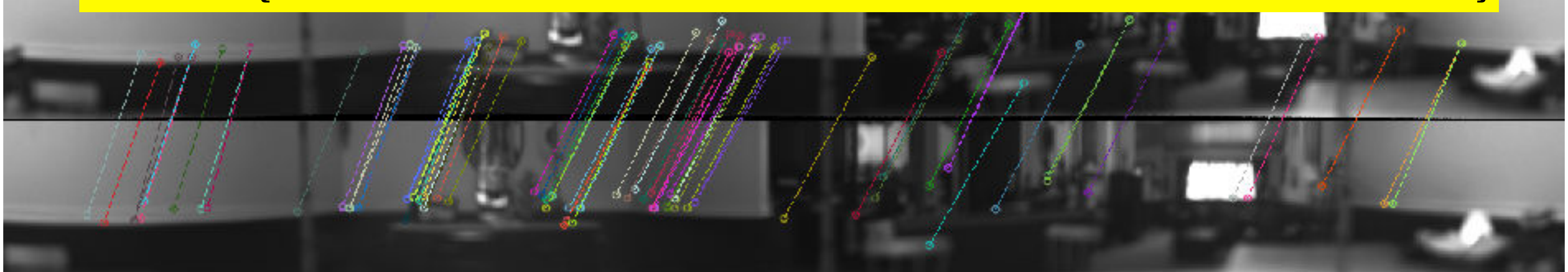
Correspondences among moving frames

Initial set of 2D point correspondences at iteration s_0

$$M_{top2top}^{(C_t,\, s_0)} := \left\{ \left({}^{[\Xi_{C_t}]}\boldsymbol{m},\, {}^{[\Xi_{K}]}\boldsymbol{m} \right)_{top,\, i}^{(s_0)} \text{ initial point-pairs between frames } [C_t] \text{ and } [K] \text{ top views} \right\}$$

Final set of 2D point correspondences fitting the model ${}^{[C_{ref}]}_{[C_k]}\mathbf{T}$ estimated via RANSAC at iter s_r

$$M_{top2top}^{(C_t,\, s_r)} := \left\{ \left({}^{[\Xi_{C_t}]}\boldsymbol{m},\, {}^{[\Xi_{K}]}\boldsymbol{m} \right)_{top,\, i}^{(s_r)} \text{ final point-pairs between frames } [C_t] \text{ and } [K] \text{ top views} \right\}$$



Feature-Based VO: 3D-to-2D

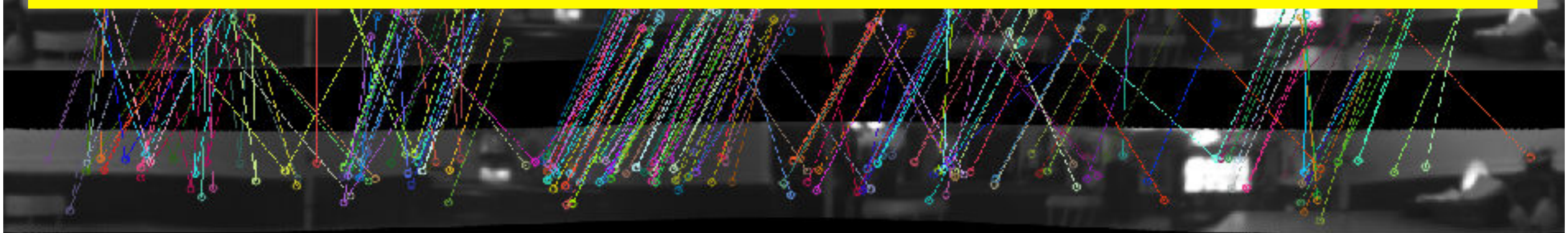
<Intro> <SOS> <GUMS> **<VO>**

Non-Central Absolute Pose Registration (among Bottom views)

Correspondences among moving frames

Initial set of 2D point correspondences at iteration s_0

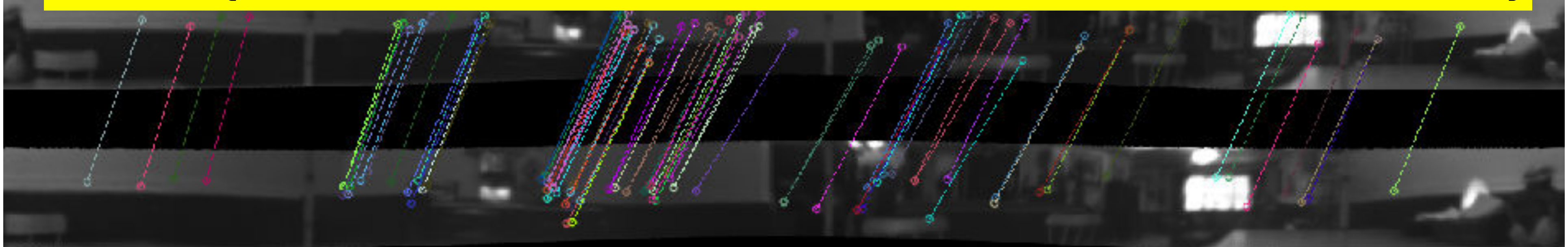
$$M_{bot2bot}^{(C_t,\, s_0)} := \left\{ \left({}^{[\Xi_{C_t}]}\boldsymbol{m},\, {}^{[\Xi_K]}\boldsymbol{m} \right)_{bot,\, i}^{(s_0)} \text{ initial point-pairs between frames } [C_t] \text{ and } [K] \text{ bottom views} \right\}$$



With search constraint: Arbitrary distance match threshold on the horizontal coordinates:

