PEOPLE TRAJECTORY PREDICTION APPLIED ON SOCIAL ROBOTICS SCENARIOS

Alexandra Ștefania Ghiță[1], Adina Magda Florea[2], Mihai Nan[3], and David-Traian Iancu[4]

Social robotic platforms became a recent trend in research. More and more systems and techniques are developed considering the possible robotic applications that can help people in their daily activities. Computer vision methods tend to focus on real-time scenarios, with approaches that can generate fast and precise results for robots interacting with people. The prediction of people trajectories represents the capacity of a system to estimate the trajectory followed by a person considering their previous steps. The integration between a social robotic application with a people trajectory prediction module can significantly improve the behaviour of a robot, by providing more information, for more robust behaviours. This research paper presents the integration of a people trajectory prediction module into an existing socially robotic framework. It details the advantages of combining the information generated by a trajectory prediction module with the capabilities of a robot by presenting experimental performed scenarios.

Keywords: social robots, people trajectory prediction, robotic framework, computer vision

1. Introduction

Social robotic applications are a recent trend in the research area considering the multitude of places where they can be deployed. Assistive robots for people in need, tourist guides in museums, checkpoint information in shops, data collectors in hospitals, are few of the possible examples where such a system can be used. People tend to depend more and more on devices that can make their life easier, by reducing a part of their daily tasks. In this context,

[1]Teaching assistant, Computer Science Department, University POLITEHNICA of Bucharest, 060042 Bucharest, Romania, e-mail: stefania.a.ghita@upb.ro

[2]Professor, Computer Science Department, University POLITEHNICA of Bucharest, 060042 Bucharest, Romania, e-mail: adina.florea@upb.ro

[3]Teaching assistant, Computer Science Department, University POLITEHNICA of Bucharest, 060042 Bucharest, Romania, e-mail: mihai.nan@upb.ro

[4]Teaching assistant, Computer Science Department, University POLITEHNICA of Bucharest, 060042 Bucharest, Romania, e-mail: david_train.iancu@upb.ro

social robots, and in particular assistive social robots, become more intelligent, with more capabilities.

The challenges that arise in the context of social robotics come from a variety of factors: unforeseen events, sensors malfunctions, limited information, user understanding, etc. One of the most challenging factors for a social autonomous robot is the human himself. Humans have a very complex and unpredictable nature, so such a robot must be able to cope with the sudden changes to be able to react accordingly when an unexpected situation appear. Besides the complex behavioural nature, humans have a lot of variation in terms of physical aspect (facial features, height, clothing, etc.) or voice (tone, accent, speed), making the interaction far more difficult.

To obtain a robust functioning of the system in a variety of situations, the robotic platform must be provided with a large volume of data. In terms of visual understanding, the system must process and extract all the possible information from the acquired images, and moreover, it should infer additional data. Besides the standard people and object detection and recognition, systems can derive estimated trajectories, that can help the robot have a better understanding of the behaviour of the user.

People trajectory prediction is the method that estimates the possible followed paths for every tracked person. Every estimated trajectory is computed based on previous observations, which are represented by the coordinates of the person in the previous moments of time. The information about the possible path a person might follow, can notably improve the behaviour of a robot: the robot can move in the direction of the estimated trajectory to meet a person, can model the movements of the people and extract patterns, can ensure the safety of the participants by avoiding collisions, etc.

In this research, we introduced a people trajectory prediction component into the AMIRO robotic framework [1], to increase the robustness of the behaviour of the Pepper robot [2]. We combined the existing computer vision modules with the trajectory prediction module to gather more data from the detected people. We tested several experimental scenarios and made an analysis on the obtain results, by adding new behaviours for the robot.

2. **Related Work**

The domain of social robotics is gaining increasing interest considering the number of related developing applications. Multiple social robotic platforms are being developed [3], while other projects aim to learn people how to interact with them [4]. To be able to program and maximise the abilities of a social robots, a variety of robotic frameworks have been introduced. Most of the robotic frameworks that are being developed include environmental perception for meaningful interactions. For the main task of human-robot interaction, social robots must be aware of the people in their surroundings.

Visual perception is a very important component that facilitates the behaviour understanding and communication adaptation.

The research in [5] introduced a full system specifically designed for environmental perception in robotic applications. The system combines input from multiple sensors, respectively a camera for RGB and depth data and a laser, to map the people in the environment. It merges multiple techniques in order to track people around a robot. The proposed architecture starts with a people detection module where the system applies an upper body detector on the RGB-D data and a leg detector on the laser data. The actual tracking of the people is performed by an Bayesian tracker, which uses an unscented Kalman Filter, to associate new detections with previous ones.

