Object Reconstruction in Robotics

Introduction and Recent Approaches

University of Hamburg
Faculty of Mathematics, Informatics and Natural Sciences
Department of Informatics

Technical Aspects of Multimodal Systems

23. May 2019

Outline

What is Object Reconstruction

- Capture the shape and dimensions (and appearance, mechanical properties etc.) of real world objects with one or more sensors
- Construct a model from the obtained data

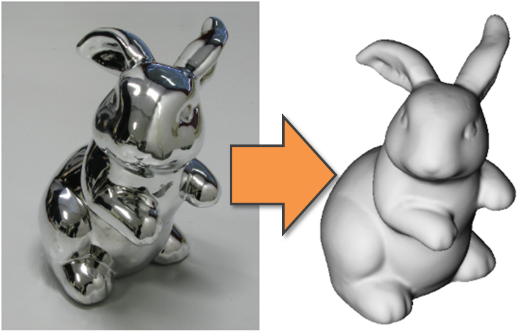


Figure: 1 Bunny figurine and reconstructed model (Weinmann, 2013)

Applications in Robotics

- Estimation for best grasping position
- Object recognition
- Object pose estimation
- Telerobotics
- Industry robotics
- Mapping
- Also related
  - Robots used in archaeology
  - Surgery robots (e.g. inner countour scanning for keyhole surgery)
  - Tomographic reconstruction
  - Virtual / augmented reality
  - ...

Depth Maps

- Typical sensors and methods for mapping depth
  - ToF (time-of-flight) cameras
  - Backscatter LiDAR (Light Detection And Ranging) systems
  - Triangulation using image cameras
  - Monocular cues
  - Triangulation using structured light and cameras
- The output is a 3D point cloud per capturing cycle

ToF Cameras

- Either light pulses or continuous wave
- Pulses: measure time of flight, continuous wave: measure phase shift

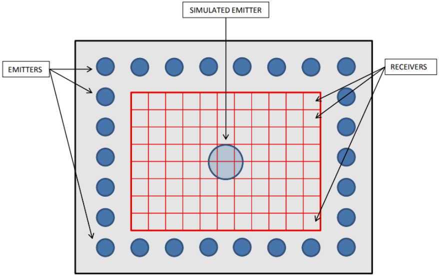


Figure: 2 Scheme of a matricial ToF camera sensor. The CCD/CMOS matrix of lock-in pixels is in red. The emitters (blue) are distributed around the lock-in pixels matrix and mimic a simulated emitter co-positioned with the center of the lock-in pixel matrix (light blue) (Mutto, 2012)

- Also based on time-of-flight or phase shift measurement

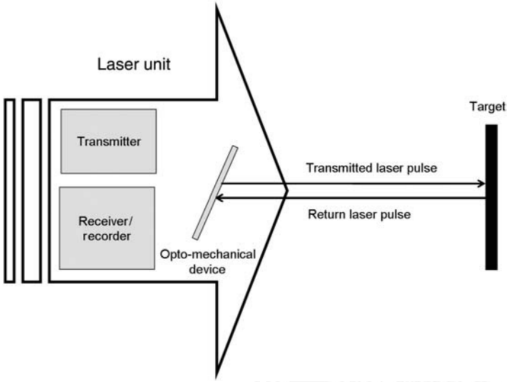


Figure: 4 Basic components of a LiDAR instrument (Heritage, 2009)

- 2 (identical) cameras, reference and target camera (Coplanar and aligned, parallel optical axis)
- Depth by triangulation of matching points in both images
- Epipolar geometry enhances search for matching points

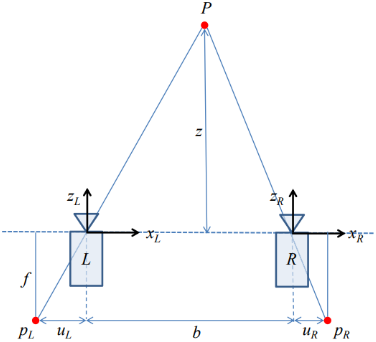


Figure: 7 Triangulation with a pair of aligned and rectified cameras (Mutto, 2012)

Mono Cameras

- "Simulate" 2 cameras with mirrors or prisms
- Track camera position and use images from different positions
  - Structure from Motion
  - Using motion sensors, e.g. inertial measurement unit

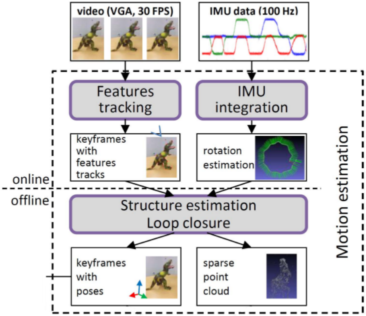


Figure: 8 System overview (Muratov, 2016)

