Dimensionality reduction in visual-inertial SLAM

Laurent Kneip
http://mpl.sist.shanghaitech.edu.cn

Macao, November 8th 2019

上海科技大学
ShanghaiTech University

MPL
Mobile Perception Lab

lkneip@shanghaitech.edu.cn

http://mpl.sist.shanghaitech.edu.cn

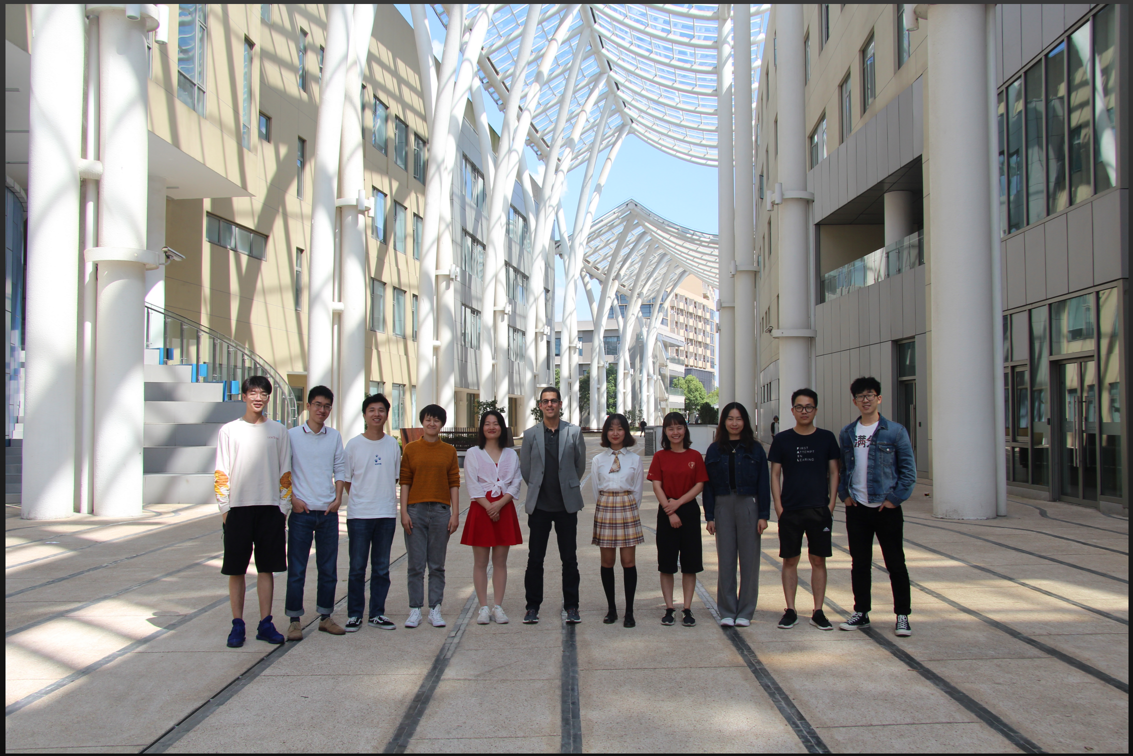
Who are we?

MPL
Mobile Perception Lab

ShanghaiTech Automation and Robotics Center · 上海科技大学自动化与机器人中心
STAR CENTER

SIST

上海科技大学
立志成才 报国裕民
2013
SHANGHAITECH UNIVERSITY

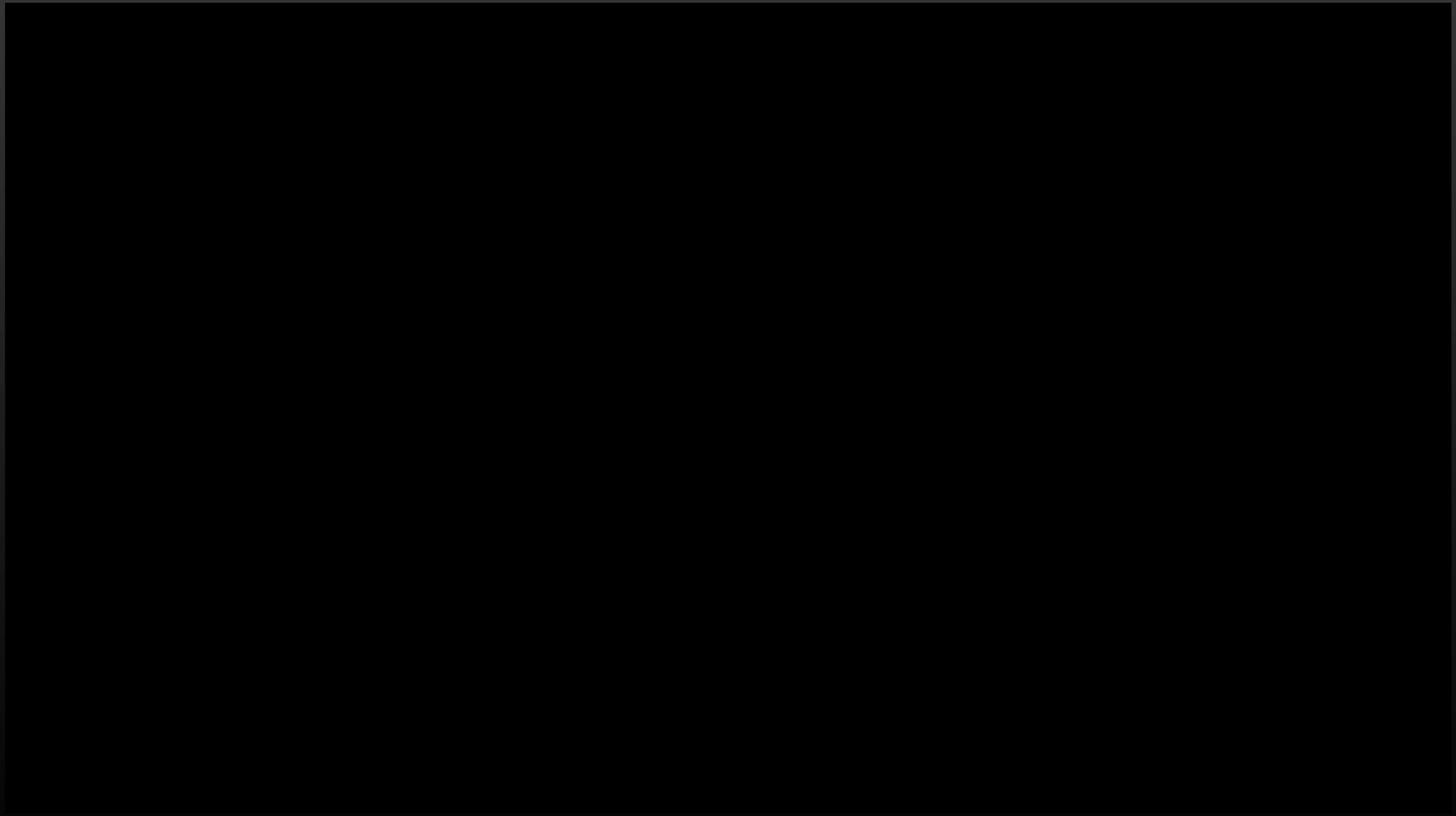


lkneip@shanghaitech.edu.cn

Research@MPL: Visual SLAM

- An enabler of new technologies
  - Factory automation
  - Service robotics
  - Augmented reality
  - Intelligent transportation

[Y Zhou, H Li, **L Kneip**, Canny-VO: Canny-VO: Visual Odometry with RGB-D Cameras based on Geometric 3D-2D Edge Alignment. *IEEE Transactions on Robotics (T-RO)*, 35(1):1–16, 2018]



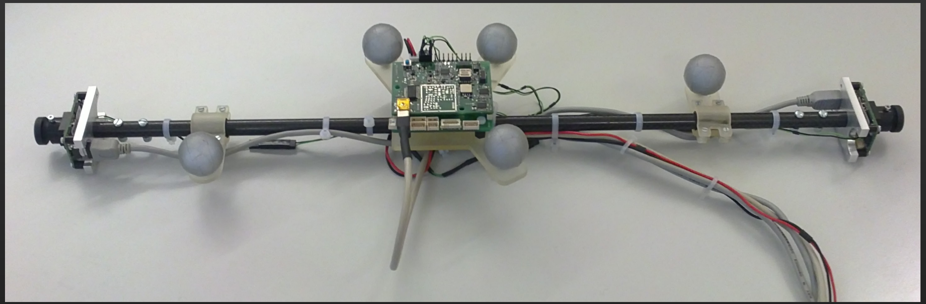
Research@MPL: Surround-view camera systems

- Origins: V-CHARGE
  - EU FP7 project
    - ETH Zurich
    - Volkswagen
  - AVP with vision only
    - Close-to-market sensors

