Robotic 3D Vision

Lecture 12: Visual SLAM 3 – Pose Graph Optimization, Place Recognition

Prof. Dr. Jörg Stückler
Computer Vision Group, TU Munich
http://vision.in.tum.de

What We Will Cover Today

- Tracking-and-Mapping
- Hybrid SLAM methods
- Pose graph optimization
- Loop closure detection and place recognition

Recap: What is Visual SLAM ?

- SLAM stands for Simultaneous Localization and Mapping
  - Estimate the pose of the camera in a map, and simultaneously
  - Reconstruct the environment map
- Visual SLAM (VSLAM): SLAM with vision sensors
- Loop-closure: Revisiting a place allows for drift compensation

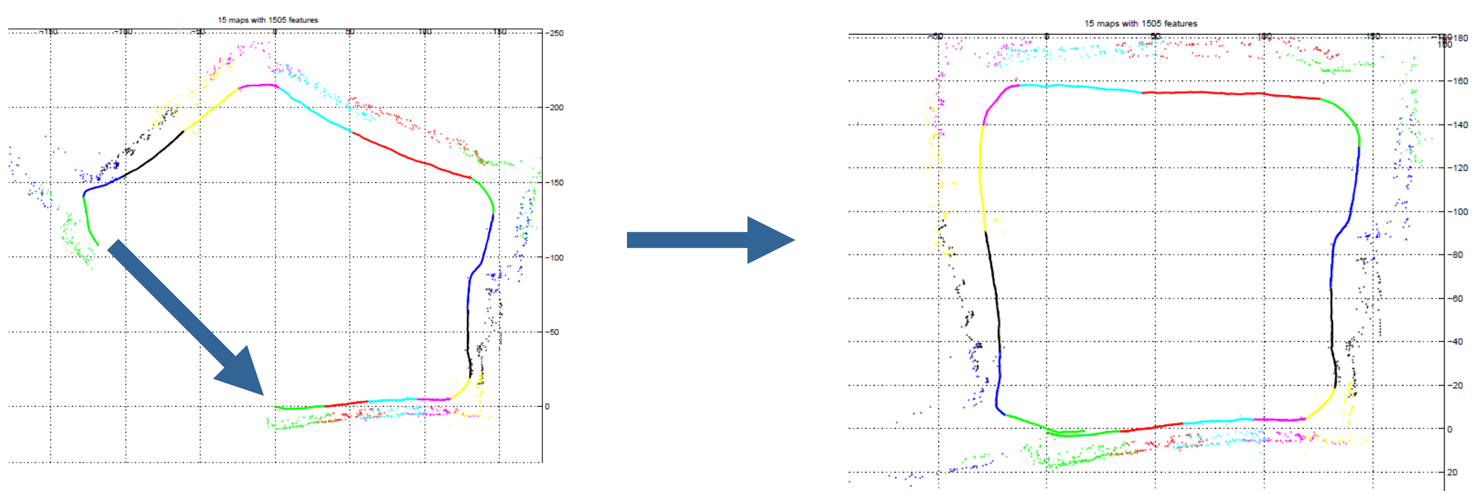
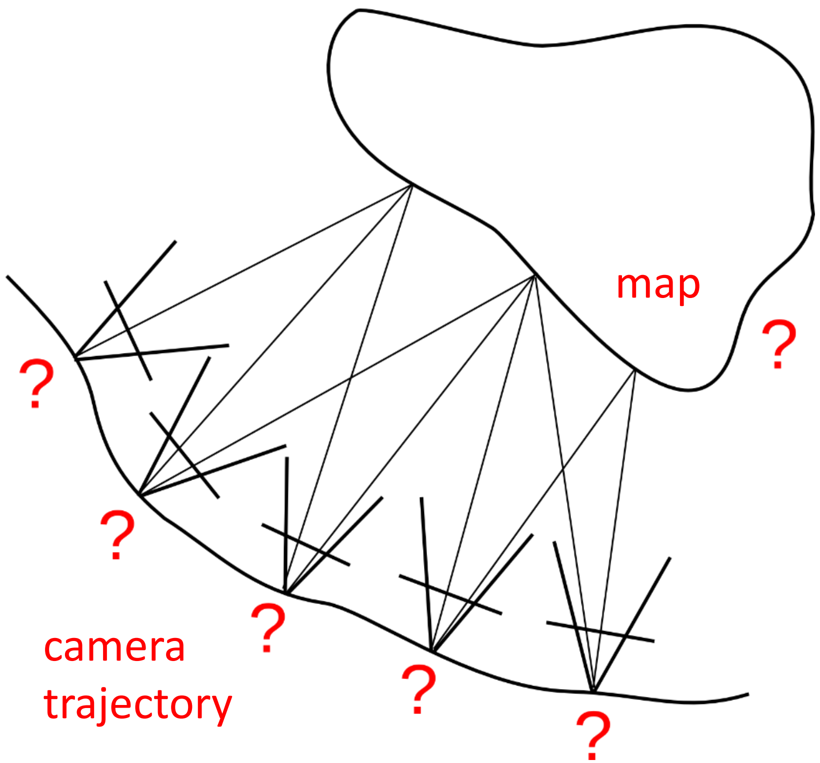


Image from Clemente et al., RSS 2007

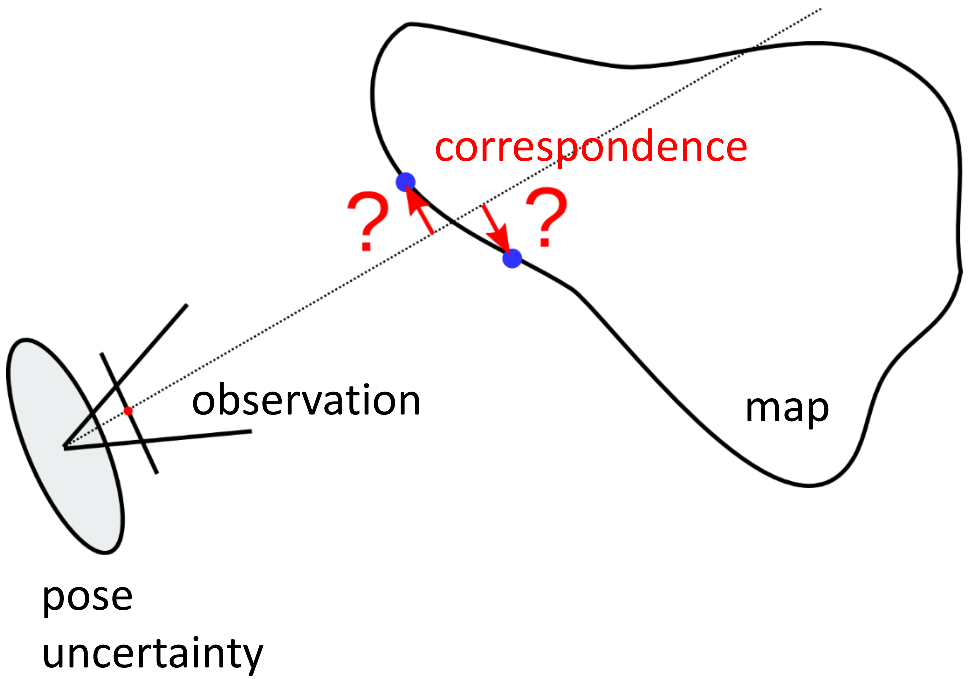
Recap: Why is SLAM difficult?

- Chicken-or-egg problem
  - Camera trajectory and map are unknown and need to be estimated from observations
  - Accurate localization requires an accurate map
  - Accurate mapping requires accurate localization

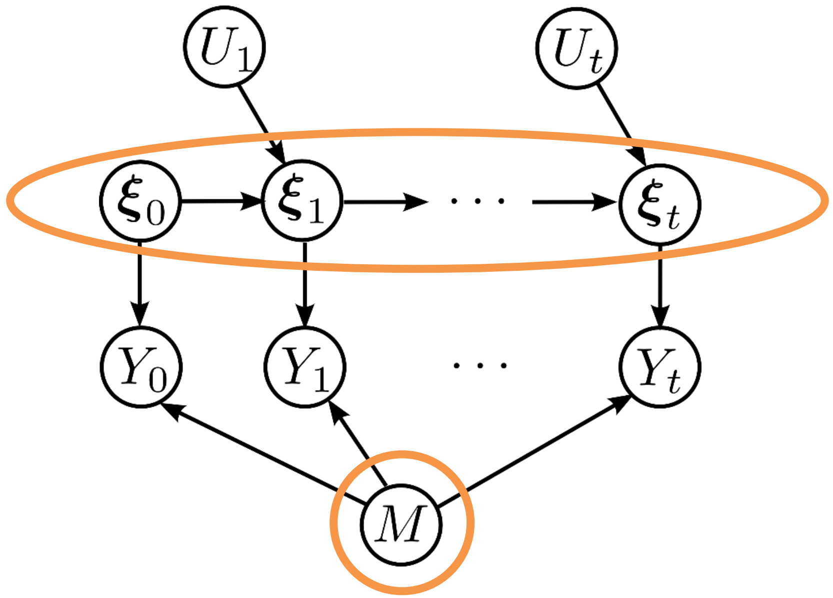


Recap: Why is SLAM difficult?