A very similar approach to [5] is presented in [6]. It introduces a full system whose purpose is to make a robot be aware of the people in its surroundings. The data handled by the system is acquired from an RGB-D camera and a laser, and processed by a leg detector and a convolutional method for people detection [7]. The detections are converted to 3D positions, which are then passed to a Kalman filter, for tracking.

An alternative development of an environmental perception system is in [8]. The challenge proposed by the system in [8] is to track and follow one target person. As in the previous case, the system uses a leg detector and a visual people detection technique to detect people around the robot, with the difference that the technique for detecting people in images is based on pose estimation, [9]. The tracking operation is performed by predicting people trajectories. The prediction is done using Support Vector Regression and it is mainly used when the target person goes out of focus.

While the previously mentioned systems focus specifically on detection and tracking of people in robotic scenarios, projects such as [10] and [11] present full systems deployed and validated on particular robots. The EnrichMe system [10] focused its development on improving the life quality for vulnerable people, such as elderly people. The system offers capabilities that can provide health or social support. It combines multiple techniques to implement relevant behaviours, such as object localization or activity monitoring.

The SocialRobot system [11] is another example of a full robotic system designed for interactions with elderly people. It implements the main capabilities for a robot (vision, navigation, speech) which are combined for simple interaction behaviours. The robot is used as an engaging companion, which can provide simple capabilities, such as taking a picture or showing an agenda. The system however cannot be used for more complex behaviours, as it has limited volume of processed data.

Compared to other existing systems, our AMIRO [1] project is proposing a framework that integrates multiple advanced capabilities which can be applied on a variety of robots. While other existing systems only implement basic functionalities which can be applied on particular scenarios, the AMIRO framework can be used to implement complex behaviours which can be extended to a multitude of applications. In terms of visual understanding, the framework combines multiple techniques to extract information from the surroundings. In particular, the people trajectory prediction module is an important addition to our framework, as it can help the robot move and perform in a more reliable and safer manner. This information, combined with the rest of the techniques, creates a more accurate visual awareness for the robot, characteristic which is not available on the other existing systems. The main advantages of the proposed framework are represented by the possibility to define complex robotic behaviours for sophisticated user interactions situations, based on the multitude various capabilities, and the characteristic to be fault tolerance. The modular architecture allows the system to continue its operation even if one of the components is not performing correctly, situation which can be handled by the behaviour defined by the programmer.

3. **Method**

The AMIRO robotic framework [1] is a platform specially created for easy deployment of social robotic applications. It has a modular structure that provides a series of capabilities exposed as function calls, which can be composed into complex behaviours. The AMIRO architecture is built on top of the Robotic Operating System (ROS) [12], to be a more general solution for multiple robotic platforms. The structure of the system is divided into local modules and cloud modules, to exploit the computational capabilities of local machines, and benefit from the precision of existing cloud systems.

The *Vision* component of the AMIRO framework consists in a composition of modules that interact and exchange information in order to process the visual information. Each module is implementing a set of functions that are then aggregated together to generate the final result of the component. The modules can be run independently, as separate applications, to prevent errors propagation in case of contingency failure. As the people trajectory prediction module generates visual information, we integrated it inside the *Vision* component. The module is estimating the future trajectories based on a neural network which requires strong computational capabilities to generate fast results. We deployed the module as a cloud-edge module inside the AMIRO framework, running on a local powerful machine.

We integrated the *People Trajectory Prediction* module by using the same process as in the case of the rest of the modules. The module is operating separately and it is exposing its own capabilities that are handled by the *Vision* manager. The integration between the existing modules and the newly added module is presented in Figure 1.

