Semantic Segmentation of Stereo Sequences

Otilia ZVORIȘTEANU
"Gheorghe Asachi" Technical University of Iasi,
Faculty of Automatic Control and Computer Engineering

Introduction

Objective: *object detection in image sequences*

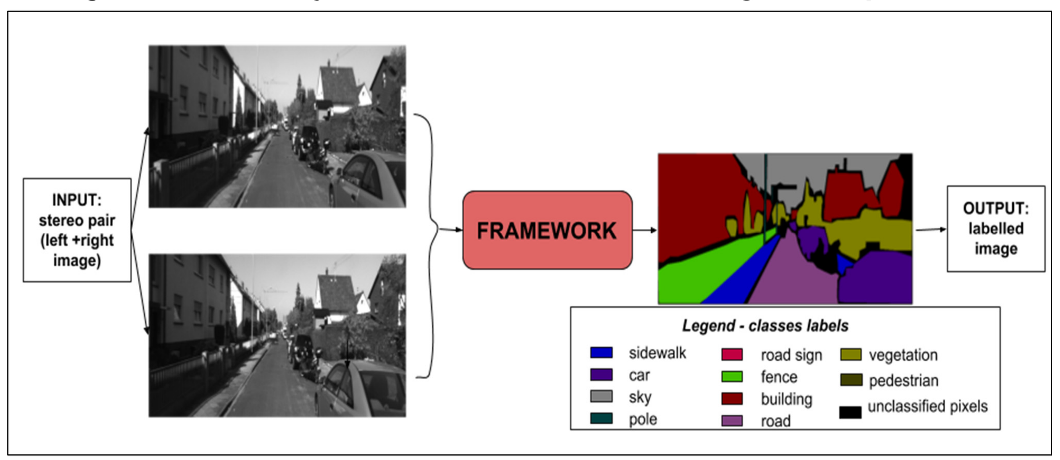

Figure 1. Semantic segmentation task

Semantic segmentation: computer vision task which aims to associate one of the pre-defined class labels to each pixel.

Keywords: *semantic segmentation, disparity, depth, point-cloud, ground-plane, Random Forest Classifier, superpixel*

Method

-use **disparity maps [1]** and **ground plane masks [2]** for **point-cloud** computation
-**extract superpixels' 3D features** from the **current reconstruction**
-**train** features using **Random Forest Classifier**
-**predict new labels** for test images