Final set of 2D point correspondences fitting the model ${}^{[C_{ref}]}_{[C_k]}\mathbf{T}$ estimated via RANSAC at iter s_r

$$M_{bot2bot}^{(C_t,\, s_r)} := \left\{ \left({}^{[\Xi_{C_t}]}\boldsymbol{m},\, {}^{[\Xi_K]}\boldsymbol{m} \right)_{bot,\, i}^{(s_r)} \text{ final point-pairs between frames } [C_t] \text{ and } [K] \text{ bottom views} \right\}$$



Feature-Based VO: 3D-to-2D

<Intro> <SOS> <GUMS> **<VO>**

Non-Central Pose Registration

RANSAC phase

Joint set of 2D point correspondences fitting the model ${}^{[\mathrm{K}]}_{[\mathrm{C}_t]}\mathbf{T}^{(s_j)}$ estimated via Non-Perspective-three-Point (NP3P) + RANSAC

$$M_{k2k}^{(t,s_j)} := \left\{ \left({}^{[\Xi_{\mathrm{C}_t}]}\boldsymbol{m}, {}^{[\Xi_{\mathrm{K}}]}\boldsymbol{m} \right)_{k,i}^{(s_j)} \;\middle|\; e_{\theta_{k,i}}\left({}^{[\mathrm{K}]}_{[\mathrm{C}_t]}\mathbf{T}^{(s_j)} \right) < \tau_\theta,\; k \in \{top, bot\},\; i \in M_{ALL} \right\}$$

where j is the RANSAC iteration index

Optimization phase

Non-linear optimization using the final set of inliers $M_{k2k}^{(\mathrm{C}_t, s_r)}$, and ${}^{[\mathrm{K}]}_{[\mathrm{C}_t]}\mathbf{T}^{(s_r)}$ as the initial pose

$${}^{[\mathrm{K}]}_{[\mathrm{C}_t]}\mathbf{T}^{*} = \underset{{}^{[\mathrm{K}]}_{[\mathrm{C}_t]}\mathbf{T}}{\arg\min}(J), \text{ where } J\left({}^{[\mathrm{K}]}_{[\mathrm{C}_t]}\mathbf{T}\right) := \frac{1}{2} \sum_{k=top,\,bot} \sum_{i=1}^{N_k} e_{\theta_{k,i}}$$

at the s_r final RANSAC iteration

Datasets and Evaluation

<Intro> <SOS> <GUMS> **<VO>**

Performance comparison against baseline 3D sensor (RGB-D camera)

RGB-D Camera **Single-camera SOS**

Ground Truth Datasets

Synthetic

- Photo-realistic simulation
  - Raytraced with POV-Ray
- Have 4 **pose** sequences (paths) in the same “office” scene
- Simulated sensors
  - Omnistereo images
  - Associated RGB-D images

Real

- **Ground-truth** from ***mocap*** system
  - Indoors only
  - Small capture volume: 6x3x2 m
  - Went in/out of room-hallway (50m)
- Hardware:
  - Hyberbolic rig (ready)

    - Pointgrey **Blackfly** Specs:
      - » Global Shutter at 30 FPS
      - » 1920x1200 pixels
  - Asus Xtion Pro Live