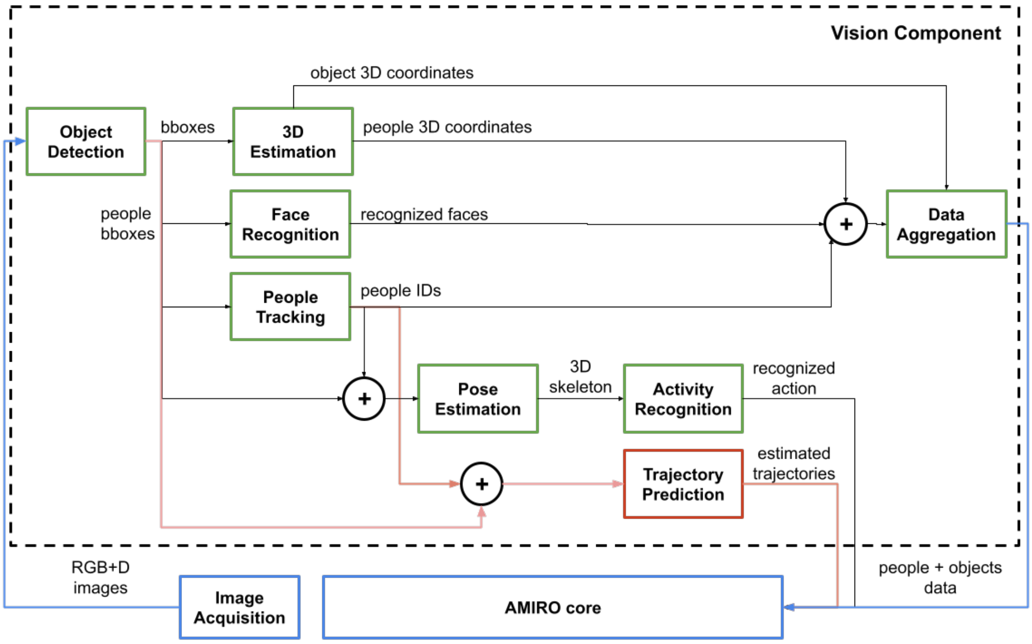

Fig. 1. Integration of *People Trajectory Prediction module* inside the *Vision* component of the AMIRO robotic framework. The green boxes represent the existing cloud-edge modules inside the *Vision* component, blue boxes represent external components and the red box represents the newly added module.

In order to generate estimated future trajectories, the module needs to receive a set of past observations. The past observations for a person represent the previous positions, in our case the image coordinates, where the person was located. After acquiring the required volume of past information, the module estimates the most probable trajectory that person can follow. Figure 1 shows that the input received by the *People Trajectory Prediction* module is composed of two pieces of information: the bounding boxes of the detected people and the identification numbers associated with the bounding boxes. We need both pieces of information as there can be multiple people in the image at a moment of time. The bounding boxes extracted by the *Object Detection* module indicate the positions of the people in the current image, while the *People Tracking* module is differentiating between the identified people. In order to predict the trajectory for a specific person, the system needs to differentiate between the detections, so it can consider only the positions relevant for that specific person. For each person identified by the *People Tracking* module, the *People Trajectory Prediction* module is storing their positions until the required number of past observations is reached. After storing the required amount of data for a person, for each new observed position, the data is updated by removing the first known position and adding the recently identified one.

In our project, we used the Social-GAN method [13] to predict the future trajectories. Social-GAN presents a method that combines a generative adversial network (GAN) with a social pooling layer. Figure 2 presents the general overview of the architecture of the Social-GAN system. The generative adversial network is trained to predict the possible future trajectories, while the social pooling is fusing information about the other people in the neighbourhood. The information about the locations of the neighbours is useful in trajectory prediction as the movement of one person is not only determined by the own behaviour of the person, but also by the movement of the people in the surroundings.

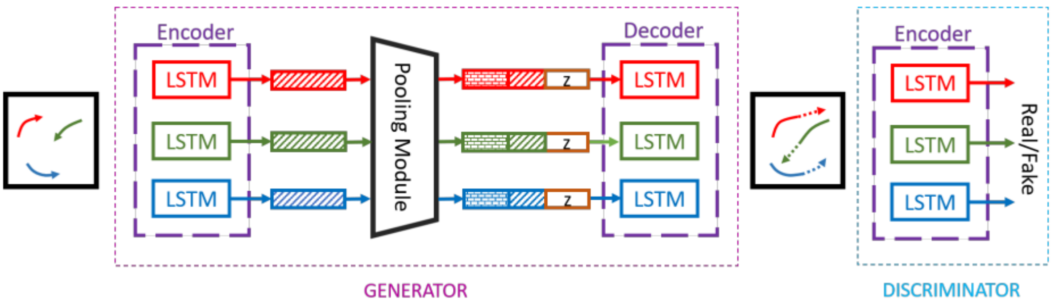

Fig. 2. Overview of the generator-discriminator architecture of the Social-GAN system.