Monocular Cues

- Infer depth from visual scene properties
  - Size of objects
  - Occlusions
  - Texture gradient and shading
  - Luminance, contrast, saturation, blur
  - ...
- "[...] by adding monocular cues to stereo (triangulation) ones, [...] obtain significantly more accurate depth estimates than is possible using either monocular or stereo cues alone." (Saxena, 2007)

Structured Light

- Project pattern on scene, various coding schemes to suit environmental conditions
- Find known pattern features in image to perform active triangulation

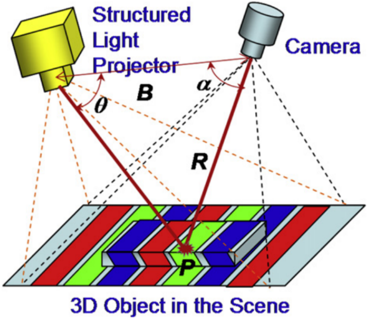


Figure: 10 Illustration of structured light (Geng, 2011)

Point Cloud Matching with ICP

- Iterative closest point algorithm
- Combine multiple incomplete (due to e.g. self occlusion) or inaccurate 3D point clouds by matching them
- Goal
  - Find corresponding points in both point clouds
  - Find a rotation R and translation T to match one to the other
- (Basic) Algorithm (Besl, 1992)
  - Find closest corresponding point (between target and reference set, for each point)
  - Compute rotation matrix R and translation vector T
  - Apply computed transformation and find distances (all corresponding points)
  - Loop unless distance is below threshold or max. iterations reached

Meshing of 3D Point Clouds

- Typical approach: Reconstruct surface with triangles (e.g. Delaunay triangulation)
- Involves steps such as noise reduction, removal of erroneous and redundant triangles, adjusting of spacing, etc.
- Can be followed by grid construction

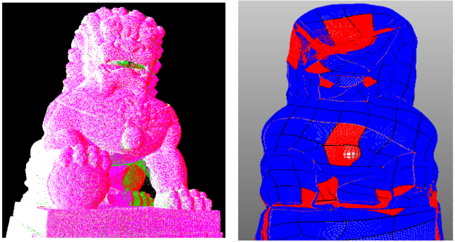
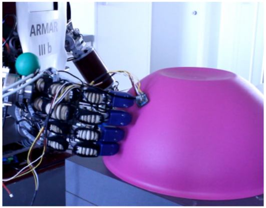


Figure: 11 Five registered point cloud scans, 3D mesh after grid construction (Nguyen, 2012)

Beyond Optical Sensing

- Many more ways to find shape information
  - Acoustic
  - Micro waves
  - ...
- Especially popular: tactile sensing
- Also allows gathering mechanical properties



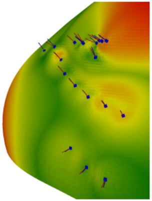


Figure: 12 Exploration of an unknown object [...] reconstructed using the gathered contact positions and normals (Ottenhaus, 2018)

Full Laser Scan

- System by Jiang et al. (2018)
  - Workbench with u-shaped gantry above, which moves across
  - Two 2D laser scanners attached to gantry pointing downwards
  - Error on 600mm diameter truck wheel:
  - 0.7mm (53s scan time), 2.6 mm (13s scan time)

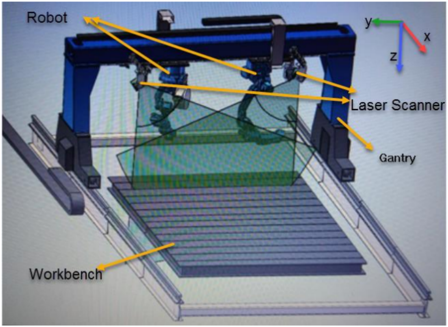


Figure: 13 Large-scale online 3D reconstruction system diagram (Jiang, 2018)

- Work of Ottenhaus et al. (2018)
  - Tactile robotic hand with soft fingertips
  - Robot's fingertips each have IMU and pressure sensor
  - Object surface is estimated through Gaussian process utilising contact positions and contact normals (local tilt)
  - Error on bowl: 4.7mm, 17°surface normal
  - Error on cardboard box: 3.5mm, 24°surface normal

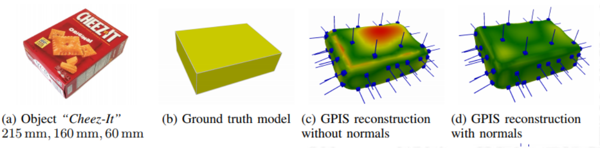


Figure: 14 Comparison of the reconstruction results with included and excluded normal information [...] (Ottenhaus, 2018)