- Correspondences between observations and the map are unknown
- Wrong correspondences can lead to divergence of trajectory/map estimates
- Important to model uncertainties of observations and estimates in a probabilistic formulation of the SLAM problem



Recap: Probabilistic Formulation of Visual SLAM



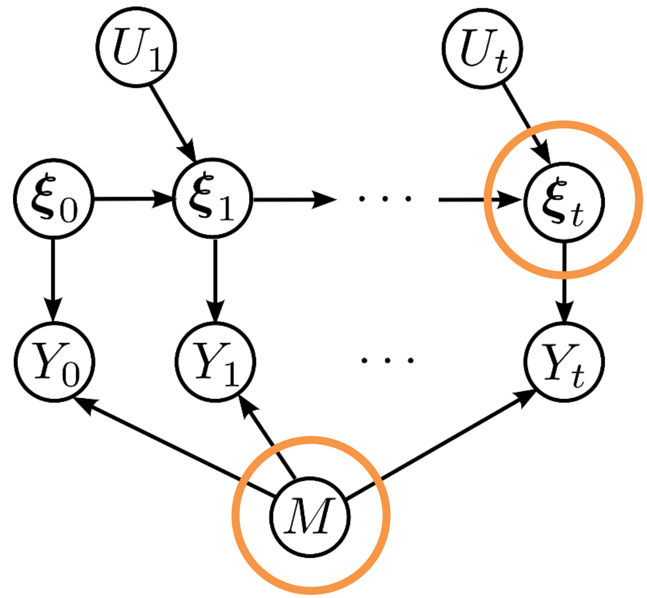
- SLAM posterior probability: $p\left(\boldsymbol{\xi}_{0:t}, M \mid Y_{0:t}, U_{1:t}\right)$
- Observation likelihood: $p\left(Y_t \mid \boldsymbol{\xi}_t, M\right)$
- State-transition probability: $p\left(\boldsymbol{\xi}_t \mid \boldsymbol{\xi}_{t-1}, U_t\right)$

Recap: Online SLAM Methods

- Marginalize out previous poses

$$p\left(\boldsymbol{\xi}_t, M \mid Y_{0:t}, U_{1:t}\right) =$$

$$\int \ldots \int p\left(\boldsymbol{\xi}_{0:t}, M \mid Y_{0:t}, U_{1:t}\right) d\boldsymbol{\xi}_{t-1} \ldots d\boldsymbol{\xi}_0$$



- Poses can be marginalized individually in a recursive way

- Variants:
  - Tracking-and-Mapping: Alternating pose and map estimation
  - Probabilistic filters, f.e. EKF-SLAM

Tracking-and-Mapping

- Alternating optimization of pose estimation and mapping

$$p\left(\boldsymbol{\xi}_t \mid Y_{0:t}, U_{1:t}, M\right) \qquad p\left(M \mid Y_{0:t}, U_{1:t}, \boldsymbol{\xi}_t\right)$$

Tracking Mapping

- F.e.,
  - Semi-dense direct visual odometry
  - KinectFusion (see lectures on dense reconstruction)
  - Parallel Tracking and Mapping (PTAM)

Parallel Tracking and Mapping (PTAM)

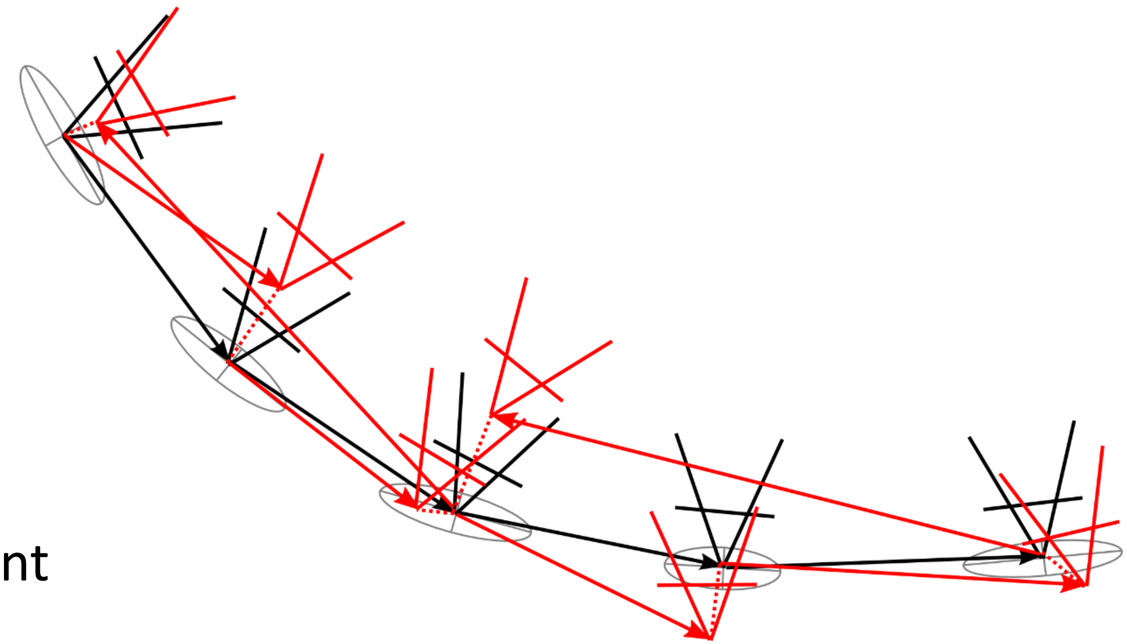
G. Klein and D. Murray, Parallel Tracking and Mapping for Small AR Workspaces, ISMAR 2007

Recap: What is Visual SLAM?

- Visual simultaneous localization and mapping (VSLAM)…
  - Tracks the pose of the camera in a map, and simultaneously
  - Estimates the parameters of the environment map (f.e. reconstruct the 3D positions of interest points in a common coordinate frame)
- Loop-closure: Revisiting a place allows for drift compensation
  - How to detect a loop closure?
- Global vs. local optimization methods
  - Global: full SLAM opt., pose-graph opt., etc.
  - Local: incremental tracking-and-mapping approaches, visual odometry with local maps. Often designed for real-time.
  - Hybrids: Real-time local SLAM + global optimization in a slower parallel process (f.e. LSD-SLAM)

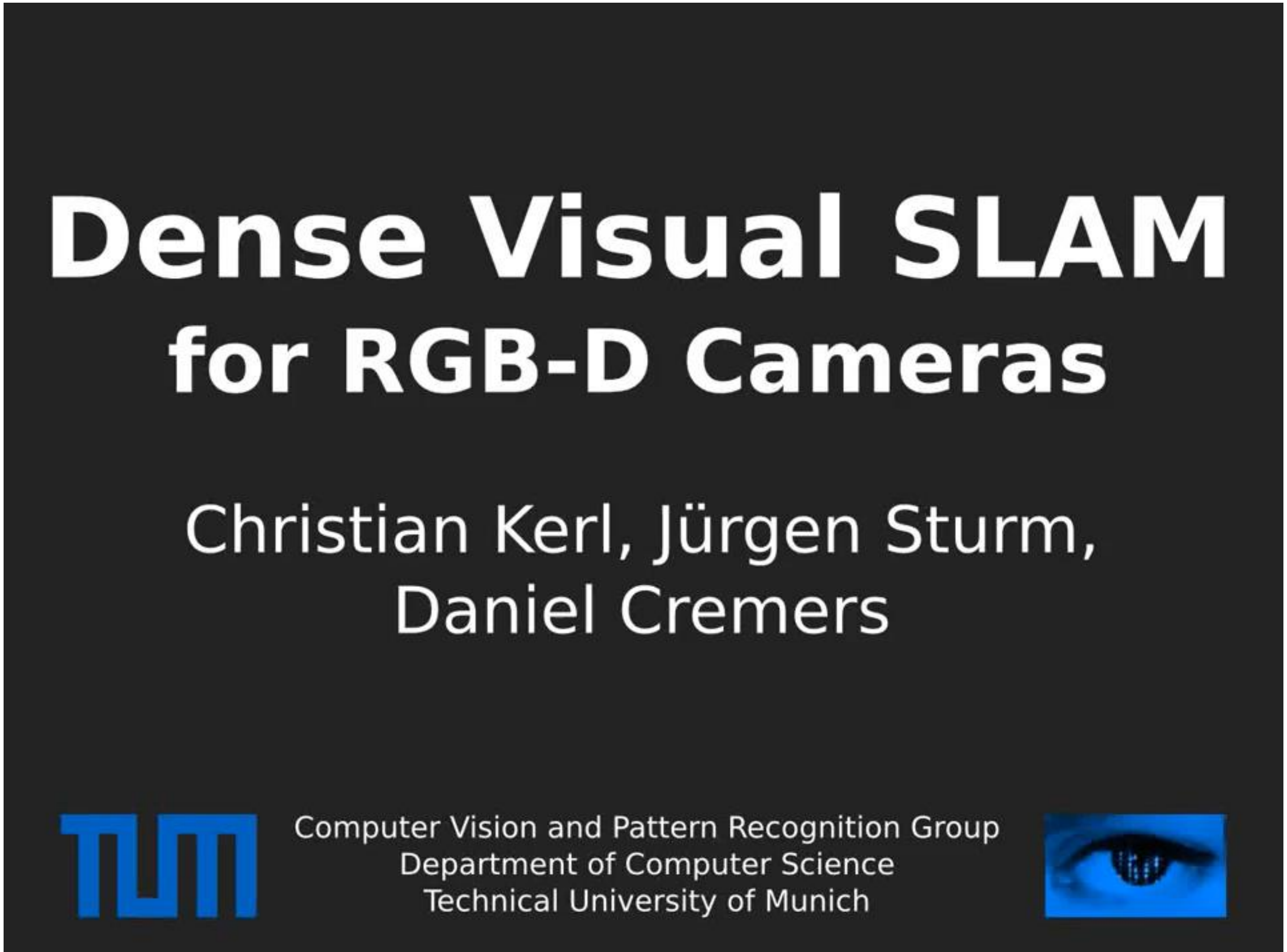
Pose Graph Optimization

- Optimization of poses from relative pose constraints, map recovered from the optimized poses



- Deduce relative constraints between poses from image observations, f.e.
  - 8-point algorithm
  - Direct image alignment

Pose Graph Optimization Example



Kerl et al., Dense Visual SLAM for RGB-D Cameras, IROS 2013

Probabilistic Formulation of Pose Graph Optim.