    - » VGA resolution 640x480 pixels

Datasets Link

Synthetic Dataset

Office – Seq. # 0

$t = 0$

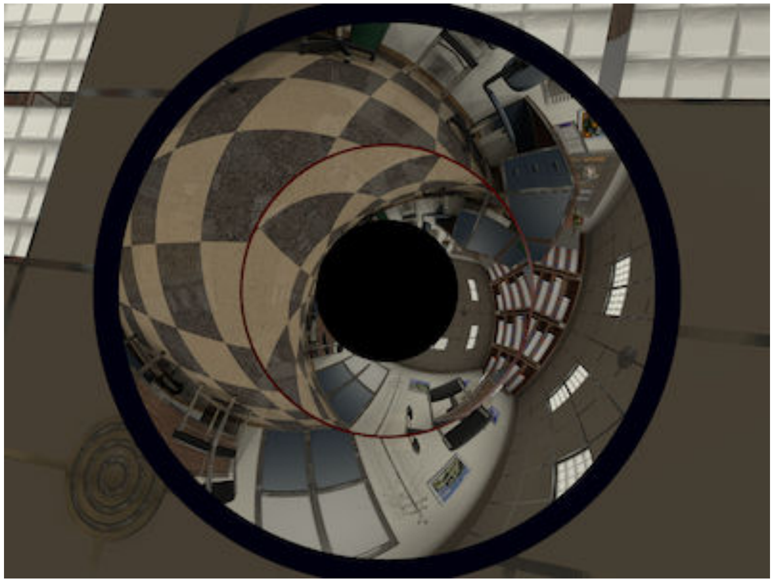
$t = 500$

Synthetic Dataset

Office – Seq. # 0

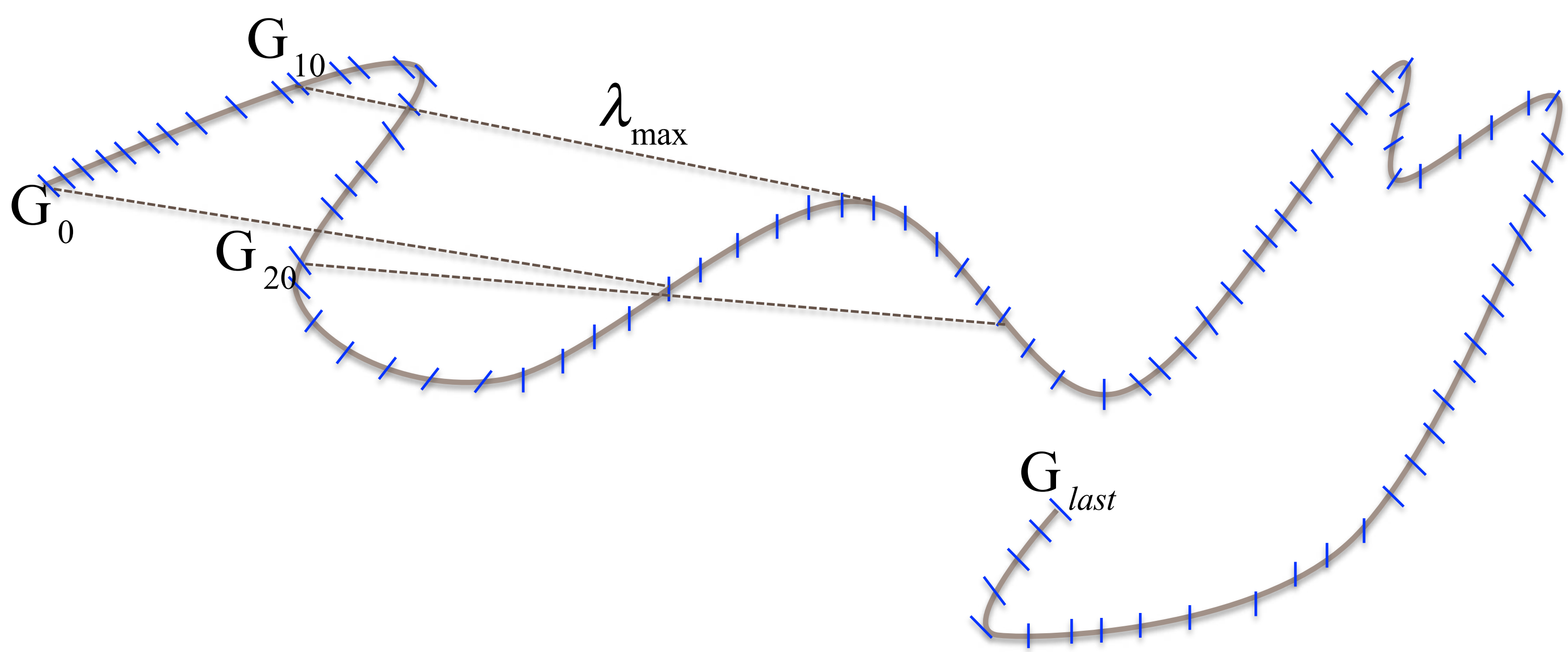
$t = 1000$

$t = 1500$



Pose Estimation Error Metrics

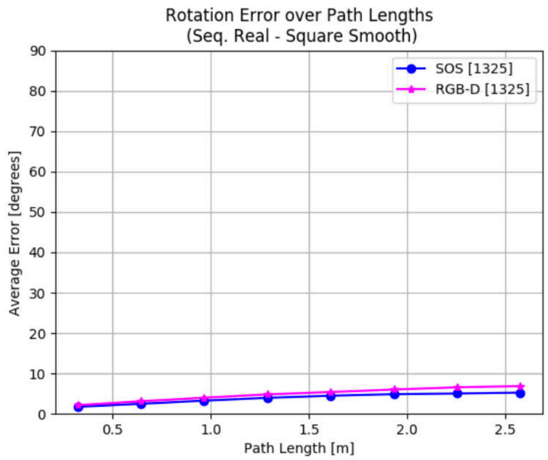
- **Absolute Trajectory Error (ATE)**
- **Relative Pose Error (RPE)**
  - 8 linearly spaced path lengths: $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6, \lambda_7, \lambda_{max}\}$
  - The **max** path length λ_{max} is **1/3** of the entire path length.

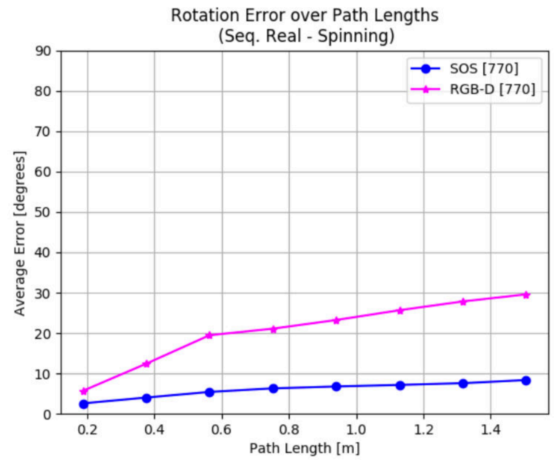


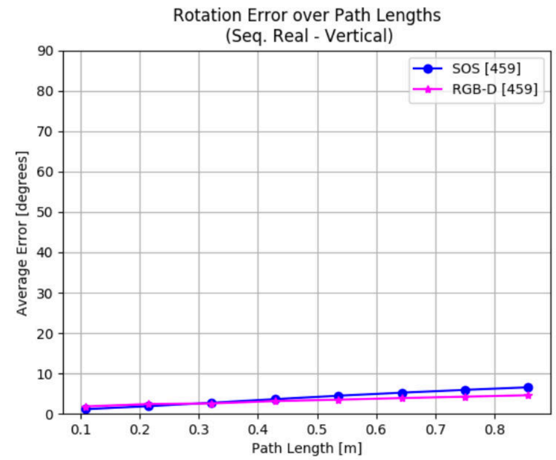
VO Quantitative Evaluation

Results for "**moving**" rigs in a **conventional** form

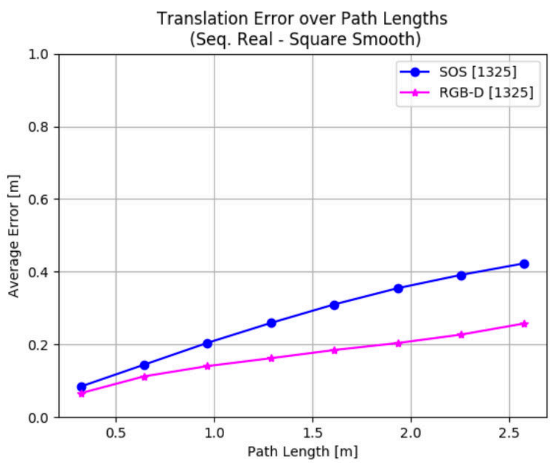
RPE (Rotation)

