**CHAPTER-01**

**STATE OF ART**

**01 - A CNN pallet detection model for an autonomous forklift**

1.1 Introduction

Automation of forklifts is one important element in industrial logistics, driven by the demand for effectiveness, precision, and safety within warehouse operations. This article, "A CNN Pallet Detection Model for an Autonomous Forklift," considers the application of a convolutional neural network in the real-time detection of pallets based on data obtained from LiDAR sensors. It, therefore, underpins the advancements in artificial intelligence, sensor fusion, and robotics with accurate detection of pallets in this ever-changing environment of the industry.

While the manuscript represents a huge step in automated material handling, the methodologies proposed therein suffer from certain limitations that can be addressed in our project. This review describes the methodology, compares it with our own in this project, and highlights how our effort overcomes these limitations to provide a better solution.

1.2 Pallet Detection Using CNN

The study uses a CNN model that is supposed to detect pallets from LiDAR-derived point cloud data. The main elements of their method are as follows:

1.Neural Network Structure:

* One-dimensional CNN processes LiDAR readings, detecting pallet positions using a set of consecutive point cloud patterns.
* The architecture comprises five convolutional layers, integrates dropout techniques, and utilizes max-pooling to mitigate overfitting and enhance generalization capabilities.

2.Dataset Generation:

* The dataset consists of 2000 randomized scenes created in CoppeliaSim with obstacles and various pallet positions.
* LiDAR data is processed to identify the center and orientation of the pallet relative to the forklift.

3.Performance Indicators:

* The model does achieve an acceptable accuracy; however, MSE for test data does not mean it cannot be better.
* It tackles overfitting by reducing the number of neurons and adding dropout layers.

4.Limitations Acknowledged by the Authors:

* Dependence on a synthetic dataset limits its applicability to real-world scenarios.
* Overfitting persists even after making architectural changes.
* The model assumes the pallet is always visible and unobstructed, which is impractical in cluttered environments.

1.3 Overview

This research presents a convolutional neural network (CNN) based model for pallet detection in autonomous forklifts using data from LiDAR sensors to process point clouds. The model successfully identifies pallets in moving industrial environments, which enables precise localization and navigation. The study uses simulation-based data generation and discusses the integration of convolutional neural networks (CNNs) to address challenges around real-time pallet detection.

1.3.1 Main Contributions

1.Neural Network Design:

* Uses a 1D CNN architecture for LiDAR reading processing.
* It implements five convolutional layers with dropout and max-pooling to prevent over-fitting and to maximize generalization.

2.Simulated Dataset:

* Utilizes the CoppeliaSim platform to generate 2,000 randomized scenarios with variability in the position and angle of pallets and obstacles.

3.Performance Appraisal:

* Metrics like Mean Squared Error (MSE) depict the model's ability to correctly predict the position of pallets.
* Includes overfitting-prevention methods, such as fewer neurons in the layers and dropout layers.

4.Background and Assumptions:

* The forklift operates in controlled factory settings with stationary obstacles.
* The pallet remains continuously observable to the LiDAR scanner without any impediments.

5.Results:

* Demonstrates ability to locate pallets in structured environments.
* Highlights overfitting as a major problem in model training.

1.3.2 Limitations

1.Sensor Dependency:

* Relies solely on LiDAR sensors, limiting robustness in scenarios where point cloud data is incomplete or noisy.

2.Simulated Dataset:

* The model’s training on synthetic data may reduce its generalizability to real-world industrial environments.

3.Static Assumptions:

* Presumes the perpetual visibility of the pallet while considering the environment to be unchanging, thereby neglecting dynamic components such as mobile obstacles.

4.Less Pallet Variability:

* Restricted primarily to standard pallets and not readily accommodating to odd or damaged configurations.

5.Environmental Limitations:

* Engineered for controlled environments, with little consideration for chaotic or unorganized warehouse conditions.

6.Overfitting:

* Despite mitigation efforts, overfitting persists, impacting the model’s ability to generalize beyond training data.

