BigATV 2024 Team Description Paper

Main-author: Austin Fu   Co-author: LincongLi   Team Members: Lincong Li, Edward Zhang, Allen Feng, Jackson Su, Xiaoping Zhu, Iris Yu, Yifan He

Abstract

Ageing societies and the associated pressure on the care systems are major drivers for new developments in socially assistive robotics. To understand better the real-world potential of robot-based assistance, we offer potential solutions with the aspect of Humanoid robots in numerous social and health care sector challenges, such as Softbank's Pepper, are mainly used as eye-catcher during events. Businesses would like to see such robots being used more functionally. To accomplish these tasks, our team try our best to make Pepper work robustly in the real world.

In this paper we present a solution of older adults' experiences with Pepper humanoid robot. The proposed solution is modular and relies on the proper integration of existing Pepper functionalities. Our research presents the Pepper robot to be able to recognize faces, communicate and interact with humans, navigation in dynamic environments, object recognition, NLP and ASR to improve the features of HRI, robot application and ambient intelligence with the medical equipment.

The aim is to implement: 1. The best solutions integrated to offer elderly in daily activities. 2. Content is provided by Pepper, who presented personal digital stories about each senior (using facial recognition), free discussion with Pepper and having Pepper demonstrate its skills, such as dance moves and telling jokes. 3. Information is collected by subjective measures, creating an optimization conceptual model aimed at maximizing the efficiency of assigning robots to server the elderly, giving multi-criteria optimization model is applied in the area of optimization for robot assignment for the elderly with robot utilization level and caregiver stress level.

1. Introduction

As a consequence of an ongoing demographic change and various associated social transformations there is an increased pressure on the care systems, which is forecasted to grow significantly. Industrialized countries are confronted with a decreasing number of professional caregivers and an increasing number of people in need of care. Providing high quality care for a rising number of elderly people will become a challenge. The shortage of skilled workers not only has a negative impact on the quality of professional care and thus on the physical and emotional wellbeing of caregivers, it also affects the wellbeing as well as physical and cognitive integrity of those who are in need of care. In the view of the complexity of the problem, governments, academia, and industries are analyzing various strategies, including robot-based assistance for the elderly, to support their health and safety, and foster social participation.

Robots in the healthcare and social sector exist in many forms, serving various purposes and supporting numerous tasks. They are applied in surgeries, rehabilitation, elderly care facilities or the home environment as assistants with housework and domestic chores. One commonly used type is a social robot. Social robots were developed for the interaction of humans and robots to support a humanlike interaction on an emotional level. They can be humanoid, with shape being important, as this type of interaction is grounded in visual and tactile perception in addition to verbal communication. The solutions of social robots focus on generic health and quality of life issues, as well as issues related to social care in hospital and home care environments. Apart from socializing and communication, these kinds of robots can provide play scenarios that may reduce the effect of certain disabilities or improve wellbeing through entertainment and companionship.

In this paper, we present a robotic-based application utilizing the robot Pepper that was especially designed to support older adults and their caregivers in care homes to increase physical and cognitive activity and initiate social interaction. We are aimed to create a system working in group settings with residents over a period of time and conducted interviews and observations with residents, caregivers and a facility worker. In this way we are able to elicit these people's attitudes, social and organizational practices, expectations towards the robot, individual and group-based performances as well as social communication among the participants. We will distill our findings from the long-term field deployment into lessons learned for future use of robots in aged care.

2. RELATED WORK - SYSTEM OVERVIEW

2.1 Technical Infrastructure - HRI

Pepper is a humanoid robot that is 1.2 meter high and weighs 28 kilograms. He is equipped with 2D and 3D cameras and has microphones as well as a tablet for human robot interaction. He has two arms that can separately be programmed to do movements. he is anchored on three rollers and can move around with them on firm ground. The robot can also move this head and hip, and the face looks a bit like human child with very big blinking eyes.

In order to make Pepper knowledgeable, we provide a comprehensive explanation of all the technical steps involved in developing the NewsGPT-based robot expert system. At the first stage, the user initiates an interaction with the Pepper robot by speaking a few words in the form of either a greeting or a question. Subsequently, the Automatic Speech Detector (ASD) system activates the speech recognition service, aiming to procure the transcription of the user's verbal utterance. This transcribed information is then relayed as a requestion to the AI agent housed on the cloud infrastructure of the proposed system, which operates using the GPT-3.5 model.