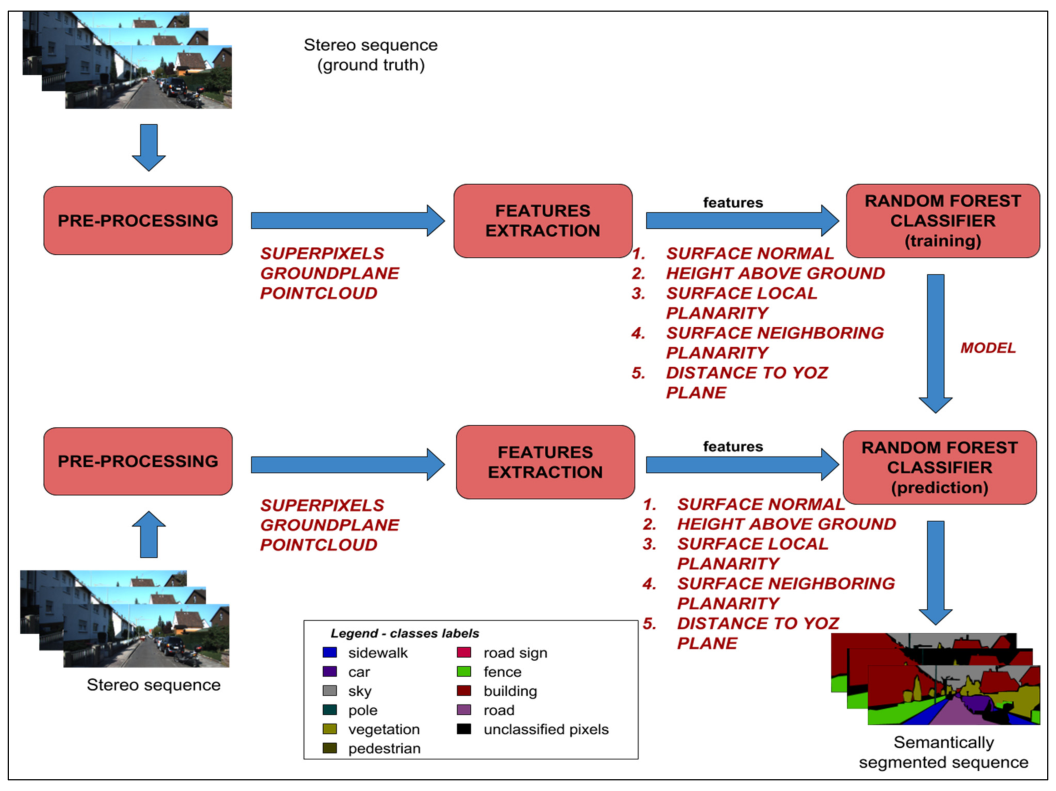

Figure 2. Framework pipeline

Pre-processing modules

I. GROUND-PLANE

Figure 3. a) Left image, b) Ground-plane mask

II. GENERIC SEGMENTATION-SUPERPIXELS

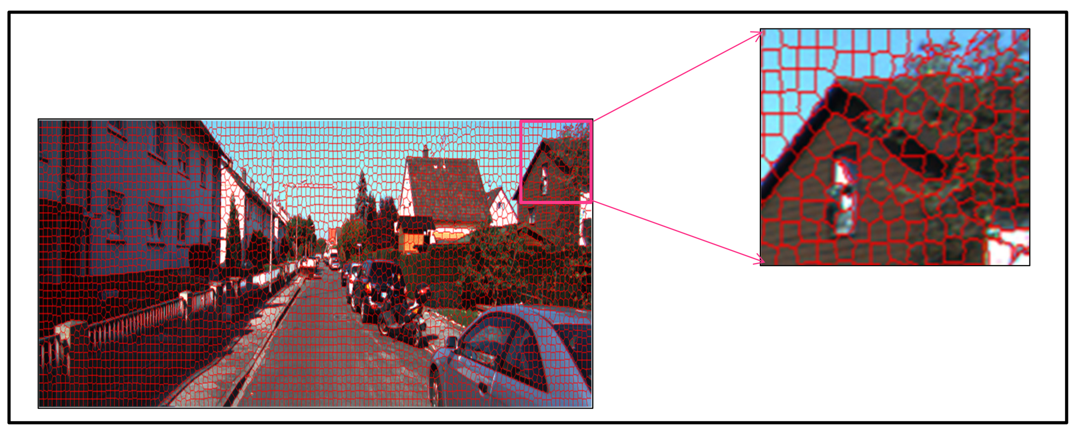

Figure 4. Generic segmentation - **superpixels**

Pre-processing modules (cont)

III. POINT-CLOUD COMPUTATION

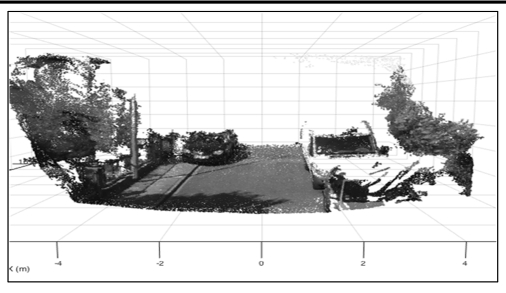

Figure 5. Point-cloud

$\mathbf{x} = (u-c_u)*depth/f$
$\mathbf{y} = (v-c_v)*depth/f$
$\mathbf{z} = f*B/depth$

where:

- **f** – focal length
- **B** – baseline
- **(c_u, c_v)** – camera central point
- **(u, v)** – current point (pixel)

***Observation**: **f**, **B** and **(c_u, c_v)** represent stereo camera parameters.*

Trained features

The following features were trained using the **Random Forest Classifier**:

1. **surface normal [3]:** fit least square plane to the superpixels 3D corresponding points
2. **height above ground [3]:** average distance to the ground
3. **local planarity [3]:** average of square distances from points to the least square plane found
4. **neighboring planarity [3]:** average difference of a superpixel's surface normal with respect to its neighbors' surface normals
5. **distance to camera path [3]:** camera path is approximated with OZ axis

Results

Tests were performed with datasets specific for two domains: **automotive** and **assistive technologies**.