- Variants of pose graph optimization
  - Full SLAM reduced to trajectory optimization
    - Corresponds to marginalization of the map
    - Alternating optimization of reduced pose-graph problem and map
  - Approximation to SLAM posterior distribution

$$p\left(\boldsymbol{\xi}_{0:t}, M \mid Y_{0:t}, U_{1:t}\right) = p\left(\boldsymbol{\xi}_{0:t} \mid Y_{0:t}, U_{1:t}\right) p\left(M \mid Y_{0:t}, U_{1:t}, \boldsymbol{\xi}_{0:t}\right)$$

$$\Downarrow$$

optimize poses directly: $p\left(\boldsymbol{\xi}_{0:t} \mid \left\{\widetilde{\boldsymbol{\xi}}_i^j\right\}, U_{1:t}\right)$

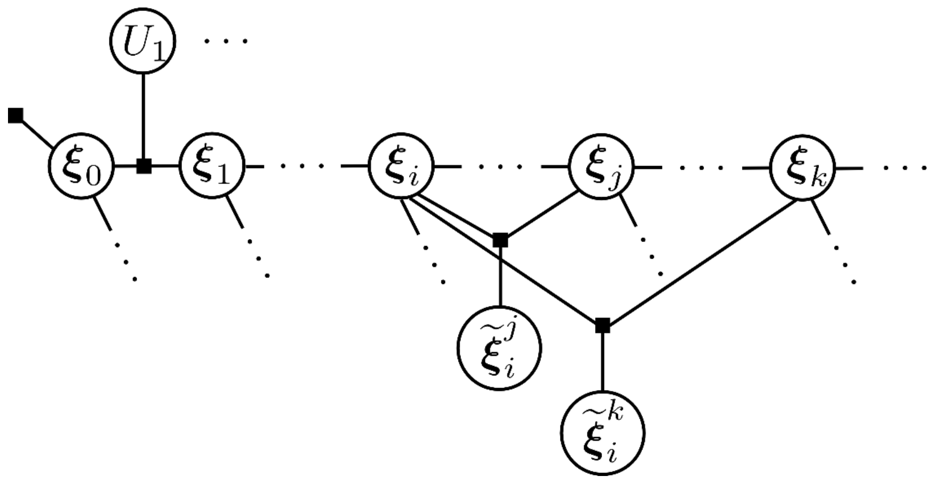
using probabilistic observations of relative poses that are estimated from the image observations Y_i, Y_j : $p\left(\widetilde{\boldsymbol{\xi}}_i^j \mid \boldsymbol{\xi}_i, \boldsymbol{\xi}_j\right)$

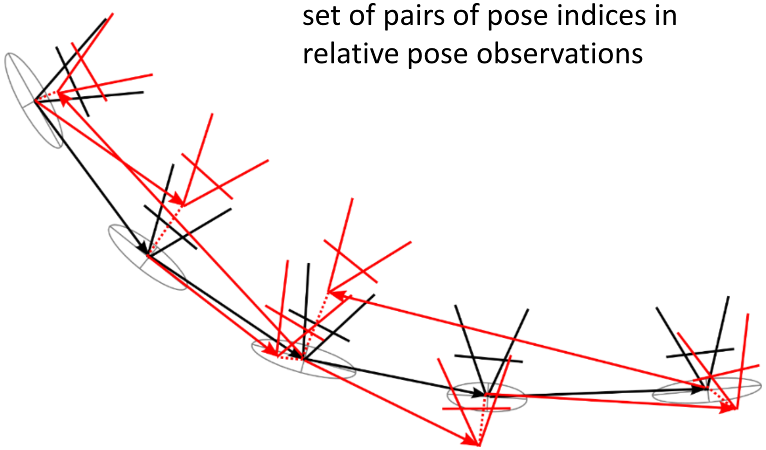
Factor Graph of Pose Graph Optimization

- Factor graph representation of the relative pose graph formulation

$$p\left(\boldsymbol{\xi}_{0:t} \mid \left\{\widetilde{\boldsymbol{\xi}}_i^j, U_{1:t}\right\}\right) = \eta p\left(\boldsymbol{\xi}_0\right) \prod_\tau p(\boldsymbol{\xi}_\tau \mid \boldsymbol{\xi}_{\tau-1}, U_\tau) \prod_{(i,j)\in\mathcal{C}} p\left(\widetilde{\boldsymbol{\xi}}_i^j \mid \boldsymbol{\xi}_i, \boldsymbol{\xi}_j\right)$$

set of pairs of pose indices in relative pose observations





An Explicit Model for Pose Graph Optimization

- Noisy observation of relative motion between camera poses

$$p\left(\widetilde{\boldsymbol{\xi}}_i^j \mid \boldsymbol{\xi}_i, \boldsymbol{\xi}_j\right) = \mathcal{N}\left(\left(\boldsymbol{\xi}_i \ominus \boldsymbol{\xi}_j\right) \ominus \widetilde{\boldsymbol{\xi}}_i^j; \mathbf{0}, \boldsymbol{\Sigma}_{i,j}\right)$$

- No control inputs available / no state-transition model

- Gaussian prior on pose $\xi_0 \sim \mathcal{N}\left(\xi^0, \Sigma_{0,\xi}\right)$

Pose Graph Optimization as Energy Minimization

- Optimize negative log posterior probability (MAP estimation)

$$E(\boldsymbol{\xi}_{0:t}) = \frac{1}{2}\left(\boldsymbol{\xi}_0 \ominus \boldsymbol{\xi}^0\right)^\top \boldsymbol{\Sigma}_{0,\xi}^{-1}\left(\boldsymbol{\xi}_0 \ominus \boldsymbol{\xi}^0\right) + \frac{1}{2}\sum_{(i,j)\in\mathcal{C}}\left(\left(\boldsymbol{\xi}_i \ominus \boldsymbol{\xi}_j\right) \ominus \widetilde{\boldsymbol{\xi}}_i^j\right)^\top \boldsymbol{\Sigma}_{i,j}^{-1}\left(\left(\boldsymbol{\xi}_i \ominus \boldsymbol{\xi}_j\right) \ominus \widetilde{\boldsymbol{\xi}}_i^j\right)$$

- Non-linear least squares…

Pose Graph Optimization as Energy Minimization

- Let's define the residuals on the full state vector $\mathbf{x} := \begin{pmatrix} \boldsymbol{\xi}_0 \\ \vdots \\ \boldsymbol{\xi}_t \end{pmatrix}$

$$\mathbf{r}^0(\mathbf{x}) := \boldsymbol{\xi}_0 \ominus \boldsymbol{\xi}^0$$

$$\mathbf{r}^{i,j}(\mathbf{x}) := \left(\boldsymbol{\xi}_i \ominus \boldsymbol{\xi}_j\right) \ominus \widetilde{\boldsymbol{\xi}}_i^j$$

- Stack the residuals in a vector-valued function and collect the residual covariances on the diagonal blocks of a square matrix

$$\mathbf{r}(\mathbf{x}) := \begin{pmatrix} \mathbf{r}^0(\mathbf{x}) \\ \mathbf{r}^{i,j}(\mathbf{x}) \\ \vdots \\ \mathbf{r}^{i',j'}(\mathbf{x}) \end{pmatrix} \qquad \mathbf{W} := \begin{pmatrix} \boldsymbol{\Sigma}_{0,\xi}^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{i,j}^{-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{\Sigma}_{i',j'}^{-1} \end{pmatrix}$$

- Rewrite energy as $E(\mathbf{x}) = \frac{1}{2}\mathbf{r}(\mathbf{x})^\top \mathbf{W}\mathbf{r}(\mathbf{x})$

Structure of the Pose Graph Optimization Problem

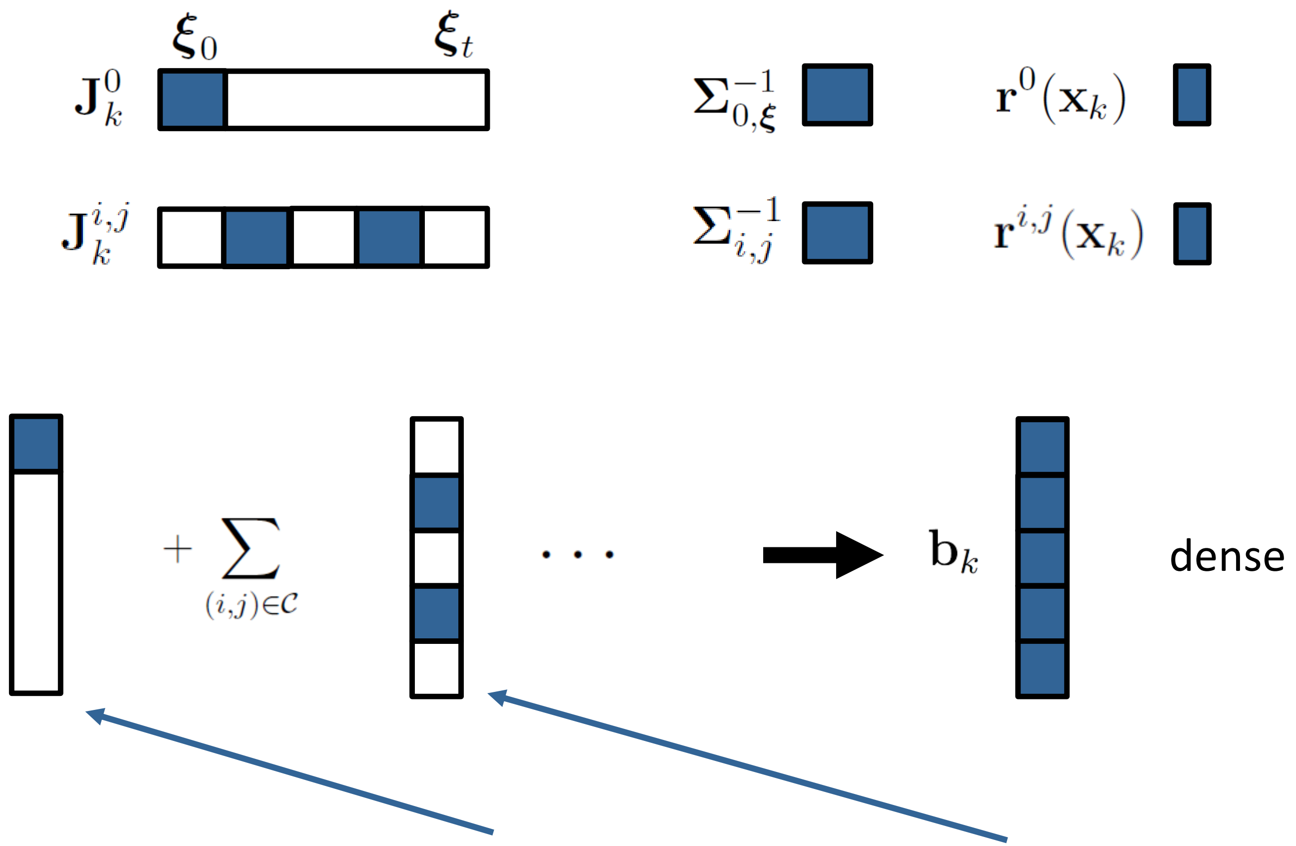
- Leads to $\mathrm{H}_k \Delta \mathbf{x} = -\mathrm{b}_k$ with