1.4 How Our Project Overcomes These Limitations

| **Limitation in Paper** | **Our Approach** |
| --- | --- |
| **Sensor Dependency**: Relies on LiDAR alone. | Integrates advanced 3D cameras (Orbbec Gemini 2) and LiDAR for enhanced accuracy and robustness. |
| **Simulated Dataset**: Synthetic data reduces generalizability. | Combines real-world and simulated data for robust model training across diverse industrial scenarios. |
| **Static Assumptions**: Environment lacks dynamic adaptability. | Incorporates SLAM for real-time mapping and navigation in dynamic, cluttered environments. |
| **Limited Pallet Variability**: Focused on standard pallets. | Utilizes CNN models trained on diverse datasets, handling irregular and damaged pallet designs. |
| **Environmental Constraints**: Optimized for static conditions. | Adapts to unstructured and semi-structured settings with multi-sensor fusion. |
| **Overfitting**: Persistent challenges during training. | Applies advanced regularization techniques, augmented data, and ensemble modeling to mitigate overfitting. |

1.5 Comparison with Our Project

| **Aspect** | **This Paper** | **Our Project** |
| --- | --- | --- |
| **Detection Method** | CNN applied to LiDAR point clouds. | Multi-sensor fusion combining CNNs, 3D cameras, and LiDAR for enhanced detection accuracy. |
| **Environment Adaptability** | Assumes clear visibility and static setups. | Adapts dynamically to cluttered, unstructured environments using SLAM and real-time mapping. |
| **Data Robustness** | Simulated data only. | Integrates real-world and simulated data for robust model training. |
| **Performance** | Acceptable, with noticeable overfitting. | Sub-centimeter localization accuracy with optimized real-time performance using NVIDIA Jetson. |
| **Scalability** | Limited to specific pallet types. | Scalable for diverse pallet types, shapes, and configurations. |

1.6 Advantages of our Project

1.Better Detection:

* Uses high-resolution 3D cameras (Orbbec Gemini 2) and LiDAR for robust detection of various pallet types and orientations.

2.Dynamic Adaptability:

* It also incorporates SLAM for real-time navigation through cluttered and dynamic environments.

3.Practical Preparedness:

* Models are trained on a mix of real and simulated datasets to ensure their applicability in industrial settings.

4.Accuracy and Effectiveness:

* Achieves sub-centimeter localization accuracy and real-time operation with optimized edge computing.

1.7 Disadvanatages of our project

1.Augmented Complexity:

* Such multi-sensor fusion and heterogeneous algorithms require greater computational resources.

2.High Upfront Costs:

* Advanced sensors and processing units increase the initial investment.

1.8 Solutions to the Identified Problems

1. Dataset Generalization:

* By combining both the empirical and simulated datasets, our project shows resilience in very diverse industrial contexts.

2.Obstruction Handling:

* Uses 3D cameras and LiDAR for partially occluded pallet detection to improve reliability in cluttered environments.

3.Overfitting Reduction:

* Employs advanced regularization techniques and large datasets to enhance generalization.

4.Dynamic Mapping:

* Employs SLAM to adapt to dynamic environments for consistent performance in dynamic setups.

1.9 Conclusion

This paper was a significant step toward achieving the goal of autonomous pallet detection using CNNs and LiDAR point clouds. However, it suffers from its dependency on synthetic data, static assumptions, and single-sensor constraints that generally limit applicability in real-world conditions; our project has addressed those challenges through advanced sensing technologies, robust algorithmic frameworks, and enhanced adaptability. By fusing CNNs, SLAM, and multi-sensor systems, our solution provides superior accuracy, versatility, and operational efficiency, setting a new benchmark for the performance of autonomous forklifts in industrial automation.