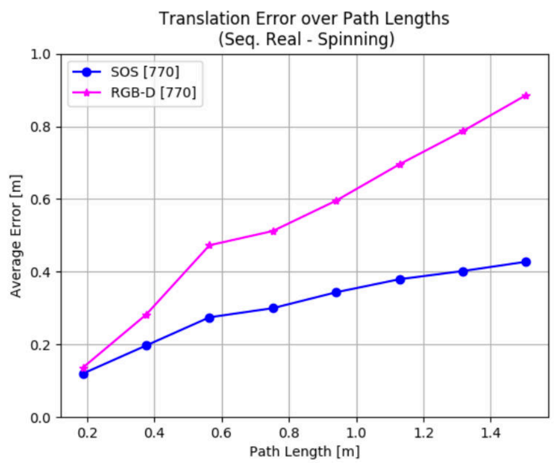


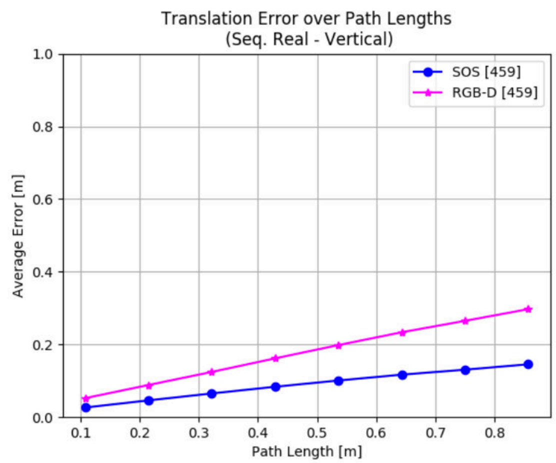




RPE (Translation)







NOTE: All frames were considered for evaluation

VO Quantitative Evaluation

Results for "**moving**" rigs in a **conventional** form

Average absolute trajectory error (ATE) and relative pose error (RPE, normalized)

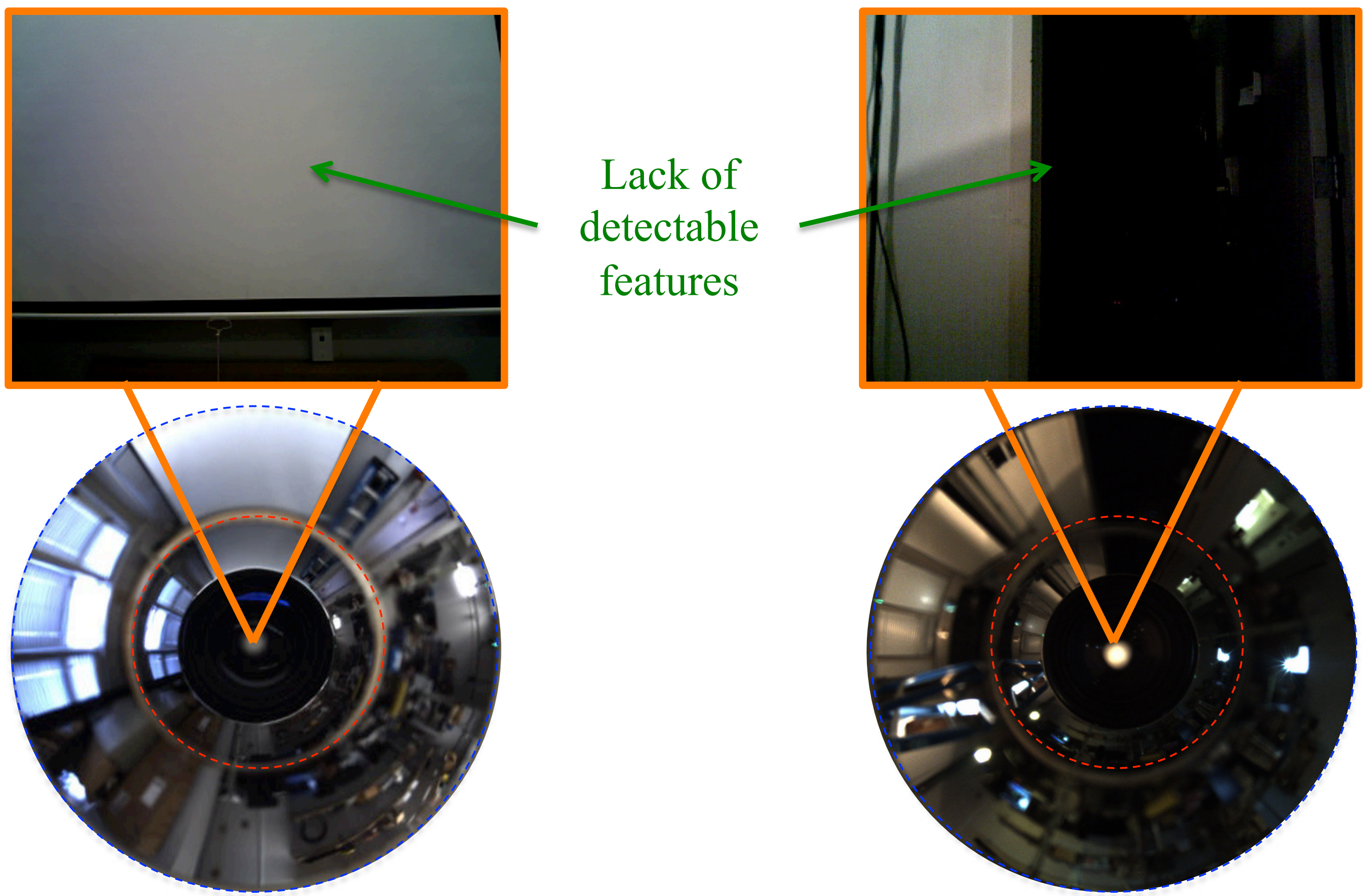
	Sequence	ATE [m]		RPE (Translation) [%]		RPE (Rotation) [°/m]	
		SOS	RGB-D	SOS	RGB-D	SOS	RGB-D
Conventional	Square Small	**0.12** ± 0.05	0.70 ± 0.23	**25.94** ± 8.18	41.76 ± 78.21	**3.46** ± 2.76	22.37 ± 45.52
	Square Smooth	**0.12** ± 0.06	0.14 ± 0.11	20.51 ± 8.54	**13.92** ± 7.89	**3.29** ± 1.97	4.11 ± 2.51
	Spinning	**0.30** ± 0.11	0.35 ± 0.08	**42.60** ± 19.59	68.30 ± 80.08	**8.64** ± 4.78	27.08 ± 36.18
	Vertical	**0.04** ± 0.02	0.14 ± 0.06	**19.86** ± 5.31	39.06 ± 16.12	**8.68** ± 3.55	8.76 ± 6.05
	Free Style	**0.14** ± 0.05	0.41 ± 0.14	**31.34** ± 13.57	45.75 ± 54.99	**9.65** ± 5.21	14.53 ± 20.35
	Hallway	0.95 ± 0.58	**0.81** ± 0.56	**262.53** ± 546.43	391.20 ± 763.28	10.14 ± 18.98	**8.54** ± 12.61

