3D object recognition used by team robOTTO

Workshop

Juliane Hoebel

February 1, 2016

Faculty of Computer Science, Otto-von-Guericke University Magdeburg

Content

1. Introduction
2. Depth sensor
3. 3D object recognition

Introduction

Introduction

A RoboCup@Work scenario consists of several elements, e.g.

- the environment,
- the objects that affect navigation,
- the objects that are to be manipulated,
- the objects with which robots interact,
- the task to be performed by a team etc.

Introduction

There are several assumptions about the kind of robots used in the competition, e.g.

> *The robots use sensors to obtain information about their whereabouts in the environment and the task-relevant objects. Often used sensors are*
>
> - *laser finders,*
> - *color CCD cameras and*
> - *3D cameras (such as the Kinect camera or the Intel Realsense camera).*
>
> *The design of the scenario should be such that the robots can solve the tasks safely and robustly using these sensors.*

Depth sensor

Structured-light 3D scanner

Figure 1: Intel® RealSense™ camera (F200).

Intel® RealSense™ camera

Technical specifications:

RGB video resolution

Full HD 1080p (1920 x 1280).

IR depth resolution

VGA (640 x 480).

Image frame rate

30 fps (RGB), 60 fps (IR depth).

Field of view

77°(RGB), 90°(IR depth).

Range

0.2 m - 1.2 m.

Laser projector

First class IR laser projector.

Input

5 V.

USB port

USB 3.0.

Intel® RealSense™ camera

Range

0.2 m - 1.2 m.

i.e. there are various ranges and accuracies possible depending on specific algorithms.

The current camera design is the following:

- The RealSense camera is mounted on the KUKA robot arm.
- The RealSense camera's minimum distance to an object (M30, M20 etc.) is approx. 0.30 m.
- The RealSense camera has a static top-down perspective to the table with objects.

Intel® RealSense™ camera

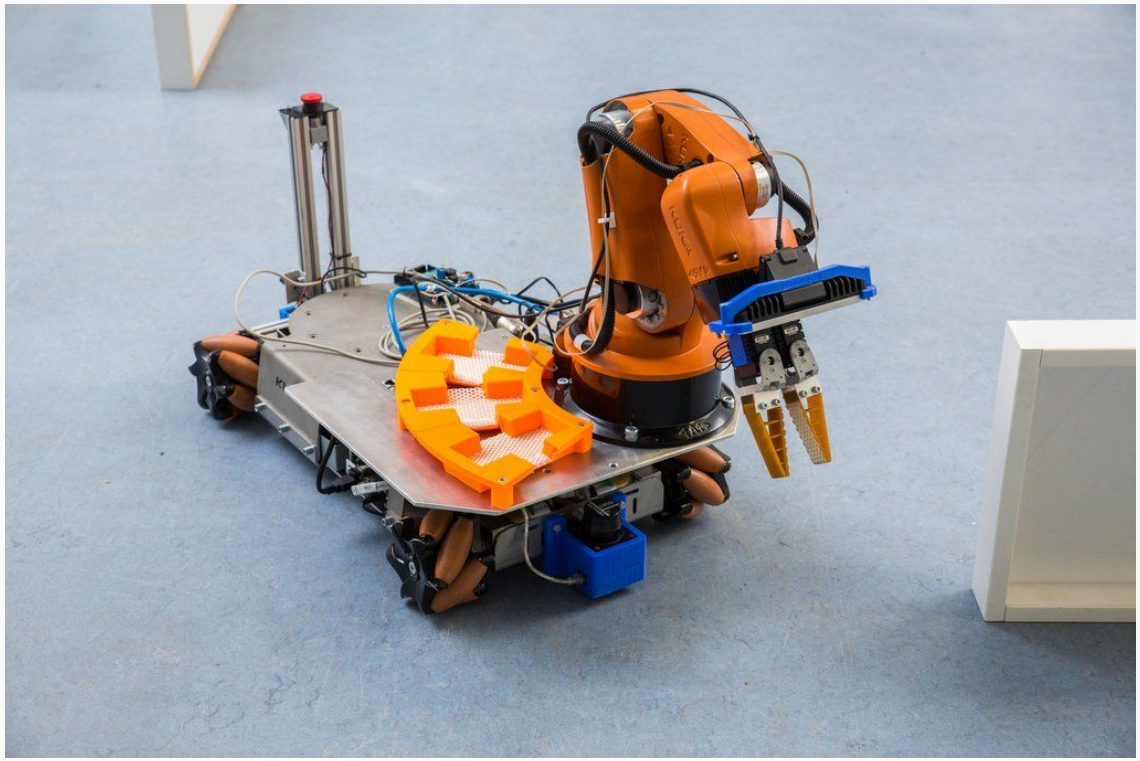


Figure 2: Robot arm with mounted RealSense camera.

Principle of a structured-light 3D scanner

- Any spatio-temporal pattern of light is projected on a surface.
- Observed patterns can be used for a reconstruction of the surface shape.
- Geometric distortions by optics and perspective must be compensated by a calibration of the measuring equipment.

