Multi-Sense-Rescuer: Multi-Target Audio-Visual Learning and Navigation in Search and Rescue Scenarios Transfer Application to Physical Robot

Nathanael Oliver, Md Amanullah Kabir Tonmoy

I. INTRODUCTION

Autonomous Navigation is a pivotal advancement in the field of robotics. It's significance derives from the ability of robots to independently operate in complex and dynamic environments. This capability allows robots to transcend their traditional limitations by harnessing technologies like computer vision, machine learning, and sensor fusion. Autonomous robots can execute tasks with precision and efficiency by adapting to obstacles and planning optimal pathing. This is especially important in fields such as search and rescue (SAR) because it reduces the need for human intervention in hazardous settings.

While visual and sensor-based data have long been primary sources for navigation, integrating audio cues as input for autonomous navigation introduces a profound dimension of significance in robotics. Sound can offer crucial information about the surroundings, including the direction and proximity of objects, potential hazards, and even human presence. In the context of SAR, audio cues are extremely important in locating human destinations. Audio-based navigation has the potential to be transformative, enabling machines to navigate with a level of awareness that was previously unattainable.

II. PROBLEM STATEMENT

The problem at hand is to extend the capabilities of a pre-existing multi sound source location model, originally designed for simulated robots, to effectively handle multi-sound source scenarios in a real world situation. Specifically, we aim to adapt and fine-tune this model using transfer learning techniques to work on a Jetson Nano, which will be controlling a simple robot. The datasets to be used for training and evaluation will primarily consist of acoustic data collected from complex indoor environments, with a focus on the Matterport3D and Replica3D datasets.

Dataset:

We will utilize acoustic data from the Matterport3D and Replica3D datasets, which represent indoor environments with diverse acoustic characteristics. These datasets offer a wide range of sound source scenarios, including varying sound source numbers, locations, and acoustic conditions. The dataset will include audio recordings, ground truth sound source locations, and environmental information to simulate real-world multi-sound source scenarios.

Expected Results:

We anticipate that the real world multi-sound source model, derived from the original theoretical model through transfer learning, will exhibit slightly lower performance compared to the theoretical model designed specifically for multi-source scenarios. Due to the added constraints of the hardware running the model as well as real world interferences, we expect a marginal decrease in accuracy, but we aim to maintain a high level of reliability in sound source localization and navigation.

Evaluation:

The evaluation of the multi-sound source model will be conducted in real-world settings by deploying it on a physical robot. The robot will be placed in environments resembling those present in the

training data, and its performance will be assessed based on its ability to identify and navigate towards multiple sound sources accurately. We will measure key performance metrics, including localization accuracy, navigation success rate, and computational efficiency on the Jetson Nano platform.

The success of the model will be determined by its practical utility in real-world scenarios, particularly in applications such as search and rescue, where multiple sound sources signify individuals in need of assistance. Ultimately, the model should demonstrate its capacity to enhance the capabilities of autonomous robots by extending their audio perception skills to multi-source environments, even if it performs slightly below the theoretical model's performance.

III. TECHNICAL APPROACH

The technical approach of extending single sound source location models to create a multi-sound source model involves several key steps. Initially, we begin with the models developed for multi sound source localization, as demonstrated by researchers such as Kartik Singhal, Mehdi Yaghouti, and Pooyan Jamshidi. These models are typically designed to identify and navigate towards a single sound source within an environment. To transform them into multi-sound source models, we employ transfer learning techniques. Transfer learning allows us to leverage the knowledge and expertise gained from training on single-source scenarios and adapt the model to handle multiple sound sources.

Transfer learning involves fine-tuning the pre-trained single-source model on a new dataset that simulates or represents multi-source scenarios. By doing so, the model can learn to recognize and respond to multiple sound sources effectively. This process reduces the need for extensive training from scratch, as the model already possesses a foundation of audio perception skills.

Once the multi-sound source model is established, the next challenge is making it runnable on resource-constrained systems like the Jetson Nano. This step often involves simplifying the model architecture and optimizing it for efficient execution. This might include reducing the number of parameters, employing quantization techniques, and optimizing inference processes. The goal is to strike a balance between maintaining model accuracy and minimizing computational resource requirements, allowing the model to operate smoothly on hardware with limited processing power and memory. This final step is crucial for practical deployment, enabling autonomous robots or devices with constrained resources to benefit from the enhanced audio perception capabilities achieved through transfer learning.

IV.PRELIMINARY RESULTS

Our training process has been focusing on single source and single goal scenarios utilizing the Replica3D dataset. The outcomes of this initial training phase will establish a foundation which we will use to assess the performance of our multi-source and multi-goal model in the future. Our evaluation will rely on success rate and SPL (Success weighted by Path Length). Currently, the model training is ongoing, with a total of 280 checkpoints recorded thus far. Following the completion of the training, we will evaluate and test the model, quantifying its performance using the specified metrics. Our next phase entails the training of the model for multi-source and multi-goal scenarios, incorporating transfer learning techniques. Additionally, we will optimize the model for deployment on the Jetson Nano platform, facilitating its execution on a physical robot.