**-----------------------------------------------------------------------**

**02 - Autonomous Pallet Localization and Picking for Industrial Forklifts Based on the Line Structured Light**

2.1 Introduction

Efficient pallet detection and handling systems are very important in modern industrial automation, especially for autonomous forklifts. The paper "Autonomous Pallet Localization and Picking for Industrial Forklifts Based on the Line Structured Light" explores a method using structured light sensors to achieve precise localization and pose estimation of the pallet. It then combines this with the analysis of geometric structure and position-based visual servoing to accomplish robust pallet detection and handling. Even though the research contributes much, it also simultaneously reveals certain limitations regarding applicability and scalability that our project aims to address.

2.2 Pallet Detection Using Structured Light Sensors

The research uses line-structured light sensors mounted on forklifts to determine the geometry and position of pallets.

2.2.1 The main components include:

1.Visual System Design:

* Uses embedded image processing boards with FPGA and DSP for real-time light stripe extraction.
* Applies Hessian matrix decomposition for robust detection of light stripe centers.

2. Pose Estimation:

* Detects pallet features using geometric matching and least squares line fitting.
* Uses ICP (Iterative Closest Point) for registration and pose estimation concerning a predefined pallet model.

3.Kinematic Control:

* Combines position-based visual servoing and instantaneous pose changes to enable forklift pallet docking in line with it.
* Facilitates consistent vehicular motion and accurate positioning in relation to pallets.

2.3 Overview

This research involves the development of a methodology for autonomous pallet handling using line-structured light sensors integrated with position-based visual servoing, both for pallet localization and forklift docking. The study provides a detailed implementation using embedded visual processing boards, hence accomplishing reliable pallet detection and pose estimation in structured industrial environments.

2.3.1 Key Contributions

1. Line Structured Light Sensors:

* Implemented a sensor that involves an FPGA-DSP combination to provide real-time processing.
* Extracts light stripe centers using Hessian matrix decomposition for accurate 3D information.

2.Pose Estimation:

* Combines ICP with model matching for precise pose alignment.
* Filters out noise and isolates collinear lines from 3D point data for pallet recognition.

3.Visual Servoing:

* Utilizes position-based visual servoing to align the forklift with the pallet, ensuring stable docking.
* Incorporates real-time visual tracking to adjust trajectory dynamically.

4.Validation:

* Demonstrated 90% success in simulations with random noise to simulate realistic deviations.
* Performance evaluated in structured environments, achieving accurate localization and docking.

2.3.2 Limitations

1.Distance Accuracy:

* Accuracy of structured light sensors decreases significantly with distance, hence limiting their effective range.

2.Environment Constraints:

* Designed for structured, uncluttered environments with minimal dynamic elements.

3.Complex Calibration:

* Requires precise calibration of laser planes and camera parameters, hence making deployment time-consuming.

4.Pallet Variability:

* Focused on standard pallets, with limited adaptability to diverse or damaged pallet designs.

5.Dynamic Scenarios:

* They fall apart in cluttered and unstructured environments because of the reliance on predefined conditions.

2.4 How Our Project Overcomes These Limitations

| **Limitation in Paper** | **Our Approach** |
| --- | --- |
| **Distance Accuracy**: Declines over 2 meters. | Utilizes 3D cameras (Orbbec Gemini 2) and LiDAR for extended range with consistent accuracy. |
| **Environment Constraints**: Limited to structured settings. | Employs SLAM for dynamic mapping, enabling operation in cluttered, unstructured environments. |
| **Complex Calibration**: Time-intensive setup. | Reduces calibration dependency using pre-trained models and multi-sensor fusion for adaptive precision. |
| **Pallet Variability**: Optimized for standard pallets. | Incorporates CNN models trained on diverse datasets to handle irregular and damaged pallets. |
| **Dynamic Scenarios**: Static assumptions limit applicability. | Adapts to dynamic setups using real-time multi-sensor feedback and advanced tracking algorithms. |