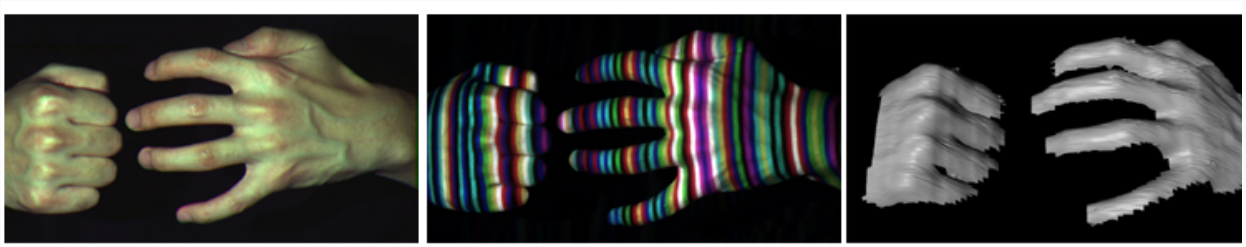


Figure 3: Complex projected light pattern on human hands.

Single light stripe scanning

Imagine the following simple case to scan a scene:

- Project a single stripe of laser light.
- Scan it across the surface of the object.

This is a very precise version of structured light scanning!

It is good for high resolution, but it needs many images and takes time.

Questions

- How to achieve faster acquisitions?

Questions

- How to achieve faster acquisitions?
  - Multiple stripes has to be projected simultaneously.

Questions

- How to achieve faster acquisitions?
  - Multiple stripes has to be projected simultaneously.
- How to solve the correspondence problem: Which stripe is which?

Questions

- How to achieve faster acquisitions?
  - Multiple stripes has to be projected simultaneously.
- How to solve the correspondence problem: Which stripe is which?
  - Binary coded light striping or gray/ color coded light striping is recommended. The correspondence problem is solved by searching the pattern in the camera image.

Challenges

During every RoboCup@Work object manipulation task the following difficulties are observed:

- Reflective or transparent surfaces. Reflections go either away from the camera or right into its optics. The dynamic range of the camera is exceeded.
- Double reflections and inter-reflections. The stripe pattern is overlaid with unwanted light. Reflective cavities and concave objects are difficult to handle.

Challenges

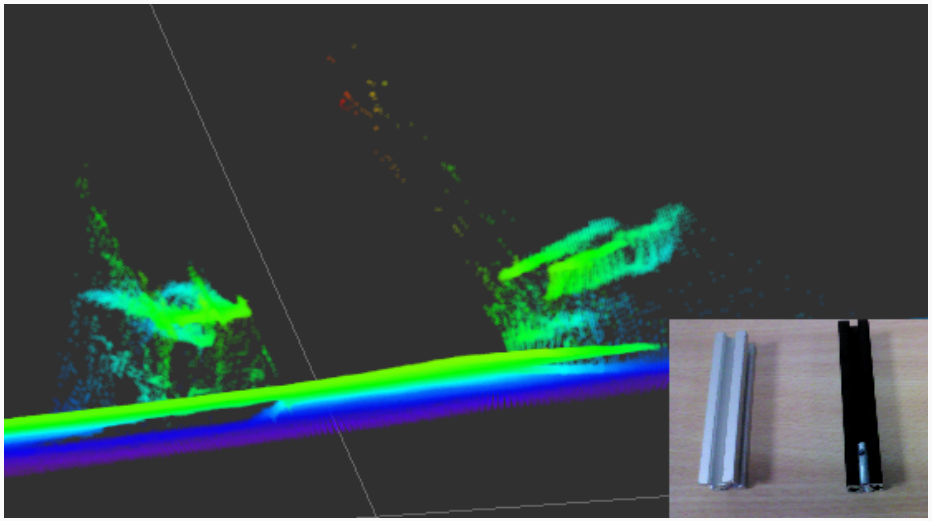


Figure 4: Point cloud with two aluminuim profiles of RoboCup@Work.

Solutions

A 1D diffuser between the projector and the object to be scanned could handle reflective objects [2].

In the computer vision community structured light patterns are designed that are resilient to individual global illumination effects [1].

