Object Reconstruction

Benjamin Scholz

Universität Hamburg
Fakultät für Mathematik, Informatik und Naturwissenschaften
Fachbereich Informatik
Technische Aspekte Multimodaler Systeme

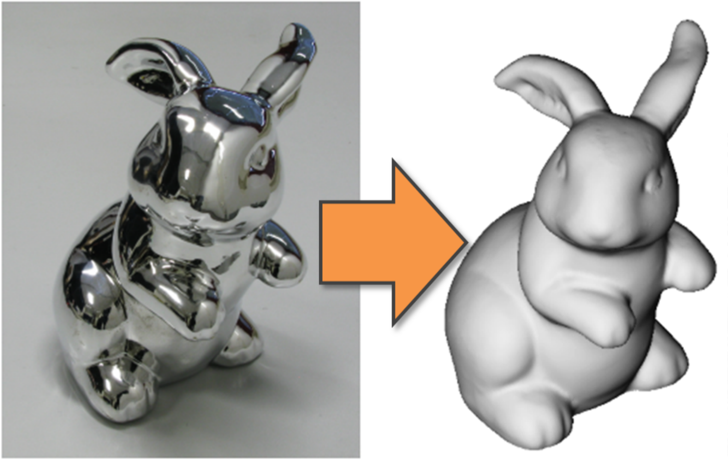
11.05.2017

Table of Contents

Motivation

- Reconstructing a real-world object
- Using vision-sensors to achieve this
- Fields where this is used:
  - Robotics
  - Medicine
  - Movies/Video Games
  - Archaeology

Motivation



From Multi-View Normal Field Integration for 3D Reconstruction of Mirroring Objects by Michael Weinmann, Aljosa Osep, Roland Ruiters, and Reinhard Klein

Structure from Motion

- Multiple images
- Different angles (motion)
- Using RGB-camera

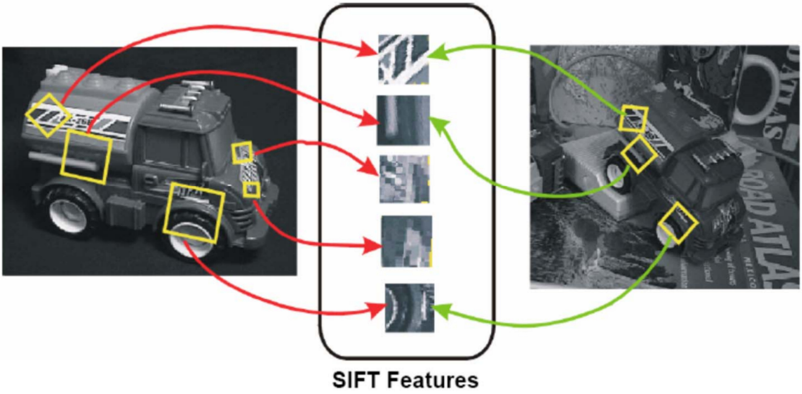
Steps

1. Collecting images
2. Feature detection (SIFT)
3. Feature Matching
4. Filtering (RANSAC)
5. Metric reconstruction (point cloud)
6. Final object reconstruction

Collecting Images

- Taking pictures from different angles by hand
- Gathering images from the Internet
- Robotic arm
- Drones

Feature Detection (SIFT)



SIFT feature matching. From CAP 5415 Computer Vision Fall 2012 - Lecture 5 by Dr. Mubarak Shah

Feature Detection (SIFT)

- **S**cale **I**nvariant **F**eature **T**ransform
- Invariance to scale and rotation
- Produces highly distinctive features
- Robust to occlusion, clutter or noise

