PH.D. SCHOOL IN INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

Final Report

September 30, 2017

Student name: Ludovico Minto

Cycle: XXX

Curriculum: Information Science and Technology

Thesis title: Deep Learning for Scene Understanding with Color and Depth Data

1 Courses, Conferences and Mobility

Courses for PhD students

- Introduction to Machine Learning (A. Sperduti).
- Applied Linear Algebra (P. Picci).
- Real Time Systems and Applications (G. Manduchi).
- Statistical Methods (L. Finesso).
- Bayesian Machine Learning (G. M. Di Nunzio). (attended only)
- Applied Machine Learning in Biomedicine (E. Grisan). (attended only)
- Introduction to GPUs and Parallel Computing (J. Pantaleoni). (attended only)

Summer schools, short courses, tutorials

- International Computer Vision Summer School (ICVSS), Scicli (Italy), July 2016.
- Advanced Topics in Machine Learning, Lyngby (Denmark), August 2015.

Seminars

- *Networks: Brain, Health and Society*, M. Corbetta, University of Padova, DEI, Padova (Italy), September 2016.
- *Dialogue on the Quantum Revolution*, S. Lloyd, DEI, University of Padova, Padova (Italy), July 2016.
- *Computational Thinking, Inferential Thinking and Data Science*, M. I. Jordan, University of Padova, DEI, Padova (Italy), June 2016.
- *Do Brains Compute?*, R. Sepulchre, University of Padova, DEI, Padova (Italy), June 2015.

Participation to international conferences and workshops

- 3D Reconstruction Meets Semantics Workshop, International Conference on Computer Vision (ICCV), Venezia (Italy), October 2017: Spotlight and paper presentation (C2). (scheduled)
- Geometry Meets Deep Learning Workshop (GMDL), European Conference on Computer Vision, Amsterdam (Netherlands), October 2016: Paper presentation (C4).
- Sixth GTTI Thematic Meeting on Multimedia Signal Processing (MMSP), Alleghe (Italy), January 2016: Demonstration (A. Memo, L. Minto, P. Zanuttigh, "Head-Mounted Gesture Controlled Interface for Human-Computer Interation").
- International Workshop on Understanding Human Activities Through 3D Sensors (UHA3DS), IEEE Conference on Automatic Face and Gesture (FG), Lubljana (Slovenia), May 2015: Oral presentation (C6).

Other learning activities

- Bachelor thesis co-supervisor, *Segmentazione Semantica di Dati RGBD Attraverso Reti Neural Convoluzionali*, G. Esposito, 2017.
- Webmaster, Sixth GTTI Thematic Meeting on Multimedia Signal Processing 2016.
- Master thesis co-supervisor, *Multi-Modal Head Mounted Vision System with Gesture Interface*, A. Memo, 2015.
- Master thesis co-supervisor, *Gesture Recognition Exploiting Deep Learning Techniques*, M. De Benedet, 2015.
- Bachelor thesis co-supervisor, *Exploiting Finger Segmentation for Gesture Recognition*, R. Aziz, 2015.
- Laboratory assistant for the course Image and Video Analysis (P. Zanuttigh), 2015 (laboratory sessions set-up, homework projects supervision).

Mobility periods

Visiting student at the Queen Mary University of London, London (United Kingdom), December 2016 – June 2017, Safe landing site detection in indoor environments.

2 Research Activity

2.1 Introduction

As a Ph.D. student at the Electrical and Computer Engineering Department (DEI) of the University of Padova I had the chance to grow and deepen many of my interests about computer vision and machine learning, exploring different aspects of both fields. In particular, I focused on the application of deep learning techniques to the solution of various problems in the computer vision field, going from well-known yet challenging problems such as scene segmentation and semantic labeling [J1, C4] to more recent ones such as 3D shape classification for object recognition [C1, C3] and stereo-ToF data fusion [C2].

2.2 Background

A series of rapid advancements both on the hardware side as well as in terms of available algorithmic tools has recently opened the path the implementation of novel and effective solutions to several classical computer vision problems.

On one hand, the introduction of depth sensors in the consumer market segment has made possible the acquisition of 3D data at a very low cost, allowing to overcome many of the limitations and ambiguities that typically affect color information. At the same time, significant improvements in the GPU computing have increased the computational power available at the hands of researchers by a considerable factor, enabling the test of time-consuming approaches on even large scales.

On the other hand, the development of more powerful machine learning algorithms including recent deep learning techniques has allowed to exploit the enormous amount of data nowadays available, and perform a variety of different tasks without the need of designing from scratch new sets of hand-crafted tools and criteria every time a specific task and input data are encountered.

2.3 Contributions

From a higher perspective, my research contribution goes in two directions.

First, it shows how deep learning techniques can be successfully applied to several computer vision problems, avoiding the design of excessively sophisticated ad-hoc methods while achieving results that are comparable or even higher than those obtained with state-of-the-art approaches.

Second, it shows the importance of depth data as a valuable source of information complementary to color data, experimentally demonstrating how jointly exploiting both types of data can lead to an improved performance.

