Bundle Adjustment with Feature Scale Constraint for Enhanced Estimation Accuracy

Vladimir Ovechkin
Under the supervision of Asst. Prof. Vadim Indelman

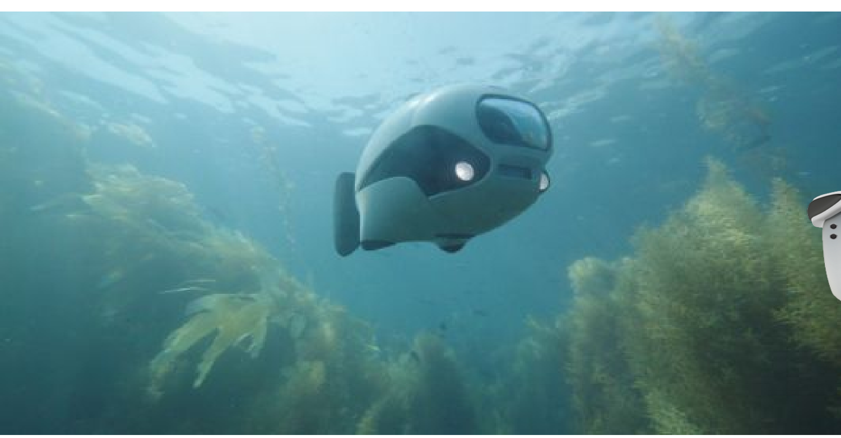
Technion
02.08.2017

TASP | Technion Autonomous System Program

ANPL | Autonomous Navigation and Perception Lab

Vision-based navigation

- Aerial unmanned vehicles
- Self-driving cars
- Augmented reality
- Underwater operations
- Space operations



ANPL | Autonomous Navigation and Perception Lab

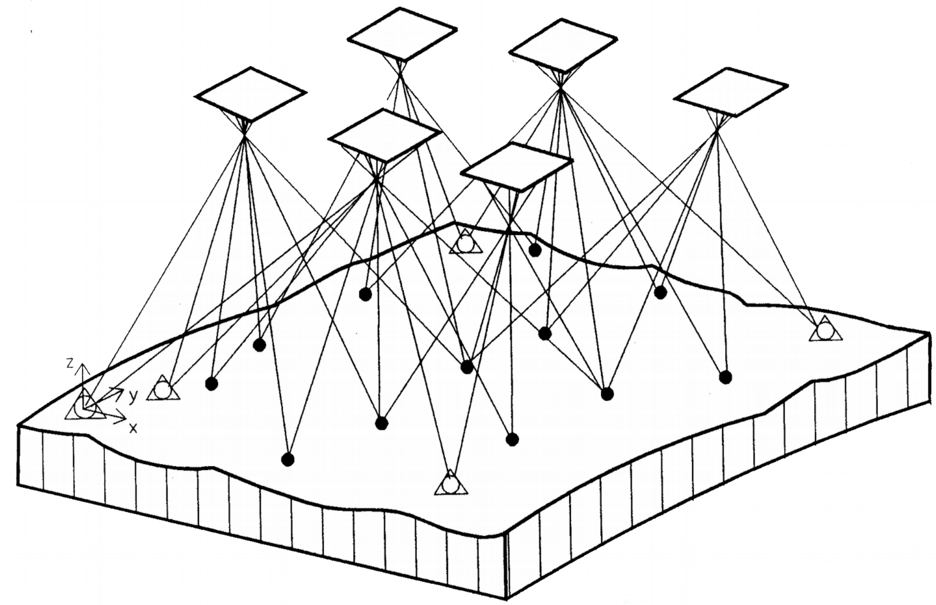
Problem description

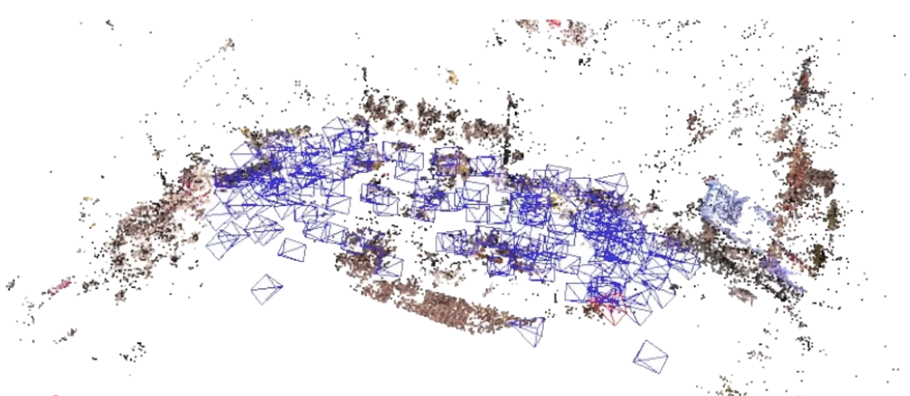
<u>Problems:</u>

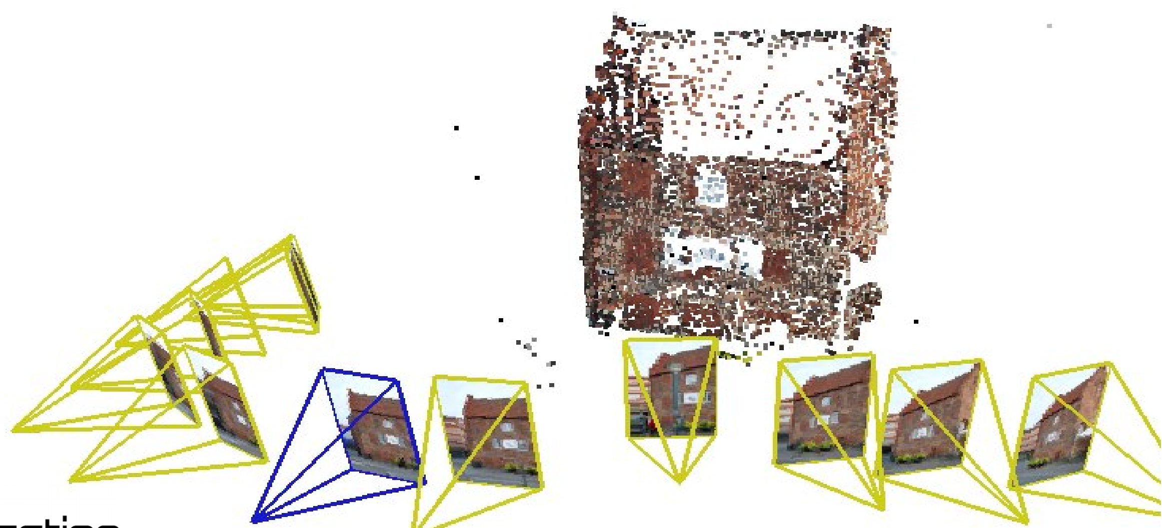
- Mapping
- 3D reconstruction
- Accurate self-pose estimation

<u>Common approach to use:</u>

Bundle Adjustment





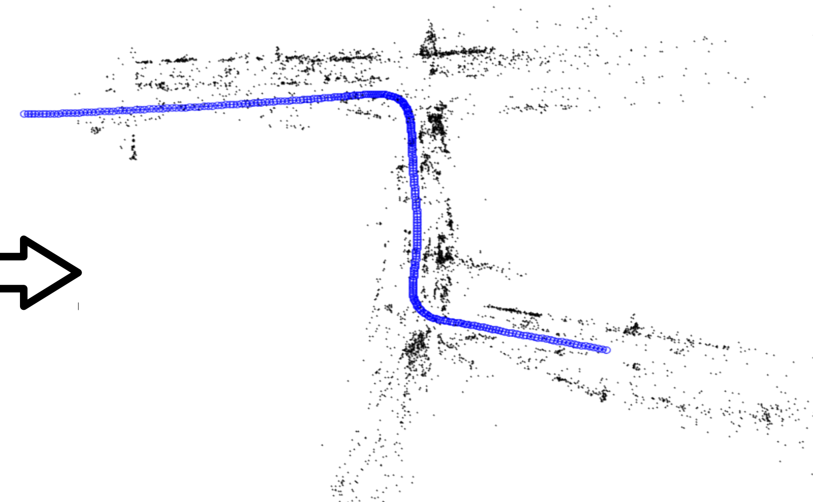


Monocular SLAM

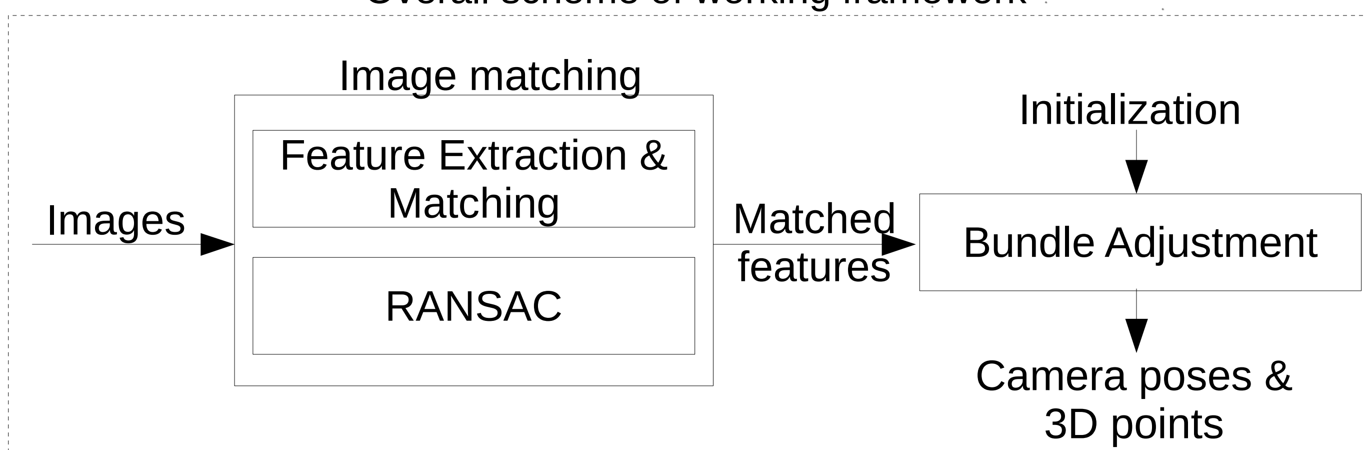
Input: sequence of images

Preprocessing: Feature matching and tracking

Output: 3D-reconstruction



Overall scheme of working framework



Assumptions:
- no initial map
- no GPS
- only a single camera sensor

Related work

- BA: minimize re-projection errors
[B. Triggs Bundle Adjustment — A Modern Synthesis]

- visual feature-based SLAM - online
[M. Kaess – iSAM2]

- using feature scale for identifying far-away landmarks
[D.-N. Ta – Vistas and parallel tracking]

- correcting monocular scale drift by placing a prior on the size of the objects
[D. P. Frost – Object-Aware Bundle Adjustment]

Monocular SLAM scale drift problem

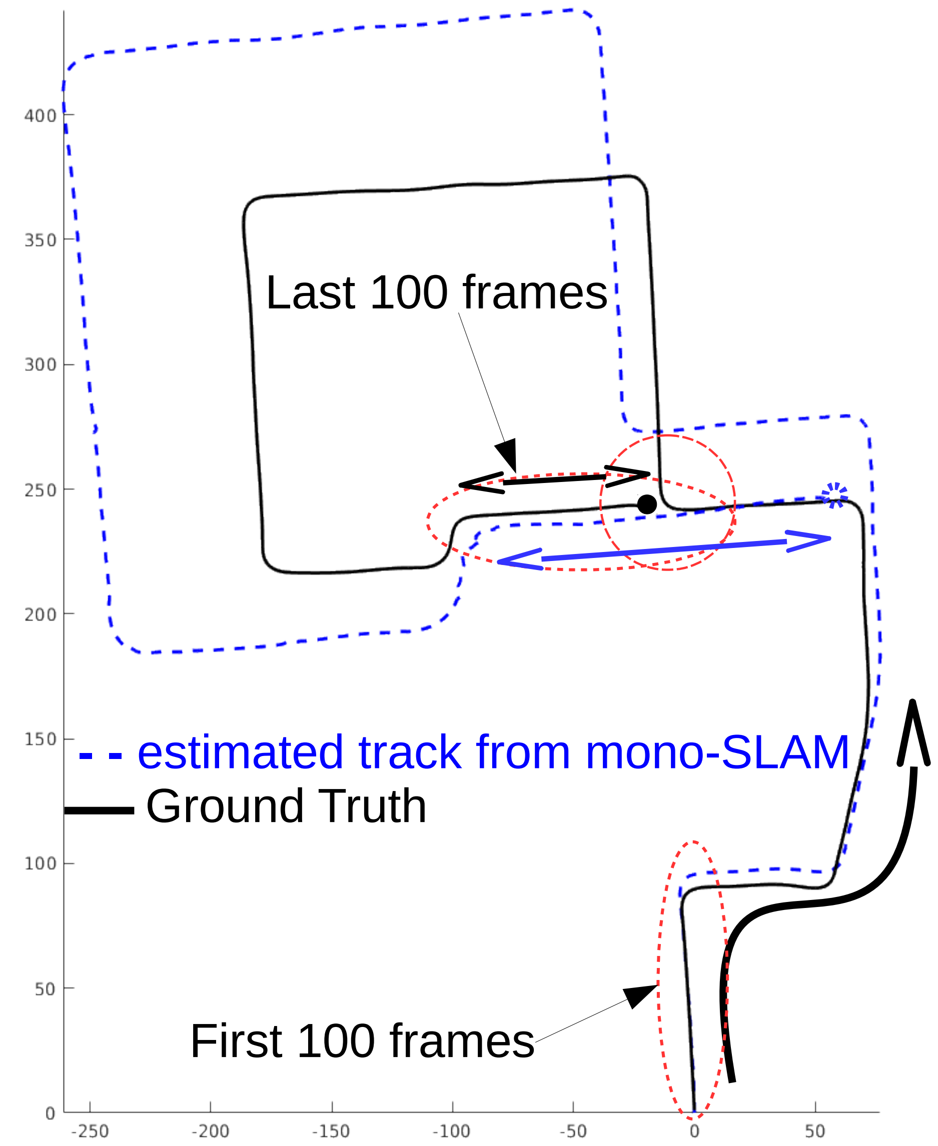
Known problem is **scale drift along optical axis** with time.