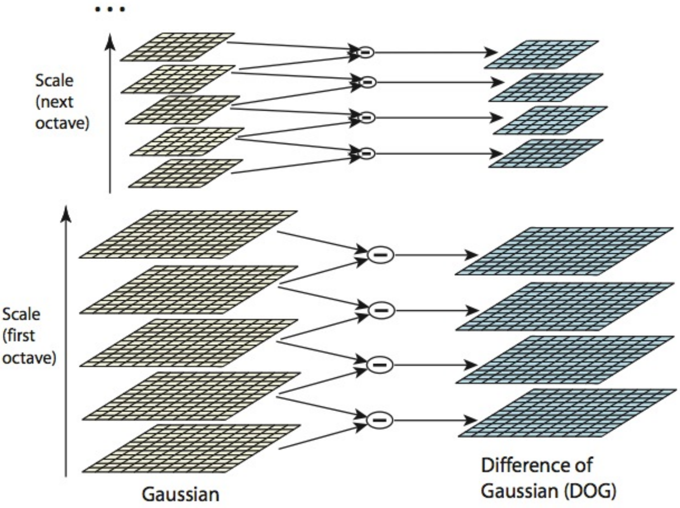
SIFT - Scale-Space Extrema Detection

Scale space representation at different scales. Retrieved from https://en.wikipedia.org/wiki/Scale_space

SIFT - Scale-Space Extrema Detection

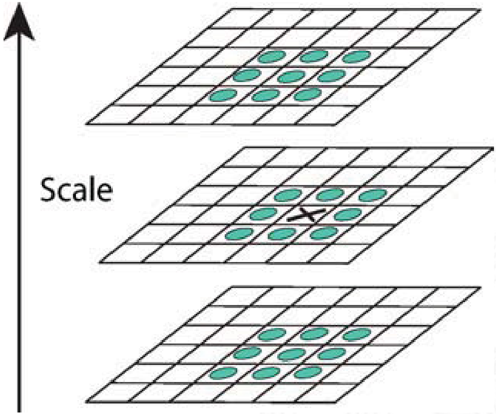
- Using Gaussian Filter
- $L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$
- Different values of σ result in different scales
- Difference of scales
- $D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma)$

SIFT - Scale-Space Extrema Detection



Building octaves. From Distinctive Image Features from Scale-Invariant Keypoints by David G. Lowe

SIFT - Scale-Space Extrema Detection



Local extrema detection. From Distinctive Image Features from Scale-Invariant Keypoints by David G. Lowe

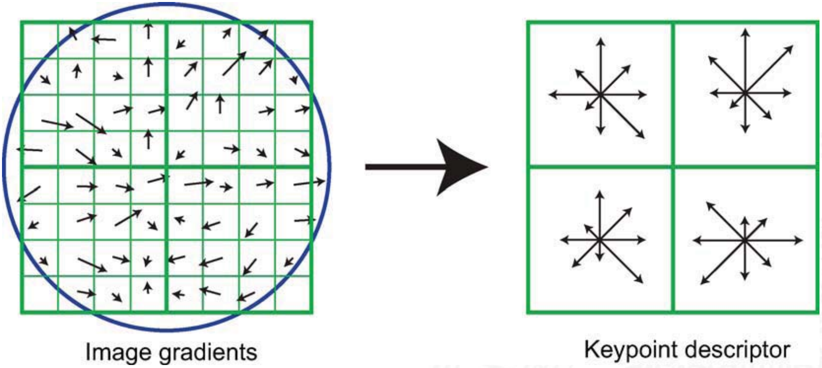
SIFT - Keypoint Localization and Orientation Assignment

- Removing low-contrast keypoints
- Removing unstable features on edges
  - Using gradient to detect edges
- Orientation assignment
- Gradient magnitude and orientation are computed using pixel differences
- 36 bins for orientations from 0° to 360°

SIFT - The Local Image Descriptor

- Sample gradient magnitude and orientations around keypoint
- Achieving scale invariance
- Achieving rotation invariance
- Gaussian weighting function to assign importance