To match the requirements of our project, which is social robotics, we trained the generative adversial network from [13] on the JRDB dataset [14], which contains videos acquired from the point of view of a robot. Before the training phase, we pre-processed the dataset to match the framerate provided by our robot, and we reduced the framerate of the JRDB dataset from 15 frames per second to 2.5. Figure 3 presents a clearer understanding about how the trajectory prediction module generates the final result on an image extracted from the JRDB dataset. In the figure the blue dotted line represents the observed positions of the identified person, the red dotted line represents the predicted positions, and the green dotted line represents the real trajectory followed by the person. The person is observed for 8 positions and the trajectory is generated for the next 8 positions.

The module is defined as a continuously running unit. We decided upon this deployment of the module considering the situations when the data would be required by the AMIRO framework. In the event that the system requires the trajectory information for one or multiple people, the module can provide the information instantly (if the person was previously detected), without

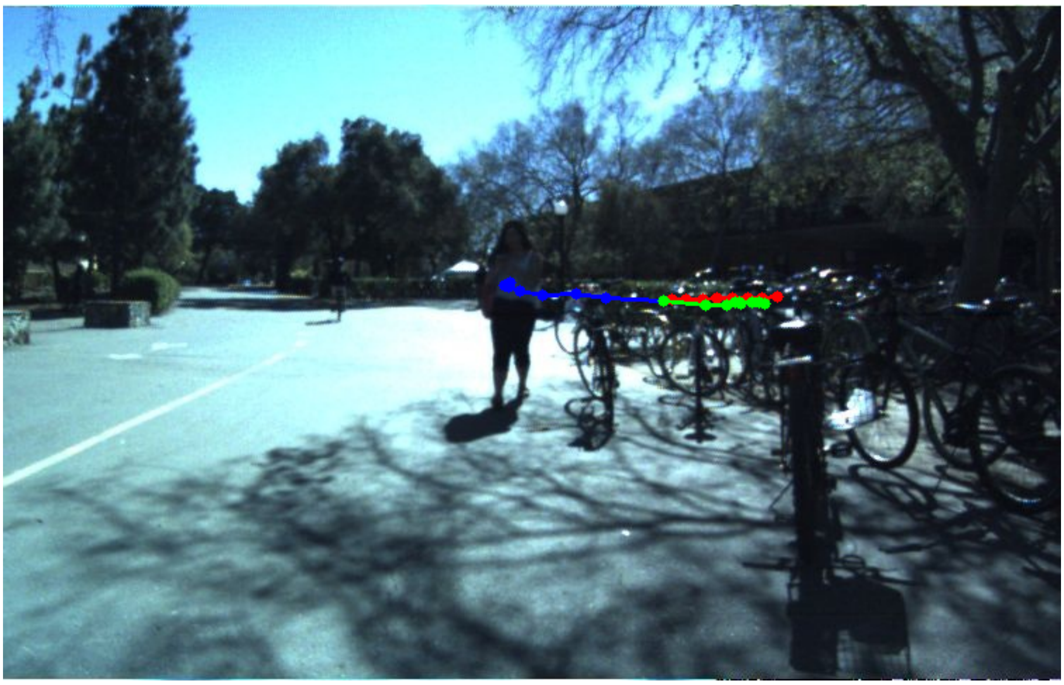

Fig. 3. Result example of the of *People Trajectory Prediction* module. The blue line represents the previous observed trajectory, the red line represents the estimated future trajectory predicted by the module based on the observations and the green line represents the real trajectory followed by the person.

waiting to observe the required number of positions. This will speed up the interaction process and would make the communication more meaningful. The module receives the required processed data constantly, for every image acquired by the robot. The result of the module is generated on request, when required by the *Planning* component.

4. **Experiments**

To obtain an analysis on how the module interacts with the rest of the system we performed multiple experiments in the context of social robotics. We defined several plausible scenarios that are valid for a human-robot interaction application, to determine the influence of the newly introduced module on the behaviour of the Pepper robot. The scenario in Section 4.**1** was defined based on the previous use cases in which the AMIRO framework was tested. The other two scenarios, Section 4.**2** and Section 4.**3**, were introduced considering other possible use cases for a socially assistive robot. We designed the robot behaviours associated to the proposed scenarios, which were then validated on the robot.

For each experimental scenario described in the sections below, we presented the scenario flow and the schema that describes the interaction between the integrated capabilities of the AMIRO framework. In all the figures, the capabilities that come from the same component are represented by boxes of one particular colour: red for vision, purple for navigation, green for speech, blue for notification management and yellow for storage management.

