POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Tino V2, the revenge of the maze (Temporal Title)

Tesi di Laurea Magistrale in
Computer Science and Engineering
Ingegneria Informatica

Author: **Sebastian Enrique Perea Lopez**

Student ID: 10986638
Advisor: Prof. Andrea Bonarini
Co-advisors: Federico Esposito
Academic Year: 2024-25

Dedicated to my family.

Abstract

Here goes the abstract.

Keywords: key, words, go, here

Abstract in lingua italiana

Qui va inserito l'abstract in italiano.

Parole chiave: qui, vanno, le, parole, chiave

Contents

1 Introduction

1.1. Project Overview

This section will introduce the Tino V2 social robotics project, providing context for the comprehensive system redesign and modernization effort undertaken during this thesis work. The project motivation will be presented first, covering the evolution from the original Tino social robot to the need for substantial technological upgrades that address limitations in processing power, localization accuracy, and remote interaction capabilities. The thesis scope will be defined, encompassing the complete migration from legacy Raspberry Pi architecture to modern NVIDIA Orin Nano systems, the implementation of advanced SLAM and sensor fusion technologies, the development of VR integration systems for remote human-robot interaction, and the redesign of mechanical systems to support enhanced functionality. The problem statement will be outlined, identifying key challenges including the unreliable localization performance of the original system that suffered from significant drift issues, the computational limitations that prevented real-time implementation of advanced computer vision algorithms, the lack of remote operation capabilities that limited the robot's utility for telepresence applications, and the mechanical reliability issues that affected consistent operation during extended use. The research objectives will be clearly defined, covering the development of robust localization systems that combine visual SLAM with UWB positioning for drift-free navigation, the implementation of real-time human detection and pose estimation using optimized YOLOv11 models, the creation of seamless VR integration that enables natural remote control through Unity-based virtual environments, and the enhancement of mechanical systems to support reliable long-term operation with improved differential drive locomotion.

1.2. Social Robotics Context

This section will establish the broader context of social robotics research and position the Tino V2 project within current technological trends and academic research directions.

The social robotics landscape will be examined, covering the growing importance of robots designed for human interaction in domestic, healthcare, and educational environments, the role of telepresence robotics in enabling remote social connections, and the integration of virtual and augmented reality technologies to enhance human-robot interaction modalities. The technological convergence will be discussed, highlighting how advances in embedded AI processing, real-time computer vision, and wireless communication technologies have created new possibilities for sophisticated social robots, the emergence of ROS2 as a standardized framework for modern robotics development, and the increasing accessibility of high-performance sensors like depth cameras and UWB positioning systems. The research contribution significance will be established, covering how the Tino V2 project advances the state of the art in VR-integrated social robotics through practical implementation of bidirectional communication systems, the development of atomic movement architectures that ensure predictable robot behavior for remote users, and the demonstration of sensor fusion approaches that address common localization challenges in indoor environments.

1.3. Technical Challenges and Innovation

This section will detail the specific technical challenges addressed in this thesis and the innovative solutions developed to overcome them. The localization challenge will be examined first, covering the fundamental problem of indoor robot navigation where GPS is unavailable and traditional visual SLAM systems suffer from drift accumulation, the need for absolute positioning accuracy to enable reliable VR integration where virtual and physical environments must remain synchronized, and the implementation of sensor fusion combining RTABMap visual SLAM with UWB positioning to achieve robust localization performance. The real-time processing challenge will be addressed, covering the computational demands of simultaneous SLAM processing, human pose detection, and VR communication on embedded hardware, the optimization strategies employed including TensorRT acceleration for YOLOv11 models and efficient ROS2 node architecture design, and the balance achieved between system performance and power consumption requirements. The VR integration challenge will be explored, covering the development of atomic movement systems that ensure consistent mapping between virtual user actions and physical robot responses, the implementation of bidirectional audio communication for natural telepresence interaction, and the creation of Unity-based VR environments that provide intuitive control interfaces for non-technical users.

1.4. Thesis Structure and Contributions

This section will outline the organization of this thesis and summarize the key contributions made to the field of social robotics. The thesis organization will be presented, covering how Chapter 2 establishes the background and related work in SLAM, sensor fusion, human detection, and VR integration technologies, Chapter 3 details the conceptual work including technology selection criteria and system architecture design decisions, Chapter 4 presents the implementation of all major system components including hardware integration and software development, Chapter 5 provides evaluation of system performance and comparison with baseline approaches, and Chapter 6 concludes with discussion of achievements, limitations, and future research directions. The primary contributions will be summarized, covering the successful integration of UWB positioning with visual SLAM to address indoor localization drift issues, the development of atomic movement control architecture that enables predictable VR interaction with physical robots, the implementation of real-time human pose detection on embedded hardware using optimized YOLOv11 models, the creation of comprehensive VR integration systems that support bidirectional communication and natural user interfaces, and the demonstration of practical sensor fusion approaches that combine complementary positioning technologies. The broader impact will be discussed, covering how these contributions advance the field of social robotics by providing validated approaches for reliable indoor navigation, remote interaction, and human detection that can be adopted by other researchers and extended for various social robotics applications.

