AI-MAS Group

Computer Science
& Engineering
Department

Human Activity Recognition using Robot-Mounted RGB(D) Cameras

Coordinator
conf. dr. ing. Irina Mocanu
drd. ing. Mihai Trăscău (mihai.trascau@cti.pub.ro)

No. Students
2

1. Context

Recognizing human activities has been an important research topic which has generated significant interest throughout the years. Understanding what the user is doing is the first prerequisite when building a system intended to assist him. It allows building knowledge about when to activate certain actuators in the environment, when to interact with the user, what are the most probable next inputs of the user or even for detecting dangerous situation (e.g. the user tripped and fell).

Using video cameras to capture RGB data has been a common way of recognizing activities [1, 2, 3, 4]. For scenarios where the imaging sensors requires to be more robust, depth cameras have been used and the activity recognition task is performed using RGBD data directly [5, 6, 7] or by using the extracted human skeleton [8, 9, 10, 11].

Although they perform well and state of the art methods yield good results, most of the previous work has been based on scenarios with several fixed assumptions. Most of them are usually *single-user* but that is not always the case, especially in more social environments. Moreover, most of the solutions that do tackle situations with multiple users do so by considering *only joint actions* (e.g. shaking hands, hugging). In general, only *fixed cameras* are considered and these are usually mounted on walls or on higher poles. Since the cameras are fixed then the implication is that the user must perform his entire action in front of the camera such that *the action is always fully performed in view*. This set of conditions does limit the number of scenarios where previously described methods may be deployed.

In our approach, we consider a robotic platform consisting of a either a humanoid robot [1] or a mobile manipulator robot [2] which are equipped with both RGB and depth imaging sensors. The task of performing activity recognition for the users in the robot's environment presents us with several challenges that need to be addressed:

- the robot is generally deployed in *multi-user* scenarios
- when analyzing the input data we must consider *egomotion*
- the robot should *try to keep in focus the activity of a user* even if this means moving the camera or event itself around
- training data from the particular viewpoints that the robots have is limited

Starting from the challenges described above and from previous work, including research resources on activity recognition from a robot perspective [12, 13, 14] we will implement a pipeline adapted for the ROS framework [3].

2. Objectives

Our objective is to obtain a robust, state-of-the-art ROS pipeline for performing human activity recognition using RGB/RGB-D data from the robot-mounted cameras. At the end of our research we must make ROS topics available and compatible for integrating in larger projects.

The project has several high-level tasks that need to be done in order to successfully finalize the project:

- State of the Art study
    - get up to speed with Computer Vision techniques related to the topic (mostly Deep Learning)
    - split research based RGB and depth based methods
    - compile a general overview of available research on the topic
    - compile a more in-depth analysis of the most interesting work
- Gathering and analyzing available datasets
    - search and provide a general overview on popular datasets on the topic
    - identify resources more tailored to our specific needs
    - investigate possibility to combine data from multiple sources
- Implementing the training pipeline
    - get acquainted with PyTorch [4]
    - implement data preprocessors, data loaders and augmentation methods
    - implement the actual training pipeline in PyTorch
- Testing baselines
    - use or adapt previous research in order to build a baseline
- Implementing the processing pipeline
    - get acquainted with ROS
    - implement topics based on available algorithms
    - integrate topics within a larger project
    - test integrations on live robots
- Exploratory prototyping
    - search for architectures or techniques that improve the baselines
    - test viability of prototype on live robot conditions

These tasks correspond, more or less, to the project's milestones. Keep in mind, many of the tasks are parallelizable and this will be fully exploited when managing the project.

The proposed work is based on active projects running in the laboratory and their contributions will be added to these projects. The projects require teamwork and we do adopt several project management tools and procedures in order to coordinate.

3. Required and Learned Skills

Requirements

- should be comfortable with Python
- ML knowledge is appreciated
- proactive mindset
- open for working in teams

Skills learned:

- working with PyTorch
- working with ROS
- implementing and testing a machine learning algorithms
- implementing a ROS pipeline on a live robot

[1] https://www.softbankrobotics.com/emea/en/robots/pepper
[2] http://tiago.pal-robotics.com/
[3] http://www.ros.org/
[4] https://pytorch.org/

References

[1] Ji, Shuiwang, et al. "3D convolutional neural networks for human action recognition." *IEEE transactions on pattern analysis and machine intelligence* 35.1 (2013): 221-231.
[2] Donahue, Jeffrey, et al. "Long-term recurrent convolutional networks for visual recognition and description." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.
[3] Shi, Yemin, et al. "Sequential deep trajectory descriptor for action recognition with three-stream CNN." *arXiv preprint arXiv:1609.03056* (2016).
[4] Berlin, S. Jeba, and Mala John. "Human interaction recognition through deep learning network." *Security Technology (ICCST), 2016 IEEE International Carnahan Conference on*. IEEE, 2016.
[5] Hu, Jian-Fang, et al. "Jointly learning heterogeneous features for RGB-D activity recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.
[6] Jalal, Ahmad, et al. "Robust human activity recognition from depth video using spatiotemporal multi-fused features." *Pattern recognition* 61 (2017): 295-308.
[7] Liu, Zhi, Chenyang Zhang, and Yingli Tian. "3d-based deep convolutional neural network for action recognition with depth sequences." *Image and Vision Computing* 55 (2016): 93-100.
[8] Liu, Jun, et al. "Skeleton-based action recognition using spatio-temporal LSTM network with trust gates." *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2017).
[9] Cippitelli, Enea, et al. "A human activity recognition system using skeleton data from RGBD sensors." *Computational intelligence and neuroscience* 2016 (2016): 21.

[10] Yan, Sijie, Yuanjun Xiong, and Dahua Lin. "Spatial temporal graph convolutional networks for skeleton-based action recognition." *arXiv preprint arXiv:1801.07455* (2018).

[11] Thakkar, Kalpit, and P. J. Narayanan. "Part-based Graph Convolutional Network for Action Recognition." *arXiv preprint arXiv:1809.04983* (2018).

[12] Xia, Lu, et al. "Robot-centric activity recognition from first-person rgb-d videos." *Applications of Computer Vision (WACV), 2015 IEEE Winter Conference on*. IEEE, 2015.

[13] Gori, Ilaria, et al. "Multitype activity recognition in robot-centric scenarios." *IEEE Robotics and Automation Letters* 1.1 (2016): 593-600.

[14] Gori, Ilaria, et al. "Multi-Type Activity Recognition from a Robot’s Viewpoint." *Proceedings of the 26th International Joint Conference on Artificial Intelligence*. AAAI Press, 2017.