SIFT - The Local Image Descriptor



From Distinctive Image Features from Scale-Invariant Keypoints by David G. Lowe

Matching Features from Different Images

- Nearest neighbor matching
- Creates mismatches
- Method needed for finding correct correspondences

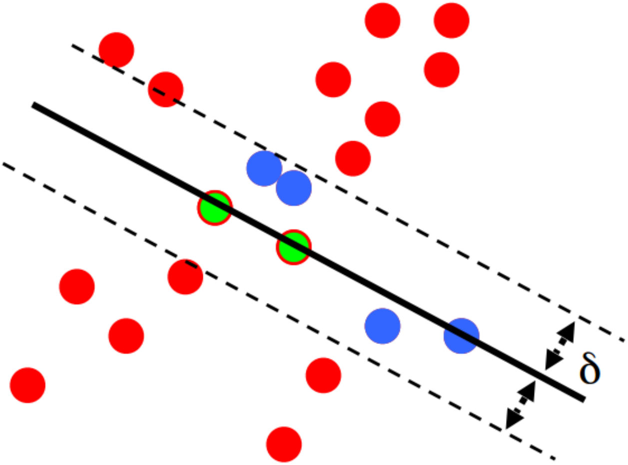
RANSAC

- **Ran**dom **Sa**mple **C**onsensus

  1. Select random subset of data
  2. Fit model to subset
  3. Count inliers
  4. Repeat

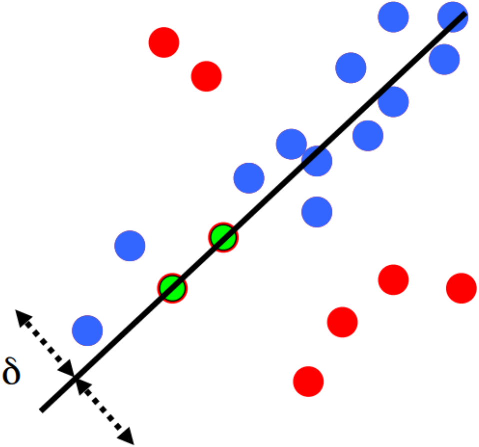
- Highest number of inliers

RANSAC - Example

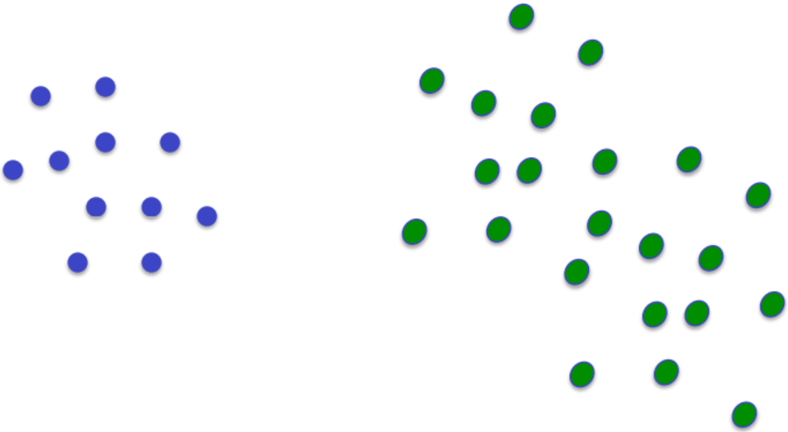


From Stanford CS231A, Computer Vision: from 3D reconstruction to recognition, Winter 2015 lecture 9 by Prof. Silvio Savarese