Example:

Error along optical axis is growing with time

For the first 100 frames:
Error: 8.6%

For the last 100 frames:
Error: 94.6%

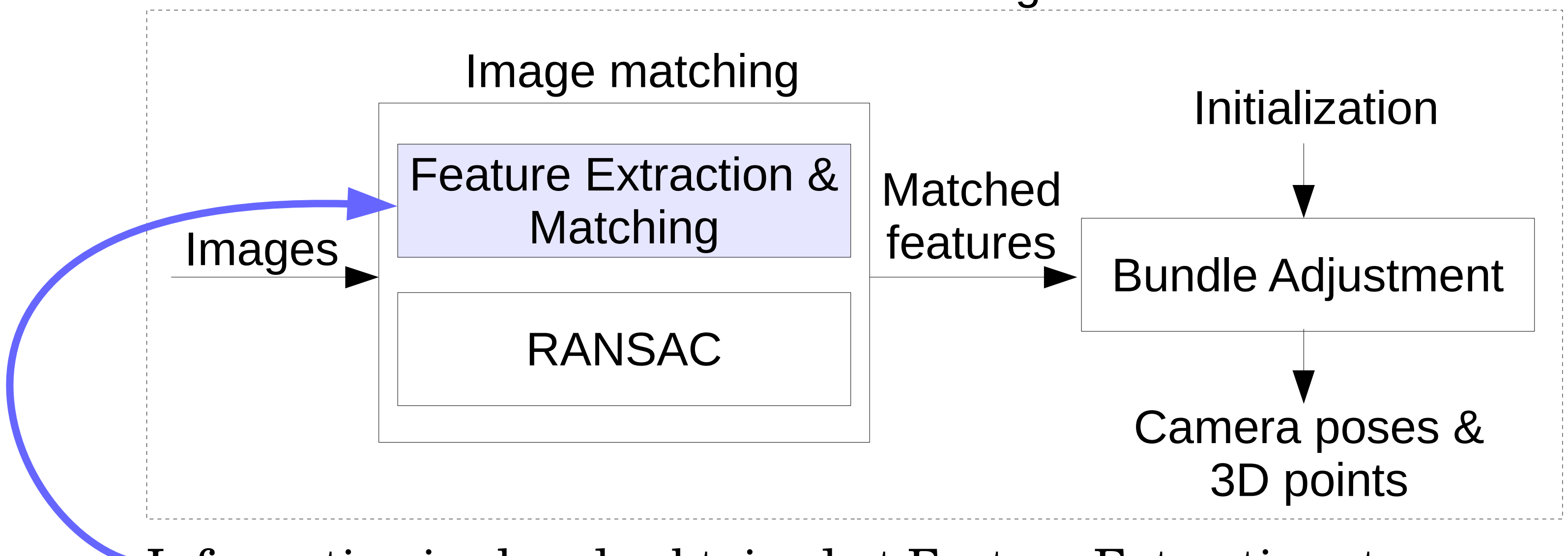


ANPL Autonomous Navigation and Perception Lab

Contribution

We introduce scale constraints into Bundle Adjustment aiming to improve accuracy along optical axis direction.

Overall scheme of working framework

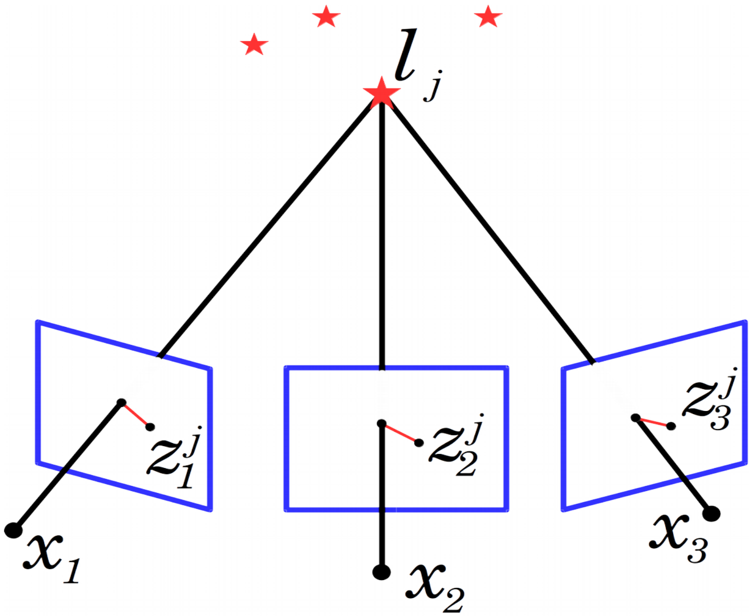


Information is already obtained at Feature Extraction step.

Outline

- Background:
  - Bundle Adjustment (BA)
  - SIFT features
- Our approach:
  - Adding new constraints to BA
  - Improving accuracy of detected feature scale
- Results
- Variations with new constraints
- Conclusions

BA notations



Landmark set

$$L \doteq \{l_1, \ldots, l_j, \ldots, l_M\}$$

Pose set

$$X \doteq \{x_1, \ldots, x_i, \ldots, x_N\}$$

Single observation of landmark ***j*** from pose ***i***: z_i^j

All measurements set

$$\mathcal{Z} \doteq \{z_i^j\}$$

ANPL | Autonomous Navigation and Perception Lab

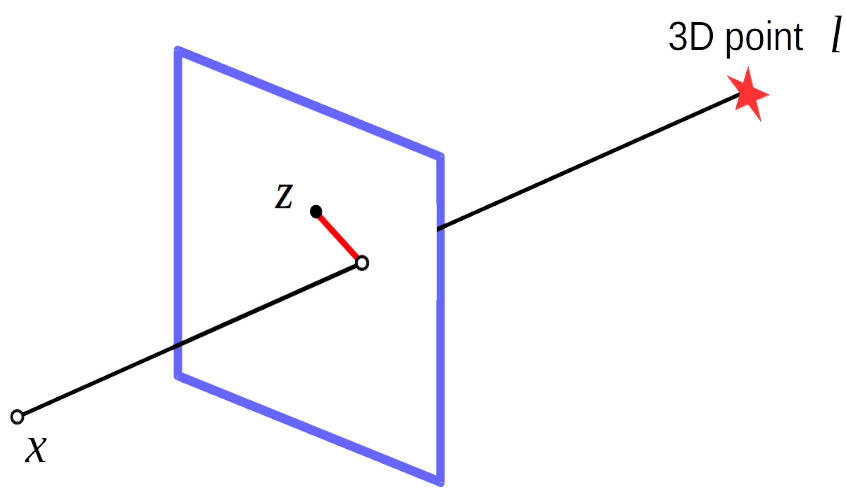
BA Probabilistic Representation

Measurement model: $z = \pi(x, l) + v \qquad v \sim N(0, \Sigma_v)$

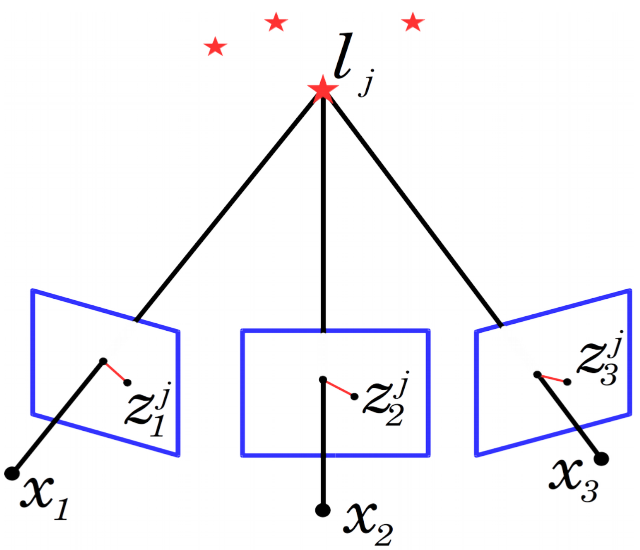
Projection operator: $\pi(x, l)$

Measurement likelihood:

$$p(z|x, l) = \frac{1}{\sqrt{det(2\pi\Sigma_v)}} exp\left(-\frac{1}{2}\underbrace{||z - \pi(x, l)||^2_{\Sigma_v}}_{\text{re-projection error}}\right)$$



BA Probabilistic Representation



Using Bayes rule we have:

$$p(x,l|z) = \frac{p(x,l)p(z|x,l)}{p(z)} \propto \underbrace{p(x,l)}_{\text{Prior}}\underbrace{p(z|x,l)}_{\text{Measurement likelihood}}$$

Objective: calculate the joint posterior distribution

$$p\left(X, L|\mathcal{Z}\right)$$

ANPL Autonomous Navigation and Perception Lab

BA Probabilistic Representation

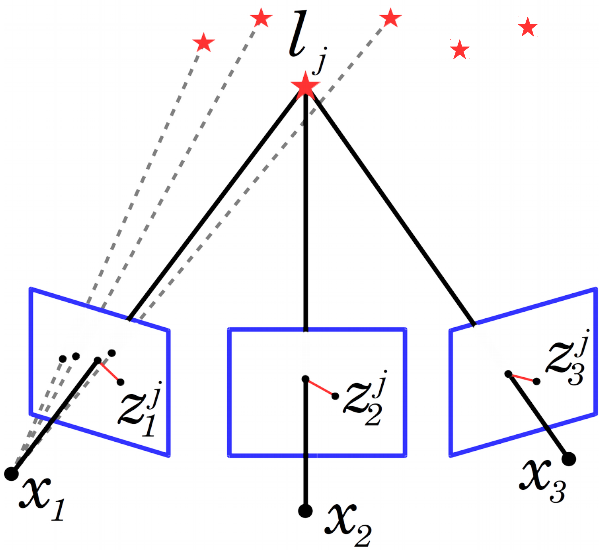
For N images:

$$p(X, L|\mathcal{Z}) = priors \cdot \prod_{i}^{N} \prod_{j \in \mathcal{M}_i} p\left(z_i^j | x_i, l_j\right)$$

Where $\mathcal{M}_i$ is a landmark subset observed from camera i

$$p\left(z_i^j | x_i, l_j\right) \propto \exp\left(-\frac{1}{2}\left\|z_i^j - \pi\left(x_i, l_j\right)\right\|_{\Sigma_v}^2\right)$$

Example:
For 1st camera pose all observed landmarks belong to subset $\mathcal{M}_1$



ANPL | Autonomous Navigation and Perception Lab

BA Probabilistic Representation

Maximum a posteriori solution:

$$X^*, L^* = \arg\max_{X,L} p\left(X, L | \mathcal{Z}\right)$$

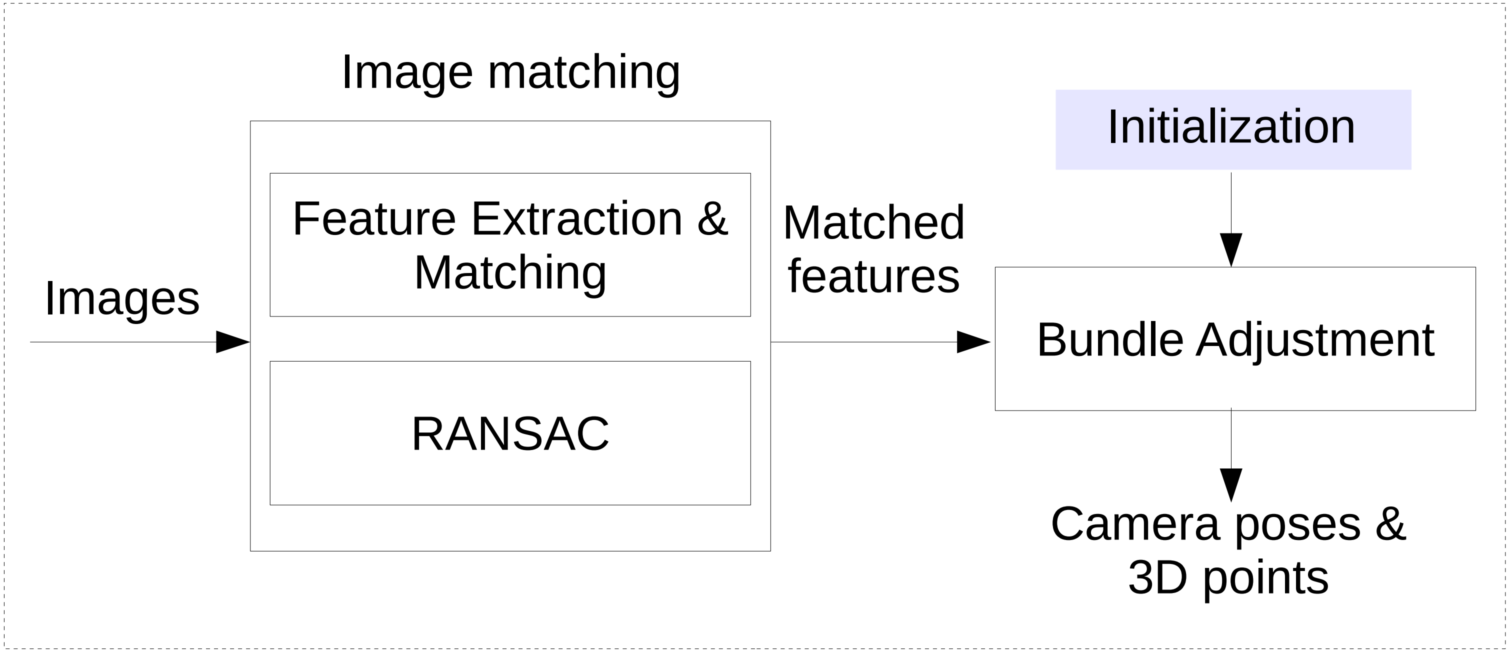
After omitting prior terms cost function to minimize is:

$$J_{BA}\left(X, L\right) = \sum_{i}^{N} \sum_{j \in \mathcal{M}_i} \left\| z_i^j - \pi\left(x_i, l_j\right) \right\|_{\Sigma_v}^2$$