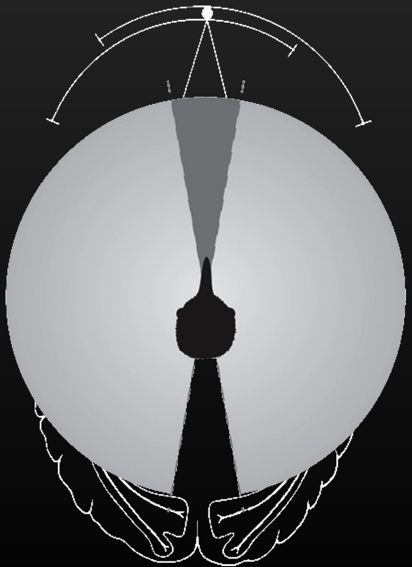
V-CHARGE

Research@MPL: Surround-view camera systems

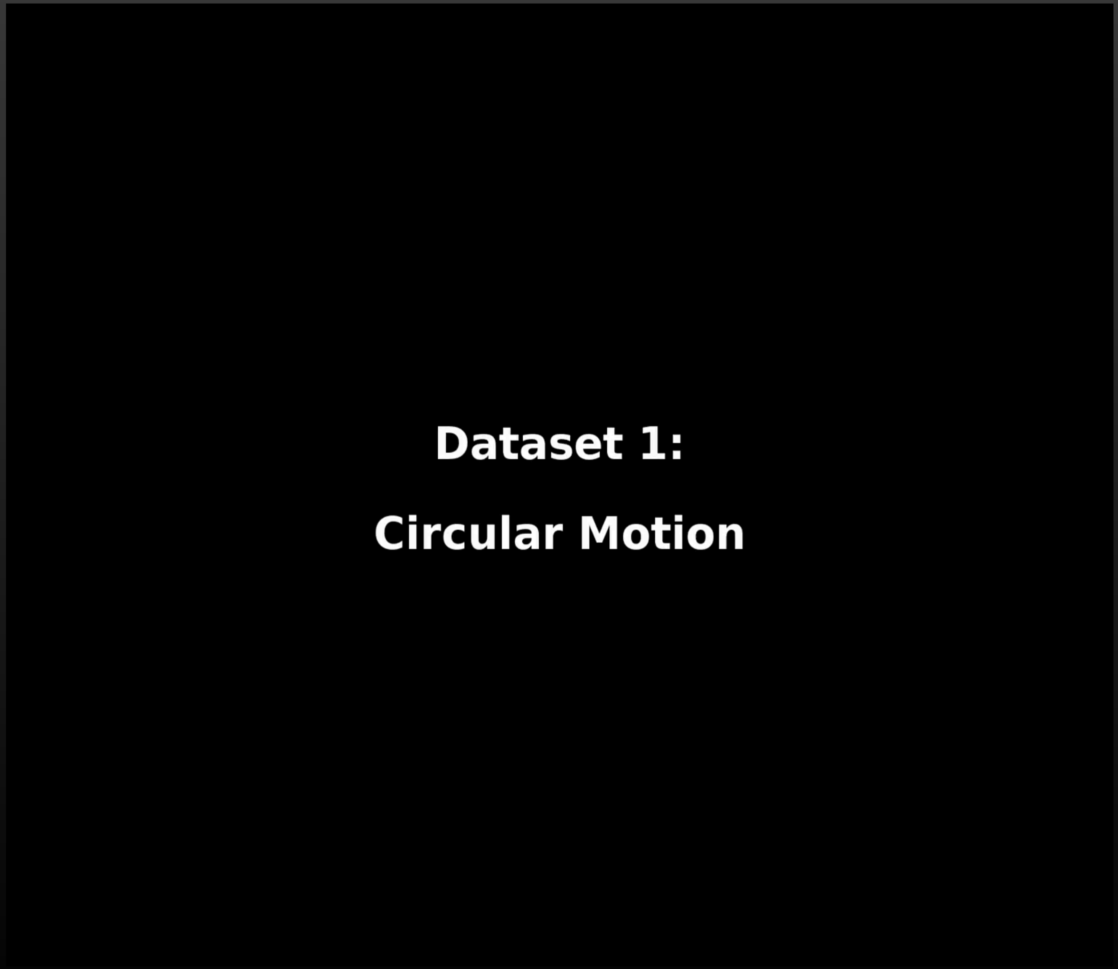
- Early research (2012):
  - Non-overlapping stereo



  - Inspired by nature
    - Field of view of humans
    - Field of view of pigeons

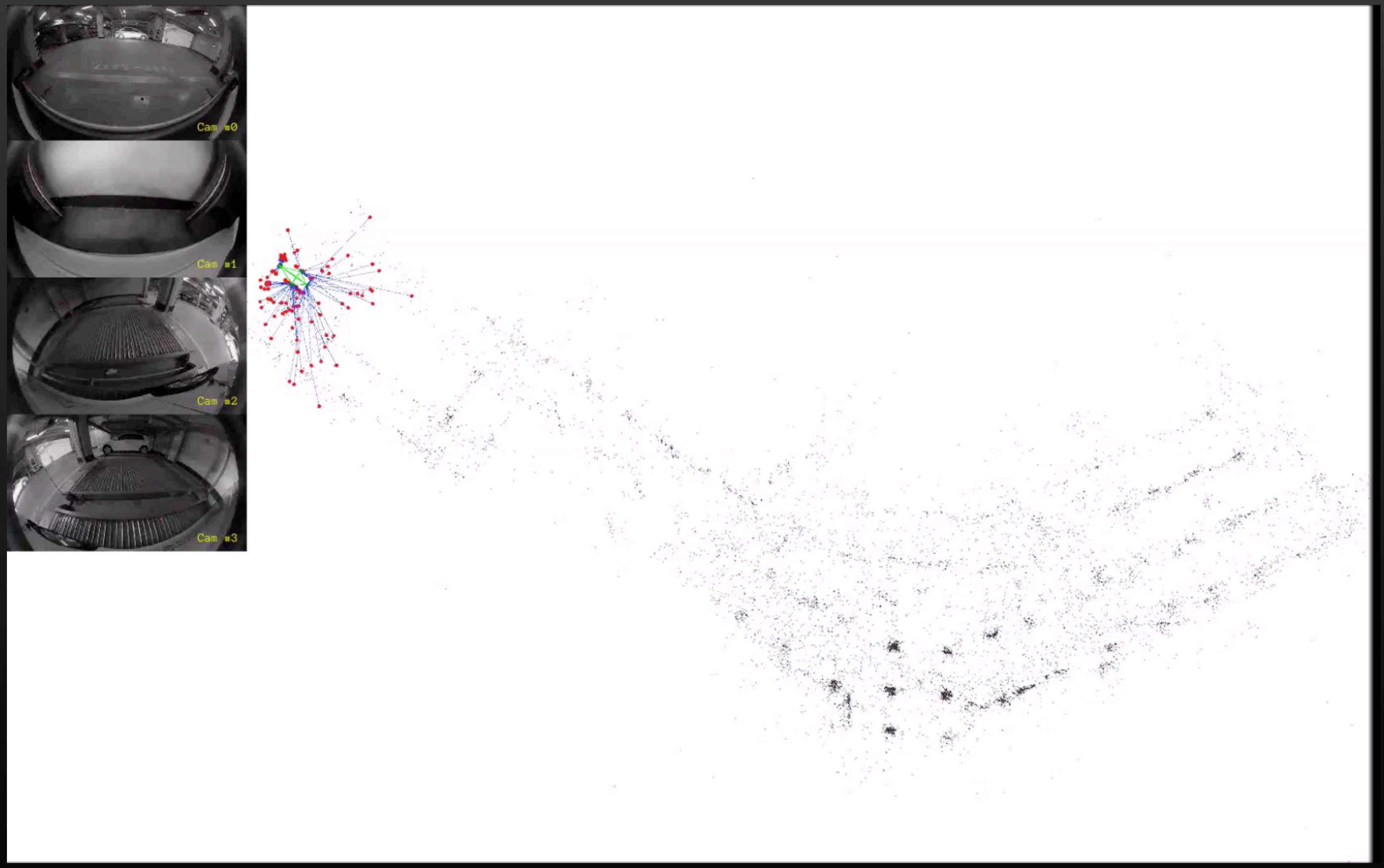


[Kazik, **Kneip**, Nikolic, Pollefeys, Siegwart, Real-Time 6D Stereo Visual Odometry with Non-Overlapping Fields of View, CVPR'12]



Research@MPL: Surround-view camera systems

- Now: Joint work with Motovis Intelligent Technologies, Co Ltd. (Shanghai)