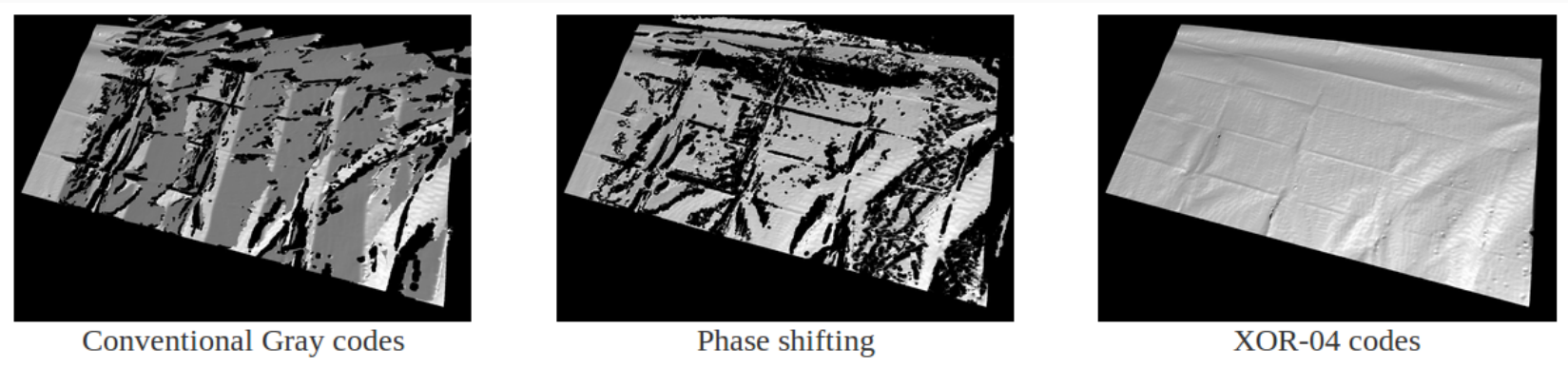
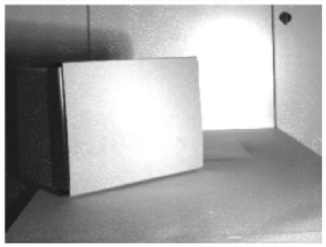


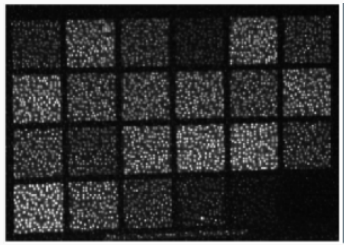
Figure 5: Reconstruction of a shower curtain with different light patterns.[1]

[1] Image from graphics.cs.cmu.edu

Questions

What could be further challenges?






Challenges

Figure 6: Dataset with sparse outliers.[2]

[2]Image from pointclouds.org

3D object recognition

Overview

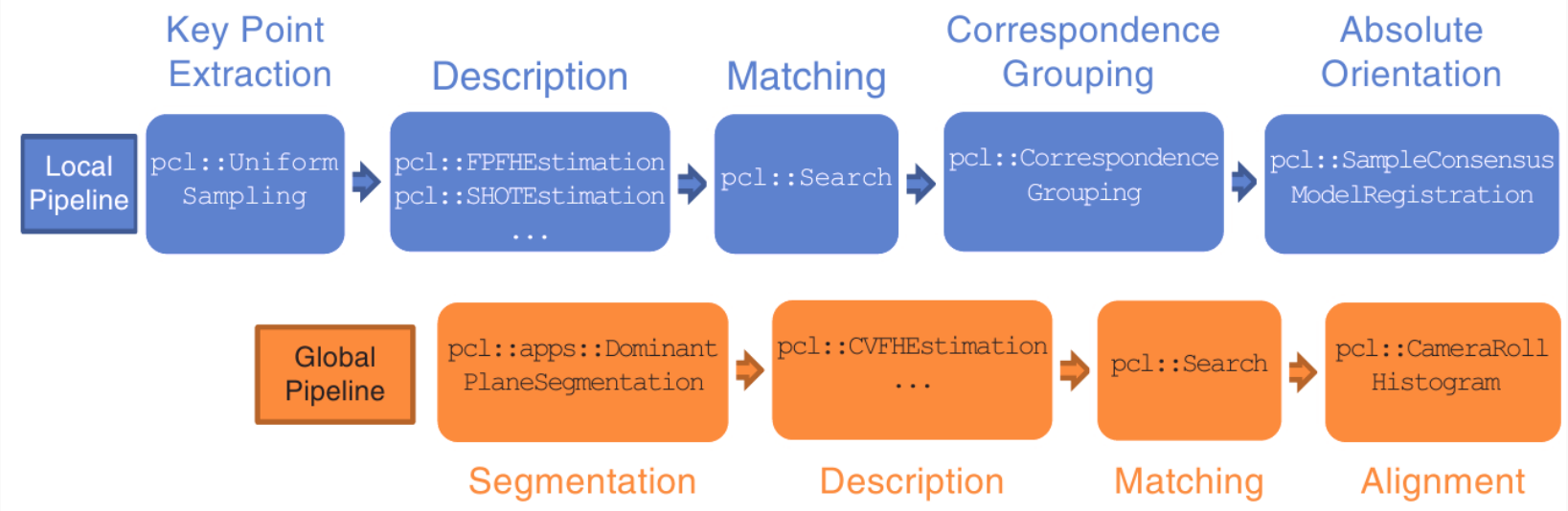


Figure 7: Example of local and global 3D recognition pipelines in PCL.[3]

[3] Image from pointclouds.org

Segmentation

Segmentation consists of breaking the cloud apart in different groups of points that share certain characteristics.

Our 3D plane segmentation is based on a plane model fitting. We use an iterative method called RANSAC (Random Sample Consensus).

The objects' cluster extraction happens in an Euclidean sense.