2.5 **Comparison to Our Project**

| **Aspect** | **This Paper** | **Our Project** |
| --- | --- | --- |
| **Detection Method** | Line structured light sensors. | CNN-based detection with multi-sensor integration (3D cameras and LiDAR). |
| **Localization Accuracy** | Decreases with distance. | Maintains sub-centimeter accuracy over a broader range using robust pose estimation algorithms. |
| **Environment Adaptability** | Designed for structured environments. | Handles dynamic, cluttered industrial setups with real-time adaptability through SLAM. |
| **Scalability** | Limited to standard pallet designs. | Supports diverse pallet configurations, including non-standard and damaged pallets. |
| **Real-Time Operation** | Relies on position-based servoing for adjustments. | Leverages NVIDIA Jetson for efficient real-time detection, navigation, and task execution. |

2.6 Advantages Our for Project

1.Enhanced Sensing:

Combines LiDAR and 3D cameras for increased accuracy and dependability under various circumstances.  
2.Dynamic Adaptability:

Makes use of SLAM to navigate in real time through crowded and shifting situations.  
3.Greater Applicability:

CNN models are trained on a variety of datasets, making it possible to detect non-standard pallets.  
4.Efficiency:

NVIDIA Jetson's real-time computing capabilities speed up operations.

2.7 Disadvantages of our project

1.Cost Increase

* More sophisticated sensors and processing systems necessitate a larger upfront expenditure.

2.Complex System

* Dynamic mapping and multi-sensor fusion increase system and computational complexity.

2.8 Similarities

| **Aspect** | **Analyzed Paper** | **Our Project** |
| --- | --- | --- |
| **Goal** | Autonomous pallet localization and picking for industrial forklifts. | Autonomous pallet detection, localization, and handling for industrial automation. |
| **Application** | Focused on improving warehouse logistics through automation. | Geared towards enhancing industrial logistics and material handling efficiency. |
| **Pose Estimation** | Utilizes ICP and geometric feature matching for pose estimation. | Incorporates RANSAC, DBSCAN, and ICP for robust pose estimation and accuracy. |
| **Sensor Technology** | Employs structured light sensors for 3D data acquisition. | Uses multi-sensor fusion with 3D cameras (Orbbec Gemini 2) and LiDAR for comprehensive data capture. |
| **Real-Time Processing** | Enables real-time trajectory correction during docking. | Processes data in real-time for detection, navigation, and path planning. |
| **Industrial Focus** | Aimed at structured warehouse environments with industrial forklift setups. | Optimized for dynamic, cluttered, and unstructured warehouse environments. |
| **Automation Level** | Offers automated pallet detection and alignment for forklifts. | Provides end-to-end automation, including load recognition, obstacle avoidance, and pallet handling. |
| **Data Processing** | Analyzes geometric features of 3D point clouds for localization. | Processes point clouds and RGB data with CNNs for enhanced precision and versatility. |

2.9 Conclusion

The study emphasizes geometric feature extraction and visual servoing while offering an effective method for pallet localization and docking utilizing structured light sensors. Scalability and robustness are, however, limited by its drawbacks, which include environmental restrictions, distance precision, and dependence on organized conditions. Through multi-sensor fusion, sophisticated computational frameworks, and improved flexibility, our initiative fills these gaps and guarantees increased accuracy, adaptability, and operational efficiency. Our proposal greatly exceeds the capabilities of current systems and establishes a new standard in industrial automation by utilizing cutting-edge technologies.

**----------------------------------------------------------------------**

**03 - Robotic Forklift for Stacking Multiple Pallets with RGB-D Cameras**

3.1 Introduction

Robotic Forklifts for Pallet Stacking Using RGB-D Cameras - Naturally, one way is handling pallet stack in warehouse logistic automation, since this task needs accurate localization and alignment of pallets. Here we conduct a literature review of one paper specifically titled, by Iinuma et al. that designs a robotic forklift system for multiple mesh pallet stacking via RGB-D cameras and compares to the one proposed that integrates independent algorithms and technologies into autopiloting forklifts.