Research@MPL: OpenGV

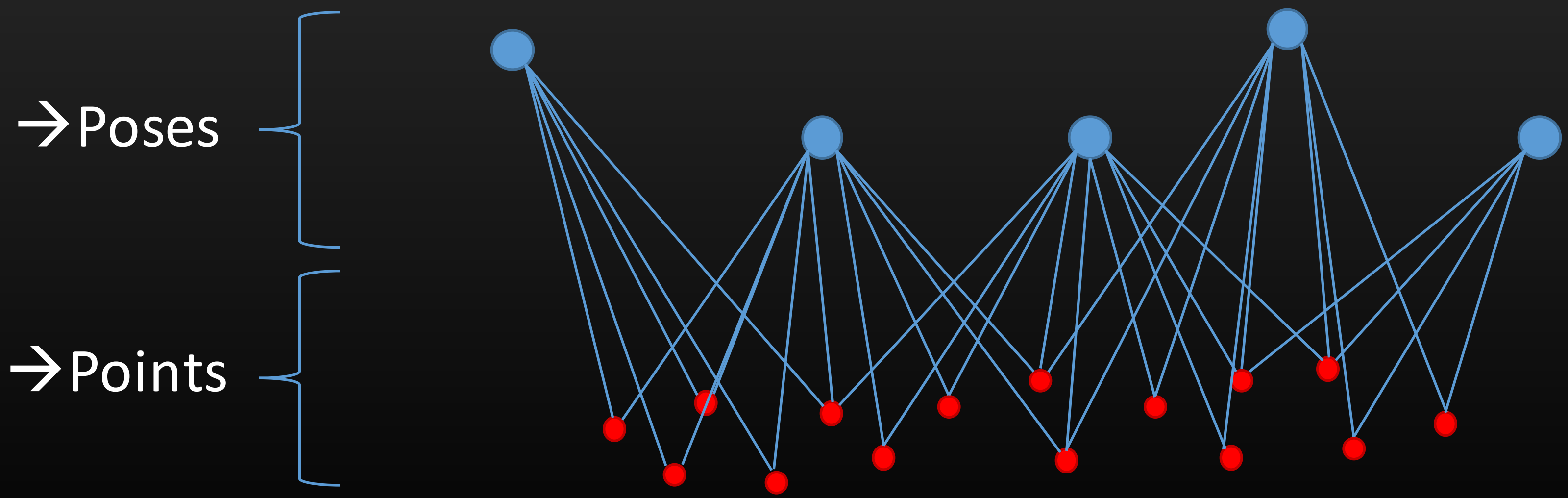
- [**L Kneip** and P Furgale. OpenGV: A unified and generalized approach to calibrated geometric vision. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Hong Kong, China, May 2014
- Open-source, hosted on github
- Widely used in both academia and industry

GitHub

https://github.com/laurentkneip/opengv

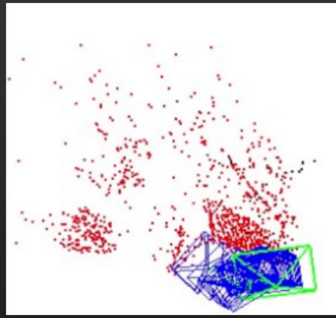
Is visual SLAM a solved problem?

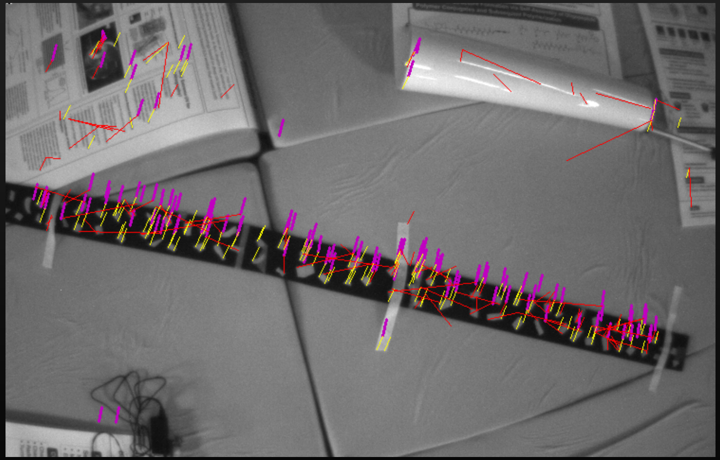
- Traditional SLAM is solved as a graph optimization problem using sparse feature correspondences
  - 2D-2D correspondences for bootstrapping
  - 2D-3D correspondences for tracking
  - Entire SfM view-graph for mapping

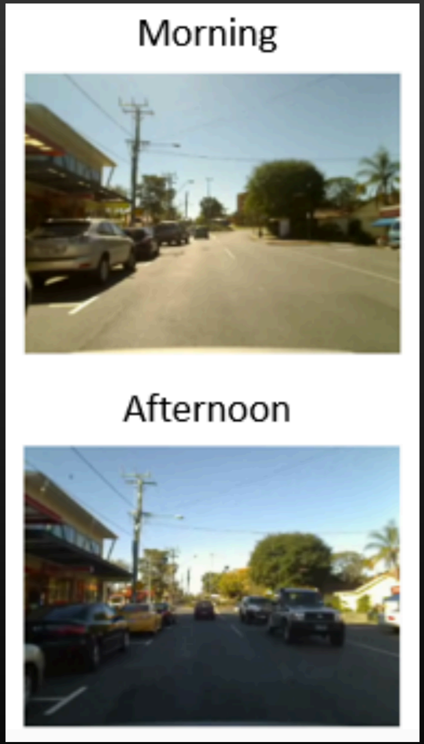


Is visual SLAM a solved problem?

- Traditional SLAM is solved as a graph optimization problem using sparse feature correspondences
- Issues:
  - Feature-poor scenarios
  - Bad feature distributions
  - Blur/high disparity
  - Meaningless maps
  - Poor long-term stability



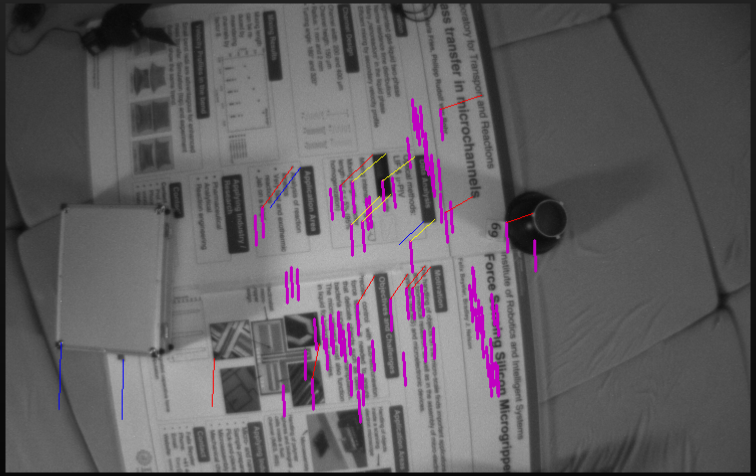


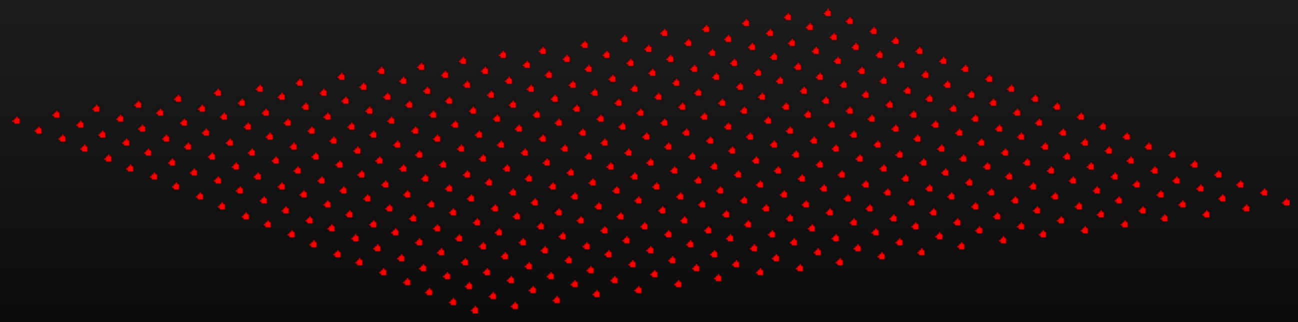


lkneip@shanghaitech.edu.cn

Is visual SLAM a solved problem?