2 | Background

2.1. Social Robotics Foundations

This section will establish the theoretical and practical foundations of social robotics that underpin the Tino V2 project development. The evolution of social robotics will be traced, covering the progression from industrial automation to human-centered robotic systems designed for interaction, companionship, and assistance in social environments, the emergence of telepresence robotics as a solution for remote communication and presence, and the integration of artificial intelligence and machine learning technologies to enable more natural and responsive human-robot interactions. The key characteristics of effective social robots will be examined, including the importance of predictable and reliable behavior that builds user trust and confidence, the need for intuitive interaction modalities that accommodate users with varying technical expertise, and the critical role of robust localization and navigation systems that enable autonomous operation in human environments. The current research landscape will be surveyed, covering major social robotics platforms and their capabilities, ongoing research in human-robot interaction psychology and user experience design, and the technological challenges that continue to limit widespread adoption of social robotics systems.

2.2. Simultaneous Localization and Mapping (SLAM)

This section will provide comprehensive coverage of SLAM technologies and their application to mobile robotics, with particular focus on the challenges and solutions relevant to indoor social robot operation. The SLAM problem formulation will be established, covering the fundamental challenge of simultaneously building a map of an unknown environment while determining the robot's location within that environment, the mathematical foundations including probabilistic approaches and uncertainty representation, and the trade-offs between computational complexity and accuracy that affect real-time implementation. Visual SLAM approaches will be examined in detail, covering feature-based methods like ORB-SLAM that rely on distinctive visual landmarks for tracking and mapping, direct methods that use raw pixel intensities for dense reconstruction, and hybrid approaches that combine the advantages of both feature-based and direct techniques. The specific capabilities and limitations of RTABMap will be analyzed, including its strength in loop closure detection that enables long-term mapping without drift accumulation, the RGB-D processing pipeline that leverages depth information for robust feature matching, and the memory management strategies that allow operation in large environments without excessive computational overhead. Common challenges in indoor SLAM will be addressed, covering the impact of dynamic environments with moving objects and people, the difficulties posed by textureless surfaces and repetitive patterns that provide insufficient visual features, and the accumulation of odometry drift that can lead to mapping inconsistencies over extended operation periods.

2.3. Ultra-Wideband (UWB) Positioning Technology

This section will examine UWB positioning technology as a complementary localization solution that addresses the limitations of visual SLAM in indoor environments. The UWB technology fundamentals will be established, covering the physics of ultra-wideband radio signals and their propagation characteristics that enable precise time-of-flight measurements, the anchor-based positioning architecture that provides absolute coordinate references, and the centimeter-level accuracy achievable under ideal conditions. The advantages of UWB for indoor robotics will be detailed, including immunity to visual challenges like poor lighting or featureless environments that can disrupt camera-based systems, the ability to provide absolute positioning without drift accumulation, and the low power consumption that makes UWB suitable for battery-powered mobile robots. The integration challenges will be addressed, covering the calibration requirements for anchor placement and coordinate system alignment, the impact of non-line-of-sight conditions and multipath effects in indoor environments, and the need for sensor fusion techniques to combine UWB position data with orientation information from other sources. Practical implementation considerations will be examined, including the setup and maintenance requirements for UWB anchor infrastructure, the trade-offs between positioning accuracy and update rates, and the strategies for handling temporary signal loss or degraded accuracy conditions.

2.4. Computer Vision and Human Detection

This section will cover the computer vision technologies employed for human detection and pose estimation in the Tino V2 system. The evolution of object detection algorithms will be traced, covering the progression from traditional computer vision techniques to modern deep learning approaches, the development of real-time detection frameworks like YOLO (You Only Look Once) that enable practical implementation on embedded hardware, and the specific advances in YOLOv11 that provide improved accuracy and efficiency for human detection tasks. Human pose estimation fundamentals will be established, covering the mathematical representation of human skeletal structure using key body joints, the challenges of 2D to 3D pose reconstruction from single camera views, and the integration of depth information to provide accurate 3D human tracking capabilities. The TensorRT optimization framework will be examined, covering the model conversion and optimization processes that enable real-time inference on NVIDIA hardware, the trade-offs between model complexity and inference speed that affect practical deployment, and the memory management strategies required for efficient operation on embedded systems with limited resources. Practical challenges in human detection will be addressed, covering the impact of varying lighting conditions and camera perspectives on detection accuracy, the handling of multiple people in the scene with potential occlusions, and the integration of detection results with robot control systems for responsive human-robot interaction.