4.1. Social Robot in Assistive Care Scenario

The robot receives a question about the location of a specific target. If the robot knows the location, it provides the required information and observes the path followed by the person. If the person is not moving in the right direction, the robot will alert the person and will repeat the information.

One of the end goals of our project is to develop a robust system for assistive care scenarios. To be able to program a social robot to help people in need is a major concern in our research. One of the proposed scenarios involve a person in need interacting with the Pepper robot.

The proposed scenario assumes that the Pepper robot is helping a person by providing directions to a specific location, as described below. Figure 4 presents the capabilities combination schema for this behaviour.

(1) The user initiates the interaction with the robot and asks about the position of a target.
(2) The robot answers the question about the position of the target and provides the directions if known.
(3) The user goes towards the direction of the target. The robot is observing the followed trajectory, to estimate the motion.
    (a) If the predicted trajectory matches the expected route the robot goes back to the passive state.
    (b) If the predicted trajectory is different from the expected route the robot will inform the person that the followed path is wrong. If the followed path still does not match the expected one, the robot will inform the healthcare supervisor.

4.2. Social Robot in Tourist Assistance Scenario

The robot is placed in a museum and is in an active state, looking for people to help with information. When the robot detects a person going in the direction of an exhibit, it will go to that person to ask if they want to know more information about it. In case of a positive answer, the robot will provide the required information, otherwise it will look for other people to interact with.

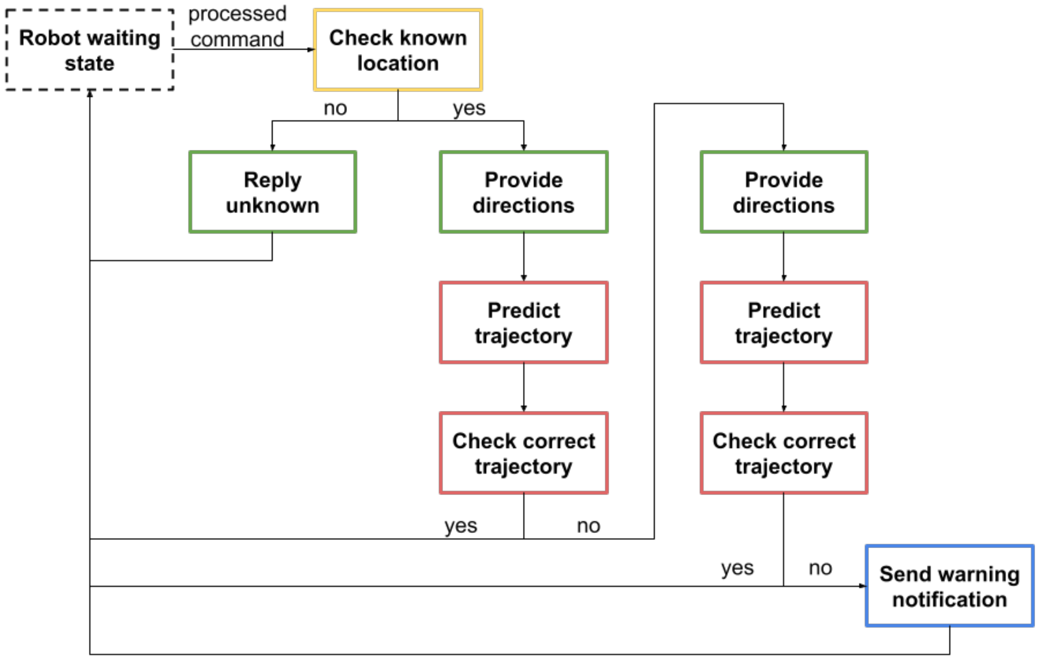

Fig. 4. Behaviour composition for an assistive care scenario using the AMIRO framework.

In many socially assistive applications, the robotic platform is used as an information point. The robot must be able to understand commands and provide the required information.

The proposed scenario assumes that the robot is working as a tourist guide in a museum. The tourists go to different exhibits and Pepper can provide information, as in the scenario described below. Figure 5 shows the interaction between the capabilities integrated in this behaviour.

(1) The robot is in active state looking for people to interact with. It observes the trajectories of the people in the neighbourhood.
(2) The robot detects a person that is walking towards a specific piece of art.
(3) The robot navigates towards the person.
(4) The robot asks the person if wants more information about the exhibit.
    (a) If the person agrees the robot will provide the information.
    (b) If the person denies the robot will go back into the state to look for another person.