RANSAC - Example

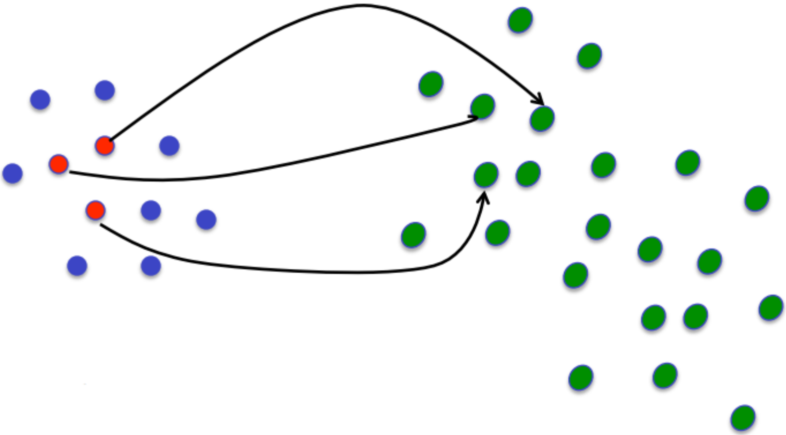


From Stanford CS231A, Computer Vision: from 3D reconstruction to recognition, Winter 2015 lecture 9 by Prof. Silvio Savarese



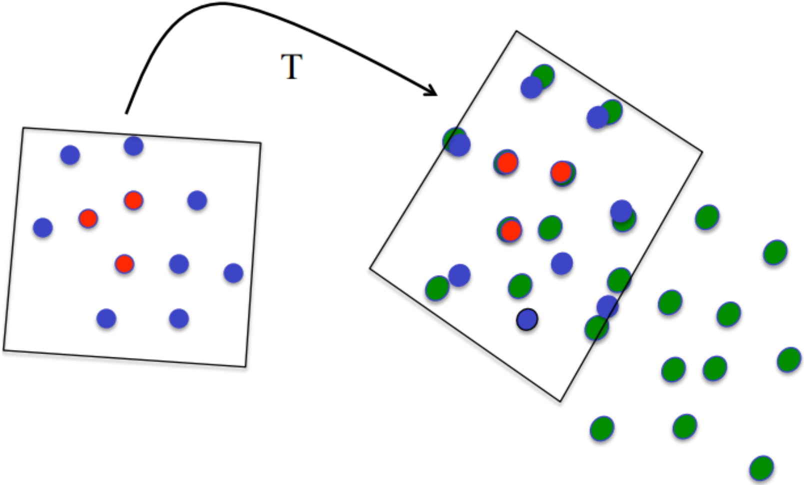
From CS 390: ST in Computer Science: Image and Video Understanding, Matching SIFT by Hao Jiang

RANSAC - Example



From CS 390: ST in Computer Science: Image and Video Understanding, Matching SIFT by Hao Jiang

RANSAC - Example



From CS 390: ST in Computer Science: Image and Video Understanding, Matching SIFT by Hao Jiang

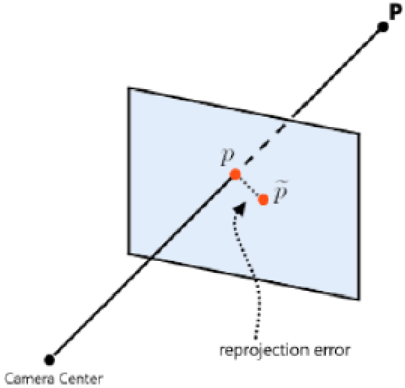
Matching Features from Different Images

- Stitching images together
- Still no 3D structure
- Rough estimation of camera location
- Using robots advantageous

Bundle Adjustment

- Features form point-cloud
- Refining 3D structure, camera motion
- Complex problem

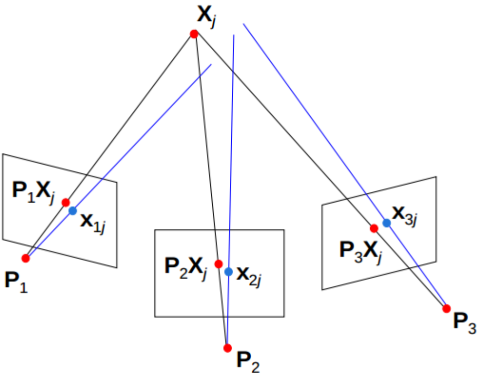
- Minimizing reprojection error



From 3D Models from the Black Box: Investigating the Current State of Image-Based Modeling by Hoang Minh Nguyen, Burkhard Wünsche, Patrice Delmas and Christof Lutteroth