Segmentation

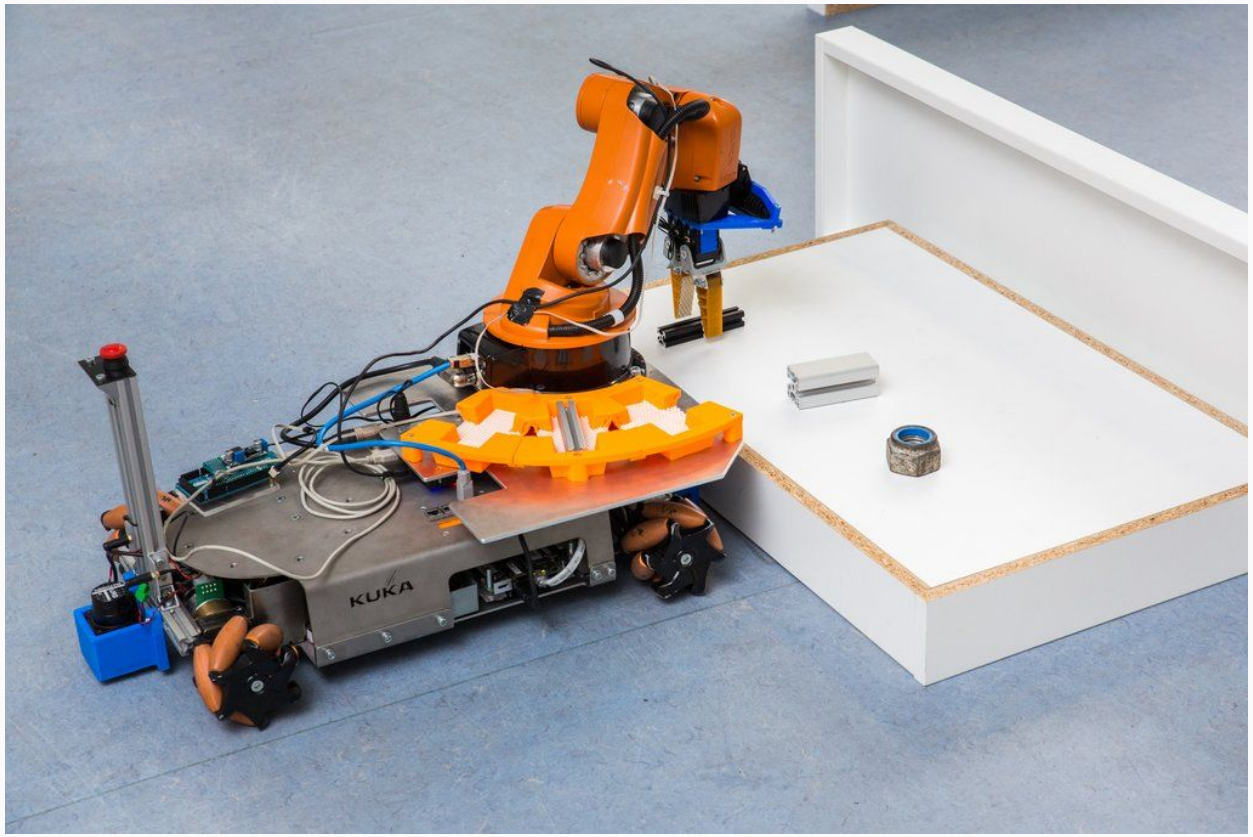


Figure 8: Firstly, we segment and remove the 3D plane. Then, the objects' points are separated.

Model fitting (RANSAC)

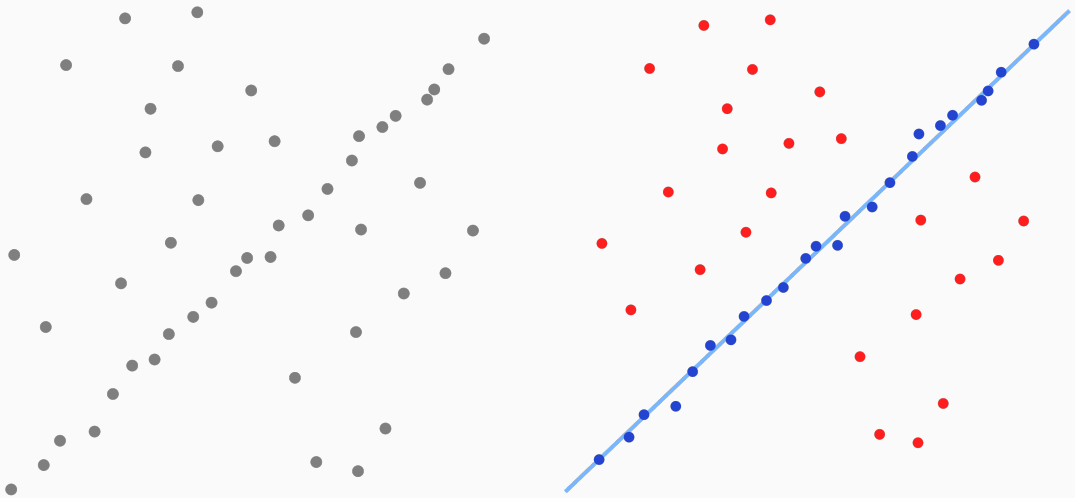


Figure 9: Data points before and after fitting of a line model.[4]

[4] Image from robotica.unileon.es

Descriptor

Descriptors are complex and prezise signatures of a point, that encode information about the surrounding geometry.