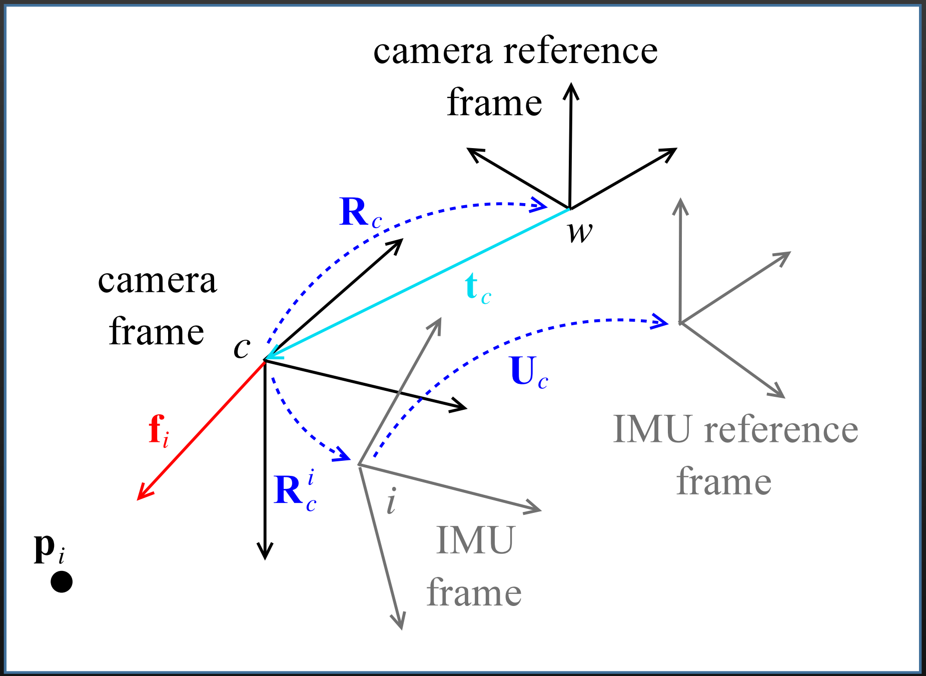
- Traditional SLAM is solved as a graph optimization problem using sparse feature correspondences
- Issues for geometric relative pose computation (used in bootstrapping):
  - Planar/degenerate point distributions
  - Pure rotation, rotation/translation ambiguity





Inertial-assisted visual odometry

- Idea: Use relative rotation priors from IMU
  - Short-term integration of gyroscopic signals for full 3D orientation change
  - Integration typically accomplished inside IMUs
    - Orientation drifts only slowly
    - Short-term relative rotations recoverable from consecutive IMU measurements



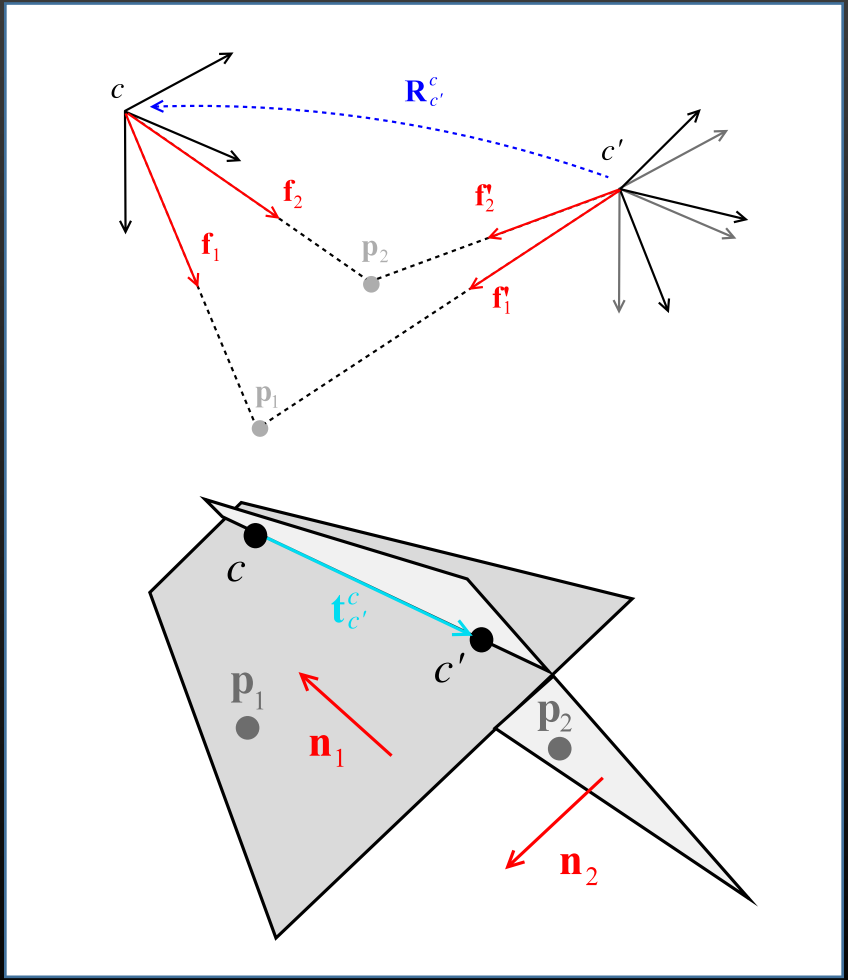
- Assumption
  - Extrinsic transformation parameters are known

- Delta-rotation prior given by: $\mathbf{R}_{c'}^{c} = \mathbf{R}_c^{iT}\mathbf{U}_c^T\mathbf{U}_{c'}\mathbf{R}_c^i$

[**L Kneip**, M Chli, and R Siegwart. Robust real-time visual odometry with a single camera and an IMU. In *Proceedings of the British Machine Vision Conference (BMVC)*, 2011]

Inertial-assisted visual odometry

- Relative pose: Computation of translation using 2-point algorithm
  - Epipolar plane normals: $\mathbf{n}_i = \mathbf{f}_i \times \mathbf{R}^c_{c'} \mathbf{f}'_i$
  - Translation direction: $\mathbf{d}^c_{c'} = \mathbf{n}_1 \times \mathbf{n}_2$
  - Translation vector: $\mathbf{t}^c_{c'} = +/- \dfrac{\mathbf{d}^c_{c'}}{\|\mathbf{d}^c_{c'}\|}$
- Absolute pose:
  - Becomes a 1 ½ point algorithm!



[**L Kneip**, M Chli, and R Siegwart. Robust real-time visual odometry with a single camera and an IMU. In *Proceedings of the British Machine Vision Conference (BMVC)*, 2011]

Inertial assisted visual odometry

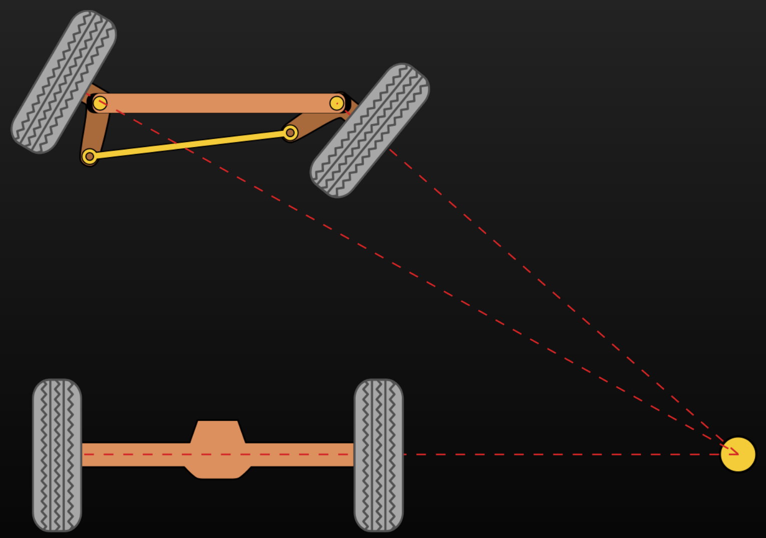
- Results on MAV dataset
  - Challenging motion (human pilot, high velocity, full 3D)
  - Challenging structure (moving, planar degeneracy, specularity)
  - FOV 100, 2.8 GHz machine, ~80 Hz

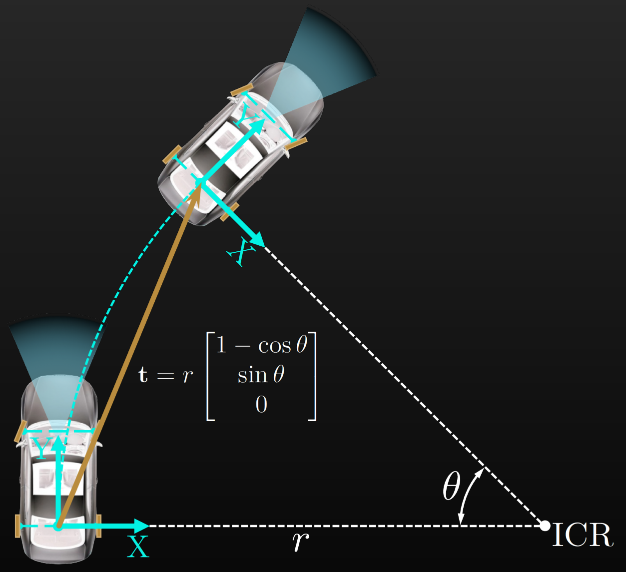
Dimensionality reduction in SLAM

Representing ***motion*** using $\left\{ \begin{array}{c} \text{higher-order} \\ \text{low-dimensional} \\ \text{robust} \\ \text{implicitly smooth} \end{array} \right\}$ prior models