NOTE: All frames were considered for evaluation

Special Issue: Featureless Scenes

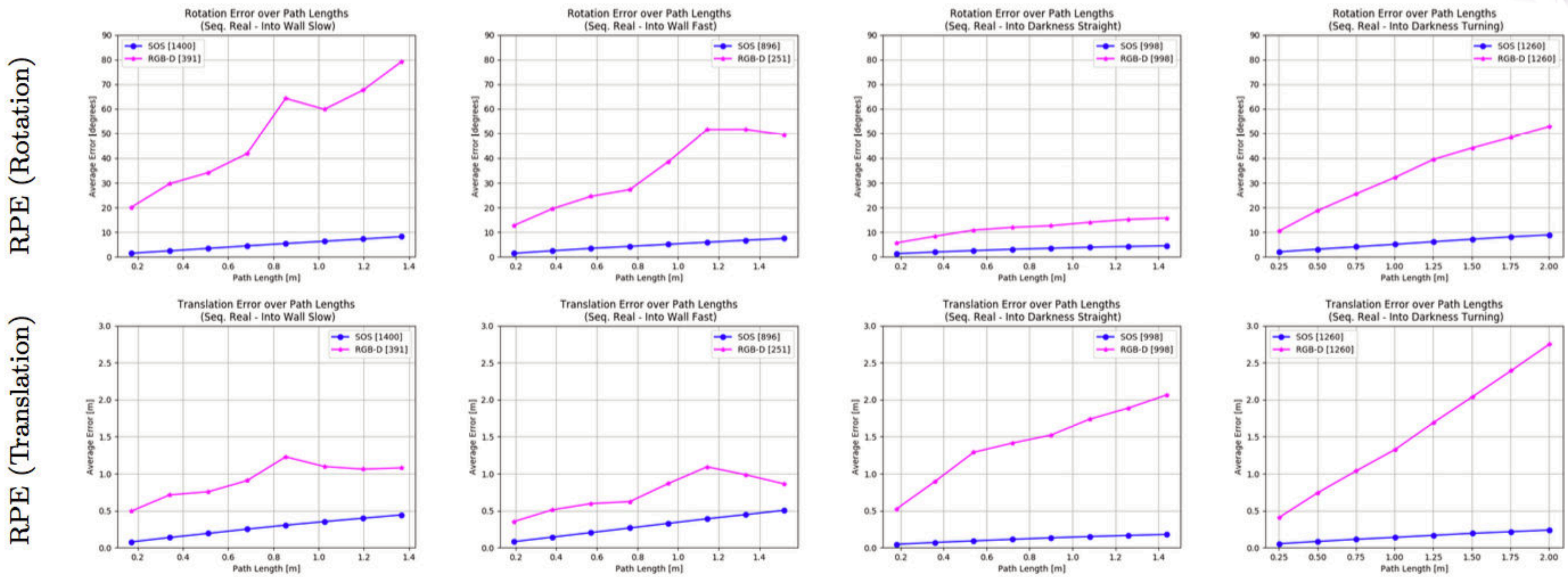
<Intro> <SOS> <GUMS> **<VO>**

Into a Wall **Into the Dark**



Special Issue: Featureless Scenes

Results for "**moving**" rigs in **special** feature-based circumstances



Sequence	ATE [m] SOS	ATE [m] RGB-D	RPE (Translation) [%] SOS	RPE (Translation) [%] RGB-D	RPE (Rotation) [°/m] SOS	RPE (Rotation) [°/m] RGB-D	Frames [#] SOS	Frames [#] RGB-D
Special: Into Wall - Regular	**0.13** ± 0.04	0.22 ± 0.08	**37.70** ± 11.00	130.41 ± 79.68	**6.70** ± 1.86	57.27 ± 42.25	**1041**	315
Into Wall - Slow	**0.09** ± 0.03	0.19 ± 0.09	**37.21** ± 10.64	165.29 ± 106.22	**6.72** ± 1.99	76.63 ± 65.46	**1400**	391
Into Wall - Fast	**0.09** ± 0.04	0.18 ± 0.09	**36.02** ± 9.61	115.06 ± 79.37	**5.83** ± 1.91	47.08 ± 36.90	**896**	251
Into Wall - Curvy	0.28 ± 0.08	**0.22** ± 0.11	**35.45** ± 14.35	136.35 ± 123.50	**6.39** ± 2.82	77.65 ± 81.61	**838**	309
Into Dark - Straight	**0.06** ± 0.03	0.53 ± 0.24	**16.87** ± 6.31	213.43 ± 218.62	**4.54** ± 1.99	19.38 ± 14.02	**998**	554
Into Dark - Turning	**0.13** ± 0.06	0.73 ± 0.23	**14.92** ± 6.07	141.20 ± 177.83	**5.50** ± 2.36	32.95 ± 37.47	1260	1260

NOTE: All frames were considered for evaluation

VO in Dynamic Environments

<Intro> <SOS> <GUMS> **<VO>**

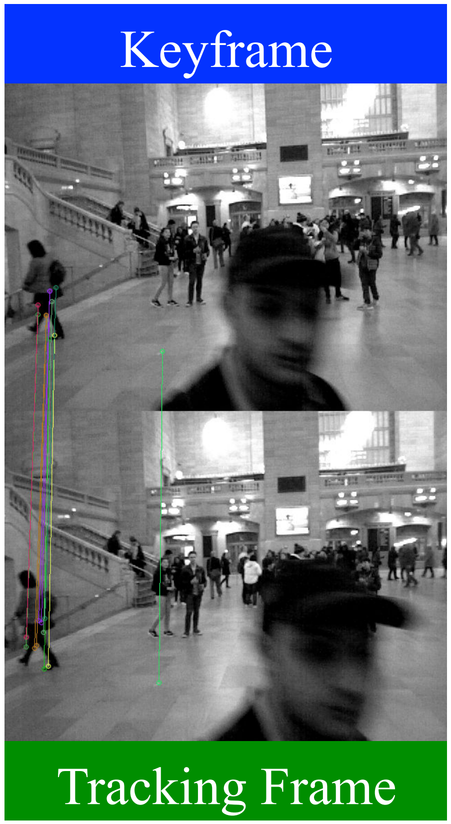
Results for "**static**" rigs in dynamic environments

Prox [m]	Peop [#]	Translation Error [m] SOS	Translation Error [m] RGB-D	Rotation Error [°] SOS	Rotation Error [°] RGB-D
1	1	**0.021** ± 0.008	0.279 ± 0.207	**0.310** ± 0.120	5.230 ± 3.160
1	2	**0.021** ± 0.006	0.600 ± 0.348	**0.380** ± 0.080	8.720 ± 6.000
1	4	**0.047** ± 0.017	1.614 ± 0.768	**0.610** ± 0.240	21.320 ± 10.820
2	1	**0.015** ± 0.005	0.586 ± 0.330	**0.230** ± 0.070	7.940 ± 4.220
2	2	**0.017** ± 0.007	1.049 ± 0.404	**0.250** ± 0.120	17.300 ± 9.310
2	4	**0.030** ± 0.008	2.247 ± 0.982	**0.570** ± 0.270	13.400 ± 6.370
3	1	**0.013** ± 0.004	1.029 ± 0.632	**0.140** ± 0.040	11.940 ± 8.900
3	2	**0.022** ± 0.005	1.728 ± 0.598	**0.340** ± 0.090	24.240 ± 13.100
3	4	**0.028** ± 0.007	1.854 ± 0.681	**0.460** ± 0.110	13.710 ± 12.780
Var	2	**0.049** ± 0.021	0.481 ± 0.260	**0.900** ± 0.310	17.980 ± 10.900
Var	Var	**0.374** ± 0.193	4.487 ± 1.261	**5.630** ± 3.600	39.390 ± 11.120