The purpose is to identify a point across multiple point clouds, no matter the noise, resolution or transformations.

The result of a descriptor is binned into an histogram.

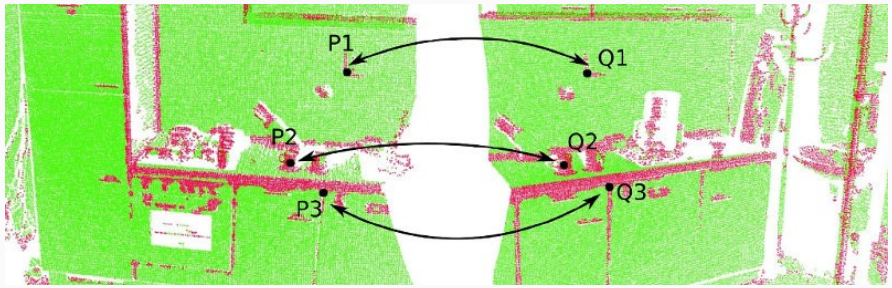


Figure 10: Finding correspondences between point features of two clouds.[5]

[5] Image from pointclouds.org

Local descriptor

Local descriptors ...

- are computed for keypoints that you give as input.
- have no notion of what an object is, they just describe how the local geometry is around that point.
- are used for object recognition and registration.

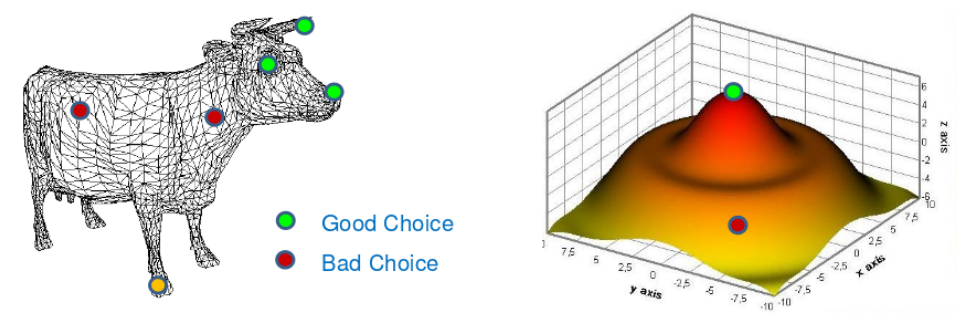


Figure 11: 3D keypoints are distinctive and repeatable.[6]

[6] Image from pointclouds.org

Local descriptor

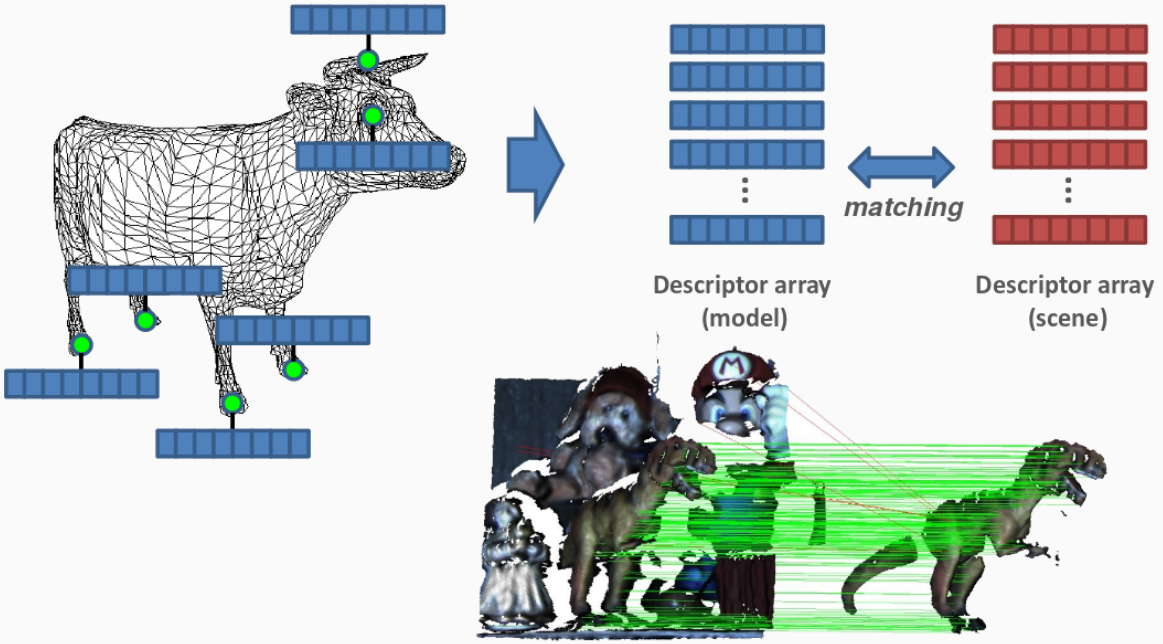


Figure 12: Matching descriptions yields point-to-point correspondences.[7]

[7] Image from pointclouds.org

Global descriptor

Global descriptors ...