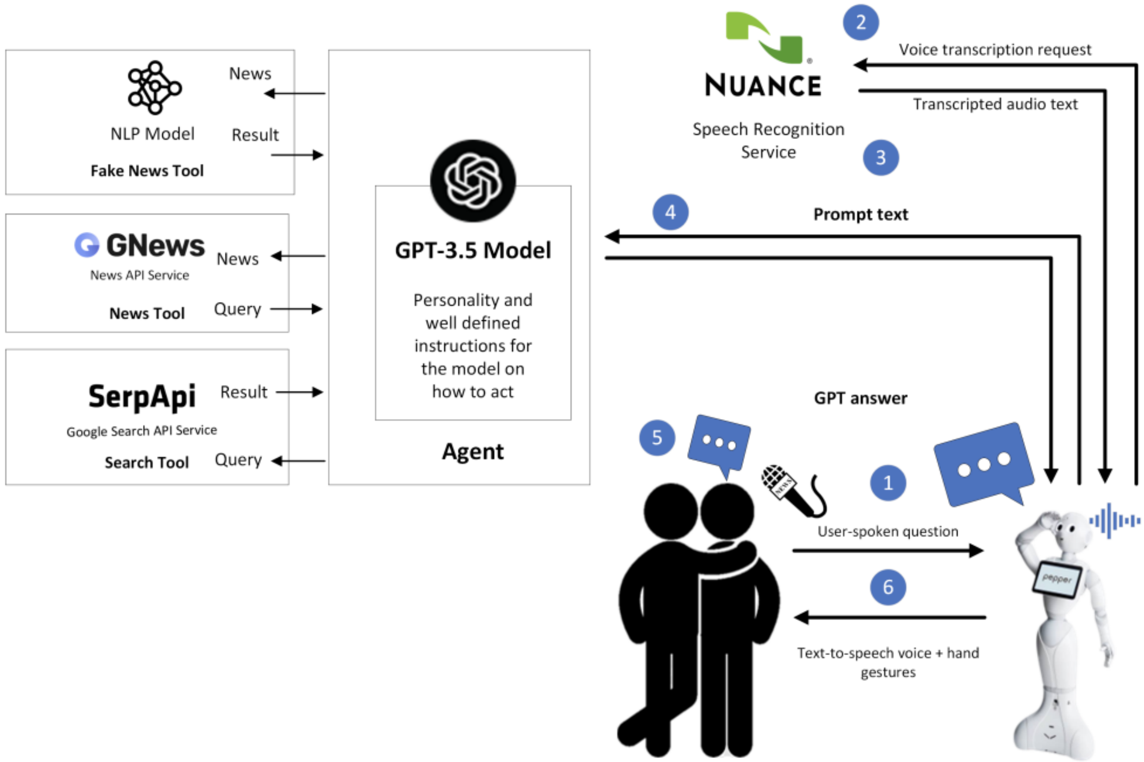

Fig. 1 Steps for NewsGPT integration with Pepper at indoor settings: 1) The Pepper robot's blue lights indicate its listening state during the interaction. 2) The robot requests the Nuance Automatic Speech Recognition service to transcribe the user's speech. 3) The Pepper robot sends a request to a cloud-hosted AI agent powered by the GPT 3.5 model. 4) The AI agent used is a React agent, which combines reasoning traces and task-specific actions for interactive and synergistic interactions. 5)The agent incorporates tools such as the Question Answer Tool and the News Tool. 6) The Question Answer Tool retrieves relevant information to respond to user queries. 7) The News Tool uses the GewsAPI to gather recent article headlines based on search terms provided by the model. 8) The retrieved response data from the News Tool is utilized by the language model to construct a news report. 9) The constructed news report is then conveyed back to the user.

2.2 Technical Infrastructure - SLAM

Softbank has created a coherent and well-thought development environment for the Nao/Pepper robots, called Choregraphe. It has the possibility to read signals from the sensor, process this information and control the robot. However, experience indicates that for the implementation of a stable and more advanced application, the dedicated SDKs should be used.

Robot Operating System (ROS) is a structured communications layer above host operating systems. It is designed to support the philosophy of modular, tool-based software development. ROS reuses code from numerous other open-source projects, such as hardware drivers, navigation system, and simulators. In each case, ROS is used only to expose various configuration options and to route data into and out of the respective software. Several hundred (or more) ROS packages exist across several publicly viewable repositories

2.2.1 Simultaneous Localization and Mapping

Mapping Techniques

Mapping algorithms can be roughly classified according to the map representation and the underlying estimation technique. One popular map representation is the occupancy grid. While these grid-based approaches are relatively computationally expensive, they are able to represent arbitrary and detailed features. Feature based representations can be attractive because of their compactness, but they rely on feature extractors which assume some knowledge of the structures in the environment. The most popular estimation techniques are based on the Extended Kalman filters (EKFs), graph-based optimization techniques or particle-based methods.

In researching, we present an improved approach to learning grid maps using Rao-Blackwellized Particle Filters. An on the fly computed improved proposal distribution is computed, instead of a previously used fixed proposal distribution. This allows for the direct usage of most of the information obtained from the sensors while generating the particles. While the computation of the proposal distribution is similar of that of others, it does not rely on predefined landmarks. This approach also draws the particles more effectively and the highly accurate proposal distribution gives a robust indicator to decide whether a resampling has to be carried out.