$$\mathbf{b}_k = \mathbf{J}_k^\top \mathbf{W} \mathbf{r}(\mathbf{x}) = \left(\mathbf{J}_k^0\right)^\top \boldsymbol{\Sigma}_{0,\boldsymbol{\xi}}^{-1} \mathbf{r}^0(\mathbf{x}_k) + \sum_{(i,j)\in\mathcal{C}} \left(\mathbf{J}_k^{i,j}\right)^\top \boldsymbol{\Sigma}_{i,j}^{-1} \mathbf{r}^{i,j}(\mathbf{x}_k)$$

$$\mathbf{H}_k = \mathbf{J}_k^\top \mathbf{W} \mathbf{J}_k = \left(\mathbf{J}_k^0\right)^\top \boldsymbol{\Sigma}_{0,\boldsymbol{\xi}}^{-1} \mathbf{J}_k^0 + \sum_{(i,j)\in\mathcal{C}} \left(\mathbf{J}_k^{i,j}\right)^\top \boldsymbol{\Sigma}_{i,j}^{-1} \mathbf{J}_k^{i,j}$$

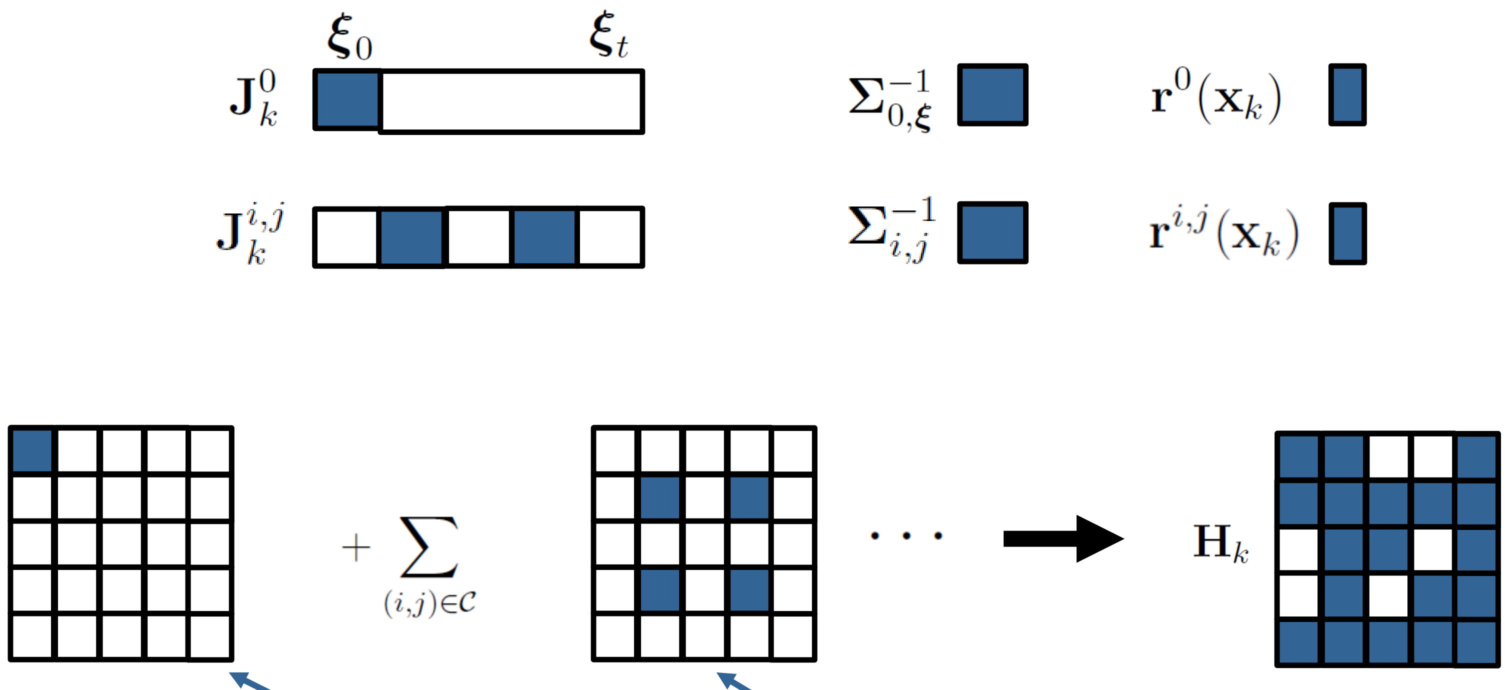
- What is the structure now?

Structure of the Pose Graph Optimization Problem



$$\mathbf{b}_k = \mathbf{J}_k^\top \mathbf{W}\mathbf{r}(\mathbf{x}) = \left(\mathbf{J}_k^0\right)^\top \boldsymbol{\Sigma}_{0,\boldsymbol{\xi}}^{-1}\mathbf{r}^0(\mathbf{x}_k) + \sum_{(i,j)\in\mathcal{C}} \left(\mathbf{J}_k^{i,j}\right)^\top \boldsymbol{\Sigma}_{i,j}^{-1}\mathbf{r}^{i,j}(\mathbf{x}_k)$$

Structure of the Pose Graph Optimization Problem



Sparse, depending on constraint connectivity

$$\mathbf{H}_k = \mathbf{J}_k^\top \mathbf{W} \mathbf{J}_k = \left(\mathbf{J}_k^0\right)^\top \boldsymbol{\Sigma}_{0,\xi}^{-1} \mathbf{J}_k^0 + \sum_{(i,j)\in\mathcal{C}} \left(\mathbf{J}_k^{i,j}\right)^\top \boldsymbol{\Sigma}_{i,j}^{-1} \mathbf{J}_k^{i,j}$$

Scale Consistency in Monocular SLAM

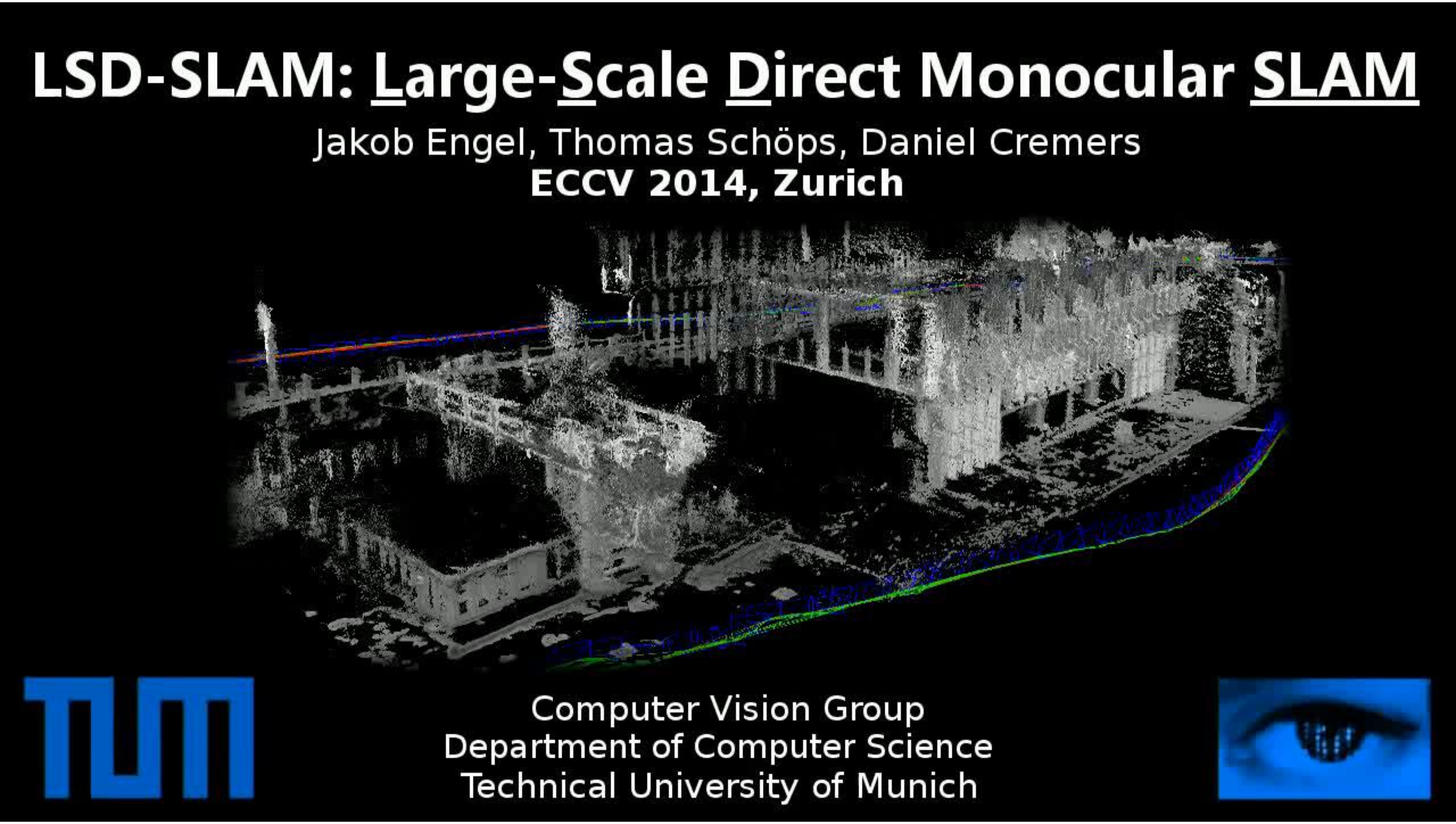
- Monocular SLAM: Scale not observable!
  - Scale as an additional degree of freedom
  - Parametrize poses in group of similarity transformations $\mathbf{Sim}(3)$ instead of Euclidean transformations ($\mathbf{SE}(3)$)
  - Optimize for globally consistent scale