2.5. Virtual Reality Integration and Telepresence

This section will examine the technologies and methodologies for integrating VR systems with physical robotics platforms to enable immersive telepresence experiences. The VR development landscape will be surveyed, covering the Unity game engine as a platform for creating interactive VR applications, the hardware requirements and capabilities of modern VR headsets for spatial tracking and user interaction, and the software frameworks that enable communication between VR applications and external systems. The telepresence robotics concept will be established, covering the psychological and technical requirements for effective remote presence that makes users feel genuinely connected to distant environments, the importance of low-latency communication for natural interaction, and the challenges of mapping human movements and intentions to appropriate robot behaviors. The communication architecture will be detailed, covering the ROS-TCP-Endpoint framework that enables Unity-ROS2 integration, the message passing protocols required for real-time data exchange, and the synchronization challenges that arise when coordinating virtual and physical systems with different update rates and latency characteristics. The user experience design considerations will be addressed, covering the development of intuitive VR interfaces that enable effective robot control without requiring technical expertise, the importance of providing appropriate feedback to help users understand robot state and capabilities, and the safety considerations that must be incorporated when enabling remote control of physical robotic systems.

3 Conceptual Work

3.1. Technology Research and Selection Methodology

This section will detail the systematic approach undertaken to research and select appropriate technologies for the Tino V2 system redesign. The research methodology will be established, covering the comprehensive literature review process used to identify state-of-the-art solutions in SLAM, sensor fusion, human detection, and VR integration, the evaluation criteria developed to assess technologies based on accuracy, computational requirements, integration complexity, and long-term maintenance considerations, and the experimental validation approach used to test promising technologies before final selection. The localization technology analysis will be presented, covering the systematic comparison of visual odometry approaches including ORB-SLAM3, SVO, and RTABMap with detailed evaluation of their computational requirements and accuracy performance, the assessment of UWB positioning systems including anchor configuration requirements and accuracy characteristics under various environmental conditions, and the analysis of sensor fusion approaches that combine complementary positioning technologies to address individual limitations. The human detection technology evaluation will be detailed, covering the comparison of different YOLO variants and their performance on embedded hardware, the assessment of depth camera technologies including Intel RealSense and Oak-D Pro systems, and the analysis of optimization frameworks like TensorRT for real-time inference on NVIDIA platforms. The VR integration technology selection will be examined, covering the evaluation of Unity versus other VR development platforms, the assessment of communication frameworks for ROS2-VR integration, and the analysis of VR hardware compatibility and user experience considerations.

3.2. System Architecture Design

This section will present the comprehensive system architecture developed for the Tino V2 platform, covering both hardware and software components and their integration strategies. The overall architecture philosophy will be established, covering the modular design approach that enables independent development and testing of subsystems, the ROS2-based communication framework that provides standardized interfaces between components, and the scalability considerations that allow for future expansion and modification of system capabilities. The hardware architecture will be detailed, covering the migration from Raspberry Pi to NVIDIA Orin Nano as the primary computing platform with analysis of performance improvements and power consumption considerations, the integration of Oak-D Pro depth camera for simultaneous SLAM and human detection capabilities, the UWB positioning system implementation including anchor placement strategies and coordinate system alignment, and the redesigned mechanical systems including the differential drive base and improved Stewart platform head mechanism. The software architecture will be presented, covering the ROS2 node structure that organizes functionality into discrete, communicating processes, the message passing interfaces that enable real-time data exchange between localization, detection, control, and VR systems, the data recording and analysis frameworks that support system debugging and performance evaluation, and the launch file organization that enables easy switching between mapping and localization modes. The integration strategy will be examined, covering the synchronization mechanisms that ensure consistent timing across all system components, the error handling and recovery procedures that maintain system stability during component failures, and the calibration procedures required to achieve accurate sensor fusion and coordinate system alignment.

3.3. Sensor Fusion Strategy

This section will detail the conceptual approach developed for fusing UWB positioning data with RTABMap visual SLAM to achieve robust indoor localization. The sensor fusion rationale will be established, covering the complementary nature of UWB and visual SLAM technologies where UWB provides absolute positioning without drift while visual SLAM provides accurate orientation and short-term relative positioning, the challenges addressed by sensor fusion including UWB signal degradation in non-line-of-sight conditions and visual SLAM drift accumulation over extended operation, and the advantages gained through fusion including improved localization reliability and reduced dependence on environmental conditions. The fusion architecture will be presented, covering the decision to use UWB for global position estimation while relying on RTABMap for orientation data, the coordinate system transformations required to align UWB and camera coordinate frames, and the filtering and smoothing strategies employed to handle measurement noise and temporary signal loss. The implementation strategy will be detailed, covering the ROS2 message structures used to combine positioning data from multiple sources, the timing synchronization mechanisms that ensure consistent sensor fusion despite different update rates, and the parameter tuning approaches used to optimize fusion performance for the specific operating environment. The validation methodology will be outlined, covering the experimental protocols developed to assess fusion accuracy compared to individual sensor performance, the test scenarios designed to evaluate system behavior under challenging conditions like UWB signal obstruction or visual feature scarcity, and the metrics developed to quantify localization accuracy and reliability improvements achieved through sensor fusion.