- encode object geometry.
- are not computed for individual points, but for a whole cluster that represents an object.
- are used for object recognition and pose estimation.

Figure 13: Segmentation is required in order to get object clusters.[8]

[8] Image from pointclouds.org

Local vs global descriptors

Local descriptors

- are well suited to handle clutter and occlusions.

Global descriptors

- need complete information concerning the surface.
- are more descriptive on objects with poor geometric structure.

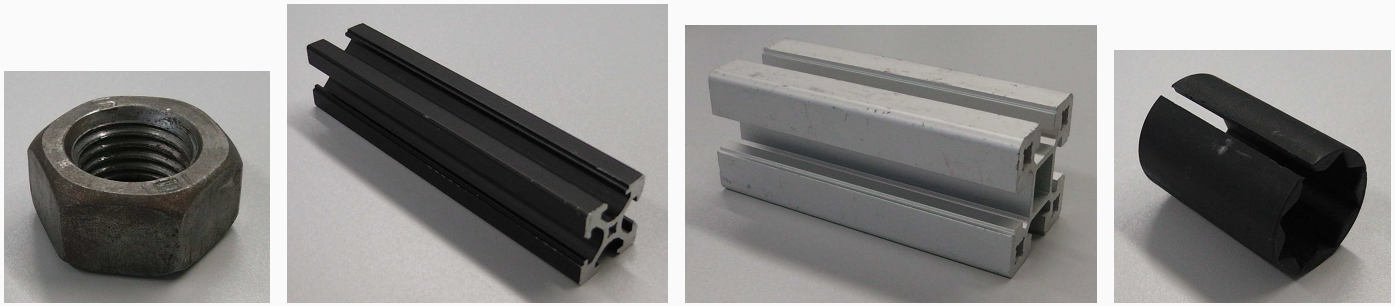
Questions

What kind of descriptor would you prefer for 3D object recognition at RoboCup@Work? Why?

Local vs global descriptors

RoboCup@Work objects are characterized by a relatively poor geometric structure.

robOTTO uses global descriptors, e.g. the ESF, to handle these objects.



(a) M30 (b) F20_20_B (c) S40_40_G (d) V20

Figure 14: A subset of the RoboCup@Work objects.

Ensemble of Shape Functions descriptor

The Ensemble of Shape Functions descriptor (ESF) ...

- describes distances, angles and area of cloud's points.
- does not require normal information.
- is robust to noise and incomplete surfaces.

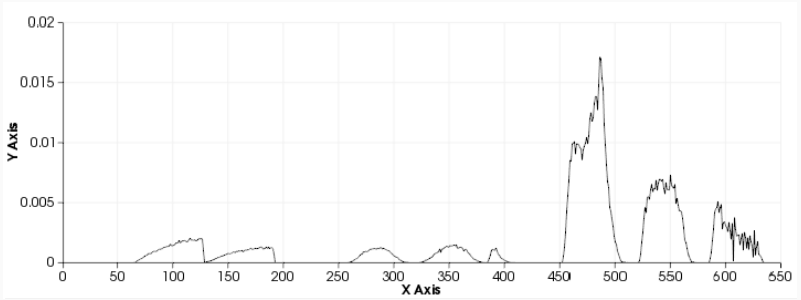


Figure 15: An ESF histogram of M30's partial view. It's an ensemble of 10 64-bin sized histograms of shape functions.

Ensemble of Shape Functions descriptor

The algorithm ...

- uses a voxel grid as an approximation of the real surface.
- iterates through all the points in the cloud, whereby 3 random points are chosen for every iteration.
- computes the shape functions for these points.

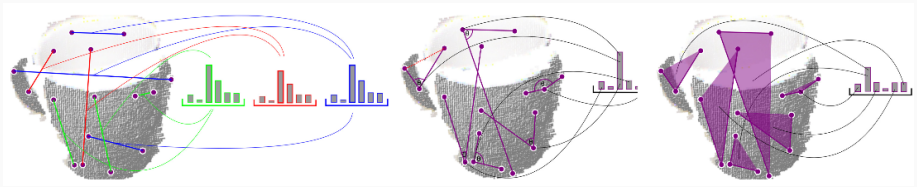


Figure 16: Calculation of the shape functions on a point cloud of a mug.[9]

[9] Image from Wohlkinger et al. [3]