- Group of similarity transformations $\mathbf{Sim}(3)$
  - Group elements now include a scale parameter

$$\mathbf{T} = \begin{pmatrix} s\mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{pmatrix} \in \mathbf{Sim}(3)$$

  - Also has an associated Lie algebra with exponential and logarithm map
  - Lie algebra elements have 7 degree of freedom, 6 for rigid motion, 1 for scale
  - See Strasdat et al., Scale Drift-Aware Large Scale Monocular SLAM, Robotics Science and Systems, 2010

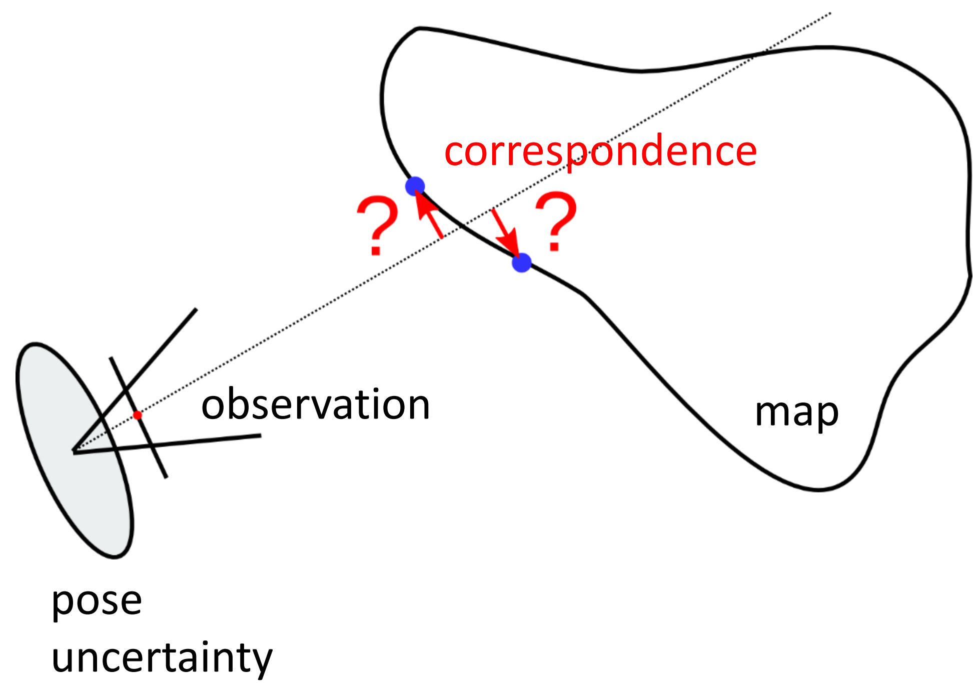
Example: Scale Consistency in Mono SLAM



Engel et al., LSD-SLAM: Large-Scale Direct Monocular SLAM, ECCV 2014

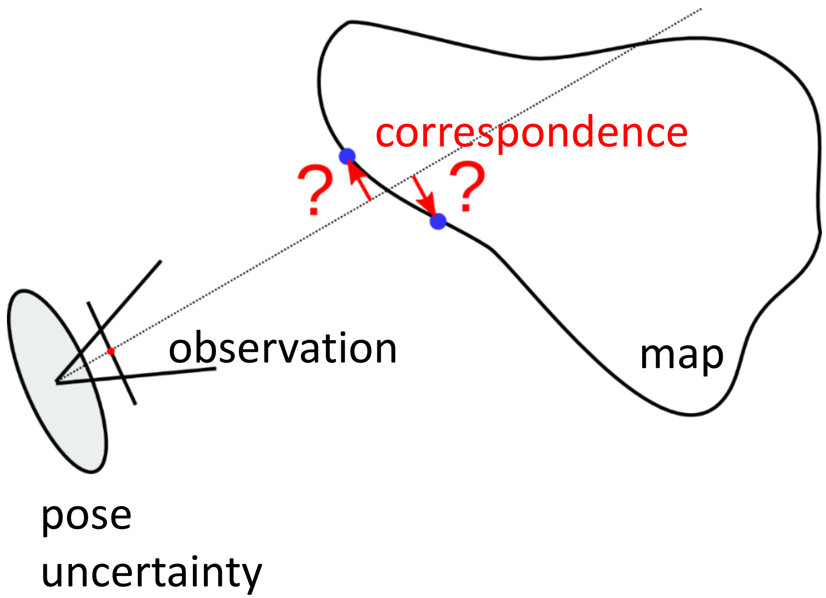
Recap: Why is SLAM difficult?

- Correspondences between observations and the map are unknown
- Wrong correspondences can lead to divergence of trajectory/map estimates
- For pose graph optimization, we need to decide which images can be matched and aligned to each others



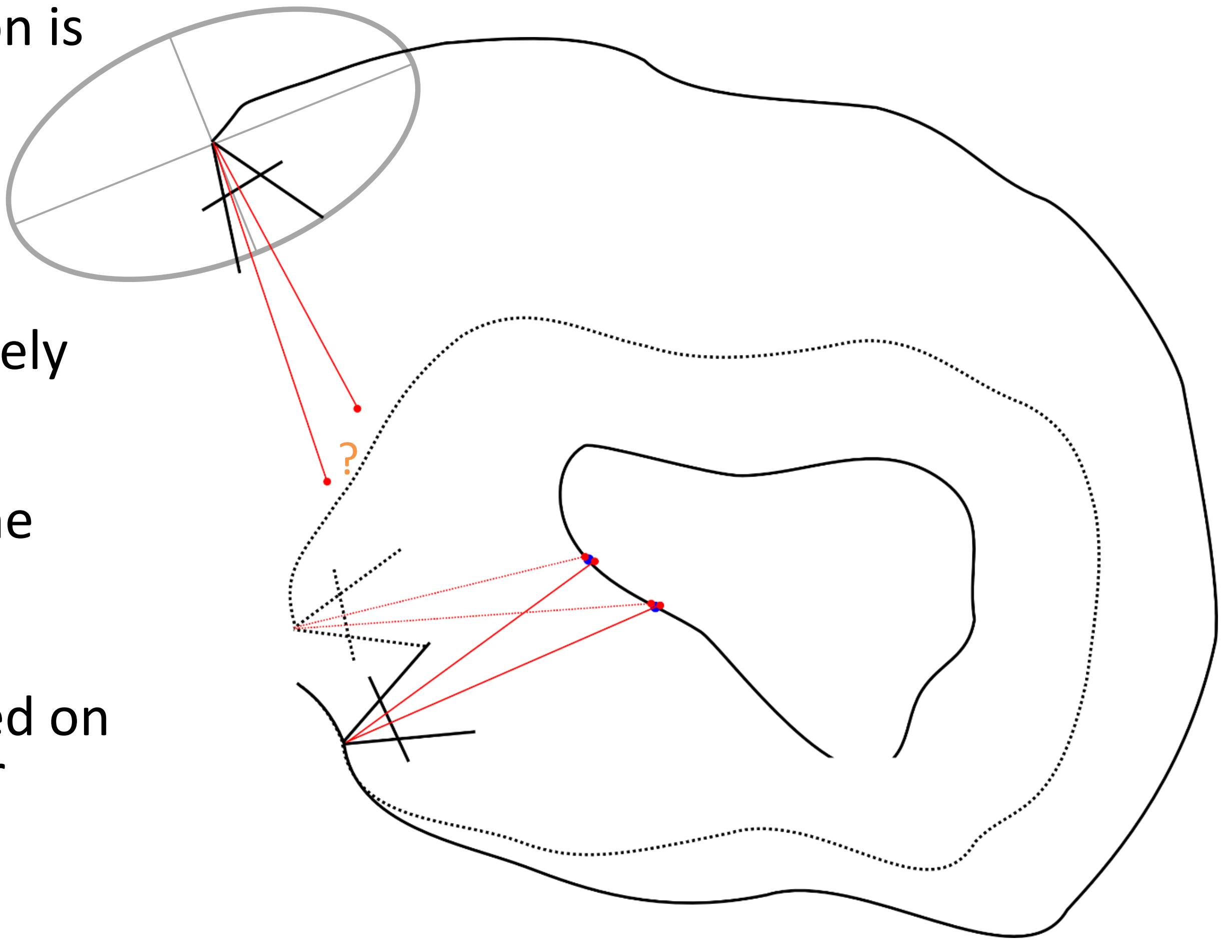
Short-Term Data Association Strategies In SLAM

- Similar to the data association problem in visual odometry with local maps
- Many approaches use interest point descriptors and RANSAC for robust association of point detections in the image with 3D point landmarks
- Also KLT at high-frame rates, or active search principles. Latter requires a pose guess (f.e. EKF-SLAM)



Loop Closure Detection

- Loop closure detection is a special case of data association
- Typically, we cannot rely on the state estimate because of the drift accumulated along the loop
- Data association based on cues such as shape or appearance needed (interest point descriptors, etc.)



