TrimBot2020 Deliverable D8.8

Proceedings of Workshop 1

Principal authors: University of Edinburgh

ETH Zurich

University of Freiburg

University of Groningen

Document type: Report

Publicity level: Public

Deliverable due: Month 24

CONTENTS

1 3D RECONSTRUCTION MEETS SEMANTICS – 3DRMS WORKSHOP

1.1 INTRODUCTION

Over the last decades, we have seen tremendous progress in the area of 3D reconstruction, enabling us to reconstruct large scenes at a high level of detail in little time. However, the resulting 3D representations only describe the scene at a geometric level. They cannot be used directly for more advanced applications, such as a robot interacting with its environment, due to a lack of semantic information. In addition, purely geometric approaches are prone to fail in challenging environments, where appearance information alone is insufficient to reconstruct complete 3D models from multiple views, for instance, in scenes with little texture or with complex and fine-grained structures. At the same time, deep learning has led to a huge boost in recognition performance, but most of this recognition is restricted to outputs in the image plane or, in the best case, to 3D bounding boxes, which makes it hard for a robot to act based on these outputs. Integrating learned knowledge and semantics with 3D reconstruction is a promising avenue towards a solution to both these problems. For example, the semantic 3D reconstruction techniques proposed in recent years, e.g., by Häne et al., jointly optimize the 3D structure and semantic meaning of a scene and semantic SLAM methods add semantic annotations to the estimated 3D structure. Learning formulations of depth estimation, such as in Eigen et al., show the promises of integrating single-image cues into multi-view reconstruction and, in principle, allow the integration of depth estimation and recognition in a joint approach.

1.2 WORKSHOP GOALS

The goal of the organized workshop was to explore and discuss **new ways for integrating techniques from 3D reconstruction with recognition and learning**. In particular, we were interested in the following questions: How can semantic information be used to improve the dense matching process in 3D reconstruction techniques? How valuable is 3D shape information for the extraction of semantic information? In the age of deep learning, can we formulate parts of 3D reconstruction as a learning problem and benefit from combined networks that estimate both 3D structures and their semantic labels? How do we obtain feedback-loops between semantic segmentation and 3D techniques that improve both components? Will this help recover more detailed 3D structures? We hoped to gained insights into the answers to these questions through a set of invited talks and contributed papers.

A second goal of the workshop was to **raise awareness of the TrimBot2020 project**. To this end, we created a semantic reconstruction challenge based on data recorded in the Wageningen garden. More details on the challenge and its outcome are given in Sec. 2 below.

1.3 WORKSHOP PROGRAM

The workshop was held in conjunction with the International Conference on Computer Vision (ICCV) 2017, which was held in Venice, Italy. ICCV is one of the top conferences in the field of

Computer Vision. Besides multiple contributed papers, we were able to attract three world-renowned experts - Jitendra Malik (UC Berkeley), Stefan Leutenegger (Imperial College London), and Raquel Urtasun (University of Toronto / Uber ATG) - as invited speakers. Due to the high number of paper submissions, we decided to present each paper in a short spotlight presentation followed by a poster session to allow for more detailed discussions. The final event of the workshop was a panel discussion with the invited speakers. Below, we provide the program of the workshop.

The organizers of the workshop were Torsten Sattler (ETH Zurich), Thomas Brox (University of Freiburg), Marc Pollefeys (ETH Zurich), Bob Fisher (University of Edinburgh), and Radim Tylecek (University of Edinburgh).

Date: Monday, October 23rd

Place: Sala Casino (Palazzo del Casino)

- 14:00 Introduction by the organizers
- 14:05 Invited Talk 1: Jitendra Malik
- 14:35 Spotlight presentations of contributed papers
    - Carl Toft, Carl Olsson, Fredrik Kahl, ***Long-term 3D Localization and Pose from Semantic Labellings***
    - Daniele De Gregorio; Tommaso Cavallari; Luigi Di Stefano, ***SkiMap++: Real-Time Mapping and Object Recognition for Robotics***
    - Joris Guerry, Alexandre Boulch, Bertrand Le Saux, Julien Moras, Aurélien Plyer, David Filliat, ***SnapNet-R: Consistent 3D Multi-View Semantic Labeling for Robotics***
    - Bo Yang, Hongkai Wen, Sen Wang, Ronald Clark, Andrew Markham, Niki Trigoni, ***3D Object Reconstruction from a Single Depth View with Adversarial Learning***
    - Nolan Lunscher, John Zelek, ***Deep Learning Anthropomorphic 3D Point Clouds from a Single Depth Map Camera Viewpoint***
    - Gianluca Agresti, Ludovico Minto, Giulio Marin, Pietro Zanuttigh, ***Deep Learning for Confidence Information in Stereo and ToF Data Fusion***
    - Andrea Romanoni, Marco Ciccone, Francesco Visin, Matteo Matteucci, ***Multi-View Stereo with Single-View Semantic Mesh Refinement***
    - Francis Engelmann, Theodora Kontogianni, Alexander Hermans, Bastian Leibe, ***Exploring Spatial Context for 3D Semantic Segmentation of Point Clouds***
- 15:20 Invited Talk 2: Stefan Leutenegger – Models and Learning for Real-Time Robotic Spatial Perception
- 15:50 Coffee Break & Poster presentations
    - Spotlight papers
    - Extended Abstract: André B. S. Guedes, Teófilo E. de Campos, Adrian Hilton, ***Semantic Scene Completion Combining Colour and Depth: preliminary experiments***
- 16:30 Discussion of the challenge
- 16:40 Results of the challenge
- 16:50 oral presentations by challenge winners
- 17:10 Invited Talk 3: Raquel Urtasun