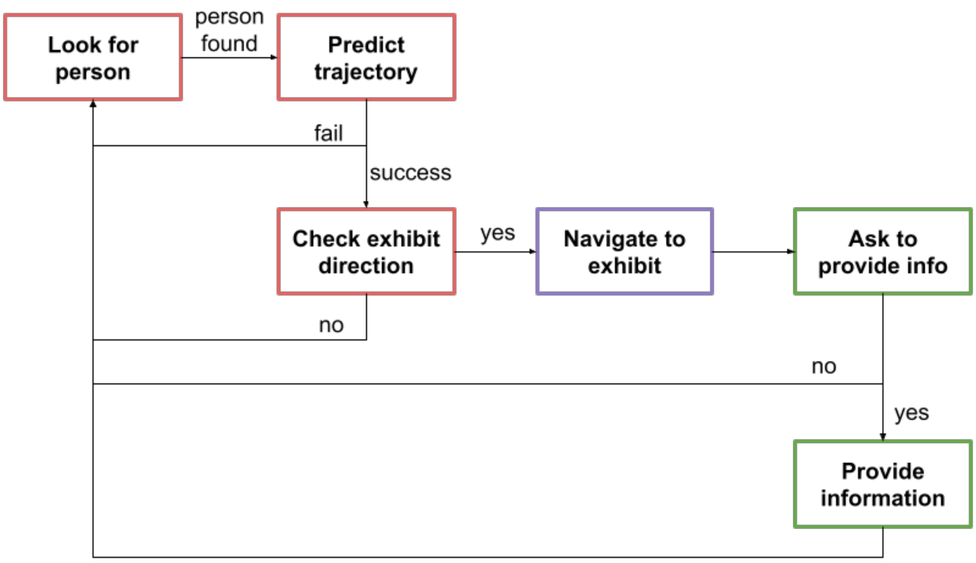

Fig. 5. Behaviour composition for a tourist assistive scenario using the AMIRO framework.

4.3. Social Robot in Emergency Evacuation Scenario

The robot is placed inside a building and receives a notification about an emergency evacuation. The robot looks for people in the surroundings to inform them about the emergency. When it detects a person, the robot will go to that direction to provide the essential information. If the robot detects a person that is going towards a dangerous part of the building, the robot will alert the person about the danger. After it detects no more people in the area, it starts looking for people in other areas to inform.

A socially assistive robot must be able to help people in emergency situations. We defined the scenario below, considering a possible emergency where Pepper can help. The robot must help people with guidance, during an evacuation. The trajectory prediction module helps the robot understand if a person is going to a dangerous area, so it can warn the person about the risk. Figure 6 presents the composition between the functions used in this behaviour.

(1) The robot receives notification about emergency evacuation.
(2) The robot searches for people in the surroundings.
(3) The robot provides information about the situation to the people in the neighbourhood.
(4) The robot starts tracking the trajectories of the people in the environment. If a person is going in the direction of a risky area the robot will warn the person about the danger.

(5) When no people are detected in the neighbourhood the robot will check other areas to see if there any people left.