Place Recognition

- Goals:
  - find additional image correspondences between non-sequential frames
  - detect when previous places are revisited

- Methods for detecting a revisit of previous places are often coined "place recognition" in the SLAM literature

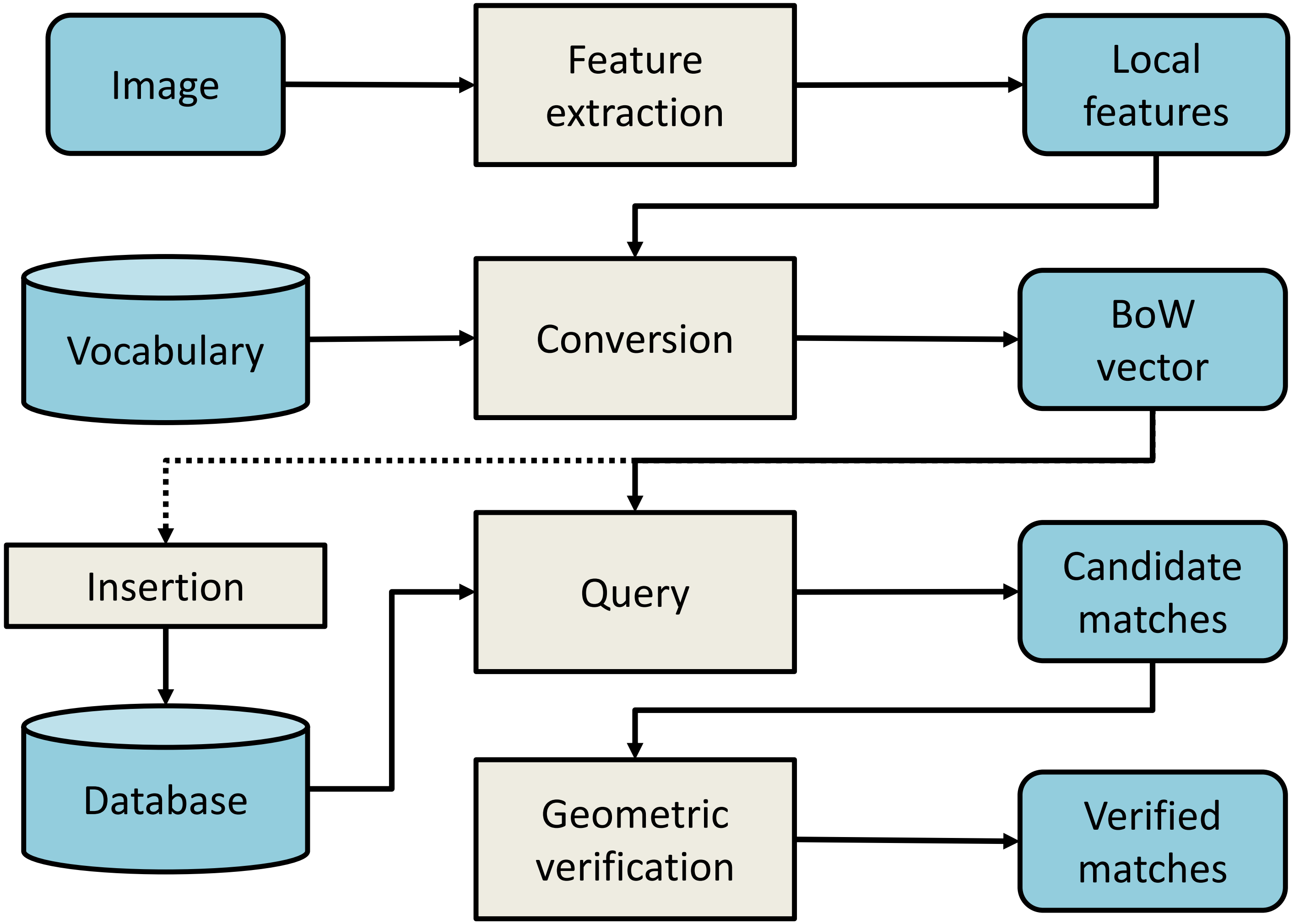
Images: Cummins and Newman, Highly Scalable Appearance-Only SLAM – FAB-MAP 2.0, RSS 2009

Place Recognition

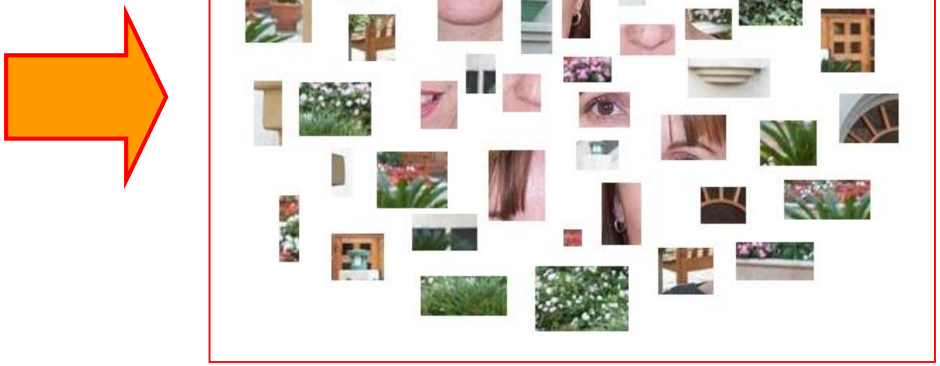

- Idea: use image retrieval techniques
- Popular approach for place recognition is to use bag-of-visual-words based image retrieval in conjunction with geometric verification (f.e. 8-point with RANSAC)

Images: Cummins and Newman, Highly Scalable Appearance-Only SLAM – FAB-MAP 2.0, RSS 2009

Bag-of-Visual-Words based Image Retrieval



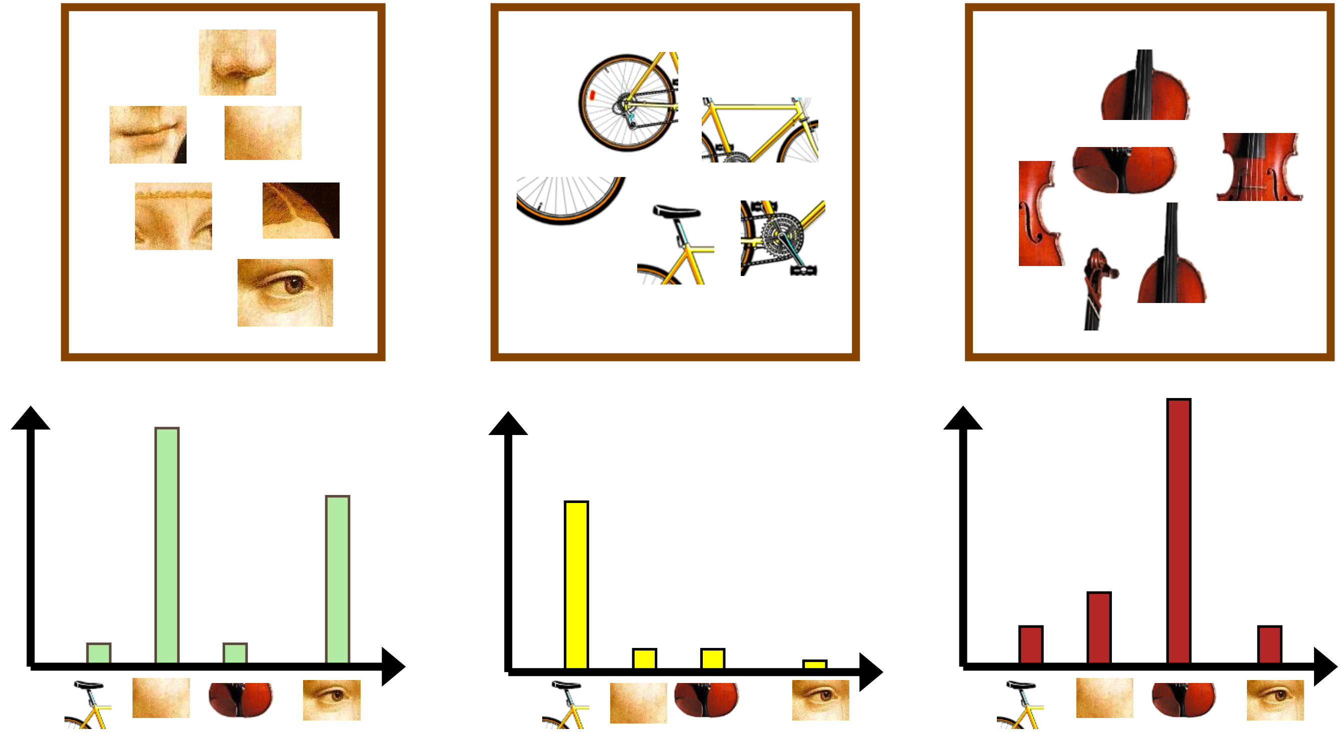
Bag of Visual Words



Slide credit: Svetlana Lazebnik

Bag of Visual Words

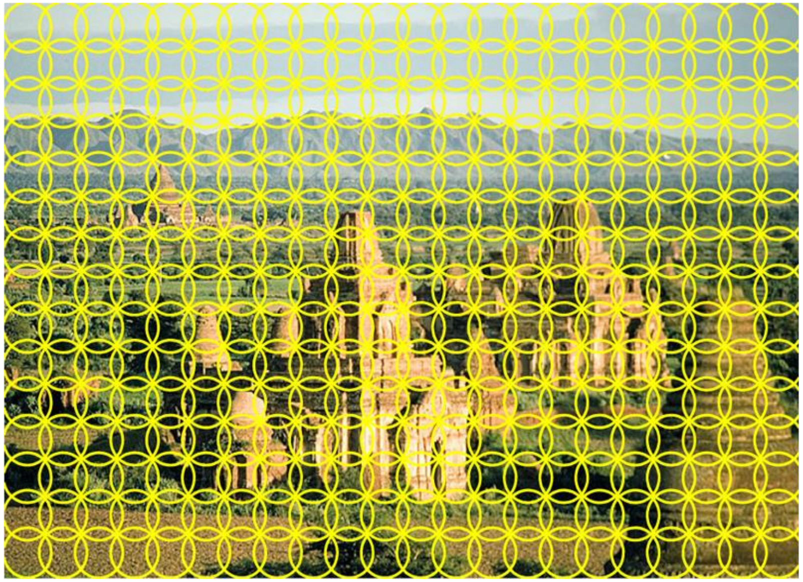
1. Extract local features
2. Learn "visual vocabulary"
3. Quantize local features using visual vocabulary
4. Represent images by frequencies of "visual words"