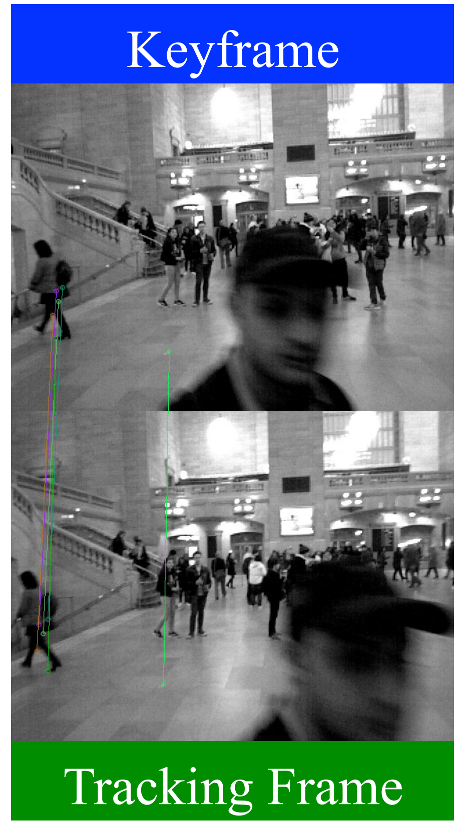
VO in Dynamic Environments

Results for "**moving**" rigs in **highly-dynamic** environments

OKAY



Next frame: LOST



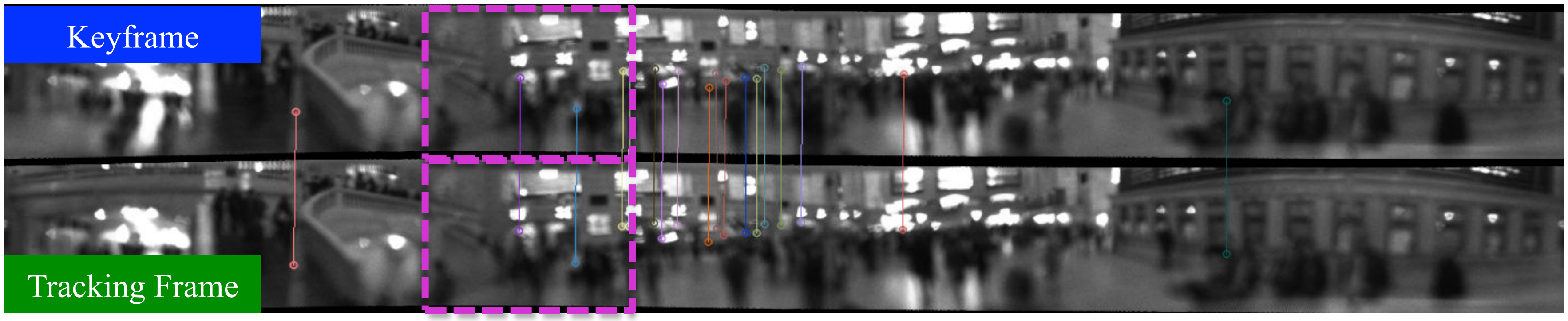
Tracking lost for **RGB-D camera** after the 191st frame

