2nd Workshop Toward Robot Avatars, IEEE International Conference on Robotics and Automation (ICRA), London, UK, 2023.

An Autonomous Mobile Robot based Tele-Presence System with Augmentation: A Pilot Trial of Virtual Museum Tour

Rushali Mohbe[1], Poornima J D[1], Srijan D[2], Shashank Goyal[1], Sridatta Chatterjee[1]and Bharadwaj Amrutur[1,2]

***Abstract*— While Robotic Avatars are gaining popularity and seeking commercialization, there arises a need for standard control mechanisms as well as perception capabilities. We propose a novel robot-agnostic communication and control middleware to operate telepresence robots. In this paper, the main components of the AHAM Teleoperation Software Stack (AHAM-TSS) are discussed. Additionally, the relevance of the AHAM-TSS is discussed with regard to solving real-world problems, such as conducting virtual tours of museums through telepresence robots.**

I. INTRODUCTION

Avatar robots enable us to get a live experience of a remote location through the use of live communication devices mounted on mobile robot platforms. As global organizations are picking up momentum in the post covid world, the unspoken agreement is to continue functioning in the “hybrid” model; this implies opening up to global audiences while keeping costs minimal [1]. Museums and art galleries are examples of some such institutions that may open up new opportunities for global collaborations by deploying avatar robots. A number of museums and art exhibitions gained popularity by offering virtual tours through avatar robots during the covid pandemic lockdown [2].

Two main problems with the application of avatar robots for virtual tours of museums are: (i) the user has no idea of the floor map of the galleries and (ii) users logging in from different countries would require an uninterrupted live walking experience of the galleries. To address these challenges, we propose a communication and control software; the AHAM Teleoperation Software Stack; that can be used to control and communicate via the robot.

II. IMPLEMENTATION

To ensure clients can use this teleoperation software on multiple robot platforms, the proposed AHAM Teleoperation Software Stack has been architected to be both robot and operating-system agnostic. The teleoperation software stack comprises three main components, a middleware, a streaming module, and an augmentation module, as shown in Fig. 1.

A. Middleware

The middleware acts as a communication broker between the two endpoints of a robotic avatar system, the robot, and its human controller. This middleware allows multiple robots and controllers to communicate through it simultaneously.

Users(human operators) and robots have to be authenticated using OAuth 2.0 thus ensuring only authorized entities can communicate via the middleware. The middleware uses SQL tables to store details of authorized users, robots, sessions created by a user, session duration, etc. The middleware was designed to be robot and operating system agnostic and hence, API endpoints are provided to communicate with it. The user uses an API endpoint to start a session with a robot using their authentication token and then uses that token again to send a control message to the robot through the middleware. After the user creates a session, the message is then routed to the appropriate robot which is received via another API endpoint.

This middleware is presented as a ready-to-use solution that is easy to integrate with existing robots and user interfaces. An alternative design approach would be to employ the AWS suite, primarily AWS lambdas or AWS IoT which has the benefit of being serverless and auto-scaling as per usage. But, developing the appropriate logic for them is often time-consuming and does not support on-premises deployments to address privacy concerns.

Aham-TSS gives a user the flexibility to host the middleware either on the cloud or on a local server. Sessions can also be created such that one controller can control multiple robots. This feature can be leveraged for monitoring a fleet of semi-autonomous or autonomous robots.

B. Streaming

The video from a robot’s onboard camera is typically directly consumed for processes like obstacle avoidance deployed on the robot. However, not every robot has a GPU or a high-powered CPU capable of running the heavy processing required for video analysis. We provide a simple streaming pipeline that uses ffmpeg or gstreamer to stream the video data from a robot’s onboard camera to a local edge server. As this server is on the same local network as the robot, latency is minimal, allowing it to handle heavy processing like video analysis for text recognition. This video stream received on the edge server can also be piped to a virtual camera which is then forwarded to a Jitsi server. The user logs into a Jitsi Meet and uses this view to understand the robot’s surroundings better.

C. Augmentation

Before the video stream is sent to the user, it is possible to augment the video feed in order to enhance the understanding and the experience of the user. The augmentation pipeline is designed to be modular and can be easily expanded to accommodate the requirements of an individual. Currently, the AHAM-TSS provides two main augmentations: zoom

[1]ARTPARK, Indian Institute of Science (IISc), India
[2]Robert Bosch Centre for Cyber-Physical Systems, IISc, India