While Mapping relies on sufficiently accurate odometry, Hector Mapping2 utilizes the high update rate of modern LIDAR systems. This approach is based on optimization of the alignment of beam endpoints with the map learned so far. The basic idea is to use a Gauss-Newton approach, similar to the idea of a Kalman Filter. Using this approach there is no need for a data association search between beam endpoints or an exhaustive search. As scans align with the existing map, all preceding scans are implicitly matched with the current one. The system is sufficiently accurate to close loops typically encountered in small scale scenarios without use of an approach for explicit loop closure, keeping computational requirements low and preventing changes to the estimated map during runtime.

Navigational Algorithms

1. Global Planners

Navigational tasks in ROS are implemented in the Navigation Stack. It takes information from odometry, sensor beams and a goal pose and outputs velocity commands to a robot base. By default, a couple of options in terms of planners are available. global_planner is built as a flexible global path finder and grid paths and more. It works in combination with of one of the local planners.

spqrel_navigation7 is a navigation system requiring very few computational resources, but still provides performance comparable with commonly used tools in the ROS environment. Using a fast global path planner that adapts the computed path to dynamic changes in the environment, it only needs to consider planning at a global level. Eventually, Dijkstra's algorithm is used to compute the minimal paths to a goal. It also works both with ROS and NAOqi middleware.

2. Local Planners

Most global planners need a local planner to provide the eventual speeds needed to follow the global path. Dynamic obstacles are taken into account as well in the local planners, with the exception of some global planners that take them into account as well.

The Dynamic Window Approach produces a circular trajectory that is optimal for a robot's local position. It maximizes an objective function that depends on the progress to the target, clearance from obstacles and forward velocity to produce the optimal trajectory. It depends on the local costmap for obstacle information, making tuning the local costmap parameters crucial.

The dwb_local_planner9 extends the DWA and aims to make it as customizable as possible. Just as DWA, the trajectories are sampled, but instead of evaluating velocities in isolation, it scores an array of sample poses that are anticipated along that trajectory.

In researching, we propose to use Timed-Elastic-Band to implement an online optimal local trajectory planner for navigation. A sparse scalarized multi-objective optimization problem is solved to obtain the optimal trajectory. Since these local planners get stuck in locally optimal trajectories as they cannot transit across obstacles, and extension is implemented. In parallel, a subset of possible trajectories is optimized. Within that candidate set, the planner is able to switch, making it possible to transit across obstacles. It is implemented in the teb_local_planner10 ROS package.

2.2.2 Technical Infrastructure - Face/Object Recognition

Environments where service and social robots operate/live are highly dynamic, in part because of the people living in those environments constantly interact with each other and carry out daily life activities. Therefore, service, and social robots require perception systems that are highly robust, but at the same time are able to operate in real-time Object detection and recognition are some of the vision tasks that require at least near real-time operation.

The Pepper robot is a social robot used to research on human robot interaction in real-world human environments (e.g., it is the official platform of the RoboCup@Home Standard Platform League), but its operation is constrained to the fact that it lacks the computational power necessary to run state-of-the-art vision algorithms.

On the other hand, the uprising of Deep Neural Networks (DNNs) has led us to use them in the development of models that can quickly recognize objects and persons in a robust manner. However, they are very expensive in terms of computational power, so they cannot be run directly on typical service robots. In particular, most of state-of-the-art DNN models cannot run on Pepper's internal computer, a situation that stands as the problem we address in this solution. Therefore, the main goal is to propose a solution to this problem that considers two components: First, the selection of YOLO, a DNN with real-time detection and recognition of objects and persons, as the most suitable DNN to this task; and second, the development and implementation of an add-on for Pepper, a backpack that permit the attachment of a single board computer onto Pepper, particularly a Nvidia Jetson

TK1, which can run YOLO at about 5 FPS when processing images of 320x320 pixels. The use of an external enhancing device, attached to the robot without modifying its structure, comes as an inspiration from similar projects developed for the NAO robot, where a fully replicable backpack was built.

In this work we provide details on how to reproduce the backpack for Pepper, and on how to install and implement YOLO on it, making the backpack CAD model, hardware, software specifications and an installation guide available for replication.

2.2.3 Technical Infrastructure - Ambient assisted living

The increase of life expectancy and the growing senior population lead to an important development of smart

equipment for healthcare and assistive services to support elderly people, especially when living at home. Elderly people living alone with degenerating physical and cognitive abilities need innovative assistance techniques. Such techniques or Internet-Of-Things (IOT) systems can be classified as wearables, smart infrastructures, and mobile IOT. All these systems aim to help people in their daily activities, especially when they are dependent and/or have mobility impairments and/or cognitive impairments. IoT systems collect data and can complete action while connected to the Internet. Moreover, robotic systems can also be deployed at home for the safety and comfort of individuals. Usually, robotics and IOT systems are used separately.