3.4. Atomic Movement Control Architecture

This section will present the conceptual framework developed for the atomic movement control system that enables predictable VR interaction with the physical robot. The atomic movement concept will be established, covering the principle that all robot movements must be discrete, completion-guaranteed operations that provide clear feedback to VR users, the four-state control architecture that separates robot behaviors into rest, attention-getting, preparation, and execution phases, and the synchronization requirements that ensure coordinated behavior between base locomotion and leg articulation systems. The VR interaction design will be detailed, covering the mapping of VR user actions to appropriate robot movement commands, the feedback mechanisms that inform users about robot state and movement completion, and the safety considerations that prevent conflicting commands and ensure predictable robot behavior. The movement coordination strategy will be presented, covering the locking mechanisms that prevent simultaneous conflicting movements, the command queuing system that handles user inputs during ongoing movements, and the timing optimization that ensures natural-looking robot behavior while maintaining system responsiveness. The state machine architecture will be examined, covering the discrete states defined for leg and base controllers, the transition conditions that govern movement between states, and the error handling procedures that maintain system stability when movements cannot be completed as planned. The pulse command system will be detailed, covering the implementation of discrete command signals that automatically return to idle state, the timing parameters that ensure complete movement execution, and the integration with the VR system that provides appropriate user feedback for each movement phase.

3.5. Real-time Performance Optimization

This section will address the conceptual approaches developed to achieve real-time performance across all system components despite the computational demands of simultaneous SLAM, human detection, and VR communication. The performance requirements analysis will be established, covering the timing constraints imposed by VR interaction that require low-latency response to user inputs, the computational load distribution across SLAM processing, pose detection, and system coordination tasks, and the memory management strategies required for sustained operation on embedded hardware with limited resources. The optimization strategy will be presented, covering the parallel processing architecture that distributes computational load across multiple ROS2 nodes, the priority scheduling that ensures critical functions like localization maintain consistent performance, and the resource allocation strategies that balance accuracy with computational efficiency. The real-time processing pipeline will be detailed, covering the image processing optimization that enables simultaneous SLAM and human detection from the same camera stream, the communication optimization that minimizes latency in VR data exchange, and the memory management that prevents performance degradation during extended operation. The scalability considerations will be examined, covering the modular architecture that allows selective activation of system components based on current requirements, the configuration management that enables performance tuning for different operational scenarios, and the monitoring systems that provide real-time feedback about system performance and resource utilization to support ongoing optimization efforts.

4 | Implementation

4.1. Hardware Platform Migration and Integration

This section will detail the complete hardware migration from the legacy Raspberry Pi system to the modern NVIDIA Orin Nano platform and the integration of new sensing technologies. The computing platform upgrade will be presented first, covering the migration process from Raspberry Pi 4 to NVIDIA Orin Nano including performance benchmarking that demonstrated substantial improvements in processing capability, the power system redesign using DC-DC converters to provide stable 19V power for the Orin Nano from 12V battery systems, and the cooling and thermal management solutions implemented to ensure reliable operation during intensive computational tasks. The camera system integration will be detailed, covering the replacement of the basic Pi camera with the Oak-D Pro depth camera system, the mechanical mounting solutions developed including the tripod-based camera support system and protective housing that shields the camera while maintaining visibility, and the camera calibration procedures required for accurate depth estimation and SLAM performance. The UWB positioning system implementation will be examined, covering the anchor placement strategy developed for the laboratory environment, the coordinate system calibration procedures that align UWB global coordinates with the robot's local reference frame, and the integration of UWB receivers with the Orin Nano through USB interfaces and custom ROS2 drivers. The mechanical system upgrades will be addressed, covering the complete redesign of the base locomotion system from omnidirectional triksta wheels to differential drive configuration, the motor and driver upgrades that provide increased power and reliability for the heavier Orin Nano system, and the Stewart platform head improvements including new arm designs with heim joints that eliminate flex and improve reliability.

4.2. ROS2 Software Architecture Implementation

This section will present the complete implementation of the ROS2-based software architecture that replaced the legacy monolithic Python scripts with a modular, maintainable system. The node architecture design will be established first, covering the development of specialized ROS2 nodes including gamepad_node.py for Xbox controller input processing, hardware_interface_node.py for Arduino communication management, robot_controller_node.py for central coordination and behavior management, and vr_interface_node.py for Unity integration and data exchange. The communication framework implementation will be detailed, covering the ROS2 topic structure that enables real-time data sharing between nodes, the service interfaces used for system configuration and control commands, the parameter management system that allows runtime configuration of system behavior, and the launch file organization that enables easy switching between mapping and localization operational modes. The Arduino integration will be examined, covering the serial communication protocols developed for reliable data exchange with head, base, and leg control systems, the device symlink configuration that ensures consistent device addressing across system restarts, and the command processing logic that translates high-level movement commands into appropriate motor control signals. The data recording and analysis infrastructure will be presented, covering the implementation of comprehensive logging systems that capture all sensor data, robot states, and user interactions for later analysis, the data extraction tools developed for processing recorded sessions, and the visualization systems that enable real-time monitoring of system performance and debugging of operational issues.