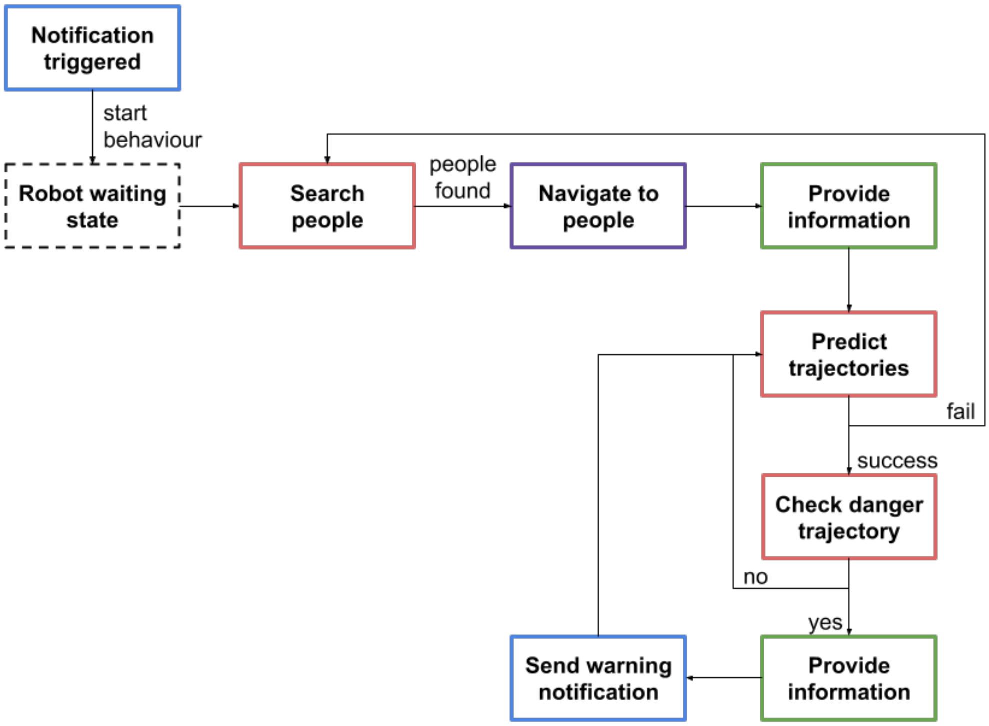

Fig. 6. Behaviour composition for an emergency evacuation scenario using the AMIRO framework.

4.4. Discussion

The experimental scenarios that we described present possible situations from different applications. Each described scenario combines existing capabilities from the AMIRO framework with the functions provided by the *People Trajectory Prediction* module. Besides the elementary function of estimating the future trajectories, the framework can detect sooner if a person is approaching a specific target, which can avert unsafe situations or can speed up the interaction process. All the presented scenarios assume environments which are not very crowded, with a medium number of people in the surroundings of the robot. This condition is required for the robot to be able to focus its attention on only one person for a meaningful human-robot interaction. When there are too many people in the environment, the robot may not be able to establish a correct interaction, given the high number of occlusions and appearance variations. In terms of lighting, all the proposed scenarios are located in indoor scenes with controlled artificial lighting conditions. For a correct functioning of the robot and, implicitly, of the trajectory prediction module, the lighting must be stable, with no additional strong sources of light that may cause the people in the images to not be visible enough. With respect to the scene settings, there are no strict requirements in terms of objects. In the proposed scenarios, the objects in the surroundings of the robot are exhibits for the second scenario, and chairs and tables for the third scenario. There should not be many objects near the robot, such that it can be able to move easily towards a target, without many detours. However, it is not a strict condition, the navigation module being able to replan a path even if it encounters an obstacle on the way.

In the first proposed scenario the robot is using the *People Trajectory Prediction* module to detect if the person has correctly understood the provided directions. This is an important element in the case of an assistive care application, because the health main supervisor can be notified to help the person in need if the interaction fails. In assistive care scenarios the robot must act as a supervisor, therefore the ability to understand the problems that may arise is a significant element. The limitations of the system in this scenario come from the required observed trajectory. If the person is moving out of the field of view of the robot, the system cannot predict the future trajectory and cannot infer if the followed path is correct. In this type of scenarios, the robot must be placed in a position that maximizes the visibility on the environment.

The second scenario presents a classical situation for a socially assistive robot, as the robot is looking for simple interactions with the people in the neighbourhood. In this scenario the *People Trajectory Prediction* module is helping the robot to understand the direction of the person and to reach the person faster. When the robot understands the direction of the person, it goes directly to that location, simplifying the intersection method. The limitations in this case may come from the behaviour of the person, if the person is briefly looking at an exhibit, then the robot may not get to start the interaction.

The last evaluated scenario shows the functionality of the *People Trajectory Prediction* module in case of an emergency situation. In this situation the trajectories are predicted to ensure the safety of the people. The robot can warn people not to go to a risky area if the estimated trajectory indicates that direction. The limitations in this scenario come from number of the people in the images and their movement speed. If there are a lot of people in the images, the system may not be able to correctly track them, and may generate wrong results. This comes from the fact that our approach is based on individual tracking instead of flow analysis. The flow analysis is outside the scope of our project, but can be investigated in a different research. Also, if the people are moving too fast, the robot may not be able to gather the required amount of data before making a prediction.

Even though it presents some limitations in particular cases, the information provided by the *People Trajectory Prediction* module can boost the understanding of the people and can improve the behaviour of the robot. In terms of processing speed the module is suitable for real-time applications, generating results in around 110 milliseconds regardless of the number of predicted trajectories.