Single View with RGBD Camera

- Work of Rodríguez-Jiménez et al. (2013)
  - Using Kinect (structured light), objects lying on table
  - RGB-based segmentation on image to fill holes in depth map
  - Assumes table top normal is extrusion axis to fill unobserved part

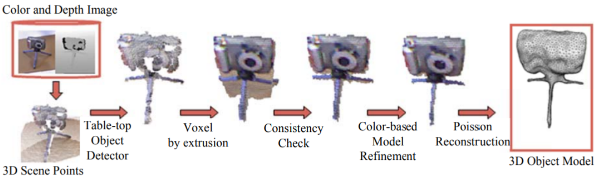
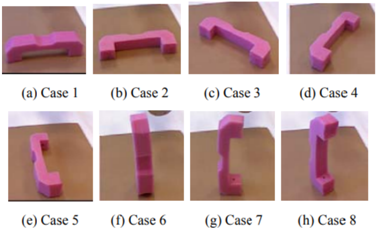


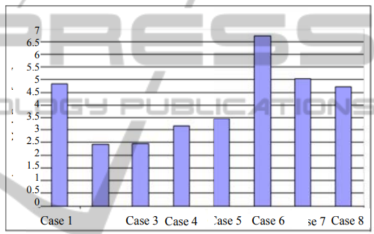
Figure: 15 Overview of [...] model acquisition process (Rodríguez-Jiménez, 2013)

Single View with RGBD Camera

- Less than 5mm error on everyday objects (book, can, camera, glue stick, cup, pen, ball, etc.)
- Higher error on asymmetric or irregular shaped objects, due to incorrect assumption



(a) Fig. 16 [...] object in the 8 evaluated orientations [...] (Rodríguez-Jiménez, 2013)



(b) Fig. 17 Evaluation of the error for 8 orientations (Rodríguez-Jiménez, 2013)

DB Based Methods

- Work of Klank et al. (2009)
  - Stereo camera based object recognition
  - Internet search and 3D object matching using ICP
  - Best search for plate on cluttered table: 201.65s (0.98 certainty)

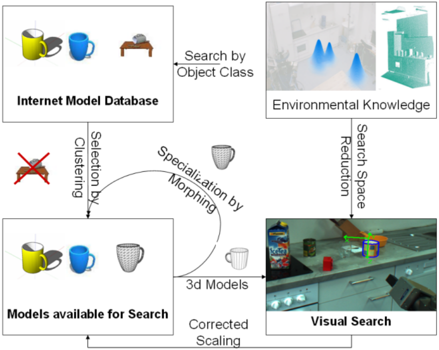


Figure: 18 Overview of the method (Klank, 2009)

Fusion of Tactile and Visual Information

- Work of Ilonen et al. (2013)
  - Kinect (RGBD) and robotic arm with gripper and tactile sensors
  - Method:
    1. Construct point cloud model from RGBD data and complete assuming symmetry
    2. Plan and execute grasp on object
    3. Compute contact points of gripper with forward kinematics
    4. Update model based on closed grasp tactile information using Iterated Extended Kalman Filter
  - Result for cylindrical container of width 53mm:
  - 45.41mm on avg. using only kinect
  - 55.43mm after fusion
  - Error of 2.43mm on avg.

Overview

- Full Laser Scan
  - Fast and very accurate
  - Not flexible, expensive hardware
- Tactile Sensing
  - Can reconstruct surface accurately through surface normals
  - Very slow, expensive hardware
- RGBD Single View
  - Operates passively, somewhat accurate, comparatively cheap hardware
  - Makes assumptions about object shape
- DB Based
  - Very accurate (if models available), can obtain other information (e.g. weight), comparatively cheap hardware
  - Classification adds dimension for errors to occur
- Visual and Tactile Fusion
  - Somewhat accurate
  - Object must be accessible, expensive hardware

Conclusion

- Broad variety of approaches (and counting)
- Different sensors / modalities to obtain data
- Often involves tradeoffs (speed, accuracy, flexibility, cost)
- In general, no "best approach", depends on application
- Common application: estimate best way to grasp object

References

[1] Weinmann, Michael and Osep, Aljosa and Ruiters, Roland and Klein, Reinhard: *Multi-View Normal Field Integration for 3D Reconstruction of Mirroring Objects*, Proceedings of the International Conference on Computer Vision, 2013

[2] Mutto, Carlo Dal, Zanuttigh, Pietro, Cortelazzo, Guido M.: *Time-of-Flight Cameras and Microsoft Kinect*, Boston, Springer US, 2012

