Low-Cost Scene Modeling using a Density Function Improves Segmentation Performance

Vivek Sharma[1], Sule Yildirim-Yayilgan[2], Luc Van Gool[1,3]

[1] ESAT-PSI, KU Leuven [2] Norwegian University of Science and Technology [3]CVL, ETH Zürich

Introduction

Goal

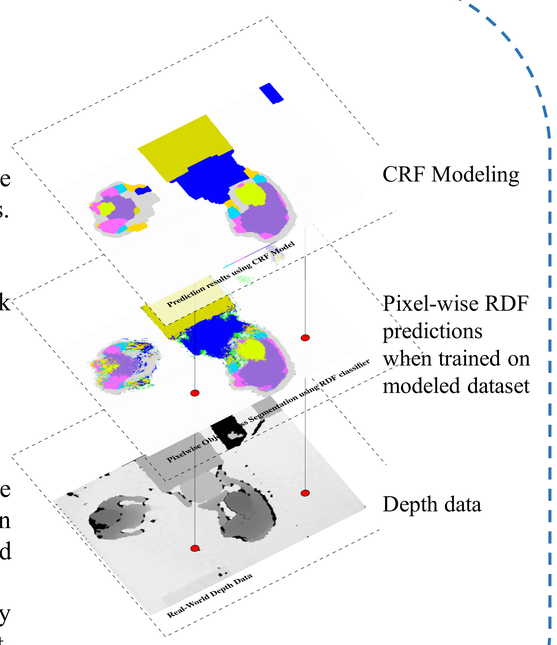
- Top-view scene analysis for safe human-robot collaboration.

Problem Statement

- In the industrial scenario humans and robots often share the same workspace posing a lot of threats to human safety issues.
- We focus on:
  - ➢ Intuitive and natural human-robot interaction.
  - ➢ Safety considerations and measures in a shared work environment.
  - ➢ Realization of cooperative process
  - ➢ Workflow optimization.

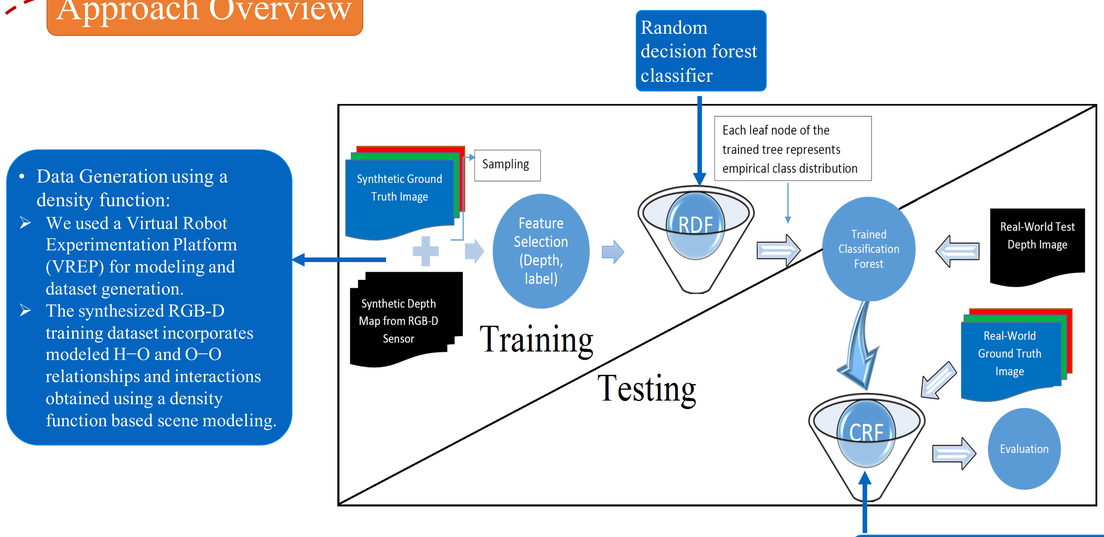
Contributions

- We model human and object interactions to learn these interactions in RGB-D data in order to improve segmentation which can be useful for safe human-robot interaction and collaboration scenarios.
- We show that synthetic data generated from a density function, that governs the human-object (H-O) and object-object (O-O) interactions, can be effectively used to train methods to achieve improved segmentation performance by ~7% in mAP and mAR over state-of-the-art methods on real-world data.



Top-view segmentation results for real-world depth data.

Approach Overview



- Data Generation using a density function:
  - ➢ We used a Virtual Robot Experimentation Platform (VREP) for modeling and dataset generation.
  - ➢ The synthesized RGB-D training dataset incorporates modeled H–O and O–O relationships and interactions obtained using a density function based scene modeling.

- The energy of the pairwise conditional random fields is minimized by alpha-expansion built on graph cuts.