4.3. SLAM and Sensor Fusion Implementation

This section will detail the implementation of the RTABMap SLAM system and its integration with UWB positioning for robust indoor localization. The RTABMap configuration and optimization will be presented first, covering the parameter tuning process that optimized performance for the Oak-D Pro camera system, the memory management configuration that enables long-term mapping without excessive resource consumption, and the loop closure detection settings that ensure reliable map consistency during extended operation. The UWB integration implementation will be detailed, covering the development of custom ROS2 nodes for UWB data processing and coordinate transformation, the sensor fusion logic that combines UWB position data with RTABMap orientation information, and the coordinate system alignment procedures that ensure consistent mapping between UWB global coordinates and robot local coordinates. The mapping and localization modes will be examined, covering the implementation of distinct operational modes for initial map creation and subsequent localization within existing maps, the map saving and loading procedures that enable persistent environment representations, and the automatic relocalization capabilities that allow the robot to recover its position after temporary tracking loss. The performance optimization will be addressed, covering the computational optimization that enables real-time SLAM processing on the Orin Nano platform, the memory management strategies that prevent performance degradation during extended mapping sessions, and the error handling procedures that maintain system stability when sensor data is temporarily unavailable or degraded.

4.4. Human Detection and Pose Estimation Implementation

This section will present the implementation of real-time human detection and pose estimation using YOLOv11 with TensorRT optimization. The YOLOv11 model preparation and optimization will be detailed first, covering the conversion process from PyTorch models to TensorRT optimized engines for the Orin Nano platform, the model quantization and optimization strategies that balance detection accuracy with inference speed, and the integration with the Oak-D Pro camera system that enables simultaneous RGB and depth data processing for 3D pose estimation. The real-time processing pipeline will be examined, covering the image preprocessing and postprocessing steps that prepare camera data for inference and extract meaningful pose information, the coordinate transformation calculations that convert 2D detection results to 3D world coordinates using depth information, and the tracking algorithms that maintain consistent person identification across multiple frames. The pose estimation implementation will be detailed, covering the 17-keypoint skeleton detection that provides comprehensive human pose information, the depth integration that enables accurate 3D position calculation for each detected person, and the filtering and smoothing algorithms that reduce noise and provide stable pose tracking despite occasional detection errors. The integration with robot systems will be addressed, covering the ROS2 message structures used to publish human detection and pose data for use by other system components, the coordinate system transformations that align human pose data with robot and world coordinate frames, and the performance monitoring systems that track detection accuracy and computational load to ensure real-time operation requirements are maintained.

4.5. VR Integration and Atomic Movement System Implementation

This section will detail the implementation of the complete VR integration system including Unity-based user interfaces and atomic movement control architecture. The Unity

VR application development will be presented first, covering the creation of immersive VR environments that provide intuitive robot control interfaces, the implementation of bidirectional communication systems that enable real-time data exchange between Unity and ROS2, and the user interface design that makes robot control accessible to users without technical expertise. The ROS-TCP-Endpoint integration will be examined, covering the implementation of reliable communication bridges between Unity and the ROS2 ecosystem, the message serialization and deserialization procedures that handle complex data structures, and the error handling and reconnection logic that maintains communication stability despite network interruptions. The atomic movement system implementation will be detailed, covering the four-state control architecture implemented in both leg and base controllers, the synchronization mechanisms that ensure coordinated movement between robot subsystems, and the command queuing and locking systems that prevent conflicting movement commands. The audio system integration will be addressed, covering the implementation of bidirectional audio communication that enables natural conversation between VR users and people near the robot, the audio processing pipelines that handle microphone input and speaker output, and the integration with the VR system that provides seamless audio transmission. The data recording and analysis implementation will be presented, covering the comprehensive logging systems that capture all VR interactions, robot responses, and sensor data for detailed analysis of human-robot interaction patterns, the data extraction and visualization tools that enable researchers to analyze user behavior and system performance, and the real-time monitoring systems that provide immediate feedback about system operation and performance metrics during VR sessions.

4.6. System Integration and Testing Infrastructure

This section will present the implementation of comprehensive system integration procedures and testing infrastructure that ensure reliable operation of all system components. The integration testing methodology will be established, covering the systematic approach used to validate individual subsystem functionality before full system integration, the interface testing procedures that verify reliable communication between ROS2 nodes, and the end-to-end testing protocols that validate complete system operation from VR input to robot response. The calibration and configuration management will be detailed, covering the sensor calibration procedures required for accurate camera, UWB, and coordinate system alignment, the parameter management systems that store and maintain optimal configuration settings for different operational scenarios, and the automated configuration validation tools that detect and report configuration errors or drift. The debugging and monitoring infrastructure will be examined, covering the comprehensive logging systems that capture detailed system state information for troubleshooting and optimization, the real-time monitoring displays that provide immediate feedback about system performance and health, and the automated error detection and reporting systems that identify and classify system issues to support rapid problem resolution. The deployment and maintenance procedures will be addressed, covering the automated setup scripts that streamline system installation and configuration, the backup and recovery procedures that protect against data loss and enable rapid system restoration, and the update and maintenance protocols that enable safe system modification and capability enhancement without disrupting ongoing research activities.

