Computer Science Department
Faculty of Computer and Information Sciences
Ain Shams University

Developing a Method for 3D Scene Understanding Using Image Sequence

Thesis submitted as a partial fulfillment of the requirements for the degree of Master of Science in Computer and Information Sciences

By

Islam Ibrahim Fouad Ahmed

Teaching Assistant at Computer Science Department,
Faculty of Computer and Information Sciences,
Ain Shams University

Under Supervision of

Prof. Dr. Mostafa Gadal-Haqq M. Mostafa
Professor in Computer Science Department,
Faculty of Computer and Information Sciences,
Ain Shams University

Dr. Sherine Rady Abdel Ghany
Assistant Professor in Information Systems Department,
Faculty of Computer and Information Sciences,
Ain Shams University

2017

Acknowledgment

I acknowledge my deep gratitude to ALLAH for providing me with the strength to complete this work on a level that I hope will please the reader.

I wish to express my sincere and heartfelt appreciation to my supervisor; Prof. Mostafa Gadal-Haqq for his supervision, resourceful ideas and suggestions that helped in enhancing my work.

Special thanks are due to my supervisor Dr. Sherine Rady for her continuous guidance and encouragement in achieving this work.

Without their assistance and dedicated involvement in every step throughout the process, this thesis would have never been accomplished. I would like to thank them very much for their support and understanding.

Finally, with all my appreciation and love that no words can describe, my deepest gratitude goes to my family for their love, prayers and endless support.

Abstract

Indoor scene understanding is a challenging problem in computer vision. To achieve an accurate solution for this task, a model that can exploit discriminating information between different scene categories and objects is necessary.

This thesis presents a framework for scene understanding which includes several components of learning models, segmentation, object recognition and tracking. A comprehensive study for supervised learning models for recognizing indoor scenes is presented. The study compares between several "Shallow Learning" models against the recent approach "Deep Learning". Furthermore, the robustness of methods is tested against environment changes such as: contrast degradation, additive blurring and additive noise.

A segmentation method is proposed for object recognition which relies on depth for accurate segmentation preprocessing before applying learning models. The process is applied on both RGB and depth images to produce two segmented images. Finally, the result of both segmented images is combined.

For object recognition, a hierarchical object recognition scheme is proposed based on multiple instances

of shallow learning models. The first level in hierarchy classifies objects based on the scene category, while the second level classifies the objects based on their occupied area, and the third and last level classifies the objects based on Histogram of Oriented Gradient descriptor of each object.

The recognized objects are tracked in image sequence by extracting Shi Tomasi's good features, then the new location for these extracted features are to be located in the next image using pyramid Lucas and Kanade tracker. Finally, a method for defining the relation between recognized objects in the scene is proposed.

All the proposed methods in the framework have been tested on the standard benchmark MIT-indoor datasets. Four experiments are presented: the first experiment compares Shallow vs Deep Learning. Experiments shows that Deep learning models outperform the shallow learning models by a huge gap with respect to classification accuracy, especially the deep architecture called "VGG-16 net" outperforms all techniques. The additional step of including a fine-tuning for a pre-trained "VGG-16 net" proved to be highly significant in improving the classification accuracy that reached 85.43%, and at a cost of 20% reduction savings in training time.

The second experiment tests the robustness of presented learning models. The HOG-SVM shows outstanding robustness against contrast degradation and blurring more than deep learning and less time and space, despite the lower accuracy performance.

The third experiment tests the proposed segmentation algorithm. Results show effectiveness with respect to average execution time that reaches 11sec per image, which is faster than other algorithms in related work. Experimentation with the RGB-D dataset consisting of 1200 images “NYU V2” of indoor scenes showed that 73.41% are correctly segmented images.

The Fourth experiment shows that the proposed hierarchical classification reaches 65.73% classification accuracy. The result of this hierarchical classification is used for training each scene category in the multiple deep learning model VGG-16 net. The achieved accuracy for scene recognition is 89.24%. Performing Tracking over an image sequence for a sequence of 10 frames proved to improve the speed performance of the indoor scene understanding framework by ten times, with the process applied on first frame only and then only on the new part of scene including the new objects.

Keywords: Scene understanding, image recognition, supervised learning, shallow learning, deep learning, RGB-D image, image segmentation, object recognition, tracking.

List of Publications

1- Fouad, I. I., Rady, S., & Mostafa, M. G., "Indoor Scene Classification: A Comparative Study of Feature Detectors and Local Descriptors," *Proceedings of the 10th International Conference on Informatics and Systems*, 2016, pp. 215-221.

2- Fouad, I. I., Rady, S., & Mostafa, M. G., "Enhancing RGB-D Image Segmentation using Edge Detection and Morphological Operations" *Proceedings of the 12th International Conference on Computer Engineering and Systems ICCES, 2017*, pp. 353-358.

Table of Contents

List of Figures