3.2 Pallet Stacking Using RGB-D Cameras

RGB-D Camera Based Pallet Stacking Iinuma et al. introduces an autonomous robotic tugger and a pallet stacking system for mesh pallets. Let Us See Their Approach —

1.RGB-D Cameras:

* The Headlines Dense Depth Prediction The system uses several RGB-D cameras (in this case we use Intel RealSense Cameras) on the side of the forklift to get dense depth data. These cameras recognize the tiny feet and edges of the mesh pallets, which are significant for precise stacking

2.Region-Growing base Algorithm:

* Authors implement region growing-based algorithm to identify pallet feet and edges. This algorithm separates pallet side, not removing feet and edges as noise that is a common failure of other detection algorithms prevailing today (like RANSAC1).

3.Path Following Control:

* The forklift is controlled with the help of path following control so that the pallets can be oriented and situated according to our needs. This allows the feet of each pallet to sit aligned with at least one edge of the other pallet, allowing for accurate stacking.

4.Experimental Verification:

* To validate the paper, authors took real-time experiment to prove system performance which shows that they can stack pallet with very high level of accuracy.

3.3 Overview

This paper presents a robotic forklift system capable of automatically stacking several mesh pallets. It uses RGB-D cameras for pallet detection and orientation estimation, together with a region-growing algorithm for accurate pallet feet and edges detection. And it deals with the problems encountered in pallet stacking, including the small detction target and the accuracy of alignment without any environmental markings.

3.3.1 Main Contributions

1.Innovative Detection Methodology:

* Uses RGB-D cameras to generate dense point clouds, which enable accurate detection of the pallet feet and edges.
* Employs a region-growing algorithm for segmentation and extraction of pallet components from noisy depth data.

2.Position and Orientation Estimation:

* Uses the Iterative Closest Point (ICP) algorithm to align and transform point clouds into one coordinate system.
* Uses least-squares optimization techniques to get resilient pose estimation of the pallet.

3.Autonomous Stacking Procedure:

* Includes regulatory models for minimizing mistakes in lateral, angular, and longitudinal leveling.
* Applies path-following control techniques to achieve precise trajectory corrections in stacking processes.

4.Experimental Verification:

* Demonstrates a 90% success rate in performing stacking operations even with initial misalignments in position and orientation of pallets.
* Reaches lateral and angular errors ≤0.02 m and ≤0.05 rad, respectively.

3.3.2 Limitations

1. Sensor Constraints:

* RGB-D cameras are prone to reflective surfaces and changes in ambient lighting, which can affect the accuracy of detection.

2.Pallet Constraints:

* Focused exclusively on mesh pallets; lacks generalizability to different pallet types or materials.

3.Dynamic Environment Challenges:

* Designed for static environments with little clutter; effectiveness in dynamic, unstructured environments has not been explored.

4.Protrusion Detection:

* It is ineffective in identifying minimal pallet protrusions, which may result in stacking inaccuracies or contribute to hazardous operational conditions.

5.Computational Overhead:

* The requirement of multiple RGB-D cameras and region-growing algorithms increases the complexity and the resource requirements.

3.4 How Our Project Overcomes These Limitations

| **Limitation in Paper** | | **Our Solution** |
| --- | --- | --- |
| **Sensor Dependency**: RGB-D cameras struggle with small features and occlusions. | | We use advanced 3D cameras (e.g., Orbbec Gemini 2) and LiDAR for precise, clutter-tolerant detection. |
| **Limited Environmental Adaptability**: Tested only in controlled indoor settings. | | Our project operates in both structured and unstructured environments, with multi-sensor fusion for robustness. |
| **Stacking Failures**: Errors due to small undetected protrusions. | | Employing CNN-based feature detection ensures the system identifies even subtle features. |
| **Computational Overhead**: High resource demand for algorithms. | | Optimized detection and pose estimation algorithms (e.g., DBSCAN, RANSAC) balance precision and efficiency. |
| **Computational Overhead**: Complex multi-camera setup. | Employs optimized algorithms and edge computing (NVIDIA Jetson) for real-time processing. | |