Ensemble of Shape Functions descriptor

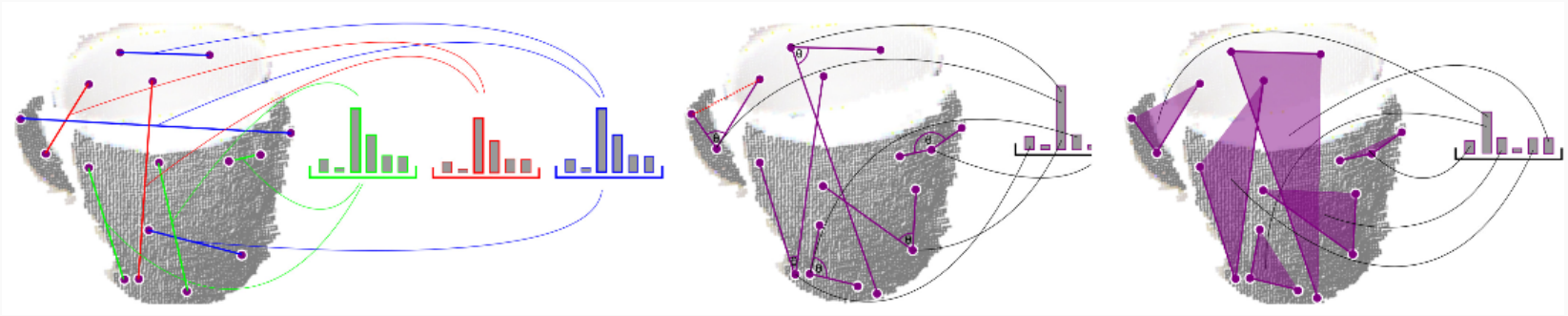


Figure 17: Left: point distance distributions; Middle: angle distributions; Right: area spanned by triplets of sampled points. Measurements are classified into "on the surface", "off surface" as well as a combination of "both", respectively depicted as green, red and blue lines in the left part of the figure [3].

Questions

- Which advantages provides the design of the ESF descriptor?

Questions

- Which advantages provides the design of the ESF descriptor?
  - ESF is invariant to translation and rotation.

Questions

- Which advantages provides the design of the ESF descriptor?
  - ESF is invariant to translation and rotation.
- Do you have a hint what could possibly go wrong during a 3D object recognition based on ESF descriptors? Why?

Training

During a training stage, objects are placed on a rotary table.

Histograms are computed for different partial views of all the objects we want to recognize and saved into *.pcd files.

A ROS node is written to summarize the training data into training_data.list and training_data.h5.

A K-d tree as a index structure is created for quick nearest-neighbor lookup (kdtree.idx).

K-d tree

A k-d tree, or k-dimensional tree, ...

- is a data structure used in computer science for organizing some number of points in a space with k dimensions.
- is very useful for range and nearest neighbor searches.

We use FLANN to perform fast approximate nearest neighbor (NN) searches in high dimensional spaces.

K-d tree

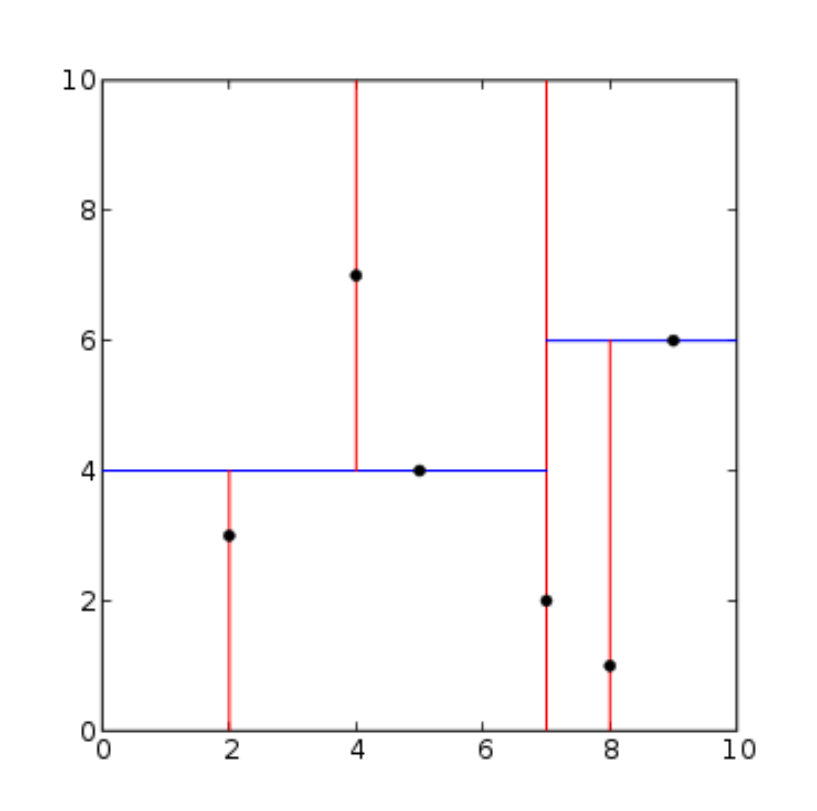


Figure 18: An example of a 2-dimensional k-d tree.[10]

[10] Image from https://en.wikipedia.org

Pose estimation

Currently, the RoboCup@Work objects are lying on its side on a table.

We get the object's **x** axis (red) from the table normal.

A Principial Component Analysis provides the eigenvector with the largest eigenvalue that is used to compute the **z** axis (blue).

The **y** axis (green) is computed by $\mathbf{x} \times \mathbf{z}$.