ANPL Autonomous Navigation and Perception Lab

Variable initialization

Overall scheme of working framework

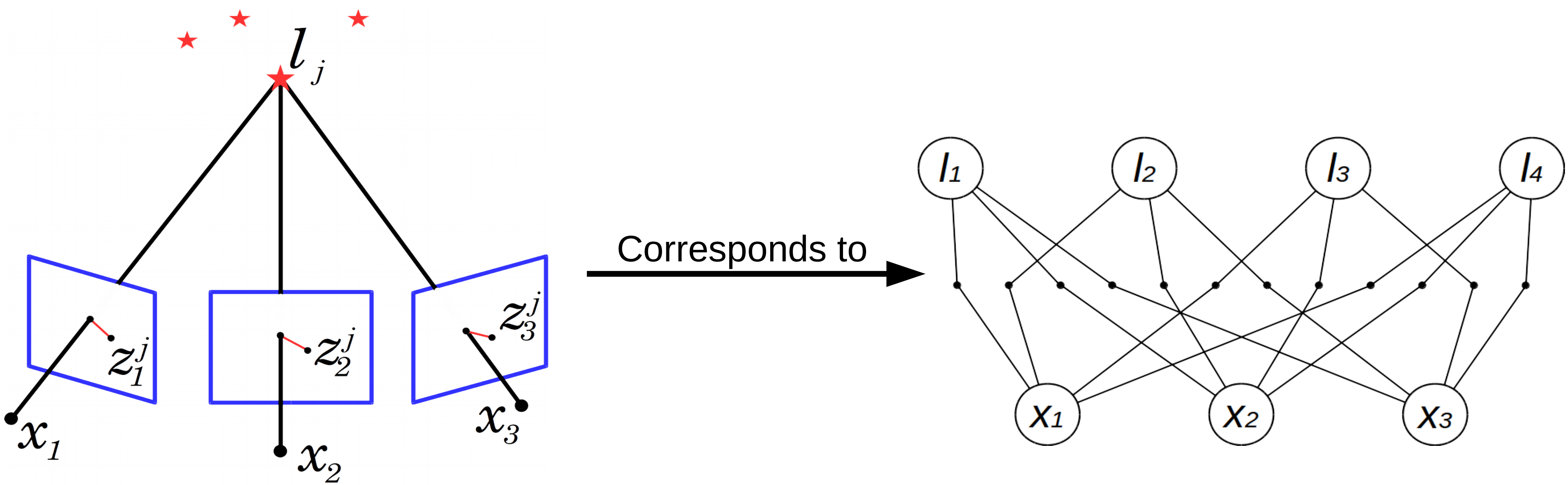


For optimization process we need to initialize the variables.

- Pose variables: via visual odometry
- Landmark variables: via triangulation

Factor Graph BA representation

<u>Example</u>: *all landmarks L are observed from all poses X.*



$$p\left(z_i^j | x_i, l_j\right) \propto \exp\left(-\frac{1}{2}\left\|z_i^j - \pi\left(x_i, l_j\right)\right\|_{\Sigma_v}^2\right) \doteq f_{proj}$$

Currently **only image feature coordinates** are involved.

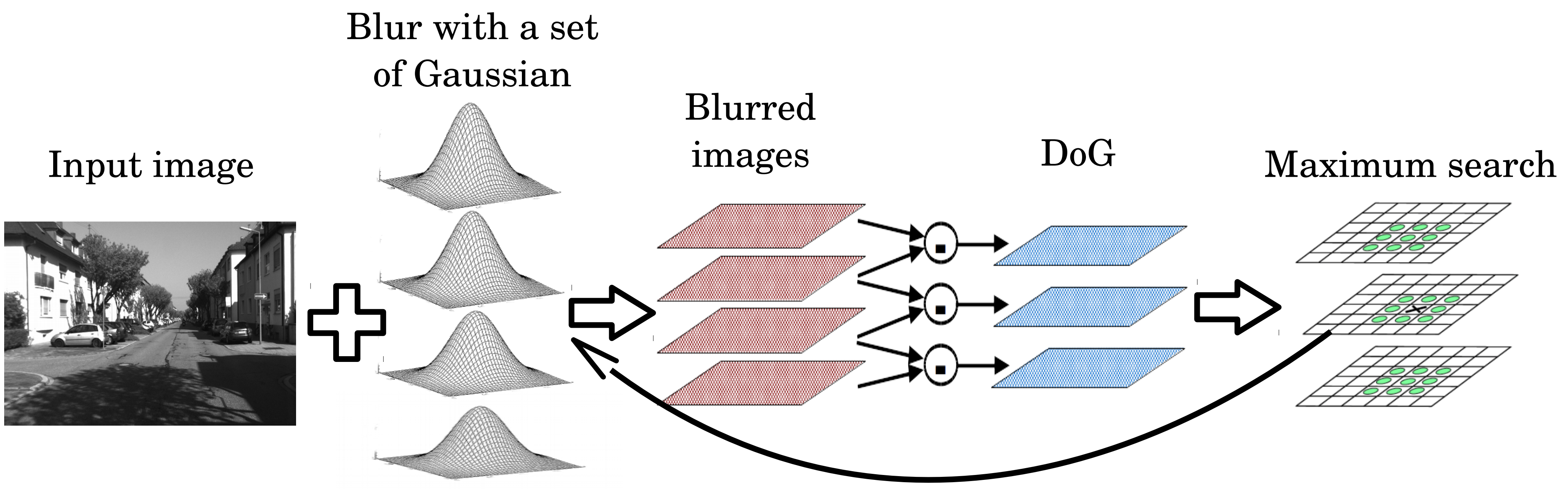
Outline

- Background:
  - Bundle Adjustment (BA)
  - **SIFT features**
- Our approach:
  - Adding new constraints to BA
  - Improving accuracy of detected feature scale
- Results
- Variations with our approach
- Conclusions

SIFT features

- Each feature is represented by image coordinates, scale and orientation.
- Only image feature coordinates are involved in BA optimization.

SIFT feature scale



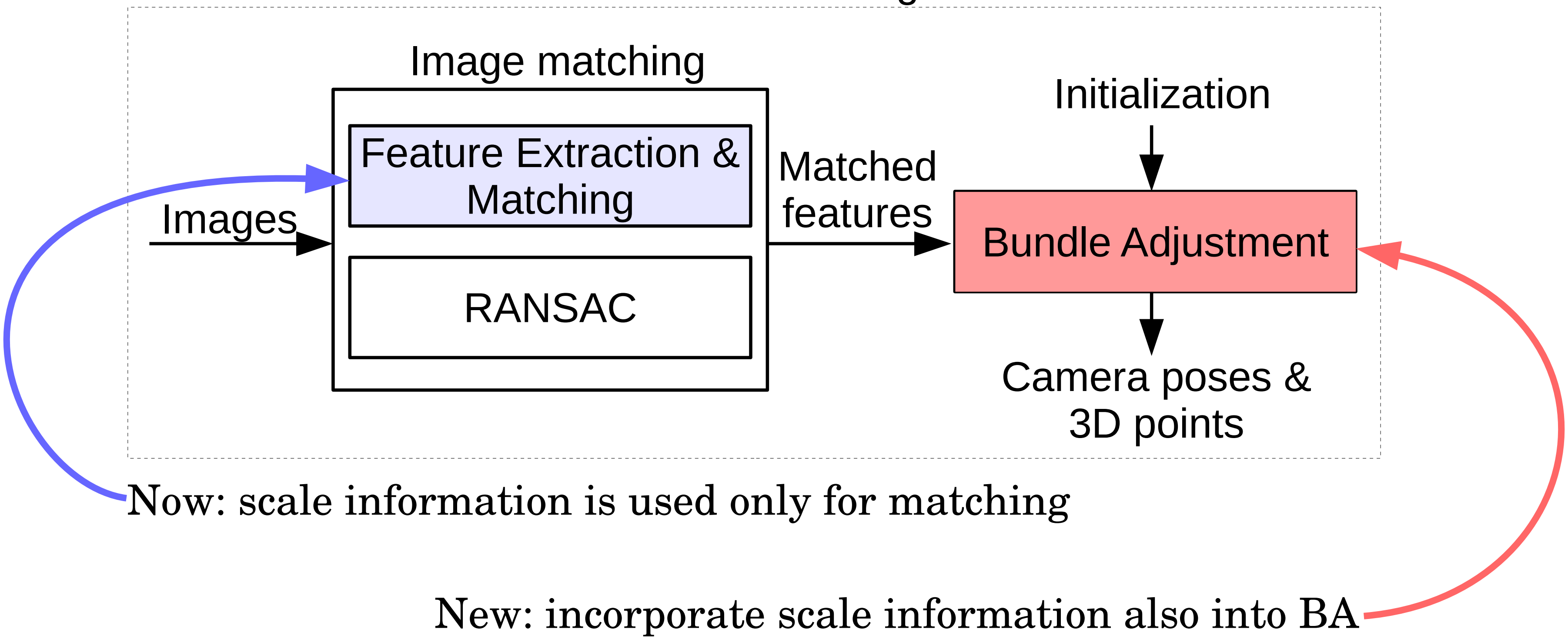
- Blur each input image with a set of Gaussian kernels with given covariance.
- Calculate Difference of Gaussians.
- Search for local maxima (features) among layers and pixels.
- Feature scale is equal to Gaussian kernel covariance corresponding to the layer where the local maxima is found.

Key observation

- Scale changes consistently across frames.
- Each feature scale is a function of environment and camera pose.
- We can use this property to predict landmark scale in the next frame!

“Big” picture

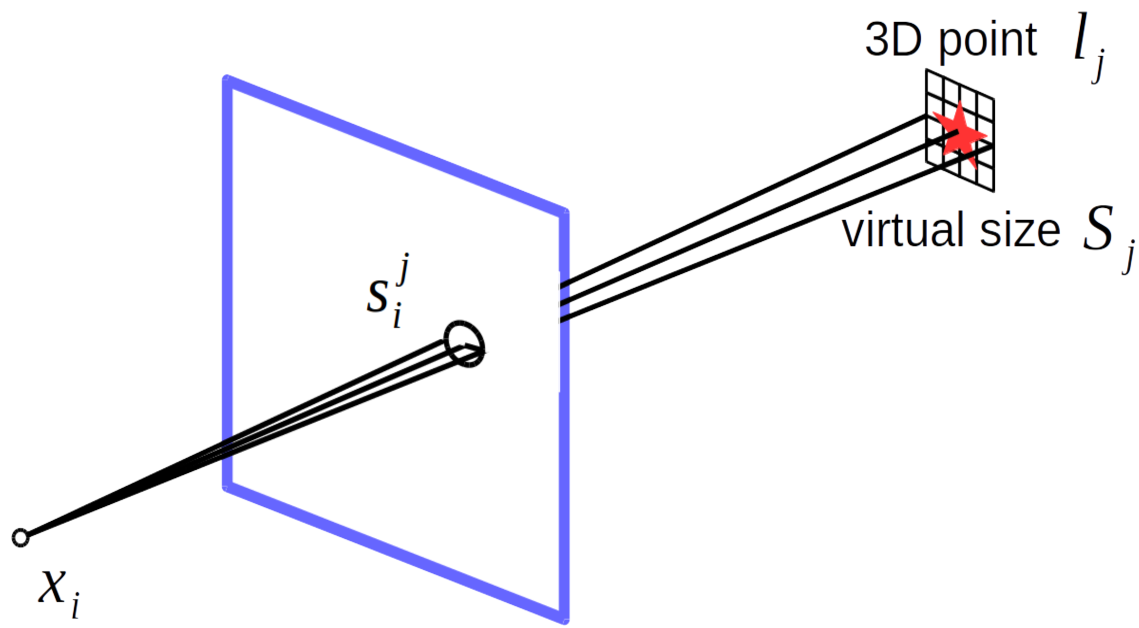
Overall scheme of working framework



Scale constraint

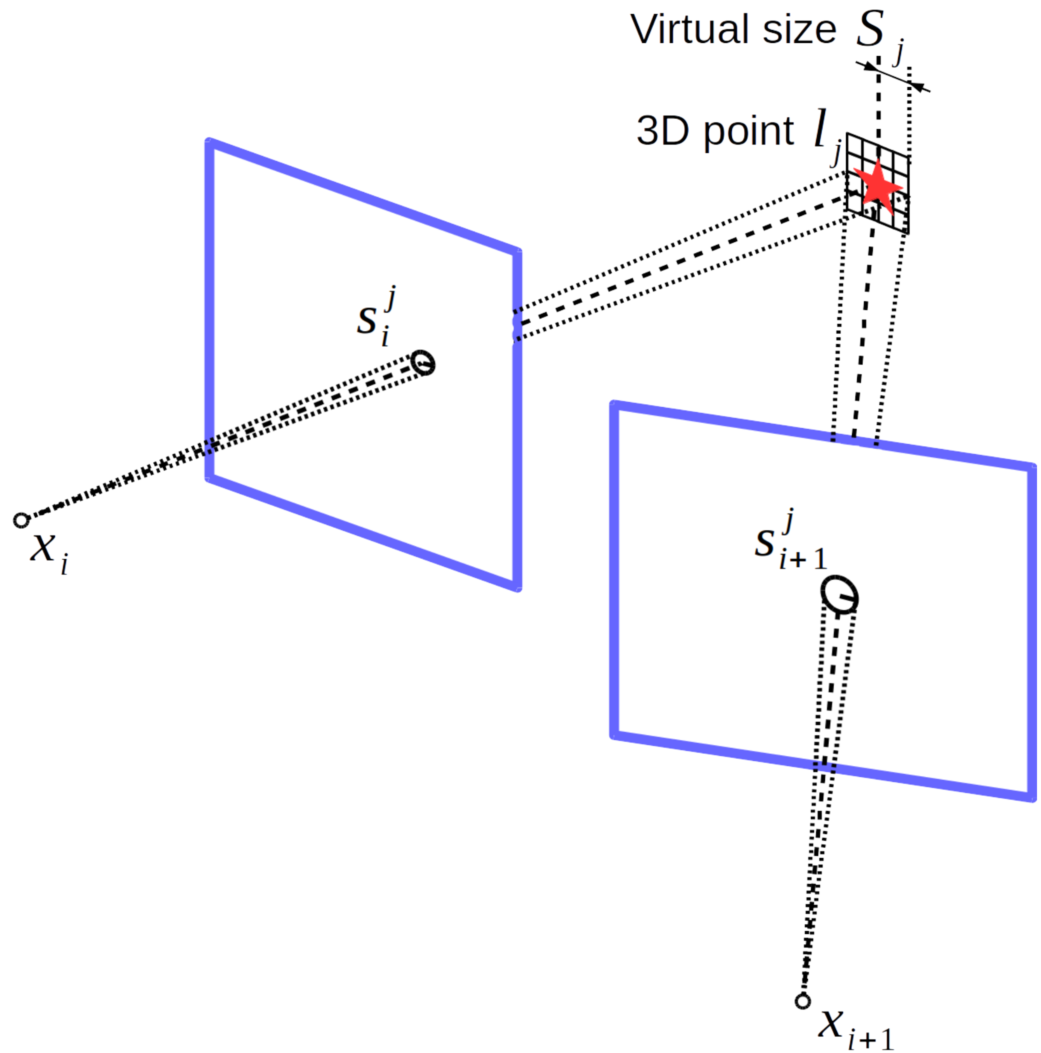
Consider an image feature in the i-th image that corresponds to a landmark with a 3D position l_j

- Denote the detected feature scale by s_i^j
- Denote by S_j the corresponding environment patch, or virtual landmark size, centered around landmark l_j



Scale constraint

SIFT is a <u>scale invariant</u> feature!