3.5 Comparison with Our Project

| **Aspect** | **This Paper** | **Our Project** |
| --- | --- | --- |
| **Detection Technique** | Region-growing algorithm for pallet features | CNNs + point cloud processing for robust feature extraction. |
| **Pose Estimation** | Least-squares minimization | Advanced algorithms (RANSAC, ICP) for sub-centimeter accuracy. |
| **Sensor Technology** | RGB-D cameras | Multi-sensor setup: 3D cameras + LiDAR for enhanced accuracy and adaptability. |
| **Environment Adaptability** | Limited to controlled indoor setups | Operates in diverse and dynamic environments with clutter. |
| **Real-Time Performance** | Limited by computational overhead | Optimized algorithms ensure real-time detection, pose estimation, and navigation. |
| **Scalability** | Designed for specific stacking tasks | Scalable to various industrial workflows, including pallet detection, stacking, and navigation. |

3.6 Advantages of Our Project

1.Combination of Multi-task:

* Our project combines many tasks including pallet detection, load shape recognition, fork orientation and real time communication which provide more integrative solution for autonomous forklift operations.

2.Scalability:

* Multiple Algorithms (DBSCAN, ICP,RANSAC), 3D point cloud. The sense of robustness and flexibility gives the computer vision systems an edge in difficult and changing contexts. Industry CompatabilityFeature:protocols (TCP/IP, OPC) so it can fit to existing industrial systems and improve the contact in new industry background.

3.7 Disadvantages of Our Project

1.Disadvantages of Our Project Heterogeneous characteristics:

* The collaborative aggregation of heterogeneous algorithms and tasks increases the complexity and computational requirements across the entire system, driving up costs.

2.Reliance on Several Technologies:

* Systems are forced to implant a number of technologies (CNNs, DBSCAN, ICP, RANSAC and so on).

3.8 Conclusion

The paper by Iinuma et al. It presents a strong method for pallet stacking using RGB-D cameras and an adaption of the recent region growing based algorithm. In this project, we expands it by using a complete set of tasks and algorithms with target robust and flexible automated forklift system. Although our project has advantages in new scope and robustness, it also suffers from complexity, computing power overheads for R&D, training duration & validation. These will be challenges that we must tackle to ensure the success; maybe most importantly, reliability of our autonomous forklift system

**-----------------------------------------------------------------------**

**04 - Automatic Visual Guidance of a Forklift Engaging a Pallet**

4.1 Introduction

The paper “Automatic Visual Guidance of a Forklift Engaging a Pallet” by Seelinger and Yoder (2006) presents a vision-guided robotic forklift system that employs Mobile Camera-Space Manipulation (MCSM) to achieve autonomous pallet engagement.

4.2 Overview of Seelinger and Yoder’s Work

4.2.1 Key Contributions:

* Proposing a vision-based mobile manipulator control method using monocular cameras with the mobile camera space manipulation (MCSM).
* Prototype forklift with 2 cameras and pallet engagement accuracy using fiducial markers and Natural feature recognition as demonstrated
* Real-time path adjustment and target tracking based on polynomial parametric trajectory generation.

4.3 overviwq

This work presents the development of a vision-based forklift system through the Mobile Camera-Space Manipulation (MCSM) method for automatic engagement of pallets. The system aims to improve the accuracy of forklift tasks without camera calibration and is directed towards both Automated Guided Vehicle (AGV) systems and forklifts operated with the assistance of human drivers.

4.3.1 Fundamental Contributions

1.MCSM Technique:

* Extends the Camera-Space Manipulation (CSM) strategy to make onboard cameras available for accurate control.
* Eliminates the need for calibration, thus allowing dynamic adaptation to changes in environmental conditions.

2.System Execution:

* Uses dual cameras to achieve 3D perception without external markers.
* Features visual updates and trajectory correction to refine movement and engagement precision.

3.Forklift Prototype:

* A retrofitted forklift with vision sensors, actuators, and motion controllers to validate MCSM.
* Provides ±1 cm accuracy for pallet engagement.

4.Performance Indicators:

* Reaches 98% success rate using fiducial markers.
* Outlines strong trajectory planning for handling steer mechanical play.