Slide credit: Svetlana Lazebnik

Bag of Visual Words

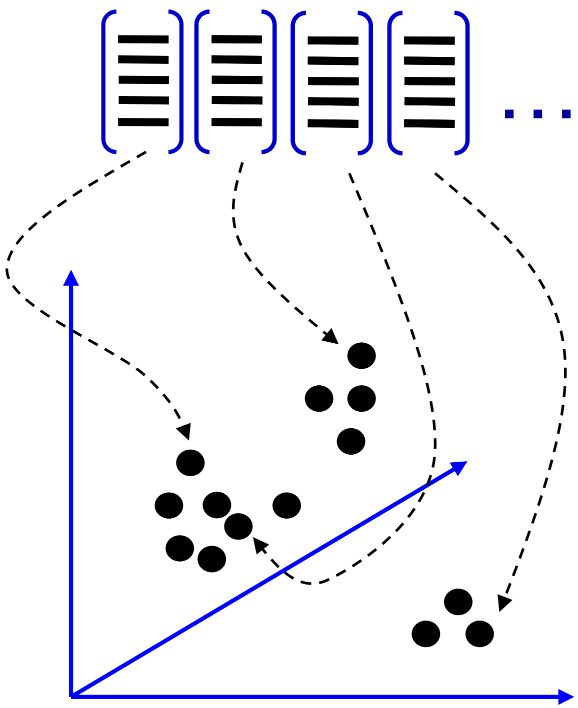
1. Extract local features
2. Learn “visual vocabulary”
3. Quantize local features using visual vocabulary
4. Represent images by frequencies of “visual words”



Slide credit: Svetlana Lazebnik

Bag of Visual Words

1. Extract local features
2. Learn "visual vocabulary"
3. Quantize local features using visual vocabulary
4. Represent images by frequencies of "visual words"

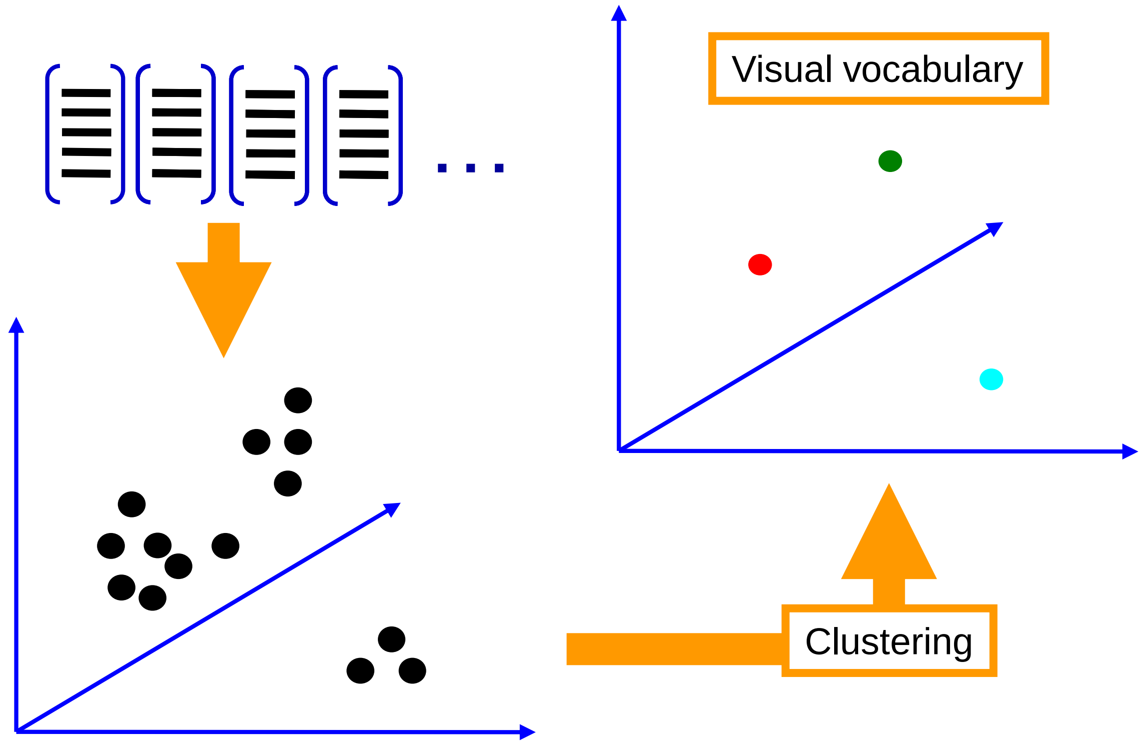


Extracted descriptors from the training set

Slide credit: Svetlana Lazebnik, Josef Sivic

Bag of Visual Words

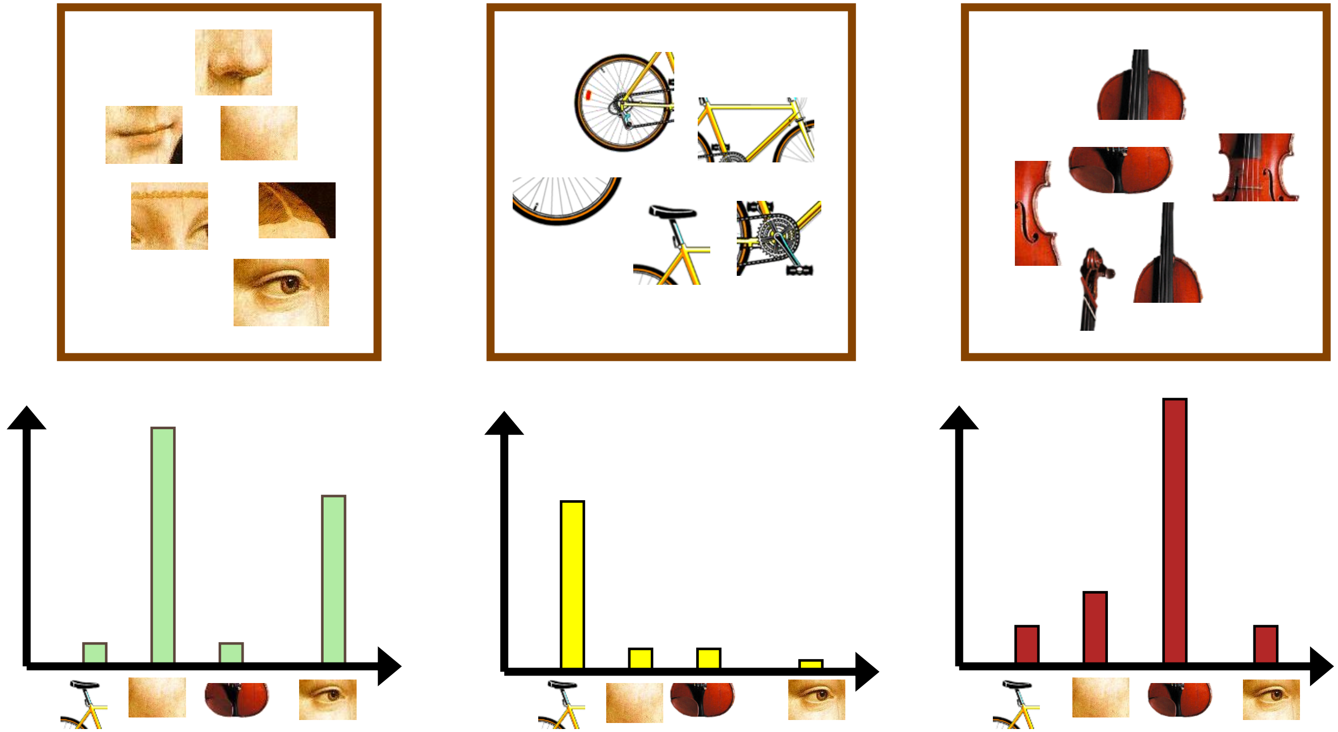
1. Extract local features
2. Learn "visual vocabulary"
3. Quantize local features using visual vocabulary
4. Represent images by frequencies of "visual words"