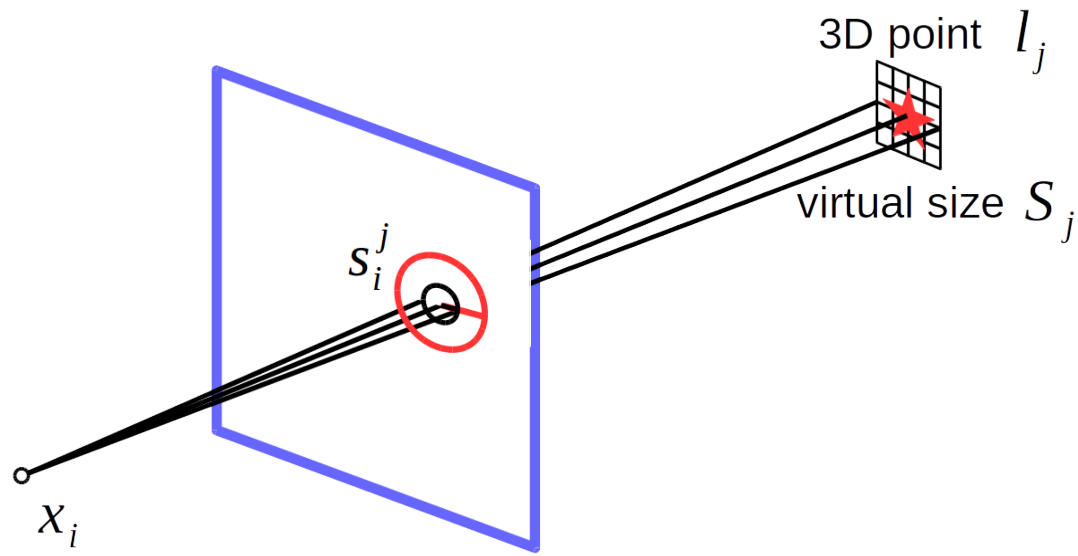


Scale constraint

Scale observation model:

$$s_i^j = f\frac{S_j}{d_i^j} + v_i$$

measured predicted



We model the noise is Gaussian

$$v_i \sim N\left(0, \Sigma_{fs}\right)$$

Scale constraint

<u>Intuition:</u>

Feature scale depends on the distance along optical axis (and not on range between camera and landmark).

Scale observation model:

$$s_i^j = f\frac{S_j}{d_i^j} + v_i \qquad\qquad v_i \sim N\left(0, \Sigma_{fs}\right)$$

Virtual landmark size:

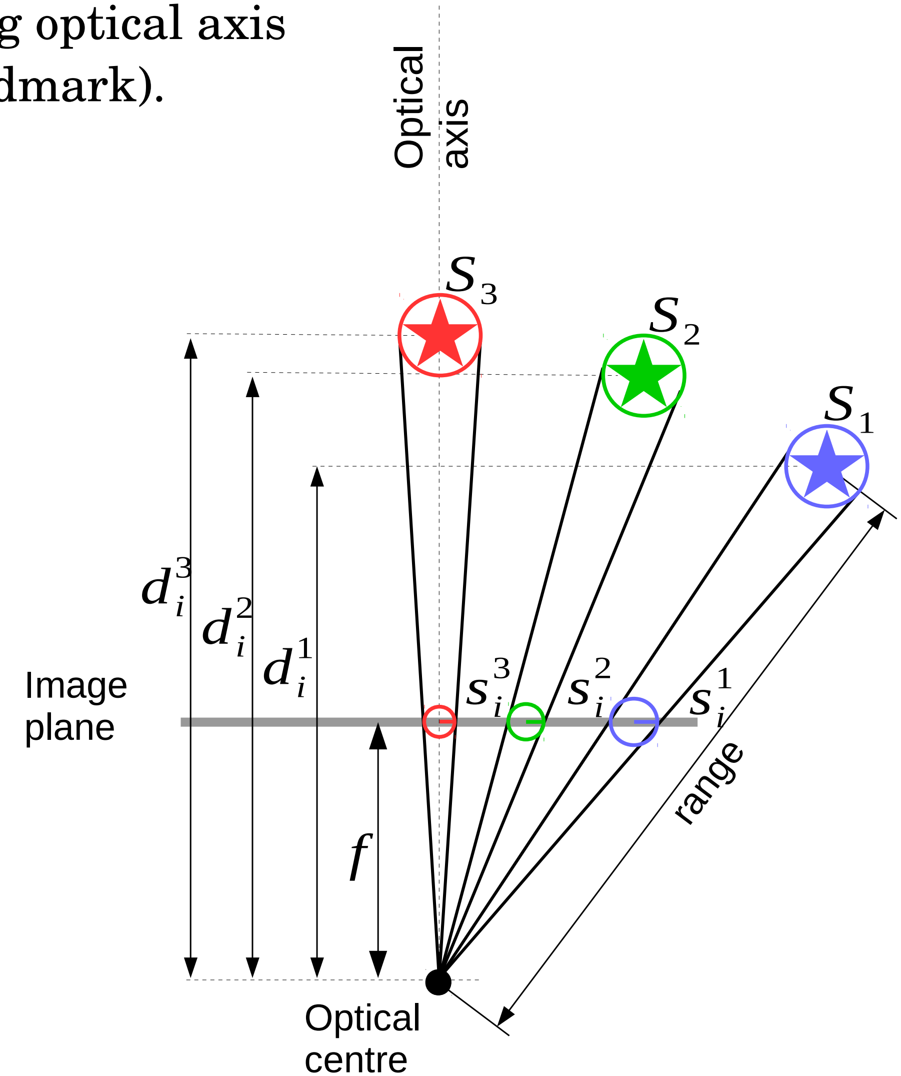
$$S_1 = S_2 = S_3$$

Feature scale:

$$s_i^1 > s_i^2 > s_i^3$$

Distance along optical axis:

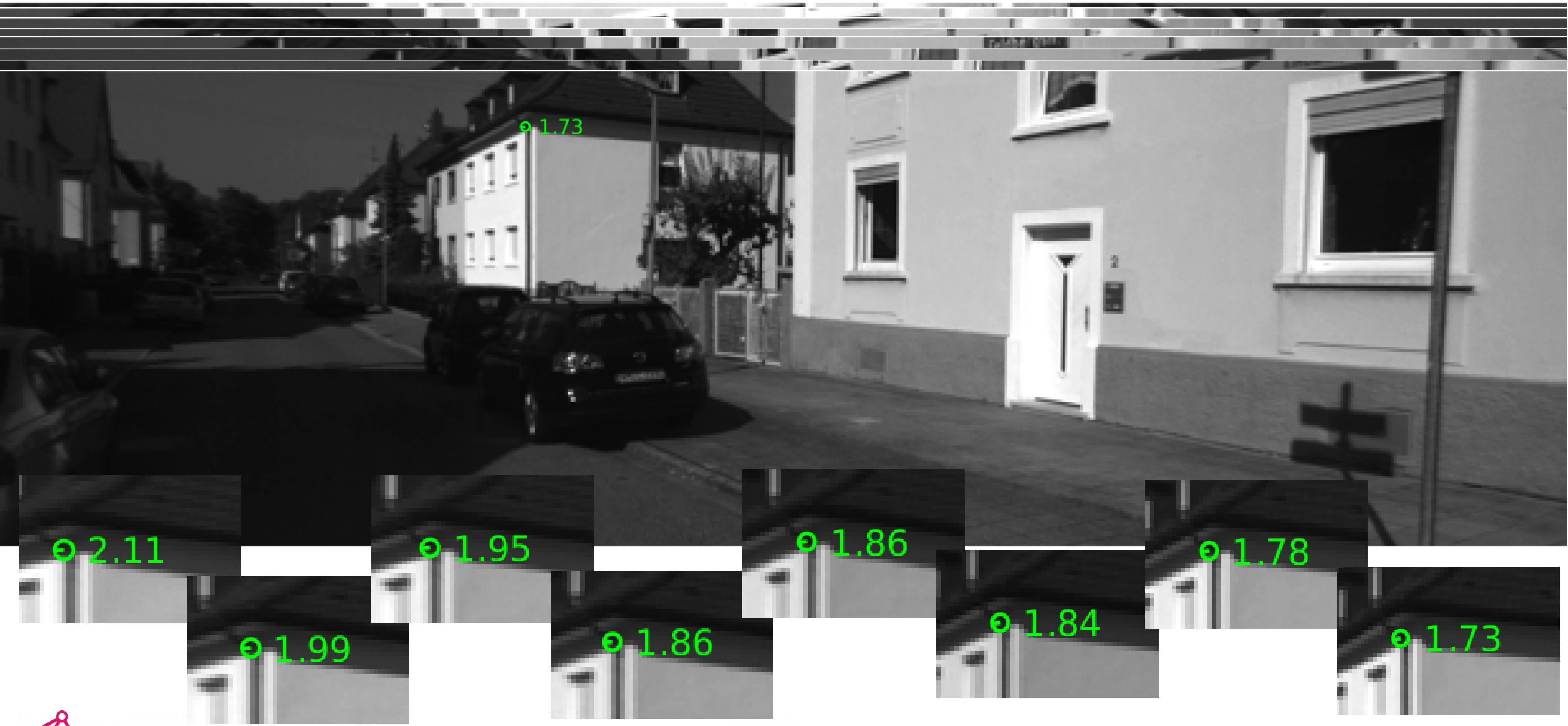
$$d_i^1 < d_i^2 < d_i^3$$



Scale constraint

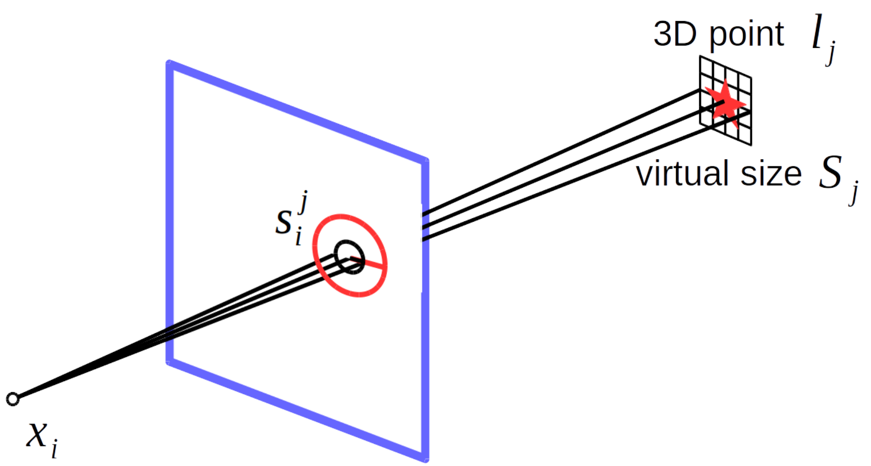
Intuition:

Feature scale is changing correspondingly to distance along optical axis from camera pose to represented environment patch.



ANPL Autonomous Navigation and Perception Lab

Scale constraints



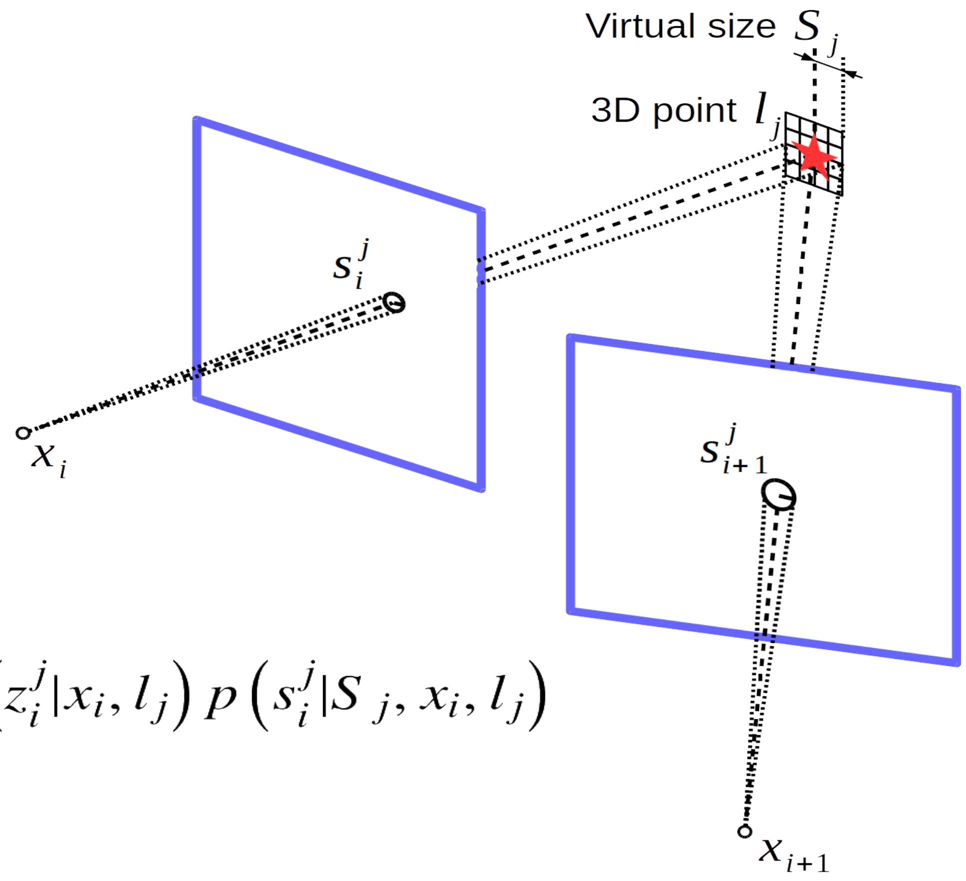
Scale observation model:

$$s_i^j = f\frac{S_j}{d_i^j} + v_i$$

Scale Measurement likelihood:

$$p\left(s_i^j | S_j, x_i, l_j\right) \propto \exp\left[-\frac{1}{2}\left\|s_i^j - f\frac{S_j}{d_i^j}\right\|^2_{\Sigma_{fs}}\right]$$