Vehicle motion is non-holonomic

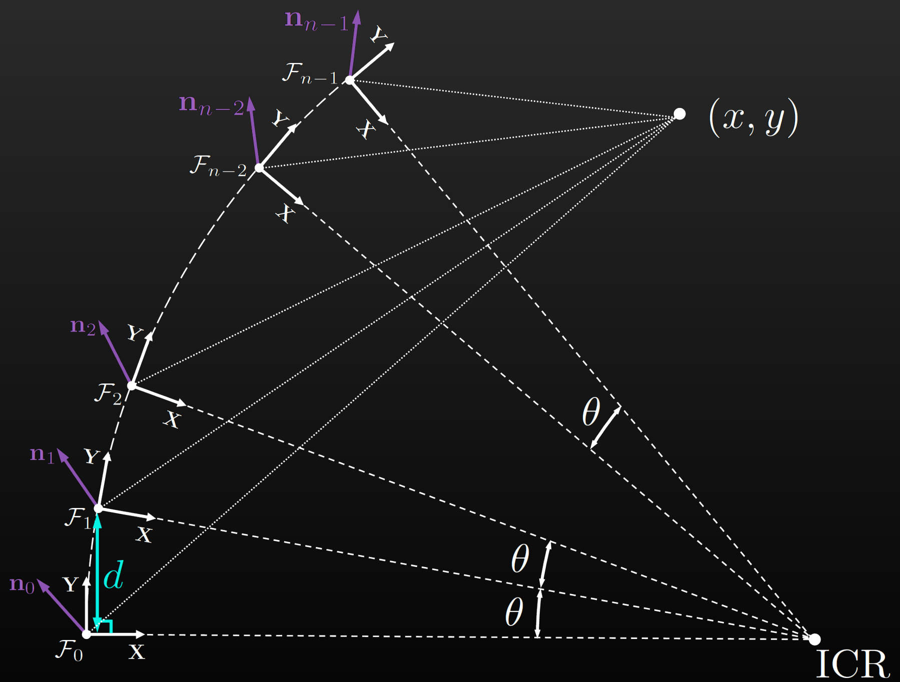
- Exploitation of the Ackermann-steering model
  - Rotation and translation of a vehicle are coupled
  - Leads to local parametrization of motion by arc of circle
  - Solution in the space of rotations only





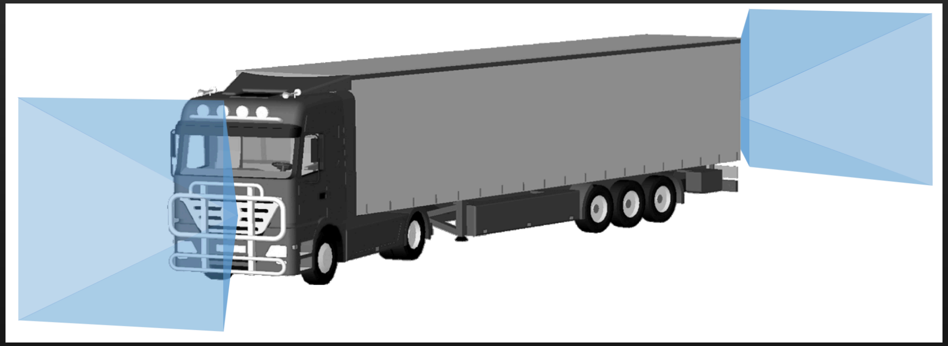
Vehicle motion is non-holonomic

- An n-view 1-point algorithm
  - [K Huang, Y Wang, and **L Kneip**. Motion estimation of non-holonomic ground vehicles from a single feature correspondence measured over n views. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, USA, June 2019]
  - [Y Wang, K Huang, X Peng, H Li, and **L Kneip**. Reliable frame-to-frame motion estimation for vehicle-mounted surround-view camera systems. Submitted to IEEE International Conference on Robotics and Automation (ICRA), 2020.]



Vehicle motion is non-holonomic

- Cameras distributed over tractor-trailer system
  - Again, a 1-point algorithm!

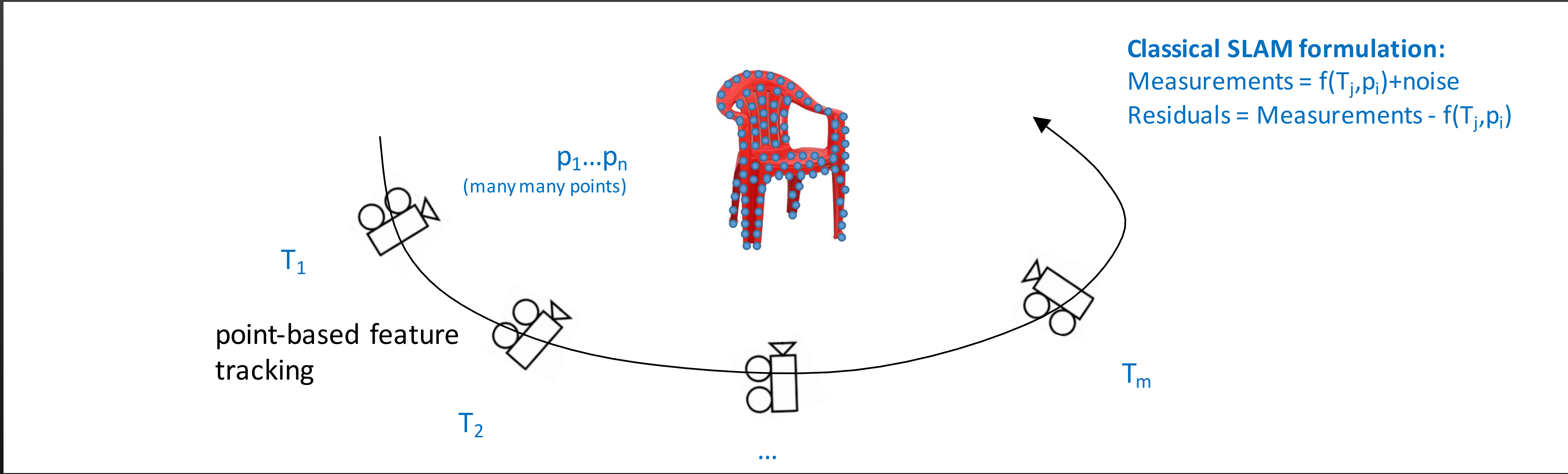


[X Peng, J Cui, and **L Kneip**. Articulated multi-perspective cameras and their application to truck motion estimation. In *Proceedings of the IEEE/RSJ Conference on Intelligent Robots and Systems (IROS)*, Macau, China, November 2019]

Dimensionality reduction in SLAM

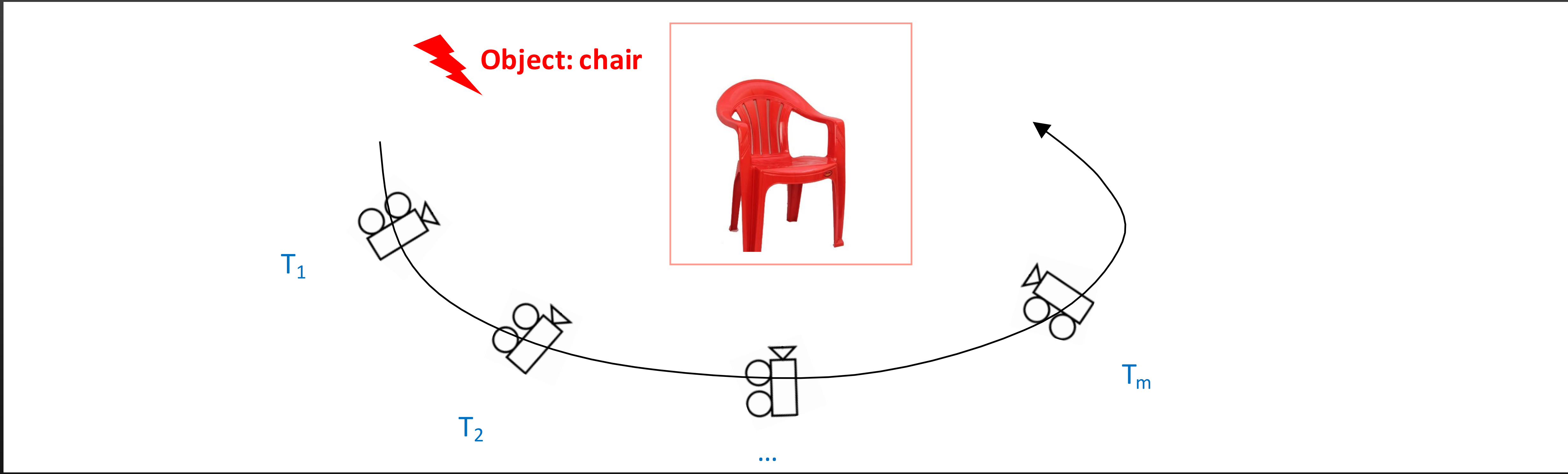
Representing ***structure*** using $\left\{ \begin{array}{c} \text{higher-order} \\ \text{low-dimensional} \\ \text{robust} \\ \text{implicitly smooth} \end{array} \right\}$ prior models

What do we mean by higher-order priors?



