Visual detection, recognition and tracking of three-dimensional objects

Author: Bc. Marek Jakab, Supervisor: Ing. Vanda Benešová, PhD.
Slovak University of Technology, Faculty of Informatics and Information Technologies

ABSTRACT

We propose an enhanced method of 3D object description and recognition based on local descriptors using RGB image and depth information (D) acquired by Kinect v2 sensor. Our main contribution is focused on a novel local depth descriptor (DD) that includes a 3D description of the key point neighborhood. Thus defined the 3D descriptor can then enter the decision-making process.

New approach has been proposed, tested and evaluated that it deals with the object recognition system using the original SIFT descriptor in combination with our novel proposed 3D descriptor, where the proposed 3D descriptor is responsible for the pre-selection of the objects. The results show an improvement in speed of the recognition system.

CONTACT

Marek Jakab
Slovak University of Technology
Email:
marko.jakab@gmail.com

1. COLOR & DEPTH STREAM

At the beginning of the process we need to capture color and depth images of the object from the depth sensors like Kinect. In our approach we use new Kinect v2 device with better depth fidelity.

2. IMAGE SEGMENTATION

Image segmentation plays a crucial step in the process of object recognition. We use the depth image to remove the background of the object using the method of growing regions. Thanks to the segmentation, we are able to filter out undesirable features which do not belong to the object, increasing the recognition speed.

3. FEATURE DETECTION

From the intensity image of the object we are now able to extract features which will be used later during creation of descriptor vector. We use the features extracted from SIFT feature detector to create Depth Descriptor in order to save computational time. Experiments show us that those features are still reliable even for depth matching.

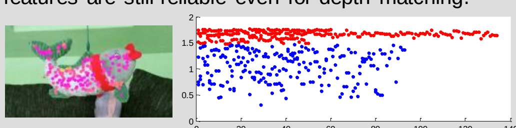


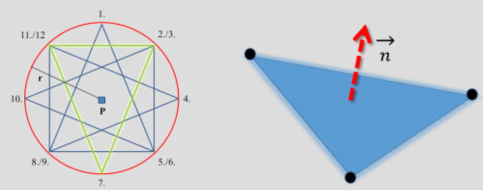
Figure: Descriptor values of flat and non-flat objects

4. DEPTH DESCRIPTOR EXTRACTION

Detected features are used to form Depth Descriptor vector. We use star pattern at each feature. Star is formed out of four triangles, where three edge points are used to obtain depth information and compute normal vector of the triangle.

Depth Descriptor is formed out of these values:

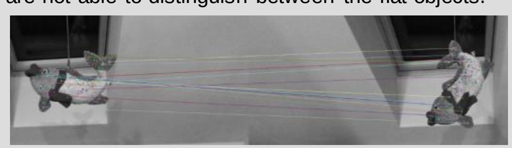
1. Average angle between normal vectors
2. Standard deviation of depth values
3. Difference of maximal and minimal depth
4. Angle between average angle and overall average angle through all normal vectors



5. FIND MATCHES (PRE-SELECTION)

Next step during the recognition process is to match the descriptors. Assuming we have stored the Depth Descriptor for our previous object (fish) we extract Depth Descriptor for newly acquired color and depth images. Now we are able to compare both descriptors. In our approach we use Euclidean distance to retrieve best matches for each descriptor value. Objects which passed pre-selection will get to the accurate recognition using the SIFT descriptor.

We can also decide, whether the object is flat. If so, we pass indices of all flat objects to the next level of recognition, as while using just depth information we are not able to distinguish between the flat objects.

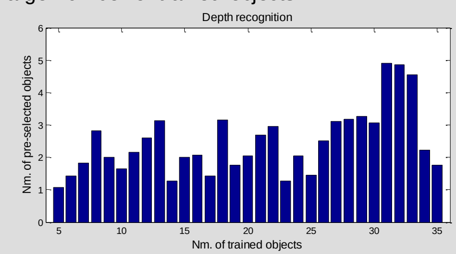


6. WRITE OUTPUT

When we recognized the object, we make an output by showing the name of the object along with the segmented image.

EXPERIMENTS

In the first experiment, we compared the number of pre-selected objects which passed to the next level of recognition. The provided figure show us that even with 4 dimensional descriptor vector we can filter out a large number of trained objects.



In next experiment, we compared the speed results of recognition with the reference model when just SIFT descriptor were used. We noticed a small (~10% to 20%) improvement increasing with the more objects trained.