5 | Evaluation

5.1. Localization System Performance Analysis

This section will present a comprehensive analysis of the localization system performance, comparing the sensor fusion approach against baseline SLAM-only operation to demonstrate the improvements achieved through UWB integration. The experimental methodology will be established first, covering the controlled testing environment setup using fixed reference points marked on the laboratory floor, the systematic data collection protocols that captured position estimates from both SLAM-only and UWB-fused systems during identical movement patterns, and the statistical analysis methods used to quantify localization accuracy and consistency. The SLAM-only baseline performance will be analyzed, covering the drift characteristics observed during extended operation where position estimates deviated by up to 1.2 meters from true positions at fixed reference points, the orientation accuracy that remained reliable even when position estimates suffered significant drift, and the specific scenarios that triggered substantial localization errors including proximity to walls and areas with limited visual features. The sensor fusion performance improvements will be detailed, covering the substantial reduction in position error achieved through UWB integration where maximum deviations were reduced to centimeter-level accuracy, the improved consistency of position estimates across multiple test runs at identical reference points, and the enhanced reliability of localization during challenging scenarios that previously caused significant SLAM drift. The coordinate system alignment results will be presented, covering the successful integration of UWB global coordinates with RTABMap local coordinate systems, the accuracy of coordinate transformations validated through systematic position measurements at known reference points, and the stability of alignment parameters that maintained consistent coordinate mapping throughout extended testing sessions.

5.2. Human Detection and Pose Estimation Performance

This section will analyze the performance of the YOLOv11-based human detection and pose estimation system, focusing on accuracy, processing speed, and reliability under various operational conditions. The detection accuracy analysis will be presented first, covering the successful implementation of 17-keypoint human pose detection that provides comprehensive skeletal tracking information, the depth integration performance that enables accurate 3D position estimation for detected humans using Oak-D Pro stereo camera data, and the tracking consistency that maintains stable person identification across consecutive frames despite variations in pose and position. The real-time performance evaluation will be detailed, covering the achievement of real-time inference speeds through TensorRT optimization that enables simultaneous SLAM and human detection processing, the computational resource utilization analysis that demonstrates efficient operation within the constraints of the Orin Nano platform, and the system responsiveness that provides immediate detection results for integration with robot control systems. The robustness testing results will be examined, covering the system performance under varying lighting conditions and camera perspectives that demonstrated reliable detection across realistic operational scenarios, the handling of multiple people in the camera field of view with appropriate tracking and identification capabilities, and the error handling performance that maintains system stability when detection temporarily fails or produces uncertain results. The integration effectiveness will be addressed, covering the successful coordination between human detection and robot behavior systems that enables responsive human-robot interaction, the coordinate system accuracy that properly aligns detected human poses with robot and world coordinate frames, and the data quality that provides meaningful input for research analysis and VR system integration.

5.3. VR Integration and User Experience Analysis

This section will evaluate the effectiveness of the VR integration system and the user experience provided by the atomic movement control architecture. The VR communication performance will be analyzed first, covering the low-latency data exchange achieved between Unity and ROS2 systems that enables responsive robot control from VR environments, the bidirectional audio communication quality that supports natural conversation between VR users and people near the robot, and the system reliability that maintains stable communication despite network variations and computational load changes. The atomic movement system evaluation will be detailed, covering the successful implementation of the four-state control architecture that provides predictable and intuitive robot behavior for VR users, the synchronization accuracy between leg and base movements that creates natural-looking robot locomotion, and the command queuing effectiveness that handles user inputs during ongoing movements without losing commands or creating conflicts. The user interface effectiveness will be examined, covering the intuitive VR control interfaces that enable effective robot operation without requiring technical expertise, the feedback systems that provide users with clear information about robot state and movement progress, and the safety mechanisms that prevent dangerous or conflicting robot behaviors during remote operation. The overall user experience will be assessed, covering the natural interaction feel achieved through careful timing and coordination of robot movements, the immersion quality that makes users feel genuinely present in the robot's environment, and the accessibility features that accommodate users with varying levels of VR experience and technical background.

5.4. System Reliability and Performance Metrics

This section will present comprehensive analysis of overall system reliability, computational performance, and operational characteristics under extended use conditions. The system stability analysis will be established first, covering the uptime performance achieved during extended testing sessions with the ROS2 architecture demonstrating robust operation without memory leaks or performance degradation, the error recovery capabilities that maintain system functionality when individual components experience temporary failures, and the graceful degradation behavior that preserves core functionality when non-critical subsystems encounter problems. The computational performance evaluation will be detailed, covering the resource utilization analysis that demonstrates efficient operation within the constraints of the Orin Nano platform while simultaneously running SLAM, human detection, and VR communication systems, the real-time performance maintenance that preserves responsive operation despite varying computational loads, and the power consumption characteristics that enable practical battery-powered operation for meaningful durations. The integration robustness will be examined, covering the communication reliability between ROS2 nodes that maintains consistent data flow despite network variations and system load changes, the sensor fusion stability that provides consistent localization performance across different environmental conditions and operational scenarios, and the modular architecture effectiveness that enables independent debugging and optimization of individual subsystems. The operational flexibility will be addressed, covering the ease of switching between mapping and localization modes that enables efficient system deployment in new environments, the configuration management effectiveness that allows system optimization for different operational requirements, and the maintenance and troubleshooting capabilities that support ongoing research and development activities with minimal downtime and technical expertise requirements.