Scale constraints



New variable: S

New Objective joint posterior distribution:

$$p\left(X, L, S \mid \mathcal{Z}\right)$$

Joint posterior distribution

$$p\left(X, L, S \mid \mathcal{Z}\right) = priors \cdot \prod_{i}^{N} \prod_{j \in \mathcal{M}_i} p\left(z_i^j \mid x_i, l_j\right) p\left(s_i^j \mid S_j, x_i, l_j\right)$$

ANPL | Autonomous Navigation and Perception Lab

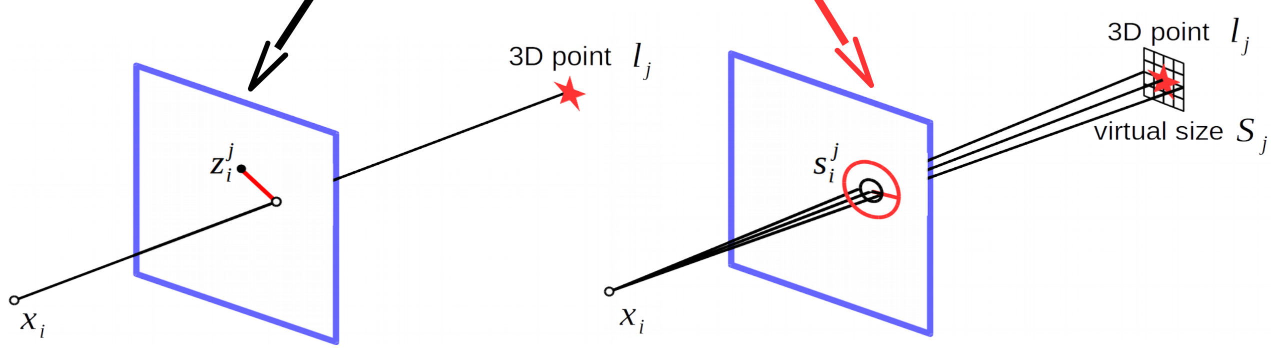
Joint cost function

Standard cost-function used in BA

$$J_{BA}(X,L) = \sum_{i}^{N} \sum_{j \in \mathcal{M}_i} \left\| z_i^j - \pi\left(x_i, l_j\right) \right\|_{\Sigma_v}^2$$

Cost-function used in BA with scale constraints

$$J(X,L,S) = \sum_{i}^{N} \sum_{j \in \mathcal{M}_i} \underbrace{\left\| z_i^j - \pi\left(x_i, l_j\right) \right\|_{\Sigma_v}^2}_{\text{re-projection error}} + \underbrace{\left\| s_i^j - f\frac{S_j}{d_i^j} \right\|_{\Sigma_{fs}}^2}_{\text{scale error}}$$



ANPL Autonomous Navigation and Perception Lab

Variable initialization

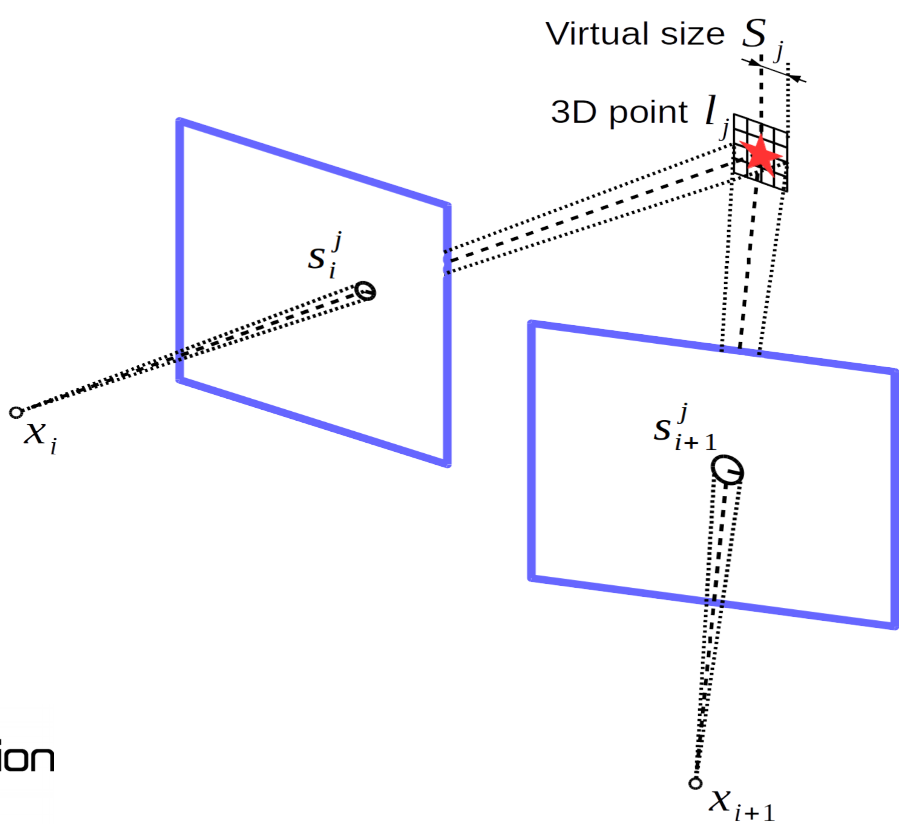
Reminder:

- Consider an image feature in the i-th image that corresponds to a landmark with a 3D position l_j
- Denote the detected feature scale by s_i^j
- Denote by S_j the corresponding environment patch, or virtual landmark size, centered around landmark l_j

Virtual landmark size initialization:

→ Landmark triangulation
→ Distance to landmark is known
→ Initialize landmark size

$$S_j = s_i^j \frac{d_i^j}{f}$$

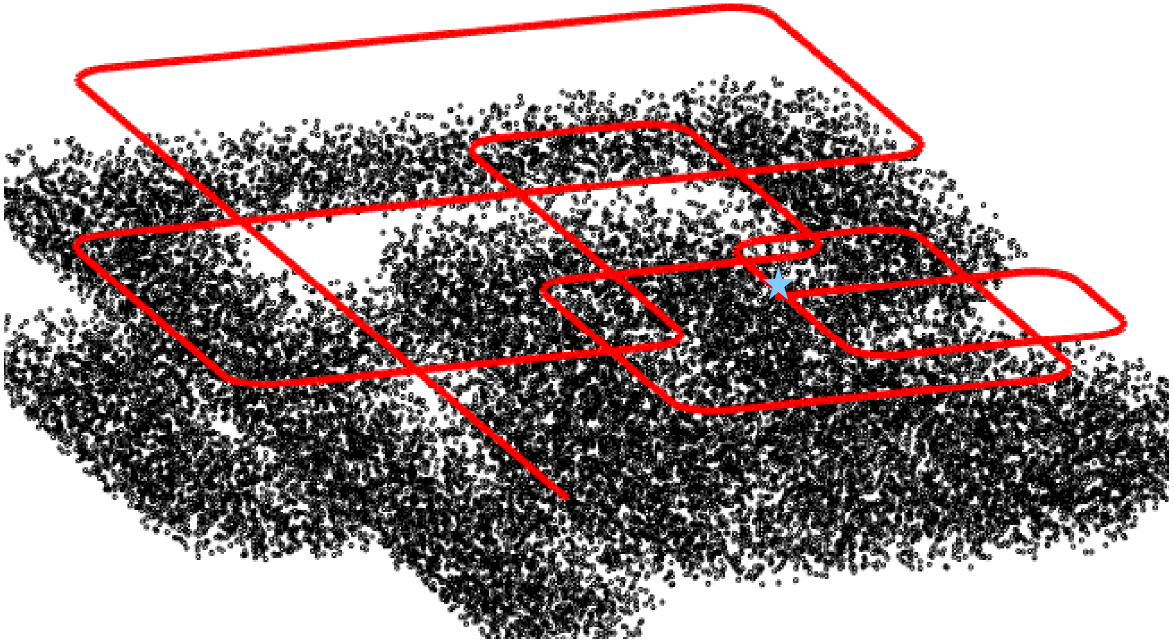


ANPL | Autonomous Navigation and Perception Lab

Simulation

Scenario:

- Downward-facing camera
- Constant height
- Landmarks scattered in 3D-space



ANPL Autonomous Navigation and Perception Lab

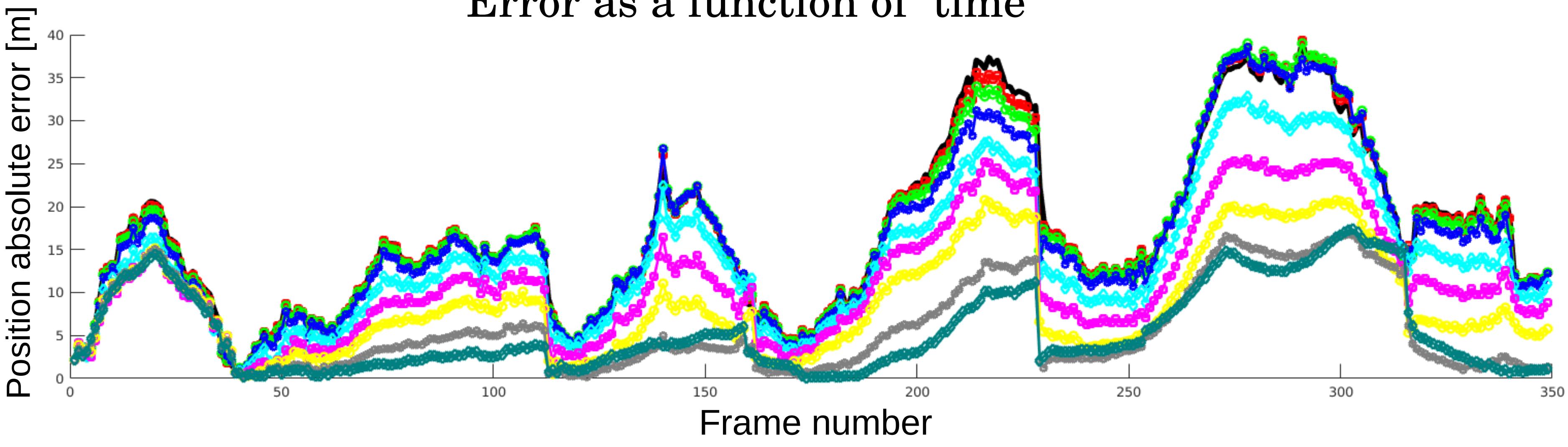
Simulation

Cost-function using projection and scale constraints

$$J\left(X,L,S\right)=\sum_{i}^{N}\sum_{j\in\mathcal{M}_i}\left\|z_i^j-\pi\left(x_i,l_j\right)\right\|_{\Sigma_v}^2+\left\|s_i^j-f\frac{S_j}{d_i^j}\right\|_{\Sigma_{fs}}^2$$

Represents noise in scale measurements

Error as a function of time

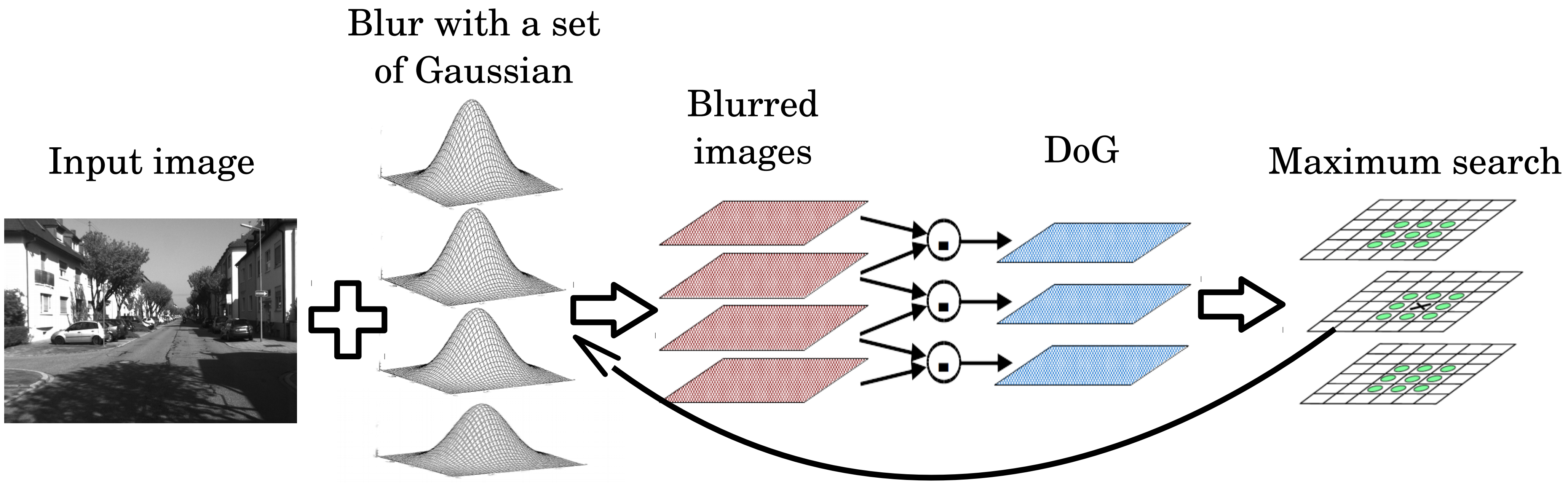


Each coloured curve corresponds to simulation with a fixed Σ_{fs}

Outline

- Background:
  - Bundle Adjustment (BA)
  - SIFT features
- Our approach:
  - Adding new constraints to BA
  - **Improving accuracy of detected feature scale**
- Results
- Variations with our approach
- Conclusions