Reason: Less than **3** consistent features for **P3P**

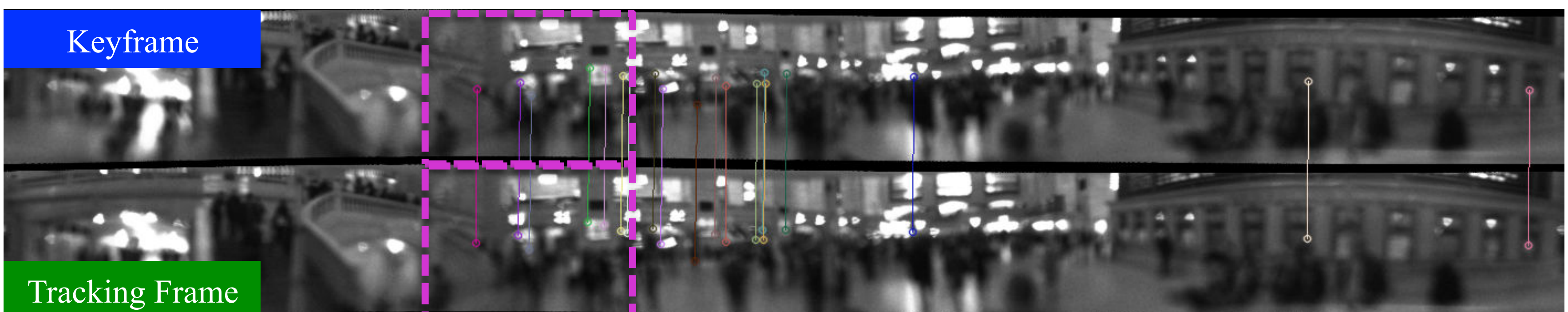
VO in Dynamic Environments

Results for "**moving**" rigs in **highly-dynamic** environments

OKAY



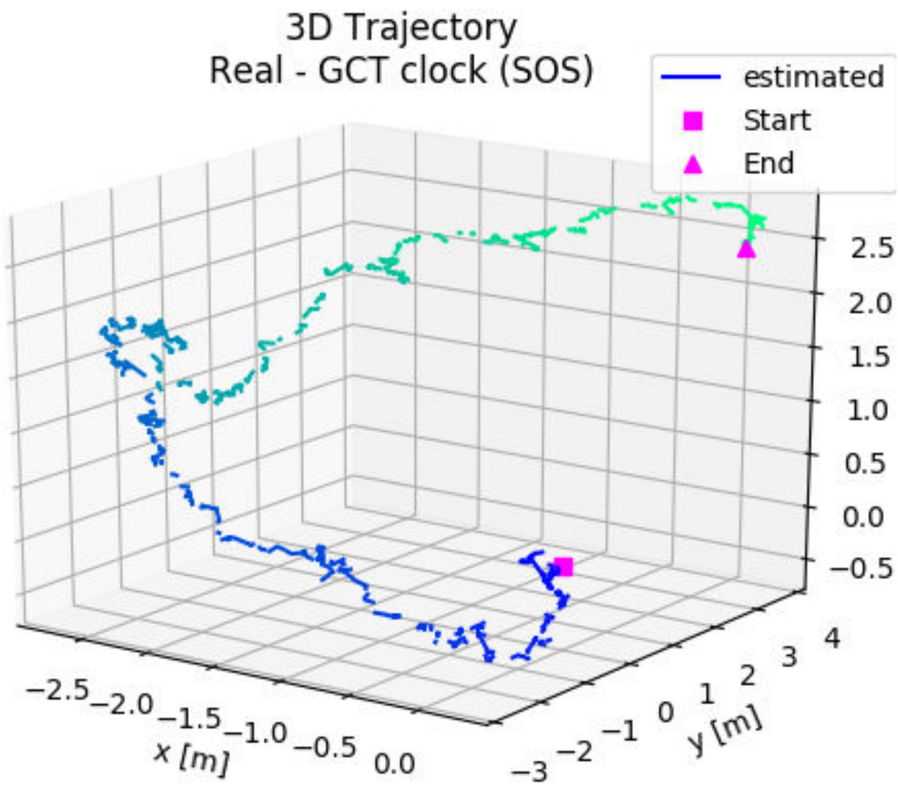
Next frame: OKAY

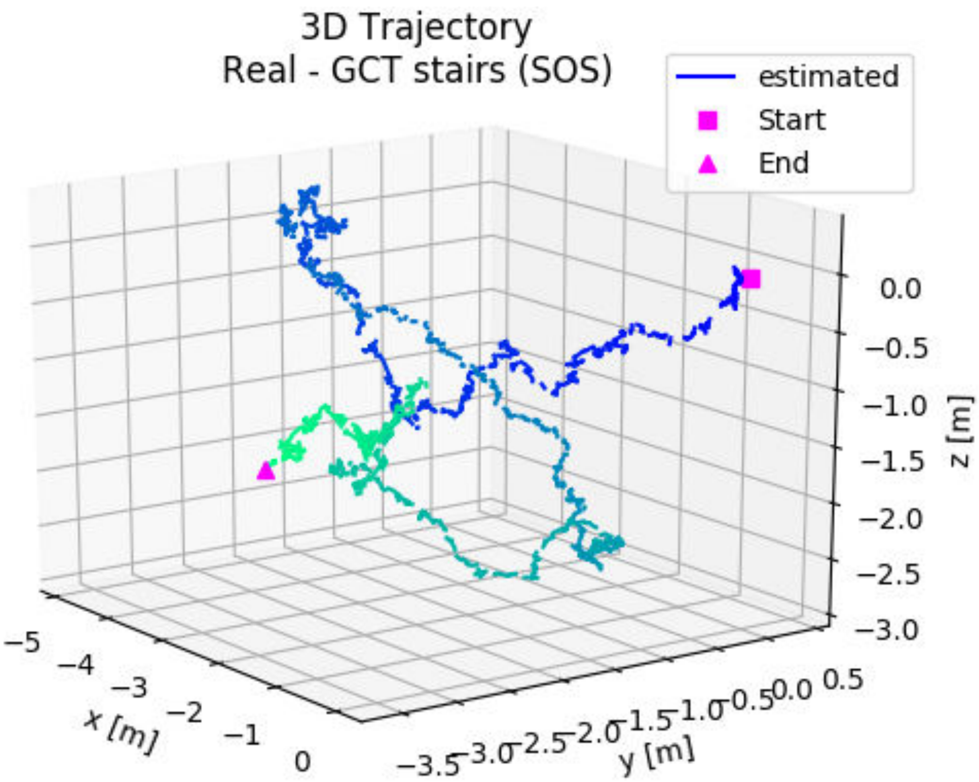


Panoramic images (**top view**)

VO in Dynamic Environments

Qualitative results for "**moving**" rigs in **highly-dynamic** environments





Video Link

NOTES: 1) Showing only keyframe positions; **2)** These trajectories are not fit to the ground plane.

Thesis Summary of Contributions

- Presented two **practical configurations** to achieve *omnistereo* vision with a single camera:
  1) Spherical SOS
  2) Hyperbolic SOS

- Introduced **GUMS** as a "Generalized Unified Model for Stereo" omnidirectional systems:
  - Simple yet effective model
  - Optimal parameters for a "**coupled**" projective function.
  - Practical calibration method performed via *control-points* imaged around the rig

- Demonstrated the 3D metric **visual odometry** (VO) capabilities of the single- camera SOS:
  - 3D pose was estimated *geometrically* via a **feature-based** tracking method
  - Compared its performance against a traditional range sensor (i.e., an RGB-D camera).
  - Generated several real and synthetic **datasets** with ground-truth pose information