I. KITTI dataset

- **58** stereo images with the corresponding **ground truth annotations**
- **11 classes** provided

II. Virtual Environment (VTE) dataset

- **280** computer generated images and their **ground truth annotations**
- **depth free errors**

TABLE I. Average accuracy on training and testing sets

	Training set	Testing set
KITTI	89.07	15.18
VTE	95.93	45.21

Results (cont)

III. Observations

-**low accuracy** on both of the testing sets
-**classes poorly** represented in images, e.g., poles (~66%), have lower accuracy than other classes better represented, e.g., buildings (~93.8%), sidewalks (~93.4%), on the training set
-**higher accuracy** on **computer generated set**

IV. Improvements

-**temporal fusion** was added and **color** was considered as a feature **[4]**

TABLE II. Average accuracy on training and testing sets after improvements

	Training set	Testing set
KITTI	90.03	76.08
VTE	95.94	92.74

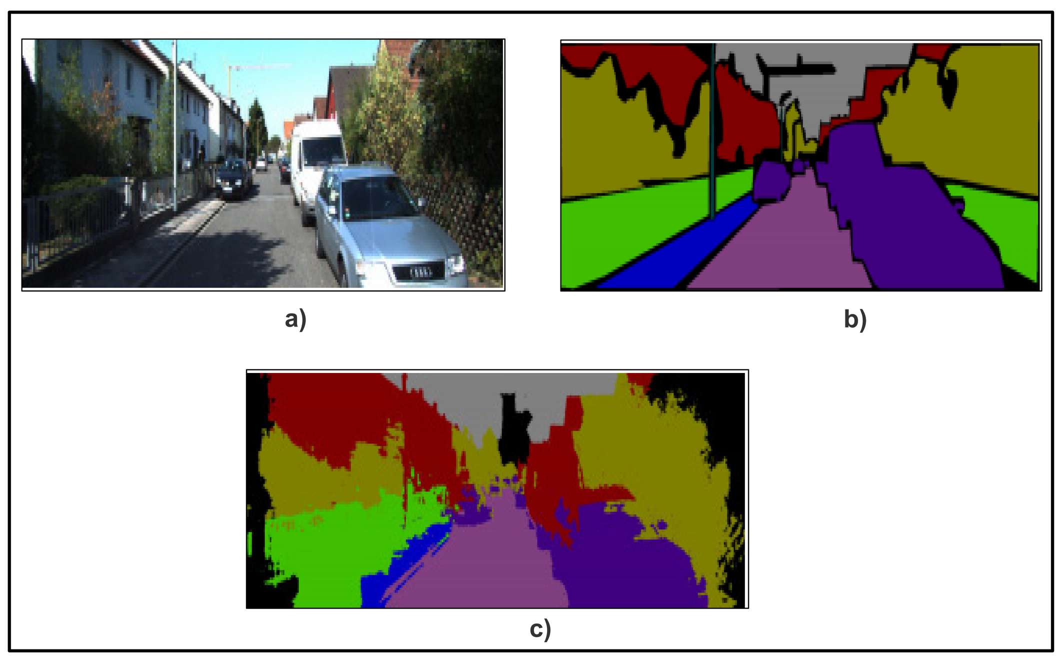

Figure 6. a) Left image, b) Annotated image, c) Prediction result

Conclusion

-to improve the results a **redesign** of the features' set was needed; **color**, was added as a feature
-**temporal fusion** significantly improved the segmentation accuracy

Future directions

-**training** and **evaluation** on larger datasets
-investigate different methods for semantic segmentation, i.e., **neural networks**
-**instance semantic segmentation**

References

[1] Andreas Geiger et al., ***Efficient LargeScale Stereo Matching***, 2010
[2] P. Herghelegiu, A. Burlacu, and S. Caraiman, ***Robust ground plane detection and tracking in stereo sequences using camera orientation,*** 2016
[3] Chenxi Zhang, Liang Wang, Ruigang Yang, ***Sematic Segmentation of Urban Scenes Using Dense Depth Maps***, 2010
[4] A. Neculai, ***Semantic Segmentation of Stereo Sequences,*** Diploma Thesis – Faculty of Automatic Control and Computer Engineering, 2016
[5] Radhakrishna Achanta et al., ***SLIC Superpixels***,2010
[6] Andreas Geiger and Philip Lenz and Raquel Urtasun, ***Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite,*** *2012*