[3] George L. Heritage, Andrew R.G. Large: *Principles of 3D Laser Scanning*, Chichester, U.K., Hoboken, N.J, Wiley-Blackwell, 2009

[4] O. Muratov, Y. Slynko, V. Chernov, M. Lyubimtseva, A. Shamsuarov and V. Bucha, "3DCapture: *3D Reconstruction for a Smartphone*, 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Las Vegas, NV, 2016, pp. 893-900

[5] J. McBride, M. Snorrason, T. Goodsell, R. Eaton and M. R. Stevens: *Single Camera Stereo for Mobile Robot Surveillance*, 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops, San Diego, CA, USA, 2005, pp. 128-128.

[6] William Lovegrove: *Single-camera stereo vision for obstacle detection in mobile robots*, SPIE 10.1117/2.1200710.0911, 2007

[7] David G. Lowe: *Distinctive image features from scale-invariant keypoints*, IJCV 60 (2), pp. 91-110, 2004

[8] Saxena, Ashutosh and Schulte, Jamie and Ng, Andrew Y.: *Depth Estimation Using Monocular and Stereo Cues*, IJCAI'07, Hyderabad, India, 2007

[9] Jason Geng: *Structured-light 3D surface imaging: a tutorial*, Adv. Opt. Photon. 3, 128-160 (2011)

[10] Ying He, Bin Liang, Jun Yang, Shunzhi Li, Jin He: *An Iterative Closest Points Algorithm for Registration of 3D Laser Scanner Point Clouds with Geometric Features*, Sensors 2017, 17, 1862; doi:10.3390/s17081862

[11] Tien Thanh Nguyen, Quang Minh Nguyen, Xiu Guo Liu1, Yao Yevenyo Ziggah: *3D OBJECT MODEL RECONSTRUCTION BASED ON LASER SCANNING POINT CLOUD DATA*, International Symposium on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences 2012

[12] Simon Ottenhaus, Pascal Weiner, Lukas Kaul, Andreea Tulbure and Tamim Asfour: *Exploration and Reconstruction of Unknown Objects using a Novel Normal and Contact Sensor*, 1614-1620. 10.1109/IROS.2018.8594272, 2018

References

[13] Jiang, Qixiang, Hou, Runshi, Wang, Shenghua, Wei, Xiuquan, Chang, B, Shan, Jiguo, Du, Dong: *On-Line 3D reconstruction based on laser scanning for robot machining of large complex components.*, Journal of Physics: Conference Series, 1074. 012166. 10.1088/1742-6596/1074/1/012166, 2018

[14] Silvia Rodríguez-Jiménez, Nicolas Burrus and Mohamed Abderrahim: *3D Object Reconstruction with a Single RGB-Depth Image*, Proceedings VISAPP, pages 155-163, 2013

[15] Ulrich Klank, Muhammad Zeeshan Zia and Michael Beetz: *3D Model Selection from an Internet Database for Robotic Vision*, IEEE International Conference on Robotics and Automation, Kobe, Japan, 2009

[16] Jarmo Ilonen, Jeannette Bohg and Ville Kyrki, *Fusing Visual and Tactile Sensing for 3-D Object Reconstruction While Grasping*, ICRA, Karlsruhe, Germany, 2013

[17] Besl, P.J.; Mckay, N.D.: *A method for registration of 3-D shapes*, IEEE Trans. Pattern Anal. Mach. Intell. 1992, 14, 239–256.

- Signal attenuation, time delay and background illumination account for the signal that is received

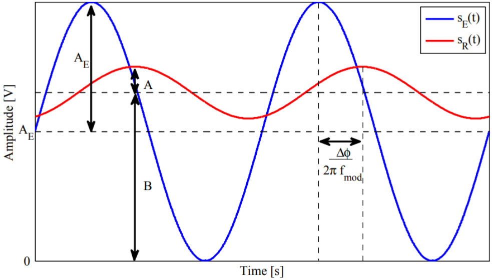


Figure: 3 Example of emitted signal sE(t) (in blue) and received signal sR(t) (in red)) (Mutto, 2012)

- Compute position in space from mirror rotation and support position

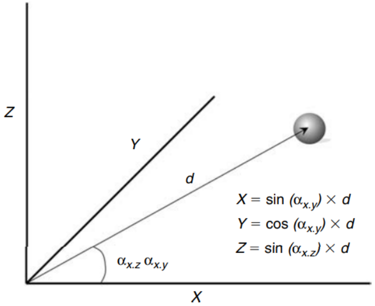


Figure: 5 Illustration of the ability of LiDAR systems to position remote objects in 3D space (Heritage, 2009)