- Improved the **tracking accuracy** of a single camera through the **direct** method:
  - In terms of high-dimensional features (channels) generated via
    a) Conventional hand-crafted features
    b) Features extracted from convolutional neural networks

Relevant Publications

- Igor Labutov, **Carlos Jaramillo**, and Jizhong Xiao. 2011. "*Generating near-Spherical Range Panoramas by Fusing Optical Flow and Stereo from a Single-Camera Folded Catadioptric Rig*." Machine Vision and Applications 24 (1).
- **Carlos Jaramillo**, Roberto G Valenti, Ling Guo, and Jizhong Xiao. 2016. "*Design and Analysis of a Single-Camera Omnistereo Sensor for Quadrotor Micro Aerial Vehicles (MAVs)*." Sensors 16 (2).
- **Carlos Jaramillo**, Roberto G. Valenti, and Jizhong Xiao. 2016. "*GUMS: A Generalized Unified Model for Stereo Omnidirectional Vision (Demonstrated Via a Folded Catadioptric System).*" In 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2528–33. Daejeon, Korea.
- **Carlos Jaramillo**, Yuichi Taguchi, and Chen Feng. 2017. "*Direct Multichannel Tracking.*" In IEEE International Conference on 3D Vision. Qingdao, China.
- **Carlos Jaramillo**, Liang Yang, J. Pablo Muñoz, Yuichi Taguchi, and Jizhong Xiao. 2018. "*Visual Odometry with a Single-Camera Stereo Omnidirectional System.*" In IEEE RA-Letters. Article ***submitted*** in May 9, 2018.

Thank You

Q

A