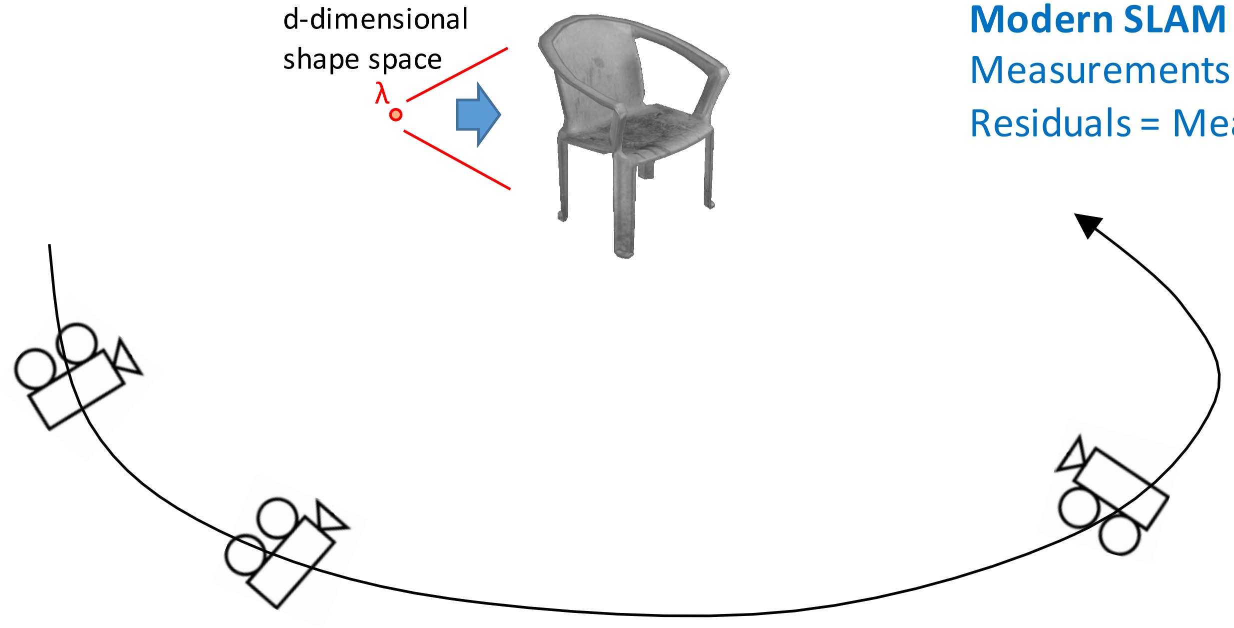
- Classical formulation ignores higher level information
- Dense representation = "many pts constrained by local smoothness"

What do we mean by higher-order priors?



- Detect object(s) in image
- Segment/annotate in 3D
- **Use semantic knowledge to reconstruct object (i.e. use shape models)**

What do we mean by higher-order priors?



Modern SLAM formulation:

Measurements = $f(T_j, \lambda)$+noise $\dim(\lambda)=d<<n$

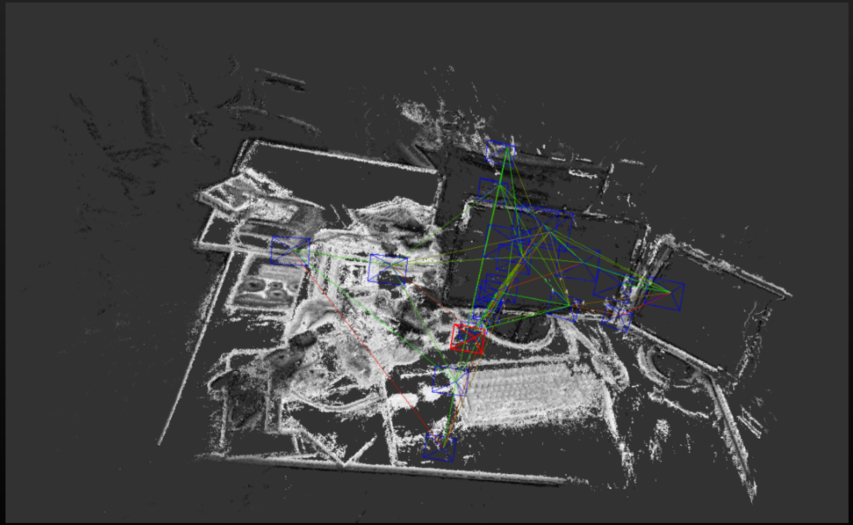
Residuals = Measurements - $f(T_j,\lambda)$

- Find the nearest shape on “a shape manifold” that agrees with the measurements!

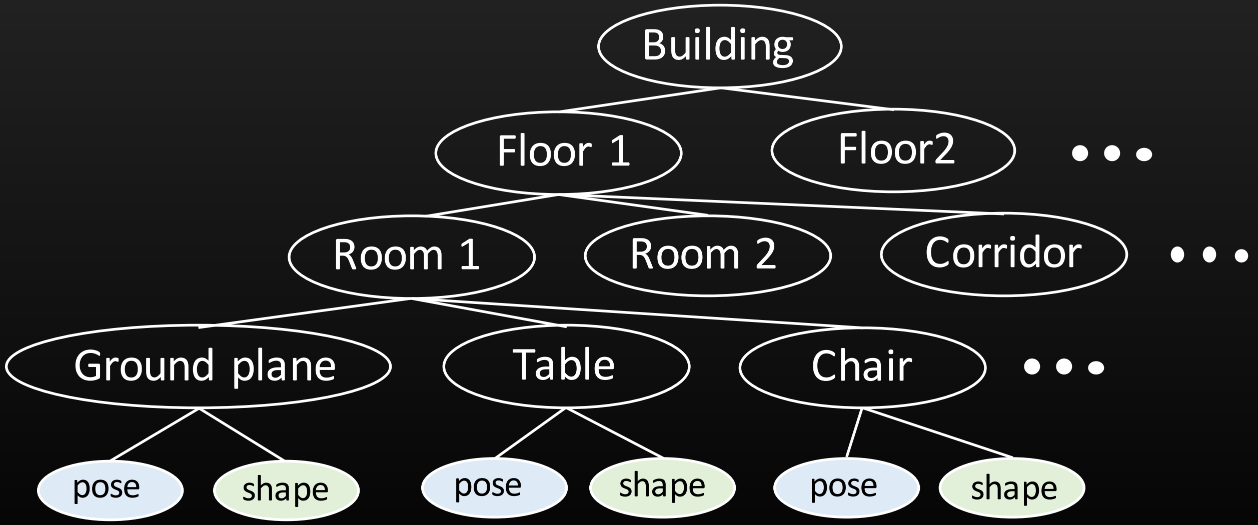
Spatial AI (Term "coined" by Andrew Davison)

- An evolution of SLAM
  - Joint geometric-semantic scene understanding at the level of objects
    - What objects? Where? **<u>What is their shape?</u>**