Some References:

[1] Sharma et al.: Improving human pose recognition accuracy using CRF modeling. CVPR Workshops'15

[1] Sharma et al.: Efficient real-time pixelwise object class labeling for SHRC in industrial domain. ICML Workshops'15

[2] Shotton et al.: Efficient human pose estimation from single depth images. PAMI'13

[3] Jiang et al.: Learning object arrangements in 3D scenes using human context. ICML'12

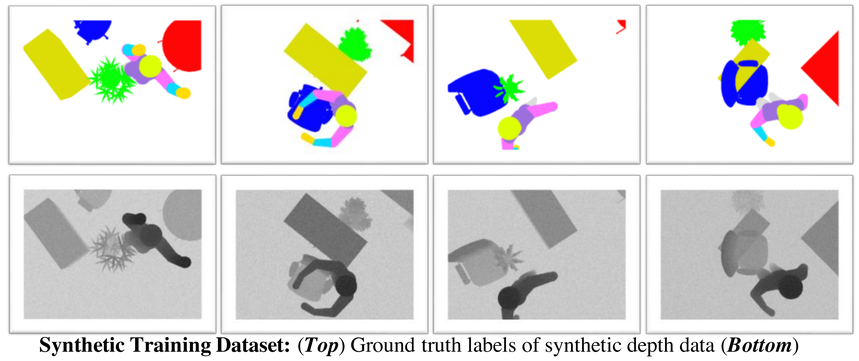
[4] Boykov et al.: Fast approximate energy minimization via graph cuts. PAMI'01

Density Function

Synthetic dataset generation using a density function

- The density function Φ capturing the context of human-object (H-O) and object-object (O-O) relationships in a scene S is defined as: $\Phi(S) = \Psi(H, O; \Theta)\Psi(O, O; \Theta)$, where Θ is threshold of preferred occlusion of boundaries.
- We chose 4 industrial objects (i.e. chair, plant, table, and storage) based on an industrial environment, and 6 localized human body-parts of the human as object classes (i.e. head, body, upper-arm, lower-arm, hand and legs).
- The density function describing the human-object and object-object relationships is defined as:

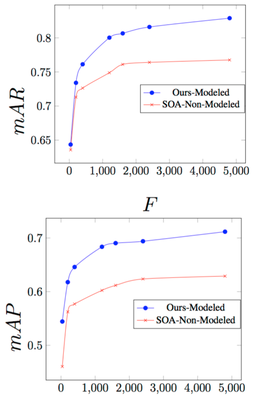
$$\Psi(\mathrm{H,O;\Theta}) = \psi(H_{height})\psi(H_{pose})\psi(H_{position})\psi(H_{orientation})\psi(O_{height})\psi(O_{position})$$
$$\psi(O_{orientation})\psi((H,O)_{\Theta})\psi\left((H,O)_{relationship}\right)$$
$$\Psi(\mathrm{O,O;\Theta}) = \psi(O_{height})\psi(O_{position})\psi(O_{orientation})\psi((O,O)_{\Theta})\psi((O,O)_{relationship})$$



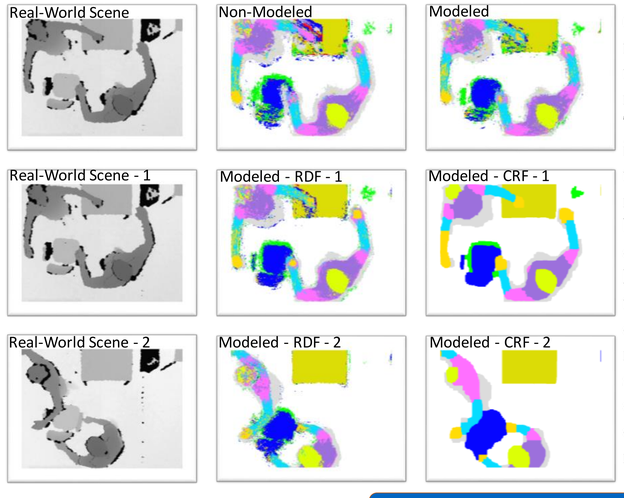
Synthetic Training Dataset: (***Top***) Ground truth labels of synthetic depth data (***Bottom***) generated using a density function with a synthetic KINECT sensor.

Experiments

F1-measure	Avg	Head	Body	UArm	LArm	Hand	Legs	Chair	Plant	Storage	Table
$CRF_{Non-Modeled}$[1,2]	0.76	0.90	0.71	0.73	0.65	0.69	0.48	0.85	0.78	0.90	0.91
$CRF_{Modeled}$	**0.84**	**0.96**	**0.84**	**0.79**	**0.70**	**0.79**	**0.52**	**0.93**	**0.90**	**0.98**	**0.97**



Comparison of the modeled and non-modeled training dataset, using mAR and mAP as a function of number of training synthetic depth frames (F).



(*Row* 1:) Prediction results for real-world test depth data using the modeled and non-modeled training dataset. The segmentation improvements can be seen in the modeled case: the misclassification around the border of the human has diminished significantly; the human hand placed on the table and the chair are classified well with reduced mislabeling. (*Row 2-3:*) shows the predictions obtained from the RDF classifier and the CRF modeling.

Please contact me at: sharma.vivek@live.in