5. Conclusions

Social robotics is a challenging research topic with multiple applications in a variety of domains. There is an increasing interest in the area of assistive robots, as they are engaging and can help people in their daily activities. Autonomous robotic platform are deployed worldwide as robotic companions, information points, health supervisors, etc. Given their possible impact on humans lives both socially and economically, the research area is becoming more focused on providing means for easy robotic applications deployment.

The AMIRO robotic framework is a platform that allows straightforward composition of modules for complex robotic behaviours. The framework combines capabilities of computer vision, navigation, speech, planning to generate robust performances.

In order to enhance the functionality of the robot, the current work integrates a trajectory prediction module inside the framework. We combined the information generated by the module with the existing information to generate more capabilities. We used a generative adversial network combined with a pooling layer (Social-GAN) trained on a robotic dataset to generate precise image trajectories. We presented several scenarios in the context of social assistive robots, to prove the enhancement of the robot capabilities.

For the future research we intend to integrate environmental information into the trajectory prediction module for more informed trajectories. We aim to add more capabilities, such as people collision detection, based on the results of the trajectory prediction module.

Acknowledgement: We would like to thank University Politehnica of Bucharest for supporting our research, by providing the necessary resources.

REFERENCES

[1] *A. S. Ghiță, A. F. Gavril, M. Nan, B. Hoteit, I. A. Awada, A. Sorici, I. G. Mocanu, and A. M. Florea*, “The amiro social robotics framework: deployment and evaluation on the pepper robot,” Sensors, vol. 20, no. 24, p. 7271, 2020.

[2] “Pepper Robot.” https://www.softbankrobotics.com/emea/en/pepper. [Online; accessed 11 July 2022].

[3] *M. Kyrarini, F. Lygerakis, A. Rajavenkatanarayanan, C. Sevastopoulos, H. R. Nambiappan, K. K. Chaitanya, A. R. Babu, J. Mathew, and F. Makedon*, "A survey of robots in healthcare," Technologies, vol. 9, no. 1, 2021.

[4] *A. Popescu, A. I. Awada, I. Mocanu, O. Cramariuc, and N. Samar Brenčič*, "A platform that aims to help people to learn how to interact with robotic platforms," in EDULEARN21 Proceedings, 13th International Conference on Education and New Learning Technologies, pp. 6342–6351, IATED, 5-6 July 2021.

[5] *C. Dondrup, N. Bellotto, F. Jovan, and M. Hanheide*, "Real-time multisensor people tracking for human-robot spatial interaction," in Workshop on Machine Learning for Social Robotics at 2015 IEEE International Conference on Robotics and Automation (ICRA), May 2015.

[6] *M. Kim, J. Lee, S. J. Jorgensen, and L. Sentis*, "Social navigation planning based on people's awareness of robots," CoRR, vol. abs/1809.08780, Sep. 2018.

[7] *J. Redmon, S. Divvala, R. Girshick, and A. Farhadi*, "You only look once: Unified, real-time object detection," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779–788, June 2016.

[8] *M. Kim, M. Arduengo, N. Walker, Y. Jiang, J. W. Hart, P. Stone, and L. Sentis*, "An architecture for person-following using active target search," CoRR, vol. abs/1809.08793, Sep. 2018.

[9] *Z. Cao, T. Simon, S. Wei, and Y. Sheikh*, "Realtime multi-person 2d pose estimation using part affinity fields," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1302–1310, July 2017.

[10] *S. Coşar, M. Fernandez-Carmona, R. Agrigoroaie, J. Pages, F. Ferland, F. Zhao, S. Yue, N. Bellotto, and A. Tapus*, "Enrichme: Perception and interaction of an assistive robot for the elderly at home," International Journal of Social Robotics, vol. 12, no. 3, pp. 779–805, 2020.

[11] *D. Portugal, P. Alvito, E. Christodoulou, G. Samaras, and J. Dias*, "A study on the deployment of a service robot in an elderly care center," International Journal of Social Robotics, vol. 11, no. 2, pp. 317–341, 2019.

[12] "Robot Operating System." https://www.ros.org/. [Online; accessed 11 July 2022].

[13] *A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi*, "Social gan: Socially acceptable trajectories with generative adversarial networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2255–2264, 2018.

[14] *R. Martín-Martín, H. Rezatofighi, A. Shenoi, M. Patel, J. Gwak, N. Dass, A. Federman, P. Goebel, and S. Savarese*, "Jrdb: A dataset and benchmark for visual perception for navigation in human environments," arXiv preprint arXiv:1910.11792, 2019.