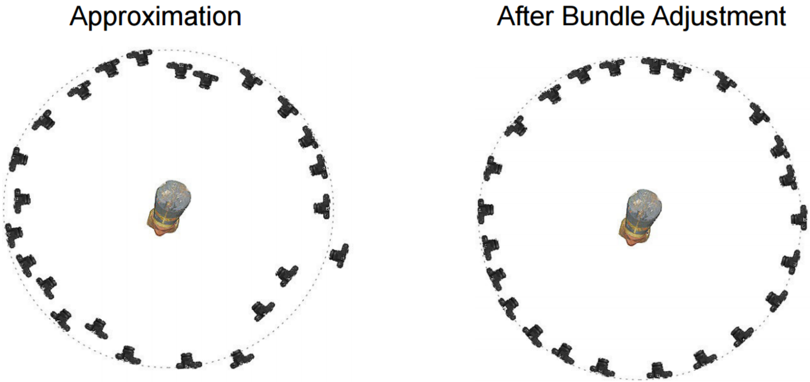
Bundle Adjustment

- Considers all images/cameras



From Lecture 6: Multi-view Stereo & Structure from Motion by Prof. Rob Fergus. Retrieved from http://cs.nyu.edu/ fergus/teaching/vision/11_12_multiview.pdf

Bundle Adjustment



From Photogrammetric Computer Vision, Lecture Notes by Volker Rodehorst

Object Reconstruction using RGB-D

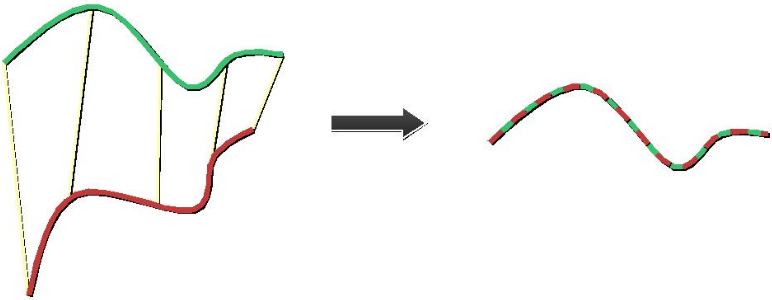
- RGB-D cameras like Kinect provide depth-data, so it does not have to be computed
- Produces Point-Clouds
- We can omit the step of feature detection
- Approximating camera location becomes easier

Iterative Closest Point (ICP)

- Matching two corresponding point clouds
  $P = \boldsymbol{p_1}, ..., \boldsymbol{p_n}$ and $Q = \boldsymbol{q_1}, ..., \boldsymbol{q_n}$
- Minimizing sum of squared error
  $E(R, t) = \frac{1}{n} \sum_{i=1}^{n} \|\boldsymbol{p_i} - R\boldsymbol{q_i} - t\|^2$
- Here R is the **rotation matrix** and t is the **translation vector**

Iterative Closest Point (ICP)

- If correspondences of points are known we can compute the solution in closed form



From Iterative Closest Point, presented by M.T. Hajiaghayi. Retrieved from http://groups.csail.mit.edu/graphics/classes/6.838/F01/lectures/IterativeAlgs/ICP/align.html