Slide credit: Svetlana Lazebnik, Josef Sivic

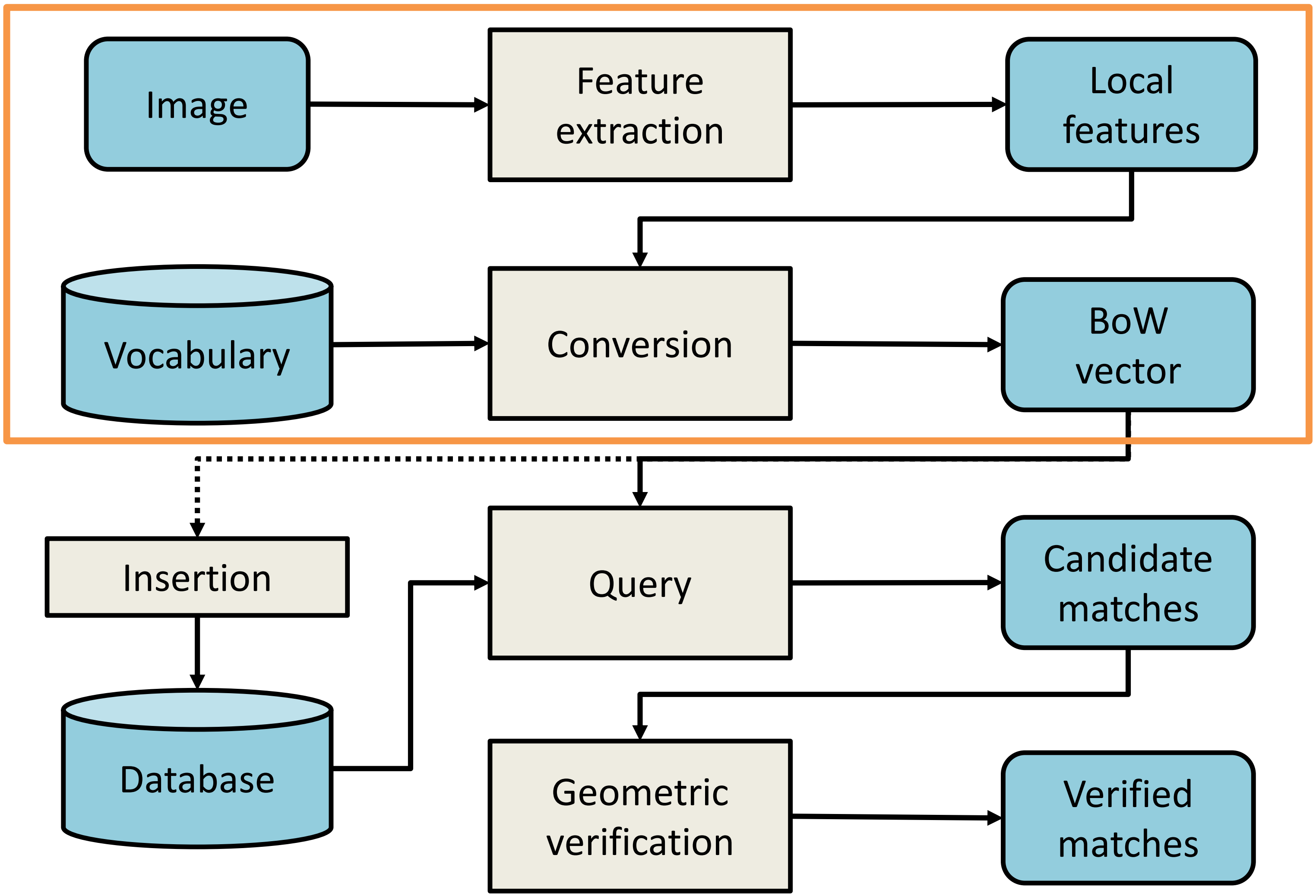
Bag of Visual Words

1. Extract local features
2. Learn "visual vocabulary"
3. Quantize local features using visual vocabulary
4. Represent images by frequencies of "visual words"



Slide credit: Svetlana Lazebnik, Josef Sivic

Bag-of-Visual-Words based Image Retrieval



Loop Closing is Difficult!

Perceptual Aliasing

Image credit: Juan D. Tardós

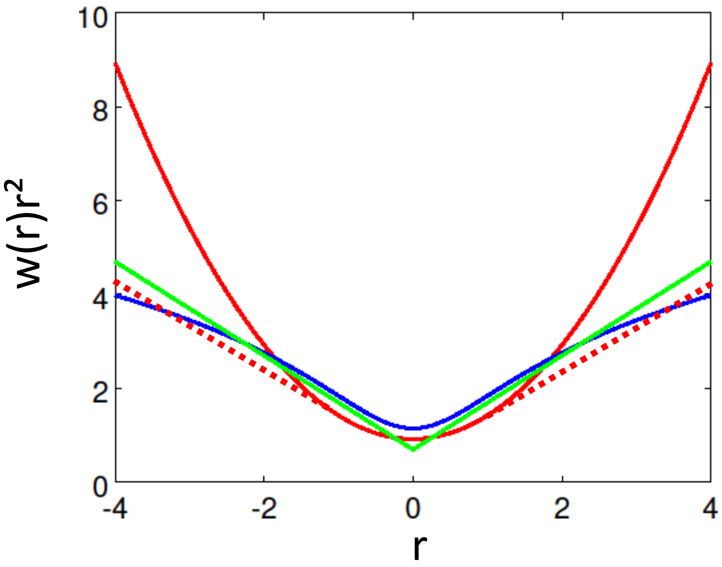
Robust Optimization

- Data association is hard
- Can we make SLAM optimization more robust to data association outliers?
- Gaussian noise assumption makes optimization sensitive to outliers
  - Use heavier-tail distributions / robust norms
  - Incorporate further random variables into probabilistic optimization problem that allow for inferring the inconsistency of measurements, f.e.: Suenderhauf and Protzel, Switchable Constraints for Robust Pose Graph SLAM, IROS 2012

Recap: Huber Loss

- Huber-loss „switches“ between Gaussian (locally at mean) and Laplace distribution

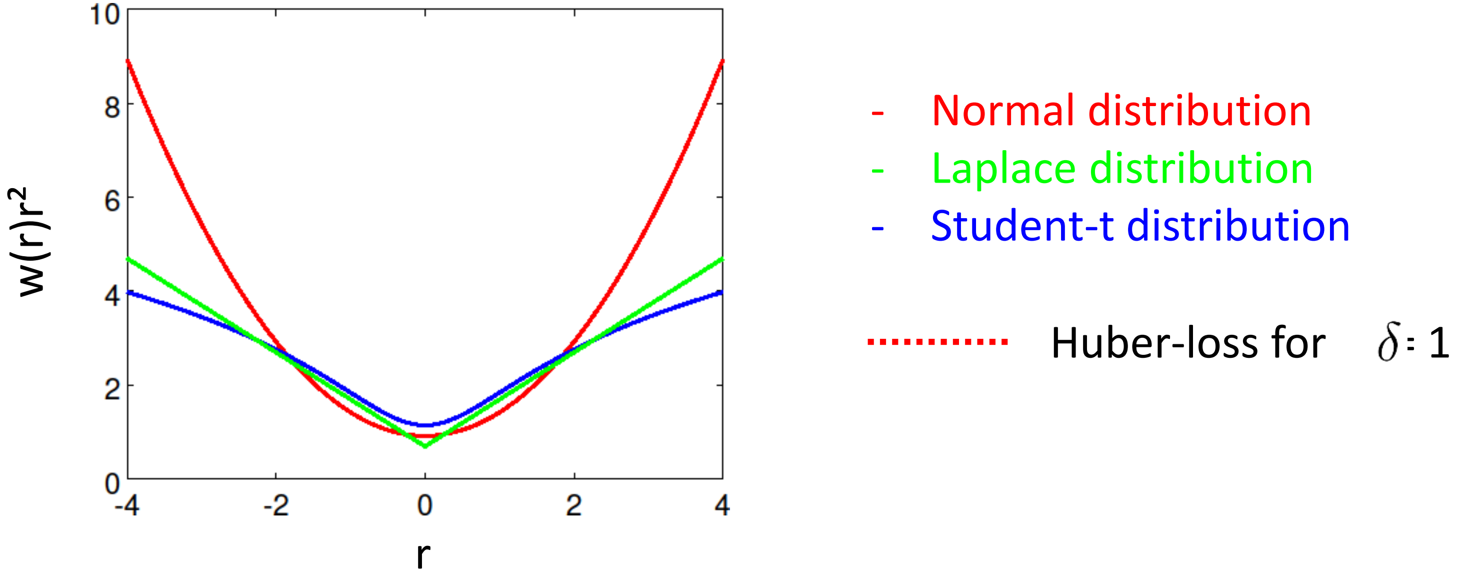
$$\|r\|_\delta = \begin{cases} \frac{1}{2}\|r\|_2^2 & \text{if } \|r\|_2 \leq \delta \\ \delta\left(\|r\|_1 - \frac{1}{2}\delta\right) & \text{otherwise} \end{cases}$$



- Normal distribution
- Laplace distribution
- Student-t distribution

Huber-loss for δ: 1

Recap: Optimization with Non-Gaussian Noise



- Can we change the residual distribution in least squares optimization?
- For specific types of distributions: yes!
- Iteratively reweighted least squares: Reweight residuals in each iteration

$$E(\boldsymbol{\xi}) = \sum_{\mathbf{y}\in\Omega} w\left(r(\mathbf{y},\boldsymbol{\xi})\right) \frac{r(\mathbf{y},\boldsymbol{\xi})^2}{\sigma_I^2}$$

Laplace distribution:

$$w\left(r(\mathbf{y},\boldsymbol{\xi})\right) = |r(\mathbf{y},\boldsymbol{\xi})|^{-1}$$

Example: ORB-SLAM

ORB-SLAM

Raúl Mur-Artal, J. M. M. Montiel and Juan D. Tardós

{raulmur, josemari, tardos} @unizar.es

Instituto Universitario de Investigación en Ingeniería de Aragón Universidad Zaragoza

Universidad Zaragoza 1542

Mur-Atal et al., ORB-SLAM: A Versatile and Accurate Monocular SLAM System, TRO 2015

Lessons Learned

- Alternating tracking and mapping to approximate online SLAM

- Pose graph optimization to approximate the full SLAM posterior with condensed relative pose measurements between frames
- Gauss-Newton approximation reveals the structure of pose graph optimization
  - Hessian is typically sparse, sparsity can be read of directly from relative pose constraints in pose graph (edge structure)
  - Loop closures introduce correlations between non-sequential poses
  - Denser structure of Hessian limits efficiency, loop closures change structure significantly

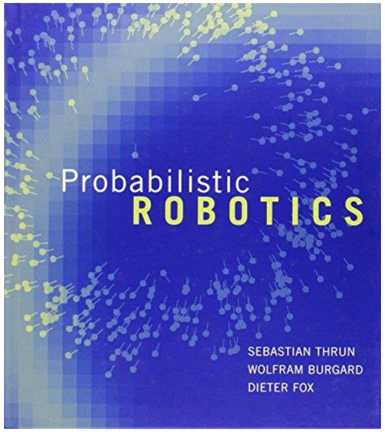
- Monocular SLAM using $\mathrm{Sim}(3)$ pose parametrization

Lessons Learned

- Matching of interest point observations in images to landmarks through descriptors and RANSAC, KLT, and/or active search
- Loop closure detection through place recognition
- Place recognition by image retrieval techniques
  - Popular: Bag-of-Visual-Words + geometric verification (RANSAC)
- Increased robustness for data association outliers:
  - Heavier-tail residual distributions
  - Switchable constraints

Further Reading

- Probabilistic Robotics textbook



Probabilistic Robotics, S. Thrun, W. Burgard, D. Fox, MIT Press, 2005

- Triggs et al., Bundle Adjustment – A modern Synthesis, Springer LNCS 1883, 2002
- Strasdat et al., Scale Drift-Aware Large Scale Monocular SLAM, Robotics Science and Systems, 2010
- R. Mur-Atal et al., ORB-SLAM: A Versatile and Accurate Monocular SLAM System, TRO 2015

Thanks for your attention!