At a lower level, a series of solutions have been implements and tested to tackle various tasks mainly related to the general scene understanding problem. More details are given in the following.

2.4 Problems and proposed approaches

Scene understanding from visual data is a long-term problem in computer vision. Although it has been subject of many research works along the years, no complete solution has been provided yet, leaving space to further improvements.

Within this context, I mainly focused on the segmentation and semantic labeling of indoor environments [J1, C4], proposing a novel scheme that makes use both of deep learning techniques and geometric cues to segment a single RGB-D image and assign each pixel to a semantic class. The approach uses normalized cuts spectral clustering to perform an initial over-segmentation, while NURBS surfaces are fitted on the segments. Color and geometry data together with surface fitting parameters are fed to a Convolutional Neural Network (CNN) trained for the semantic labeling task. Then, an iterative merging algorithm recombines the output of the over-segmentation into larger regions matching the various elements of the scene. Couples of adjacent segments with higher similarity according to the CNN features are chosen as candidates to be merged and the surface fitting accuracy is used to detect which couples of segments really belong to the same surface. Finally, a semantic labeling is obtained by combining the CNN features with the segmentation output.

In [C1, C3] I presented a solution to the 3D shape classification problem for object recognition, again exploiting deep learning and 3D data to achieve remarkable results on the Princeton ModelNet10 and ModelNet40 datasets. In particular, a multi-branch CNN is trained to classify a given 3D object model by taking as input three different data representations, namely a set of depth maps obtained by rendering the object from different views, a set of parameters describing the curvature of NURBS surfaces fitted over the object and a set of volumetric descriptors.

Finally, a deep learning approach has been used underneath the stereo-ToF fusion scheme described in [2]. Here, the confidence maps required by a fusion algorithm are computed using a 6-layer CNN trained to provide an estimate of the reliability of each data source at each pixel location.

Key topics Deep learning, depth data, scene understanding, segmentation, semantic labeling, 3D shape recognition, data fusion.

3 Publications

Publications on international journals

J1 G. Pagnutti, L. Minto, P. Zanuttigh, *Segmentation and Semantic Labelling of RGBD Data with Convolutional Neural Networks and Surface Fitting*, IET Journal Computer Vision, 2017. (accepted)

Conference proceedings

C1 L. Minto, P. Zanuttigh, G. Pagnutti, *Deep Learning for 3D Shape Classification based on Volumetric Density and Surface Approximation Clues*, International Conference on Computer Vision Theory and Applications (VISAPP), Madeira (Portugal), January 2018. (submitted)

C2 G. Agresti, L. Minto, G. Marin, P. Zanuttigh, *Deep Learning for Confidence Information in Stereo and ToF Data Fusion*, 3D Reconstruction Meets Semantics Workshop, International Conference on Computer Vision (ICCV), Venezia (Italy), October 2017. (accepted)

C3 P. Zanuttigh, L. Minto, *Deep Learning For 3D Shape Classification from Multiple Depth Maps*, IEEE International Conference on Image Processing (ICIP), Beijing (China), September 2017. (accepted)

C4 L. Minto, G. Pagnutti, P. Zanuttigh, *Scene Segmentation Driven by Deep Learning and Surface Fitting*, Geometry Meets Deep Learning Workshop, European Conference on Computer Vision (ECCV), Amsterdam (Netherlands), October 2016.

C5 A. Memo, L. Minto, P. Zanuttigh, *Exploiting Silhouette Descriptors and Synthetic Data for Hand Gesture Recognition*, Smart Tools and Apps in Computer Graphics (STAG), Verona (Italy), October 2015.

C6 L. Minto, G. Marin, P. Zanuttigh, *3D Hand Shape Analysis for Palm and Fingers Identification*, International Workshop on Understanding Human Activities Through 3D Sensors (UHA3DS), IEEE Conference on Automatic Face and Gesture (FG), Ljubljana (Slovenia), May 2015.

C7 G. Pozzato, S. Michieletto, E. Menegatti, F. Dominio, G. Marin, L. Minto, S. Milani, P. Zanuttigh, *Human-Robot Interaction with Depth-Based Gesture Recognition*, Real Time Gesture Recognition for Human-Robot Interaction Workshop, International Conference on Intelligent Autonomous Systems (IAS), Padova (Italy), July 2014.

Other publications (books, book chapters, patents)

B1 P. Zanuttigh, G. Marin, C. Dal Mutto, F. Dominio, L. Minto, G. M. Cortelazzo, *Time-of-Flight and Structured Light Depth Cameras*, Springer International Publishing, 355 p., ISBN 978-3-319-30971-2, June 2016.

B2 L. Nanni, A. Lumini, L. Minto, P. Zanuttigh, *Face Detection Coupling Texture, Color and Depth Data*, Advances in Face Detection and Facial Image Analysis, Springer, April 2016.

Student signature — Date: 29/09/2017

Supervisor signature — Date: 29/9/17