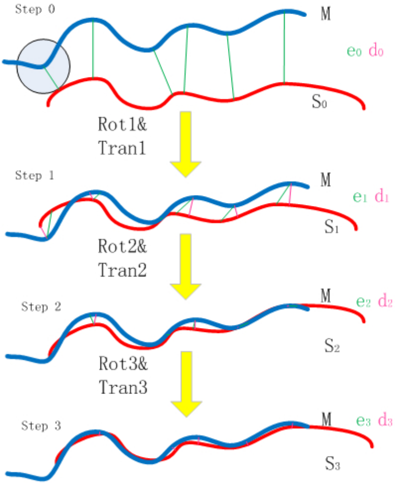
Iterative Closest Point (ICP)

- Calculate center of mass of both point clouds and subtract it from each point: $\boldsymbol{p}_i' = \boldsymbol{p}_i - \mu_{\boldsymbol{p}}$ and $\boldsymbol{q}_i' = \boldsymbol{q}_i - \mu_{\boldsymbol{q}}$
- $W = \sum_{i=1}^{N_p} \boldsymbol{p}_i' \boldsymbol{q}_i'^T$
- Using single value decomposition on W gives us the singular values σ_1, σ_2 and σ_3 of W as well as U and V
- The optimal solution is now: $R = UV^T$, $t = \mu_{\boldsymbol{p}} - R\mu_{\boldsymbol{q}}$
- With minimal error of $E(R, t) = \sum_{i=1}^{N_p} (\|\boldsymbol{p}_i'\|^2 + \|\boldsymbol{q}_i'\|^2) - 2(\sigma_1 + \sigma_2 + \sigma_3)$

Iterative Closest Point (ICP)

- Problem: We usually do not know the correspondences
- Assume correspondences
- Using method iteratively

Iterative Closest Point (ICP)



From Iterative Closest Point algorithm-point cloud/mesh registration by Taylor Wang. Retrieved from https://taylorwang.wordpress.com/2012/04/06/iterative-closest-point-algorithm-point-cloudmesh-registration/

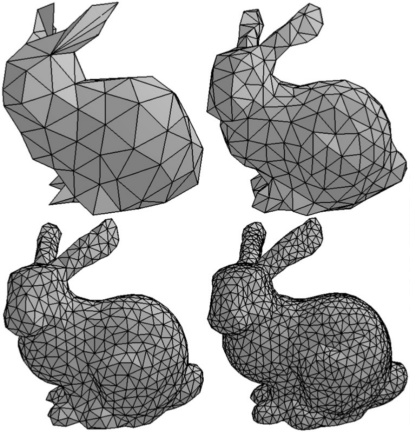
Iterative Closest Point (ICP)

- Result: large point cloud
- Representation of object
- Inferring camera locations
- Applications working in real-time are feasible (KinectFusion)

3D Object Representations

- Point cloud
- 3D mesh
- Truncated Signed Distance Function (TSDF)

3D Object Representations - 3D Mesh



Retrieved from http://www.cmap.polytechnique.fr/ peyre/geodesic_computations/

References

[1] M. A. Fischler and R. C. Bolles.
Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography.
Communications of the ACM, 24(6):381–395, 1981.

[2] K. Häming and G. Peters.
The structure-from-motion reconstruction pipeline–a survey with focus on short image sequences.
Kybernetika, 46(5):926–937, 2010.

[3] D. G. Lowe.
Distinctive image features from scale-invariant keypoints.
International journal of computer vision, 60(2):91–110, 2004.

References (cont.)

[4] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohi, J. Shotton, S. Hodges, and A. Fitzgibbon.
Kinectfusion: Real-time dense surface mapping and tracking.
In *Mixed and augmented reality (ISMAR), 2011 10th IEEE international symposium on*, pages 127–136. IEEE, 2011.

[5] O. Özyesil, V. Voroninski, R. Basri, and A. Singer.
A survey on structure from motion.
CoRR, abs/1701.08493, 2017.

[6] S. Rusinkiewicz and M. Levoy.
Efficient variants of the icp algorithm.
In *3-D Digital Imaging and Modeling, 2001. Proceedings. Third International Conference on*, pages 145–152. IEEE, 2001.