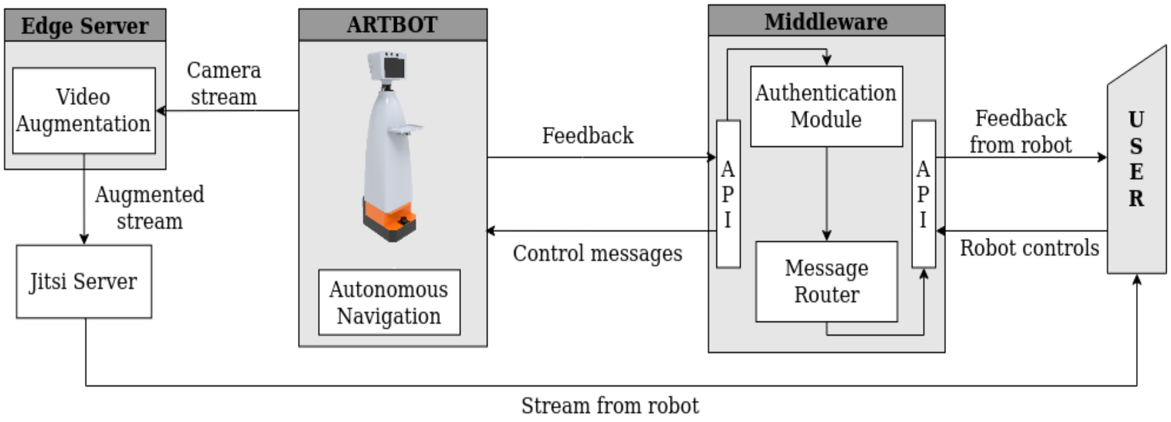

Fig. 1. The AHAM Teleoperation Software Stack.

and text recognition. The ability to zoom and still preserve the quality of the video feed received by the user is a desirable feature that is especially useful for monitoring purposes as even small items can be examined with ease. Additionally, we also perform basic text recognition on the video feed to annotate words/phrases of interest and overlay them on the video feed.

D. *Autonomous navigation*

Manually navigating robotic avatars in remote environments can be difficult, especially when obstacles or narrow lanes are present. Autonomous control is more reliable by combining onboard cameras and lidar sensors that can provide a comprehensive understanding of the robot's environment, resulting in safer and more effective navigation [3]. The footprint of objects such as chair wheels or smaller objects cannot be precisely detected in the 2D plane of the lidar; therefore, it will not be accurately represented in the navigation system's costmap. A Spatio Temporal Voxel Layer (STVL) plugin [4] has been incorporated into the navigation stack in order to circumvent the limitations of lidar-based navigation. STVL introduces novel concepts of voxel decay and decay acceleration as an alternative to the existing Voxel Layer, which can fail in high-dynamic, human-populated environments [5]. The plugin works by using laser scan data from an on-board 2D RP lidar and point cloud data from an on-board Intel realsense d455 camera model, which has been subjected to a temporal filter to reduce noise. The plugin uses the filtered point cloud to generate voxels for obstacles in the 3D environment in real time. The camera's point cloud readily detects the most protruding part of an obstacle, which is then projected onto the 2D costmap. The parameters of the STVL plug-in were fine-tuned experimentally to make the navigation robust for the environment in which the robot was deployed.

III. DISCUSSION AND FUTURE WORK

The AHAM-TSS has been deployed on a modified Bot-Sync's Volta [6]. Volta is an indoor differential drive robot that serves as the mobile base. It is based on ROS and has a Nvidia Jetson Xavier as the computing system. The body of the robot consists of 3d printed parts and has a movable neck with additional mounting for external cameras and display screens. It has a total runtime of 6 hours on a single charge.

The AHAM-TSS has been used to conduct a pilot demonstration in the Visvesvaraya Industrial and Technical Museum to conduct a virtual tour of the exhibits. An online museum guide operates the robot to take online visitors on a virtual tour of the museum. The museum tour guide is the primary operator of the telepresence robot. The guide uses an interactive webpage to control the robot and also receives a live feed of the museum from the robot's perspective. As we also allow the video feed to be augmented, the guide can use the zoom function to highlight features of interest during the tour. We also provide different levels of permissions to guests, tour guides, and service engineers.

In the future, we hope to further augment the stream with additional useful information that would allow a user to understand the dimensions of objects, automatically segment and focus on areas of interest, translate in-frame text to a different language, or even prove additional information about objects of interest. We are also working on making the source code publicly available.

ACKNOWLEDGMENT

The authors thank the Department of Science and Technology, Govt of India, Dept of IT-BT Government of Karnataka, and Nokia for funding support.

REFERENCES

[1] "What 800 executives envision for the post-pandemic workforce," Sep 2020. [Online]. Available: https://www.mckinsey.com/featured-insights/future-of-work/what-800-executives-envision-for-the-postpandemic-workforce
[2] A. Dickson, "You can't visit the museum. but your robot can." Apr 2020. [Online]. Available: https://www.nytimes.com/2020/04/15/arts/museums-robots-coronavirus.html

[3] V. De Silva, J. Roche, and A. Kondoz, “Robust fusion of lidar and wide-angle camera data for autonomous mobile robots,” *Sensors*, vol. 18, no. 8, 2018. [Online]. Available: https://www.mdpi.com/1424-8220/18/8/2730

[4] S. Macenski, D. Tsai, and M. Feinberg, “Spatio-temporal voxel layer: A view on robot perception for the dynamic world,” *International Journal of Advanced Robotic Systems*, vol. 17, 2020.

[5] Github, “Navigation issues,” https://github.com/ros-planning/navigation/issues/662, (Accessed on 03/24/2023).

[6] “Volta - indoor mobile robot for ros learning, teaching and research,” https://www.botsync.co/volta.