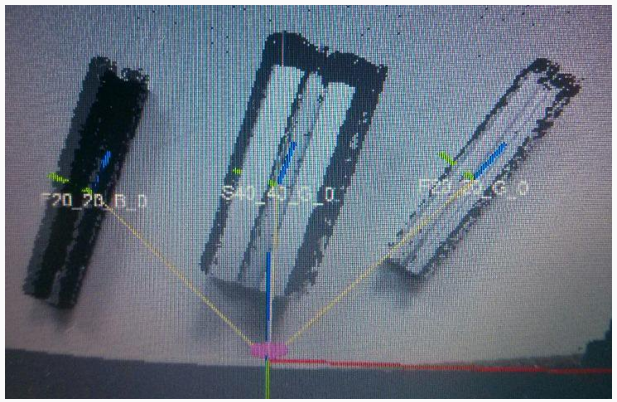
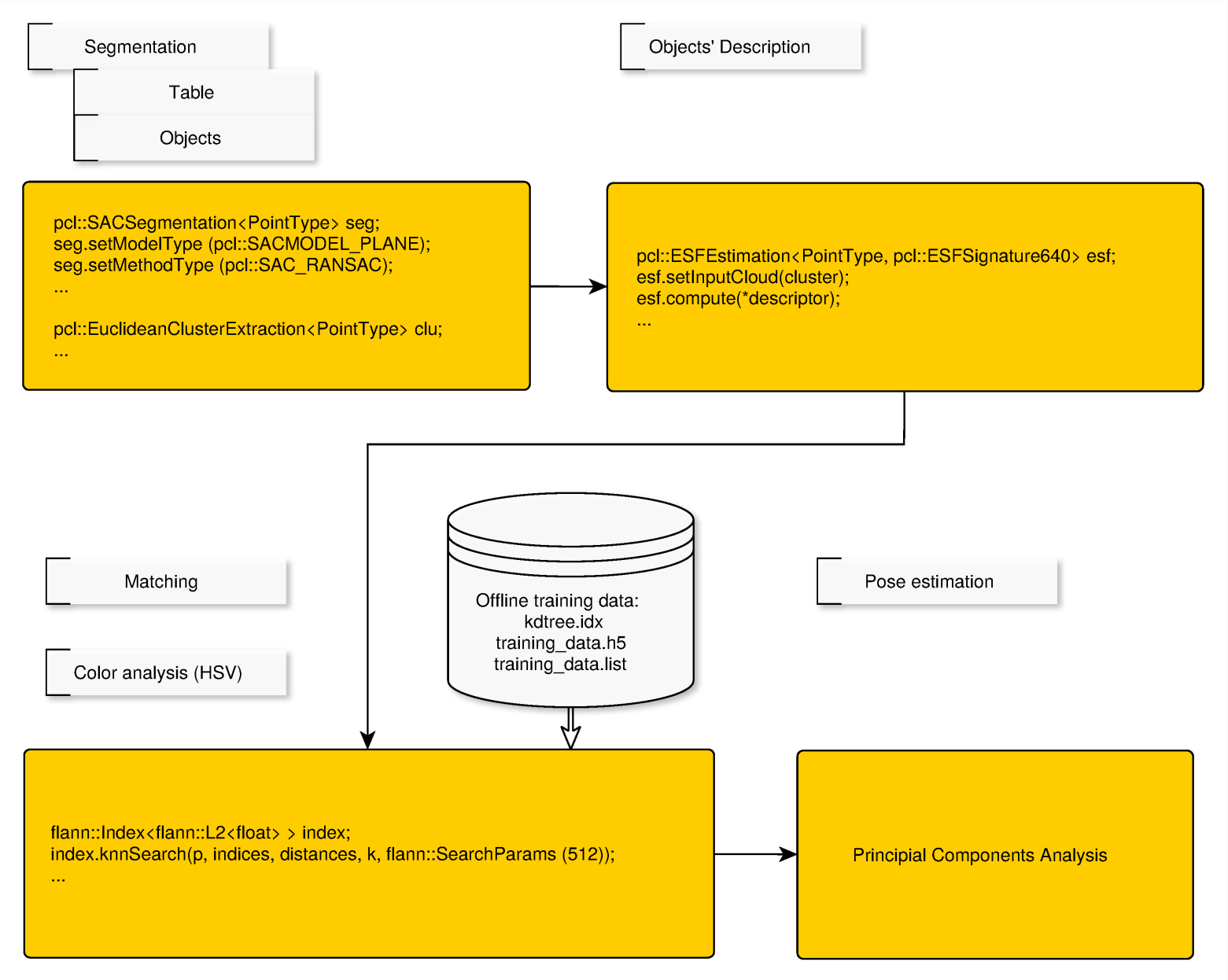


Figure 19: Objects with their coordinate frames.

Summary: Global pipeline used by team robOTTO



Questions?

References I


A. V. Mohit Gupta, Amit Agrawal and S. G. Narasimhan.
A practical approach to 3d scanning in the presence of interreflections, subsurface scattering and defocus.
International Journal of Computer Vision (IJCV), 2012.
The publication is available at http://graphics.cs.cmu.edu/projects/StructuredLight3DScanning/.


S. K. Nayar and M. Gupta.
Diffuse structured light.
Proc. IEEE International Conference on Computational Photography, 99:403–422, 2012.


W. Wohlkinger and M. Vincze.
Ensemble of shape functions for 3d object classification.
In *Robotics and Biomimetics (ROBIO), 2011 IEEE International Conference on*, pages 2987–2992, Dec 2011.