5.5. Comparative Analysis and Lessons Learned

This section will synthesize the evaluation results to provide comparative analysis against baseline systems and extract key lessons learned from the implementation and testing process. The performance improvement quantification will be presented first, covering the measurable improvements achieved through the migration from Raspberry Pi to Orin Nano architecture including processing speed increases and capability expansion, the localization accuracy improvements demonstrated through UWB sensor fusion compared to SLAM-only operation, and the system reliability enhancements achieved through modular ROS2 architecture compared to legacy monolithic implementations. The technology validation results will be detailed, covering the successful validation of RTABMap SLAM for indoor social robot applications despite challenges with visual feature scarcity and dynamic environments, the effectiveness of UWB positioning for absolute coordinate reference despite occasional signal degradation and non-line-of-sight conditions, and the practical feasibility of real-time human pose detection on embedded platforms through careful optimization and model selection. The implementation insights will be examined, covering the critical importance of systematic testing and calibration procedures for achieving reliable sensor fusion performance, the value of modular software architecture for enabling efficient debugging and system optimization, and the significance of user-centered design principles for creating effective VR interaction systems. The research contribution significance will be assessed, covering the advancement of practical sensor fusion approaches that address common indoor robotics challenges, the development of atomic movement architectures that enable predictable VR-robot interaction, and the demonstration of comprehensive integration strategies that combine multiple advanced technologies in a cohesive, reliable robotic system suitable for social robotics research and applications.

6 | Conclusions

6.1. Summary of Achievements

This section will synthesize the major achievements accomplished through the Tino V2 development project, highlighting the successful transformation from a legacy system to a modern, capable social robotics platform. The technological transformation will be summarized first, covering the complete migration from Raspberry Pi-based architecture to NVIDIA Orin Nano systems that provided substantial improvements in computational capability and real-time performance, the successful integration of advanced sensing technologies including Oak-D Pro depth cameras and UWB positioning systems that enhanced the robot's perception and localization capabilities, and the implementation of modern ROS2-based software architecture that replaced monolithic legacy code with modular, maintainable systems. The sensor fusion success will be highlighted, covering the development and validation of UWB-SLAM fusion that addressed critical indoor localization drift issues while maintaining real-time performance requirements, the achievement of centimeter-level positioning accuracy that enables reliable VR integration and precise navigation, and the demonstration of robust localization performance across various environmental conditions and operational scenarios. The VR integration accomplishments will be detailed, covering the successful implementation of bidirectional communication systems that enable natural remote interaction through Unity-based VR environments, the development of atomic movement architectures that provide predictable and intuitive robot control for non-technical users, and the creation of comprehensive data recording and analysis systems that support ongoing research in human-robot interaction. The human detection achievements will be presented, covering the successful implementation of real-time pose estimation using optimized YOLOv11 models that provide 17-keypoint skeletal tracking with depth information, the achievement of simultaneous SLAM and human detection processing on embedded hardware, and the integration of detection results with robot control systems for responsive human-robot interaction.

6.2. Technical Contributions and Innovation

This section will detail the specific technical contributions made to the field of social robotics and the innovative approaches developed during this thesis work. The sensor fusion methodology contribution will be established first, covering the practical demonstration of UWB-visual SLAM integration that addresses common indoor robotics localization challenges, the development of coordinate system alignment procedures that enable reliable fusion of absolute and relative positioning data, and the validation of sensor fusion approaches that maintain real-time performance while improving localization accuracy and reliability. The atomic movement architecture contribution will be detailed, covering the development of discrete state-based control systems that enable predictable VR interaction with physical robots, the implementation of synchronization mechanisms that coordinate multiple robot subsystems for natural movement patterns, and the creation of command queuing and conflict resolution systems that handle complex user interactions safely and effectively. The real-time optimization contributions will be examined, covering the successful implementation of simultaneous SLAM, human detection, and VR communication on embedded hardware through careful architectural design and optimization strategies, the development of parallel processing approaches that distribute computational load effectively across multiple ROS2 nodes, and the achievement of low-latency communication between VR and robotics systems that enables responsive remote interaction. The integration methodology contributions will be addressed, covering the development of comprehensive testing and validation procedures that ensure reliable multi-component system operation, the creation of modular software architectures that enable independent development and debugging of complex robotics systems, and the implementation of robust error handling and recovery mechanisms that maintain system stability under challenging operational conditions.