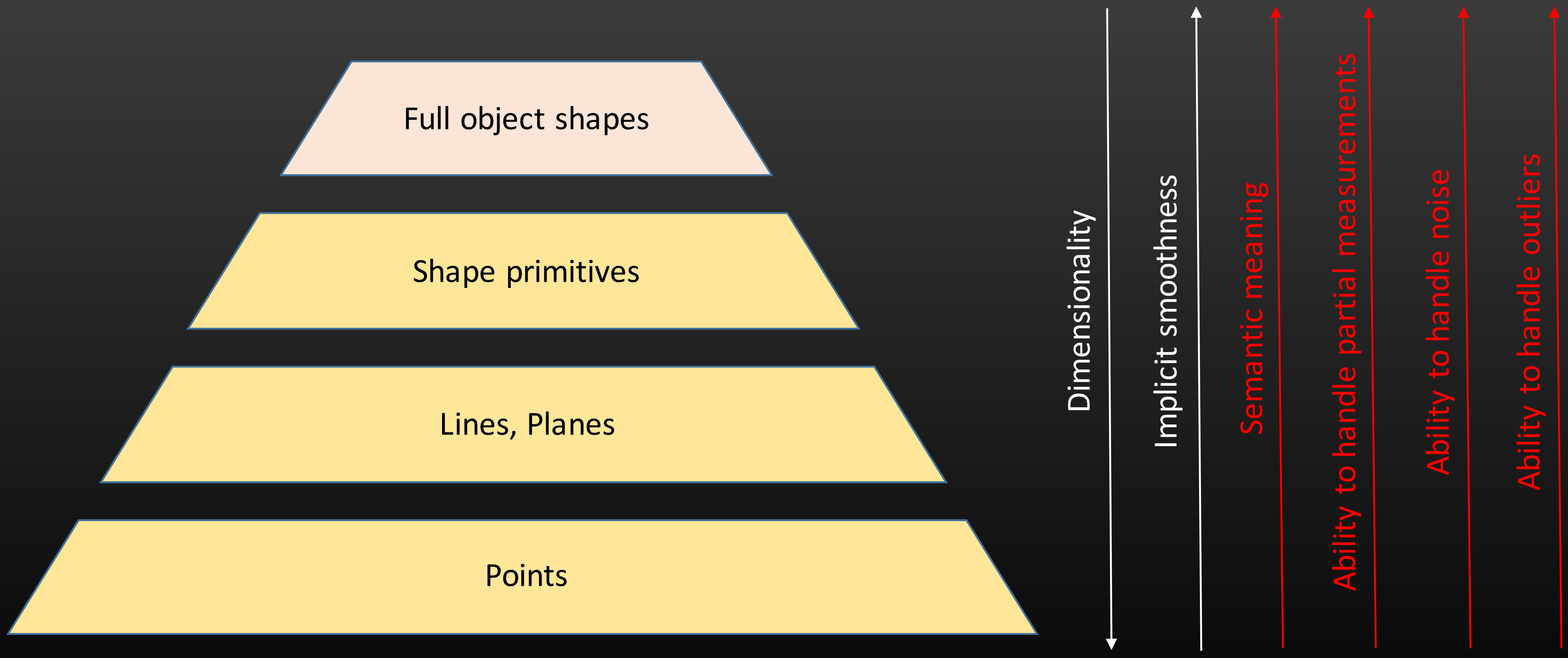
"old" style map
=a 'primitive" point cloud, mesh etc.



"new" style map
= HD map



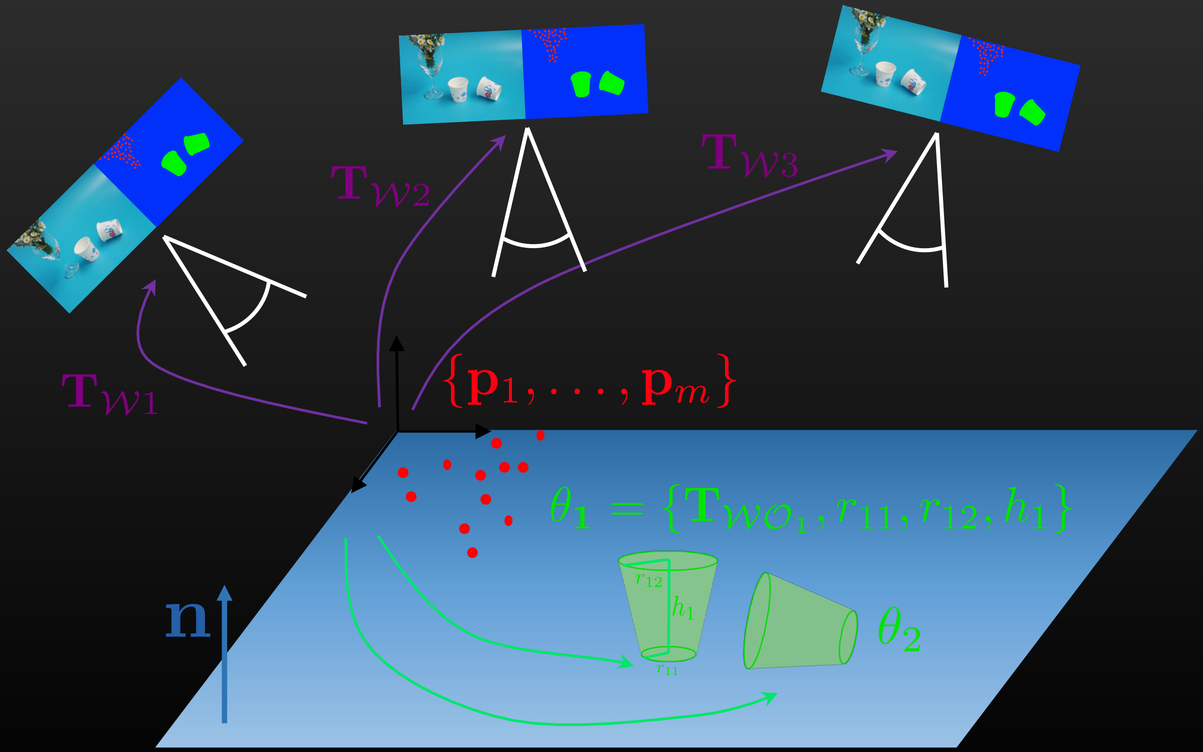
A hierarchy of scene element models

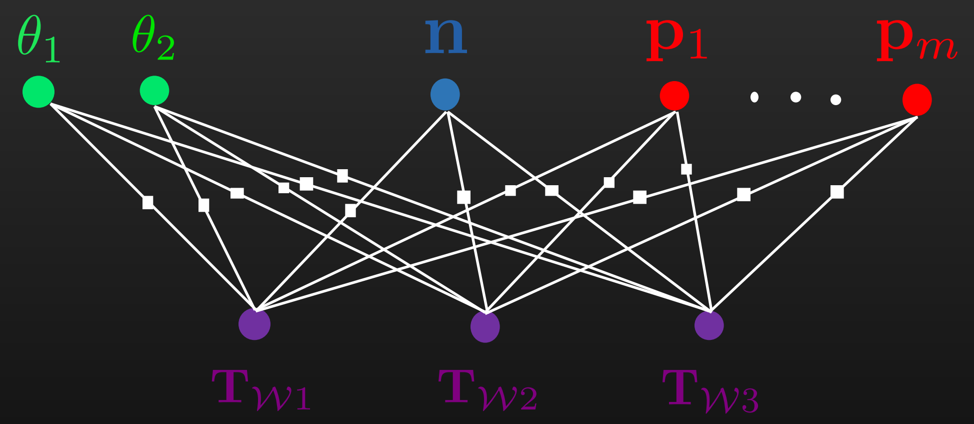


Deep-SLAM++

[L Hu, W Xu, K Huang, and **L Kneip**. Deep-SLAM++: Object-level RGBD SLAM based on class-specific deep shape priors, arXiv:cs.CV:1907.09691, 2019]

- Hybrid Graph
  - Includes both low and high-level (object-level) features
  - Has pose and shape parameters for high level landmarks