We argue that the deployment of robots in IoT environments can address the individual limitations of single devices. Indeed, robots are limited by the data recorded by the sensors embedded by default. In contrast, a system usually includes only static sensors. By sharing information and interacting together, their combination can contribute in solving the aforementioned issues. The combination of both technologies has however only been explored in a very limited number of studies until now.

1) In this system, the robot moves and uses its camera to locate people in a room, where the system has no information, thus covering potential black spots resulting from the external cameras. As a result, the elderly people can be localized in their environment. The information can then be used to react if they fall and need help.

2) A system in form of a biometric bracelet and an interaction works together with a telepresence robot. A user can interact with the interaction device to obtain general information about, e.g., the weather, or more personalized information like appointments or medication plans. In other words, the interaction device provides similar functionalities to a usual smart speaker. During a medical appointment via the telepresence robot, the biometric bracelet can send health information to the doctor. Besides, the doctor can call for an appointment with the user via the telepresence robot based on the health information collected by the biometric bracelet.

3) The system locates and analyzes the users' activities based on sound. If users need help, the robot can be navigated by the system, but the robot is not in direct contact with the users and does not recognize them.

Moreover, a static robot is connected to the system, the system includes different sensors (e.g., Kinect camera). The system can process the data from its own sensor and from the robot in order to have more information about the environment and to detect both humans and objects. Through this combination, the enhanced human and object detection helps the robot in better interacting with the users, as its gaze can be better oriented.

3. Goals

1. The best solutions integrated to offer elderly in daily activities
The combination of HRI, Vision (Face/Object recognition), SLAM/VSLAM and ambient living will be well integrated into the one-pack application.

2. Content is provided by Pepper
The materials like images, videos, audios, and documents are organized into the formatted contents that are well suited to the specific elderly groups, through the feedback, the different style of exhibitions can be applied into Pepper.

3. Multimodal for optimization
Multi-modal fusion approaches have been successful in other domains, such as wireless sensor networks, in addressing single sensor weaknesses and improving information quality/accuracy. These approaches are inherently more reliable when a data source is lost. In the different materials fused into one-pack, Pepper robot will improve itself more easily.

4. Conclusions

Overall, the solution is that Pepper will be well-accepted by the majority of elderly groups and applied to different business scenarios, in which being able to activate people, creating responses, emotions and social interactions through storytelling, free communication and jokes. Storytelling evoked additional memories in elderly groups and ideas on Pepper's application areas arose from socializing with the robot, such as easing shyness and the preference of confiding in a robot rather than another human.

Videos, audios, and texts that are the most popular materials collected from the different sensors will be collected and analyzed through multimodal algorithms and training from the big data pools. It's going to study and satisfy gradually the requirements in the real world.

References
1. Ivan Petrovich Sidorov. A universal manipulation system. The Unique Journal in Advanced Robotics, 66(51):55–90, 2099.
2. John Doe and Jane Doe. Worldwide geo-localization from outer space. The Unique Journal in Advanced Robotics, 66(51):55–90, 2099.
3. Si Li and Wu Li. A unified multilanguage recognition system. The Unique Journal in Advanced Robotics, 22(01):42–69, 2098.
4. John Doe and Jane Doe. Introduction to Robotics, volume 1. RoboCup press, Trantor, 2050.
5. Erika Mustermann and Max Mustermann. Philosophy in the Artificial Mind, volume 2. RoboCup press, Trantor, 2050.
6. Jan Jansen, Mieke Modaal, and Janneke Publiek. A Socialist Base for Robot's Rights, volume 3. RoboCup press, Trantor, 2050.
7. Sicrano da Tal and Beltrana da Tal. Endowing Consciousness into Machines, volume 4. RoboCup press, Trantor, 2050.
8. Tizio Rossi and Caio Rossi. Empathic Robots, volume 4. RoboCup press, Trantor, 2050.
9. Tarou Yamada and Hanako Yamada. The day a robot taught me how to cook. Transactions in Domestic Robotics, 1(01):1–69, 2050.

5. Software and External Devices.

We use a standard Pepper robot. No modifications have been applied.

Robot's Software Description

For our robot we are using the following softwares:

- Platform: NAOqi OS 2.5 or 2.9
- Face Recognition: The third party like Baidu or self-training.
- Object recognition: Choregraphe or self-training.
- Animation: the compiled files.
- SLAM: ROS
- HRI: ChatGPT or other LLMs
- IoT: the medical devices and network

External Devices:

- The medical devices or ambient living things.
- NVIDIA hardware for the multimodal training.

Cloud Services:

- Speech Recognition: Nuance.
- NLP: ChatGPT and Baidu.
- Face/Object recognition: Baidu