- 17:40 Panel discussion with invited speakers
- 18:00 Closing

1.4 WORKSHOP OUTCOME

As shown in the figure below, the workshop was very well-attended, especially during the invited talks and the panel discussion at the end of the program. The spotlight and poster sessions were also well-attended, even though the poster session was held during a coffee break. The room was packed during the final panel discussion and the audience asked many interesting questions. Overall, our expectations in terms of audience size were very much exceeded. Given the very positive feedback from the audience, we are considering holding a follow-up workshop at the European Conference on Computer Vision (ECCV) 2018 in Munich.

Pictures taken during the workshop can be found at http://trimbot2020.webhosting.rug.nl/3drms-workshop-iccv2017-gallery/.

1.5 SUBMISSIONS/PROCEEDINGS

Two types of submissions were allowed: Regular papers (8 pages excluding references) and extended abstract (4 pages including references). Regular papers were published in the workshop proceedings (which in turn were published together with the proceedings of the main conference) and were reviewed by the organizers of the workshop. Extended abstracts were not officially peer-reviewed and not published in the proceedings to allow authors to submit previews of ongoing work. Instead, each extended abstract was only checked for its fit with the topics of the workshop.

In total, we received 22 submissions (19 regular papers and 3 extended abstracts). After reviewing the papers, we decided to accept 8 regular papers (42%) and 1 extended abstract (33%). Overall, both the number and quality of the submissions was very high for a workshop. Initially, we had expected around 4-8 submissions with only around 2 high-quality papers.

The proceedings of the workshop will become available in IEEE Xplore. At the same time, the contributed papers are also available in open access at the url:
http://openaccess.thecvf.com/ICCV2017_workshops/ICCV2017_W13.py

2 CHALLENGE

Part of the workshop was a challenge on combining 3D and semantic information in complex scenes. To this end, a challenging outdoor dataset, captured by a robot driving through a semantically-rich garden that contains fine geometric details, was released. A multi-camera rig was mounted on top of the robot, enabling the use of both stereo and motion stereo information. Precise ground truth for the 3D structure of the garden has been obtained with a laser scanner and accurate pose estimates for the robot are available as well. Ground truth semantic labels and ground truth depth from a laser scan will be used for benchmarking the quality of the 3D reconstructions.

The data set, annotations and evaluation tools were made publicly available for benchmark purposes at the url: https://gitlab.inf.ed.ac.uk/3DRMS/Challenge2017

2.1 CHALLENGE SUBMISSIONS

Given a set of images and their known camera poses, the goal of the challenge participants was to create a semantically annotated 3D model of the scene. At the end of June 2017 we made both training and test sequences available and publicly announced a call for submissions of a semantically annotated 3D triangle mesh from the test sequence. The submission of results closed in the beginning of September 2017 and we received two submissions:

- Yuichi Taguchi and Chen Feng (Mitsubishi Electric Research Labs, USA): Semantic 3D reconstruction using depth and label fusion.
- Joris Guerry et al. (ONERA, France): SnapNet-R: consistent 3D multi-view semantic labeling for robotics (workshop paper).

Submissions were subsequently evaluated against the test part ground-truth 3D model and image annotations, as detailed in the appended report. We also compared the results to the state-of-the-art methods of project partners: Colmap 3D mapping (ETHZ) and SegNet-derived 2D semantic segmentation (UVA), which are also employed in TrimBot system. The comparison based on geometric accuracy, completeness and semantic quality was announced in October 2017 at the workshop, by Radim Tylecek, and challenge participants subsequently presented their approaches.

2.2 REPORT

We attach a report of the challenge results to the present document.