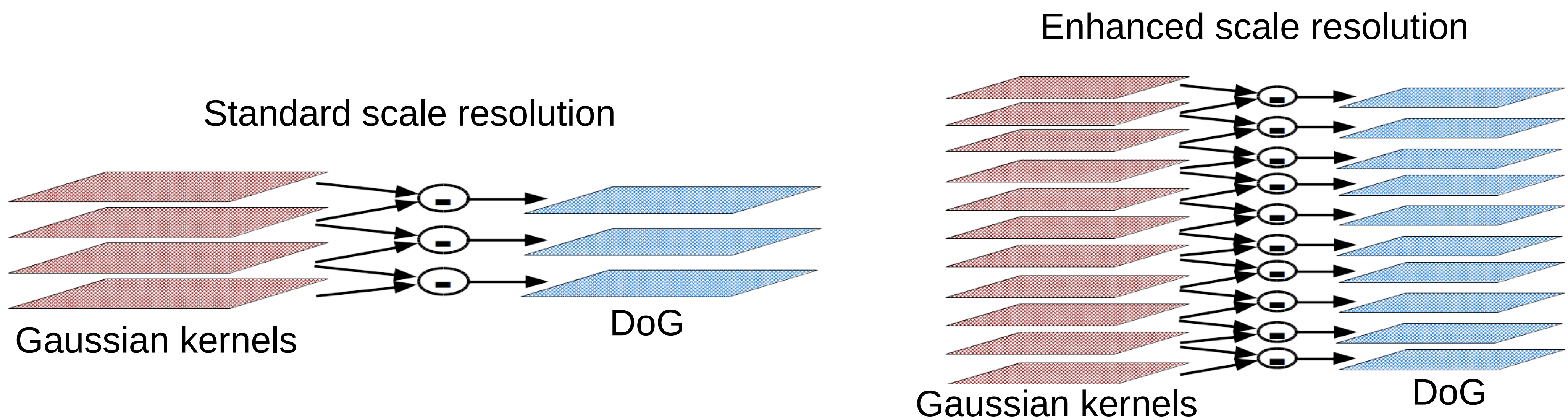
Reminder: SIFT feature scale



- Blur each input image with a set of Gaussian kernels with given covariance.
- Calculate Difference of Gaussians.
- Search for local maxima (features) among layers and pixels.
- Feature scale is equal to Gaussian kernel covariance corresponding the layer where the local maxima is found.

SIFT scale: increasing resolution

- Estimation accuracy is improved only if feature scale measurements are sufficiently accurate.
- Key observation:
  - Can get higher-accuracy scale measurements by increasing number of layers per octave
  - Noise of enhanced-resolution scale measurements can be statistically described with lower Σ_{fs}

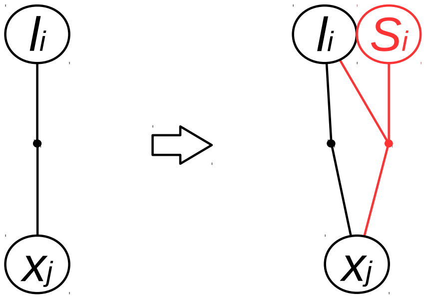


Factor graph modifications

BA + scale constraints cost-function:

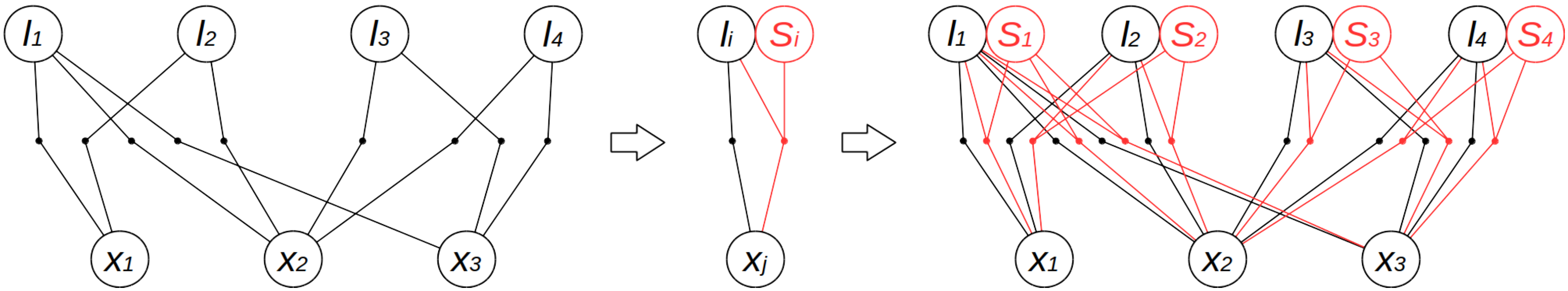
$$J(X, L, S) = \sum_{i}^{N} \sum_{j \in \mathcal{M}_i} \left\| z_i^j - \pi\left(x_i, l_j\right) \right\|_{\Sigma_v}^2 + \left\| s_i^j - f \frac{S_j}{d_i^j} \right\|_{\Sigma_{fs}}^2$$

re-projection error — scale error



Influence on optimization time

Naive implementation taking all scale constraints:



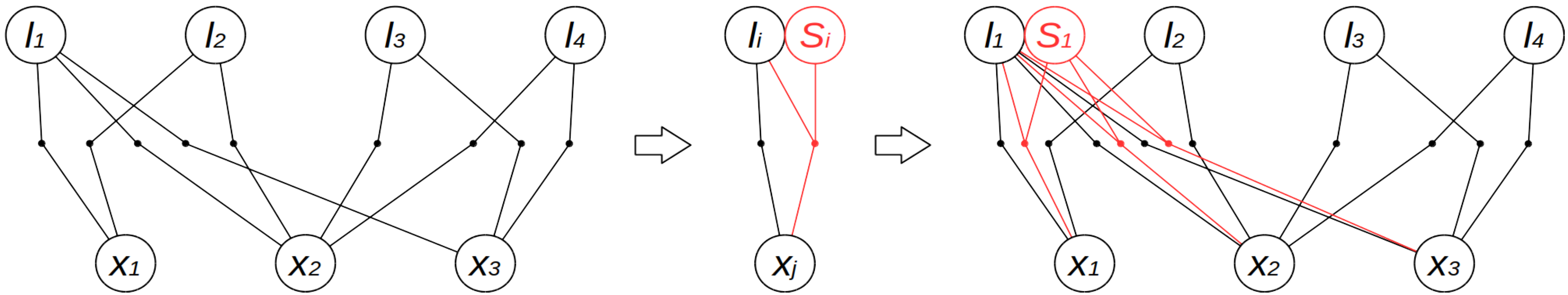
- Number of edges (variables) increased significantly
- Number of links (constraints) increased significantly

→ Optimization time increased.

ANPL | Autonomous Navigation and Perception Lab

Influence on optimization time

Heuristic: add scale constraints only for long-track features.



- Number of edges (variables) increased slightly (about 10%)

  → we keep number of optimization variables as small as possible

- Number of links (constraints) increased.

  → Optimization time increased, but much less!

Results

We tested our approach with the following datasets:

- Aerial dataset (Kagaru): a single downward-facing, flat ground surface scenario.
  - SIFT based scale approach
- Ground vehicle dataset (KITTI): single forward-looking camera, urban scenario.
  - Full set of feature scales
  - Using only scales corresponding to “long” features

KITTI dataset

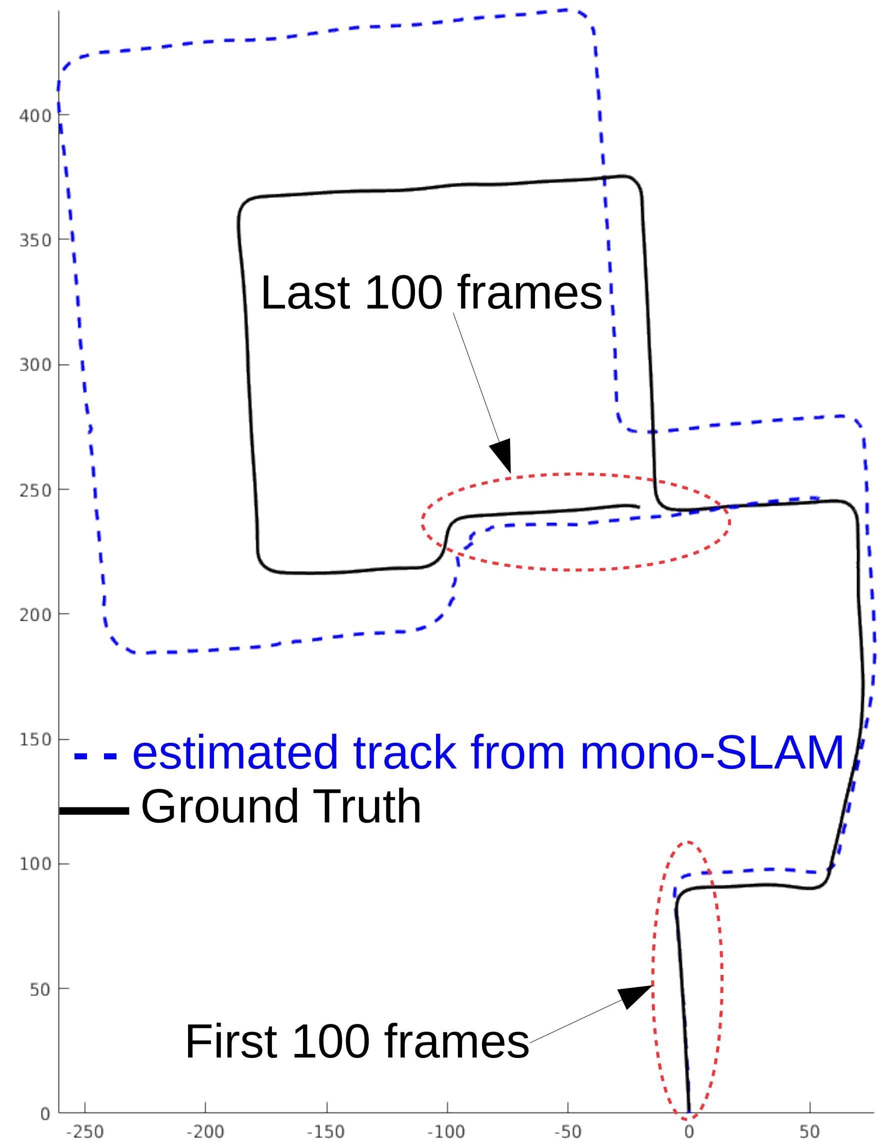
Accurate ground truth is provided by a Velodyne laser scanner and a GPS localization system.
Dataset videos are captured by driving around a city.

+ Ground truth synchronized with camera frame rate
+ Images at 10 fps
+ Camera calibration

Typical images from dataset

Monocular SLAM scale drift problem

Known problem is **scale drift along optical axis** with time.



<u>Example:</u>

Error rate along optical axis is growing with time

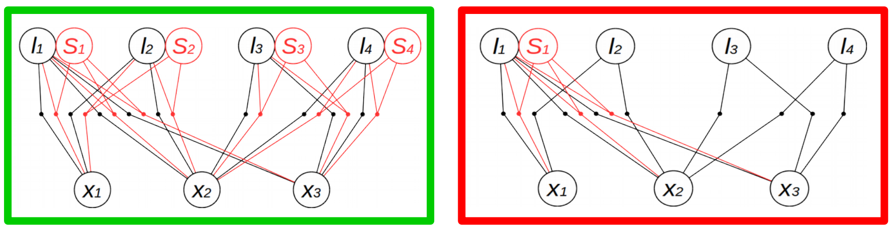
For the <u>first</u> 100 frames:
<u>Error rate: 8.6%</u>

For the <u>last</u> 100 frames:
<u>Error rate: 94.6%</u>

ANPL | Autonomous Navigation and Perception Lab

KITTI dataset

Difference between red and green track: amount of scale constraints.



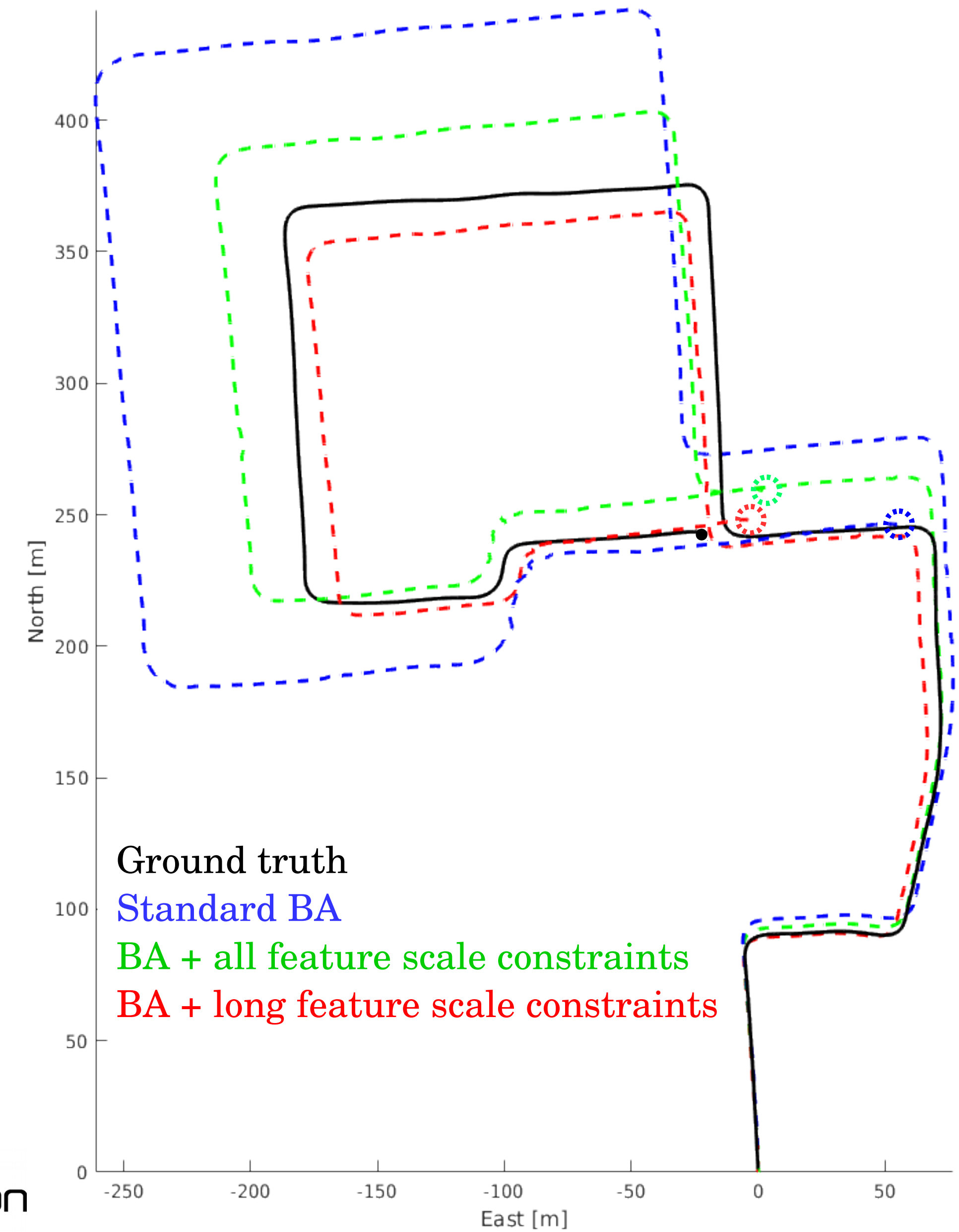
Error rate along optical axis

For the <u>first</u> 100 frames:

Error rate: 8.6% 5% 1%

For the <u>last</u> 100 frames:

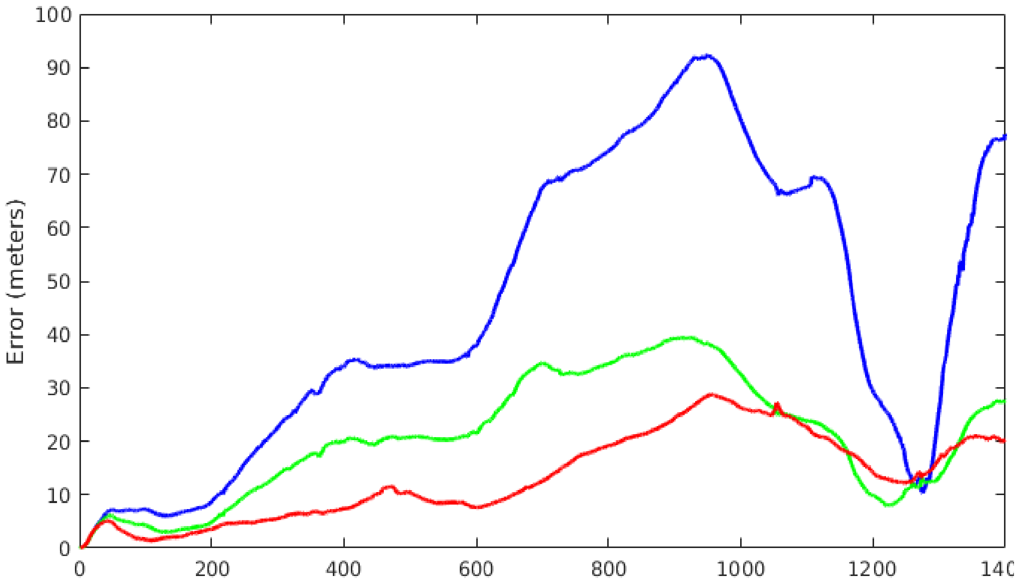
Error rate: 94.6% 27.3% 8.8%



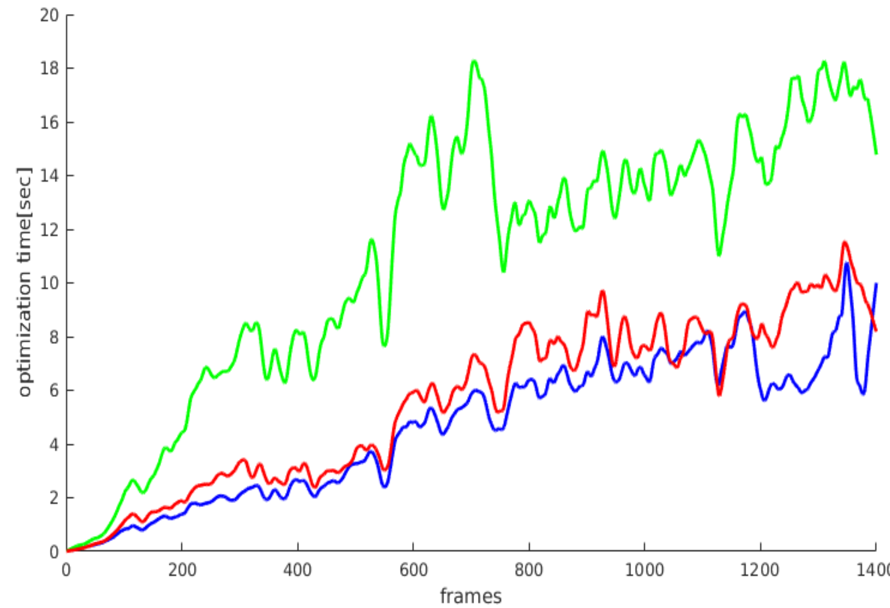
ANPL Autonomous Navigation and Perception Lab

KITTI dataset

Position error



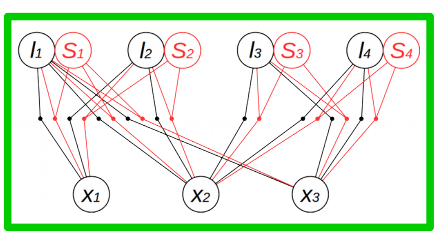
Optimization time

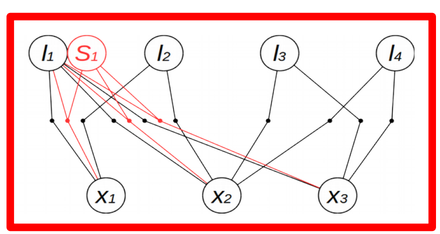


standard BA

standard BA + all feature scale constraints

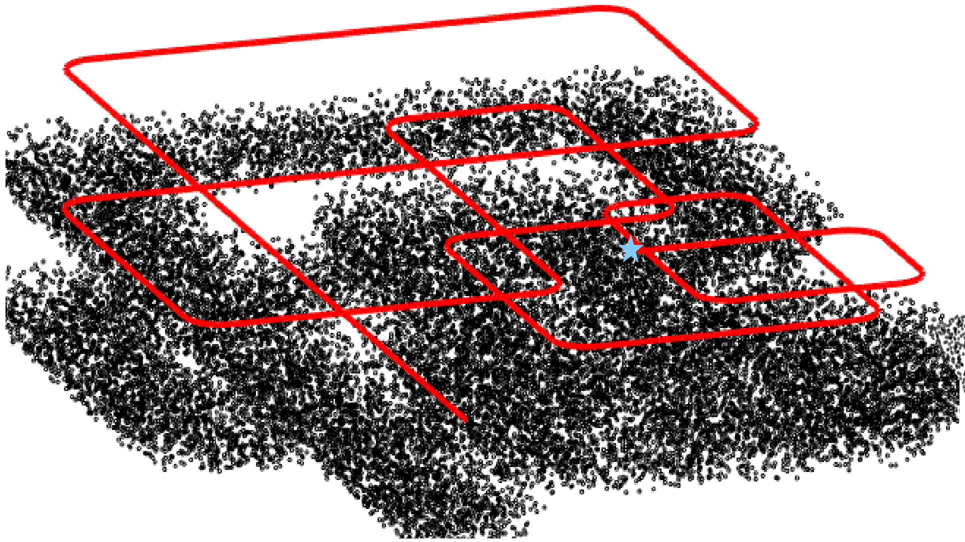
standard BA + long-term feature scale constraints



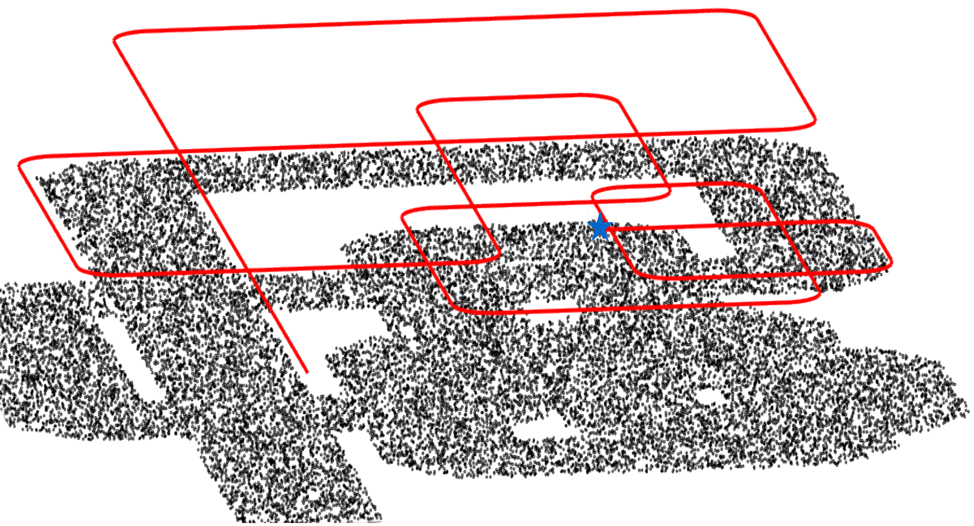


Aerial simulation

Simulation scenario with landmarks spread over height



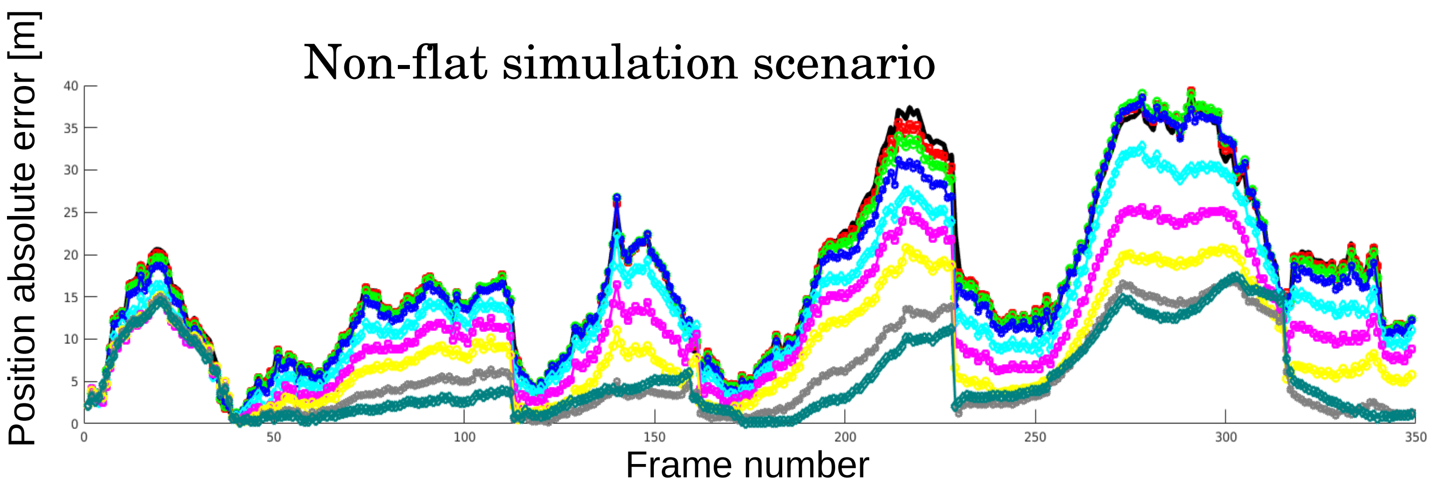
Flat simulation scenario

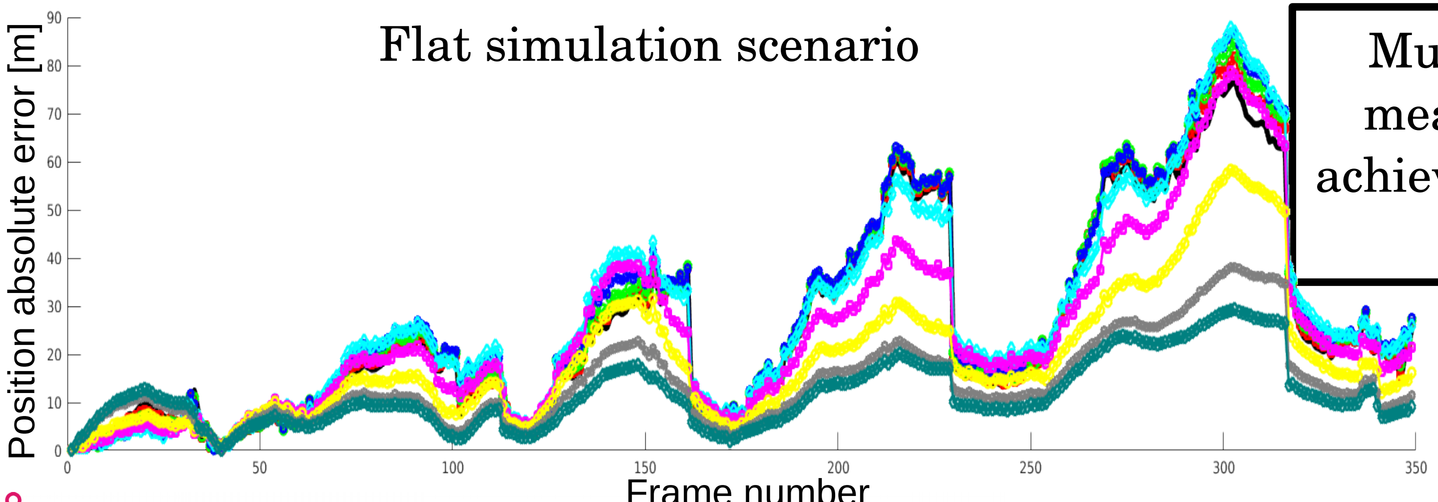


Aerial simulation

Error as a function of Σ_{fs} values.