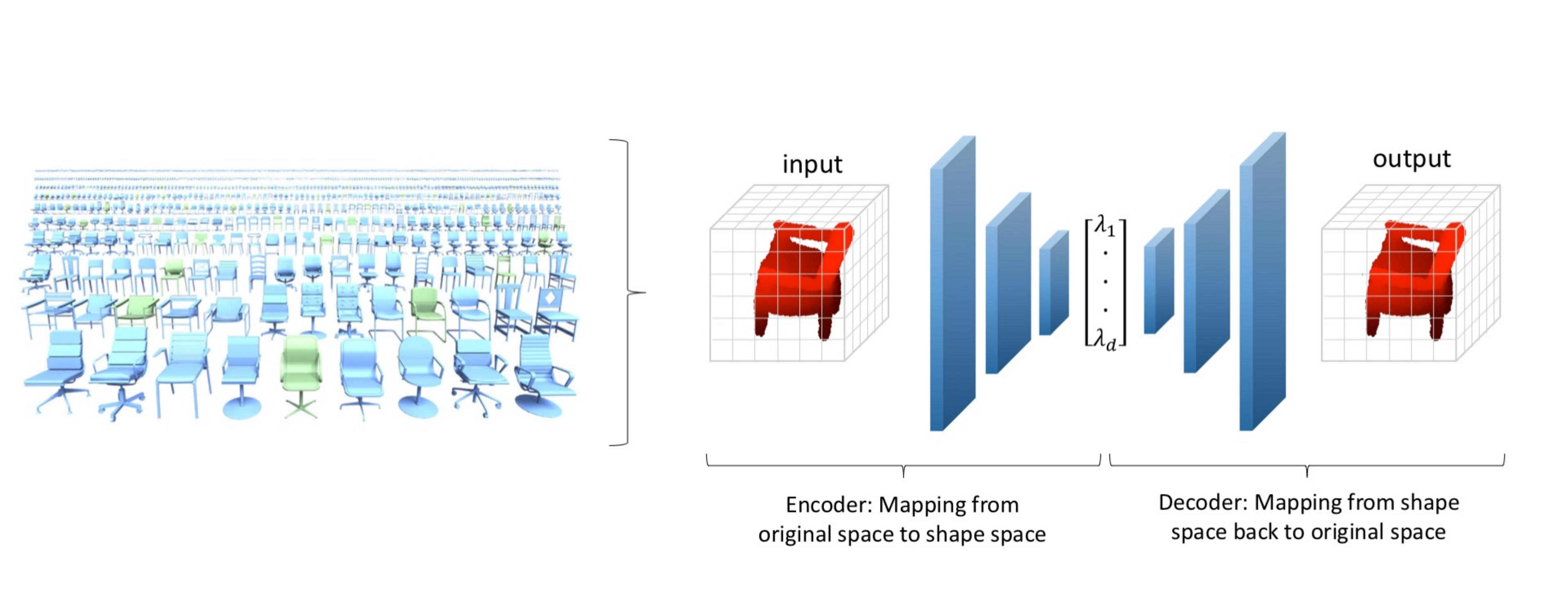


$$E = \underbrace{\sum_{j=1}^{2}\sum_{i=1}^{n} A_{ij}\left(\mathbf{T}_{\mathcal{W}i}, \mathbf{T}_{\mathcal{W}j}, r_{j1}, r_{j2}, h_j\right)}_{\substack{\text{High-level features} \\ \text{(objects with pose and shape)}}} + \underbrace{\sum_{i}^{n} B_i\left(\mathbf{T}_{\mathcal{W}i}, \mathbf{n}\right)}_{\substack{\text{Mid-level features} \\ \text{(e.g. planes, lines)}}} + \underbrace{\sum_{j=1}^{m}\sum_{i=1}^{n} C_{ij}\left(\mathbf{T}_{\mathcal{W}i}, \mathbf{p}_j\right)}_{\substack{\text{Low-level features} \\ \text{(e.g. points)}}}$$

Deep-SLAM++

[L Hu, W Xu, K Huang, and **L Kneip**. Deep-SLAM++: Object-level RGBD SLAM based on class-specific deep shape priors, arXiv:cs.CV:1907.09691, 2019]

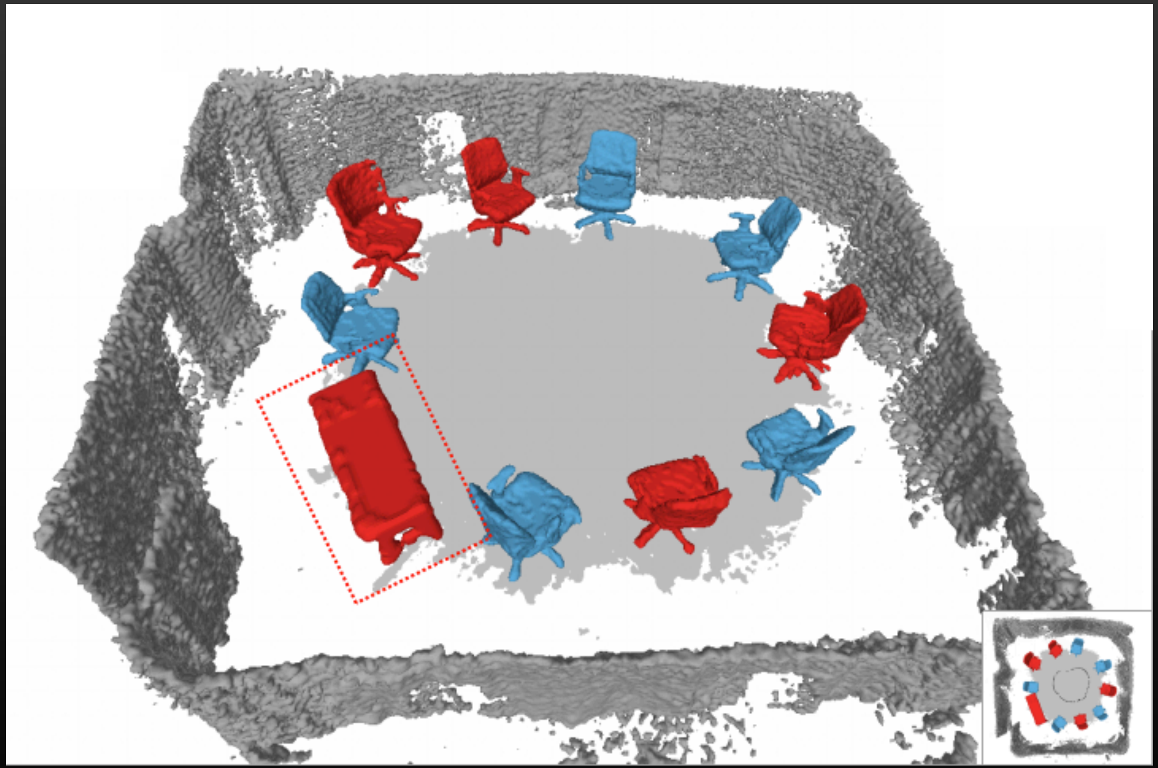
- Low-dimensional, differentiable complex shape representations?
  - Dimensionality reduction/shape manifold learning with auto-encoders



Deep-SLAM++

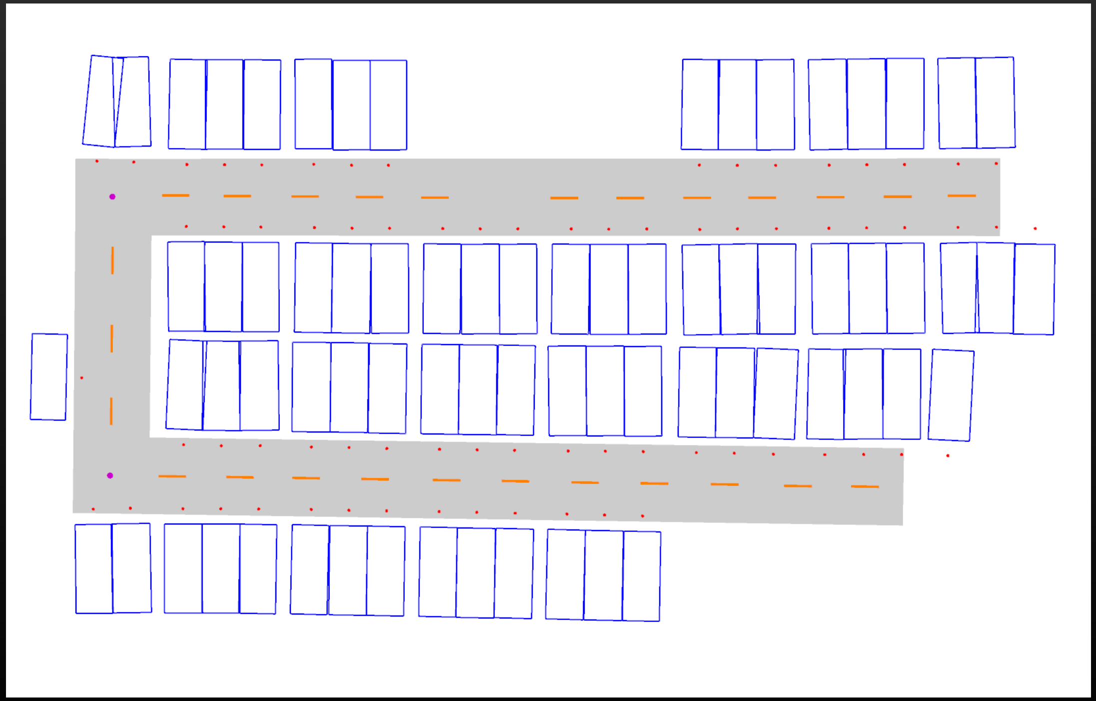
[L Hu, W Xu, K Huang, and **L Kneip**. Deep-SLAM++: Object-level RGBD SLAM based on class-specific deep shape priors, arXiv:cs.CV:1907.09691, 2019]

- Results on an indoor scenario
  - Chairs and tables are generated by a neural network
  - Reasonable geometries are obtained by differentiating latent variables w.r.t the measurements



Deep-SLAM++

- Underground parking lot mapping for AVP
  - Only surround-view fish-eye images; **AI embedded into FPGA**
  - Full optimization over pose and higher level shape parameters (lanes, parking lots, …)



Deep-SLAM++

- Towards AVP
  - Online localization and autonomous driving based on high-level feature map

Thank you!

MPL
Mobile Perception Lab

lkneip@shanghaitech.edu.cn

http://mpl.sist.shanghaitech.edu.cn