UNIVERSITY OF AMSTERDAM

Human Activity Understanding for Robot-Assisted Living

N. Hu

Summary

The thesis presents research on the understanding of human activities in the context of home companion robots. Global aging results in a shortage of caregivers, and companion robots can help alleviate the problem. At home, robots can be deployed continuously and provide daily assistance, keep watch, and relay alarms in case of accidents. All these require the companion robots to be able to sense, perceive, and understand human activities.

This thesis explores the use of multiple sensors, including an RGB-D sensor on the robot, overhead cameras on the ceiling, and small unobtrusive binary sensors that are fixed to the environment for human activity understanding. To be able to capture the large variations within each of the action classes, we propose a model that uses a sequence of latent variables which can be implicitly learned without human annotation. The results show that these latent variables can successfully capture different ways of performing actions, and outperform the state-of-the-art by over 5% percentage points in both precision and recall.

In the research, it turned out that human annotators may disagree, and that the labels they provide may contain noise. Therefore, we propose a model that incorporates the uncertainty of labeling, and introduce learning methods that can use the noisy labels for finding the optimal model parameters. The results show that when noise is present in the labels, the proposed model performs significantly better ($p < 0.05$) than the state-of-the-art.

The above two models both focus on recognizing actions, not activities. But intuitively actions and activities are tightly coupled, and knowing one of them is useful to infer the other. Therefore, thirdly, we propose to use a hierarchical model that can encode the activity labels with different complexity. In this model, the activity and action nodes are interconnected, which allows the joint estimation of both activity and action labels. The results show that by using joint modeling of actions and activities, the F-score is increased by 3 percentage points compared with the state-of-the-art, with an increase of 2.4 percentage points in terms of precision and 3.6 percentage points in terms of recall. By incorporating the layer of action labels, we can see significant improvement ($p < 0.05$) in terms of F-score when compared with the single-layered approach on activity recognition.

As in interaction it is important to not only recognize what the other is doing but also what they intend to do, in our fourth contribution we propose a novel algorithm that can predict upcoming actions. This is particularly important when the robot needs to give pro-active assistance before an actual action is complete. We incorporate a human kinematic model for estimating the cost of reaching objects, and these costs are used as features for predicting future actions. The results show an improvement of 3 percentage points in F-score compared with the state-of-the-art.

Finally, analyzing human activities and actions is quite an expensive process, and it is often limited by what the robot can sense. It often needs to navigate so that the target

person is within the range of the RGB-D sensor, sometimes leading to ineffective behavior. In the final chapter, we propose to use overhead cameras for detecting and tracking persons in the room, which can be used to efficiently approach the person for tasks such as person identification, activity recognition, and intent prediction. By fusing the data from the robot sensor and the ambient sensor, the proposed approach performs the tested tasks three times faster than using only the robot sensors.

In summary, this thesis thoroughly investigated the problem of understanding human activities, at different levels of granularity and taking into account both the variability in activities and annotator disagreement. It provides important improvement over the state of the art, and provides theoretically solid understanding to a very practical problem.

Acknowledgements

Looking back the past four years, I am very grateful for all the help that I have received, without which this would not have been possible.

First of all, I want to thank my supervisors Prof. Ben Kröse and Dr. Gwenn Englebinne. My connection to Ben and Gwenn dates back to six years ago when I was still a Master student. Gwenn was the first lecturer I met in the class, and it was the machine learning course that he taught brings me to this field and equipped me with practical skills so that I can try different approaches on my own. I met Ben when he was presenting his robots in an open lecture. It was his passion upon robotics that deeply affected me and has motivated me all along the way. They have both been very kind to students but very direct in giving feedback. They are also very open for discussions, and they encourage me to speak out different opinions. I feel very lucky to have Ben and Gwenn as my supervisors.

I am very grateful to have Prof. James Crowley, Prof. Vanessa Evers, Prof. Frans Groen, Prof. Max Welling, Prof. Marcel Worring, and Dr. Henri Bouma serving as my committee members. I would like to thank you all for taking your time to read through my thesis and attend my PhD defense.

I thank all the colleagues from the AMLab for the help and useful discussions. I also thank all my collaborators in the Accompany project. I particularly want to thank Bas Terwijn for helping me in solving technical issues at the early stage of my research. We have often worked together till night on building robots, calibrating cameras, collecting data, and talking nonsense.

I also want to thank my collaborators and friends: Zhongyu Lou, Zhaochun Ren, Yuxun Zhou and Aaron Bestick. We have spent a lot of fun time together in working on exciting projects, reading papers, brainstorming new ideas, and trying different approaches. I thank my ex-housemates: Qingwen Chen, Zhenhao Li, Fan Wang, and Xueying Ma. It is great to know you guys. There has been so many cheerful moments that I will cherish for my whole life. I also want to thank my friends Wouter Josemans, Bo Shi, Zhe Jiang, Zhangzheng Zheng, Fei Sun for their support to my family.

Finally, I would like to thank my wife Yu Tian and our parents for their unreserved support. Also thank little Evelyn for all the love and happiness that she has brought to our family.

Ninghang Hu
Amsterdam, July 10, 2016