Each coloured curve corresponds to simulation with a fixed Σ_{fs}





Much more accurate scale measurements required to achieve accuracy improvement for flat scenario.

Kagaru dataset

- Ground truth from an XSens Mti-g INS/GPS
- Downward facing camera
- The dataset traverses over farmland and includes views of grass, an air-strip, roads, trees, ponds, parked aircraft and buildings

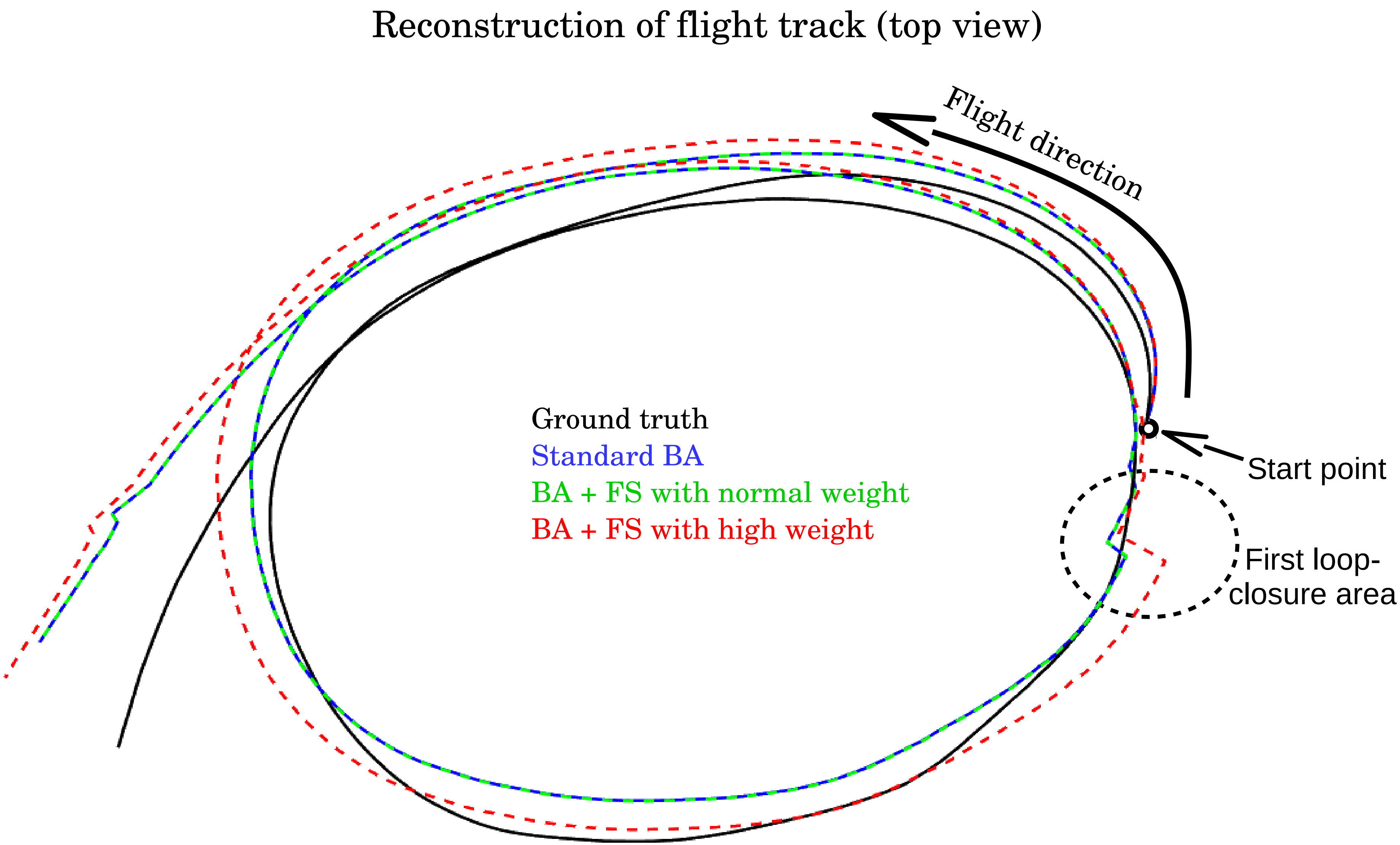
+ Ground truth synchronized with camera frame rate
+ Images at 1 fps
+ Camera calibration

Plane used for dataset creation

Typical images from the dataset

ANPL Autonomous Navigation and Perception Lab

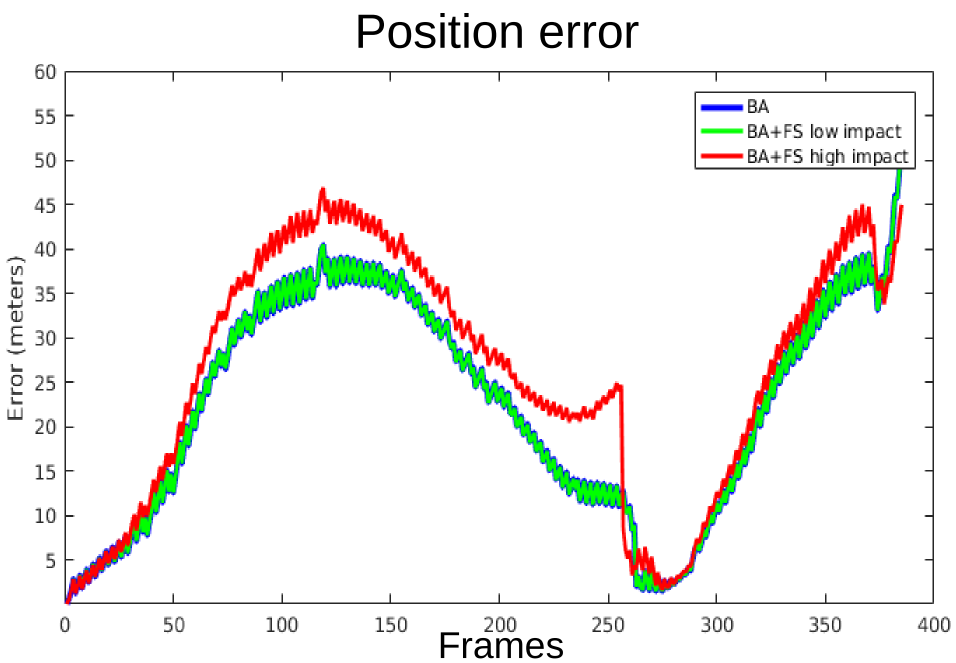
Kagaru dataset



ANPL | Autonomous Navigation and Perception Lab

Kagaru dataset

→ No accuracy improvement along optical axis
→ No position accuracy improvement



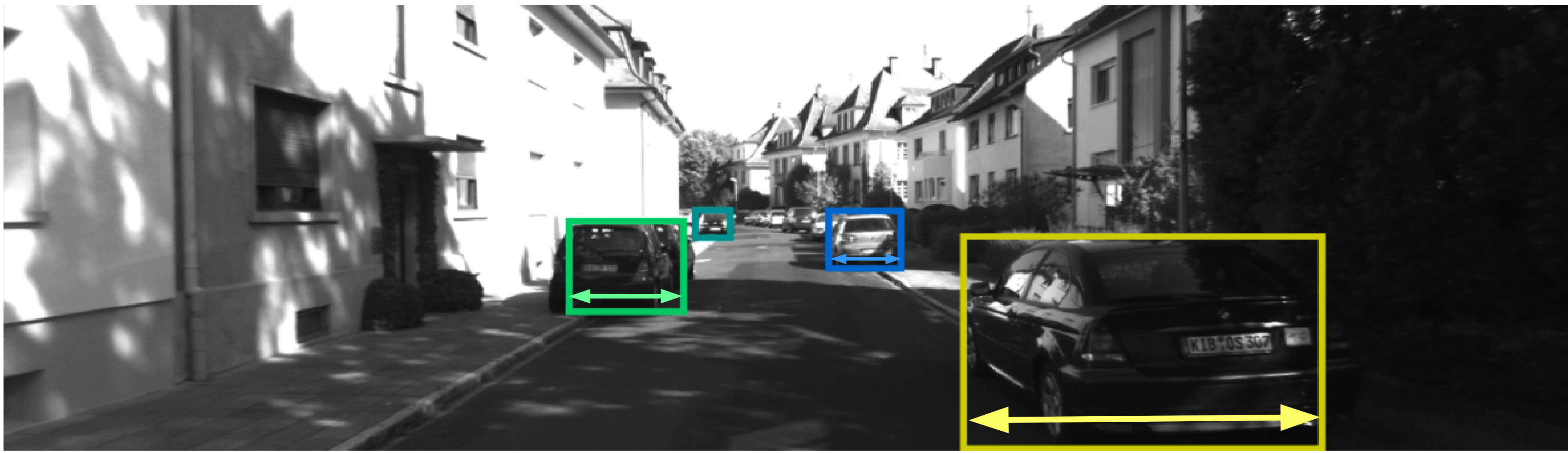
Cost-function: $$J(X,L,S) = \sum_{i}^{N} \sum_{j \in \mathcal{M}_i} \left\| z_i^j - \pi\left(x_i, l_j\right) \right\|_{\Sigma_v}^2 + \left\| s_i^j - f \frac{S_j}{d_i^j} \right\|_{\Sigma_{fs}}^2$$

Outline

- Background:
  - Bundle Adjustment (BA)
  - SIFT features
- Our approach:
  - Adding new constraints to BA
  - Improving accuracy of detected feature scale
- Results
- **Variations with our approach**
- Conclusions

Object scale

- Detect objects
- Track objects across the frames
- Use bounding boxes width as scale to create scale constraints.

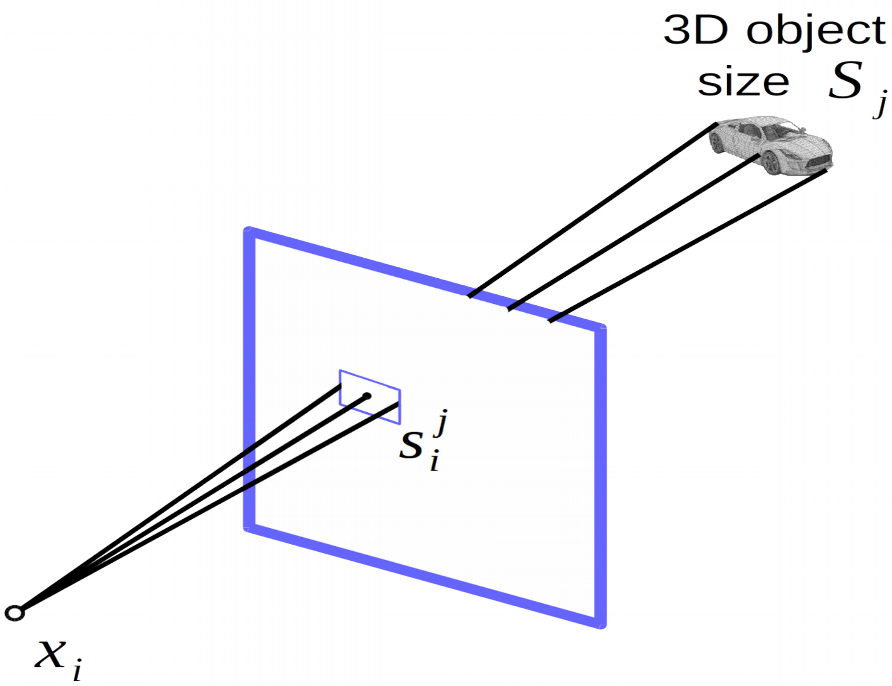


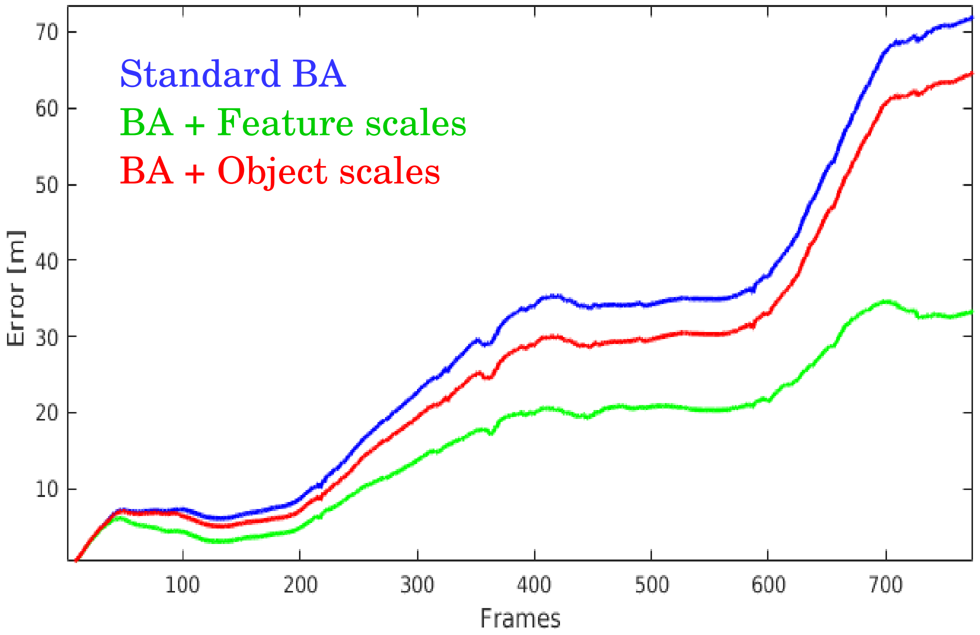
- ➢ Feature scale → object bounding box size (width or height)
- ➢ Virtual landmark size → object size
- ➢ Landmark position → object centre position

Object scale

Main idea:

Objects bounding boxes width/height instead feature scale.





ANPL Autonomous Navigation and Perception Lab

Conclusions

Improved accuracy of bundle adjustment and monocular SLAM along optical axis direction:

- Developed and introduced scale constraints within BA
- Feature scale information is already available from feature detector
- Enhanced feature scale measurement by increasing scale resolution in SIFT
- Application of feature scale constraints to object-level BA

Thank you for your attention!

ANPL Autonomous Navigation and Perception Lab