6.3. Research Impact and Broader Implications

This section will examine the broader research impact and implications of the Tino V2 project for the field of social robotics and related disciplines. The social robotics advancement will be assessed first, covering how the demonstrated VR integration capabilities advance the state of the art in telepresence robotics and remote human-robot interaction, the contribution to understanding of user experience design principles for VR-controlled robotics systems, and the validation of practical approaches for implementing complex social robotics capabilities on accessible hardware platforms. The sensor fusion research impact will be detailed, covering the practical demonstration of indoor localization solutions that address common challenges faced by mobile robots in GPS-denied environments, the contribution to understanding of sensor fusion trade-offs and optimization strategies for real-time embedded applications, and the validation of UWB positioning as a complementary technology for visual SLAM systems. The embedded AI research contributions will be examined, covering the successful demonstration of real-time human pose detection on embedded hardware through systematic optimization approaches, the contribution to understanding of model optimization trade-offs for practical robotics applications, and the validation of TensorRT and similar optimization frameworks for enabling advanced AI capabilities on resource-constrained platforms. The interdisciplinary research impact will be addressed, covering the contribution to understanding of effective integration strategies for combining computer vision, robotics, networking, and VR technologies in cohesive systems, the advancement of user-centered design principles for complex technical systems that must be accessible to non-technical users, and the demonstration of practical approaches for bridging virtual and physical environments in meaningful and reliable ways.

6.4. Limitations and Challenges

This section will provide honest assessment of the limitations encountered and challenges that remain unresolved in the current Tino V2 implementation. The technical limitations will be acknowledged first, covering the remaining dependence on UWB anchor infrastructure that limits deployment flexibility and requires careful setup and calibration procedures, the computational constraints that, while significantly improved, still impose trade-offs between system capability and real-time performance requirements, and the environmental dependencies that affect system performance under certain lighting conditions, visual feature scarcity, or UWB signal obstruction scenarios. The mechanical system limitations will be addressed, covering the ongoing reliability challenges with the Stewart platform head mechanism despite improvements with heim joint implementations, the battery life constraints that limit extended operation periods and require careful power management, and the mechanical wear issues that affect long-term system reliability and require ongoing maintenance and replacement of components. The VR integration limitations will be examined, covering the network latency and reliability dependencies that can affect user experience during remote operation, the learning curve required for users to effectively operate the VR control system despite efforts to create intuitive interfaces, and the limited environmental feedback that restricts user awareness of the robot's physical surroundings and potential obstacles. The scalability challenges will be discussed, covering the manual calibration and setup requirements that limit rapid deployment in new environments, the complexity of system integration that requires technical expertise for maintenance and troubleshooting, and the current limitation to laboratory and controlled environments that restricts broader real-world deployment and validation.

6.5. Future Work and Research Directions

This section will outline promising directions for future research and development that build upon the foundations established in the Tino V2 project. The localization enhancement opportunities will be presented first, covering the potential for implementing fully autonomous UWB anchor deployment and calibration systems that reduce setup complexity and increase deployment flexibility, the exploration of additional sensor modalities such as LiDAR or thermal imaging that could further improve localization robustness and environmental perception, and the development of adaptive sensor fusion algorithms that dynamically adjust fusion parameters based on environmental conditions and sensor reliability. The VR interaction advancement possibilities will be detailed, covering the potential for implementing more sophisticated haptic feedback systems that provide users with tactile information about the robot's environment and interactions, the development of multi-user VR systems that enable collaborative control and interaction with the robot platform, and the exploration of augmented reality integration that combines real-world robot camera feeds with virtual overlay information for enhanced user experience. The human-robot interaction research directions will be examined, covering the potential for implementing more advanced social behavior algorithms that respond appropriately to detected human emotions and social cues, the development of learning systems that adapt robot behavior based on individual user preferences and interaction patterns, and the exploration of multi-robot coordination capabilities that enable collaborative social robotics applications. The broader deployment opportunities will be addressed, covering the potential for adapting the developed technologies for healthcare applications such as elderly care and patient monitoring, the exploration of educational applications that leverage the VR integration for remote learning and interaction, and the investigation of commercial applications that could benefit from reliable telepresence robotics with advanced perception and interaction capabilities.

6.6. Final Reflections and Conclusions

This section will provide final reflections on the Tino V2 project and draw overall conclusions about the research achievements and their significance for the field of social robotics. The project success assessment will be established first, covering the achievement of all major technical objectives including successful system migration, sensor fusion implementation, VR integration, and human detection capabilities, the validation of key hypotheses about the feasibility and effectiveness of combining multiple advanced technologies in a cohesive social robotics platform, and the demonstration of practical approaches for addressing common challenges in indoor mobile robotics and human-robot interaction. The methodological insights will be reflected upon, covering the value of systematic technology evaluation and selection processes for complex multi-component systems, the importance of modular architecture design for enabling effective development and debugging of advanced robotics systems, and the critical role of user-centered design principles for creating accessible and effective human-robot interaction systems. The research contribution significance will be assessed, covering how the Tino V2 project advances the state of the art in VR-integrated social robotics through practical demonstration of reliable technologies and effective integration strategies, the contribution to broader understanding of sensor fusion approaches for indoor robotics applications, and the validation of embedded AI optimization techniques for real-time perception and interaction capabilities. The legacy and impact considerations will be addressed, covering the foundation established for future social robotics research through open-source code, comprehensive documentation, and validated system architectures, the potential for broader adoption of demonstrated technologies and approaches by other researchers and developers, and the continued evolution toward more sophisticated and accessible human-robot interaction systems that effectively bridge virtual and physical environments for meaningful social applications.

Bibliography

List of Figures

List of Tables

Acknowledgements

Here you may want to acknowledge someone.