4.3.2 Limitatins

1.Dependinig on Fiducials:

* Success is strongly dependent on the fiducial markers for proper pallet identification.
* Natural feature detection methods have limited testing and robustness.

2.Environmental Limitations:

* Designed for structured environments with minimal clutter or occlusions.
* Challenges due to changed conditions or reduced visibility.

3.Mechanical Limitations:

* Real-world accuracy is affected by prototype limitations like ±10° of steering play.

4.Computational Overhead:

* Requires significant computational power for real-time path editing and visual effects.

5.Generalizability:

* Standard pallet configurations and controlled industrial settings, in the main.

4.4 How Our Project Overcomes These Limitations

| **Limitation in Paper** | **Our Approach** |
| --- | --- |
| **Dependence on Fiducials**: Relies on artificial markers. | Utilizes CNN models and 3D point cloud analysis for marker-free pallet detection and localization. |
| **Environment Constraints**: Optimized for structured setups. | Incorporates SLAM to adapt dynamically to cluttered and unstructured environments. |
| **Mechanical Constraints**: Steering play impacts precision. | Employs precision steering mechanisms and dynamic feedback algorithms to enhance reliability. |
| **Computational Overhead**: High resource demands. | Leverages edge computing (NVIDIA Jetson) for efficient real-time processing. |
| **Generalizability**: Limited to standard pallets. | Trains detection models on diverse datasets to handle non-standard and irregular pallet designs. |

4.5 Comparison with Our Project

| **Aspect** | **Seelinger and Yoder’s Work** | **Our Project** |
| --- | --- | --- |
| **Camera Technology** | Two uncalibrated 2D cameras with fiducial markers | Orbbec Gemini 2 3D camera for depth sensing and RGB input |
| **Pallet Detection** | Fiducials and simple edge-detection algorithms | Deep learning-based object detection using CNNs |
| **Trajectory Control** | Polynomial trajectory generation | Adaptive AI-based trajectory optimization |
| **Real-Time Processing** | Open-loop visual feedback | Real-time inference with onboard processing (e.g., NVIDIA Jetson) |
| **Calibration Needs** | No calibration required for camera setup | Requires camera calibration for depth and RGB alignment |
| **Obstacle Avoidance** | Focused solely on pallet engagement | Includes dynamic navigation and obstacle avoidance |
| **System Robustness** | Limited robustness in unstructured environments | Designed for adaptability across various industrial settings |

4.6 Similarities

Vision-Guided Automation: The two projects are focused on automating forklift operations by implementing vision sensors for pallet detection and engagement. Instead, our framework uses 3D depth data from the scene which does not rely on fiducials and provides more robustness in untidy environments.

Forklift Autonomy — Both systems allow forklifts to autonomously engage pallets, minimizing the need for human interaction in material handling processes. Both are types of computer vision, but they differ on the sophistication of the algorithms and sort of vision sensor used.

Real-Time Adjustments Aesthetic notions of visual fidelity are prioritized in the simultaneous localization & mapping problem. While Seelinger and Yoder use polynomial trajectory generation, our project aims to integrate AI-driven adaptive optimization

4.7 Advantages of Our Project

1. 3D Depth Sensing: Our project makes use of Orbbec Gemini 2 3D camera which can precisely locate and detect pallets in various other types of environments without using fiducial marker. While pallet height, orientation and position, varies per payload these can be much better handled when dealing in depth data.

2. Integration of Machine Learning: Deep learning-based object detection models (CNN,point cloud segmentation) provide a strong and flexible method to achieve pallet detection in low signal to noise ratio or cluttered environments.

3. Project on Dynamic Navigation & Obstacle Detection We have implemented real-time obstacle avoidance and navigation capabilities in our project, enabling the forklift to navigate independently in a dynamic warehouse environment.

4. Edge Computing: Also great that we using NVIDIA Jetson hardware since it works processing vision and depth data fast real-time — low latency processing in the sense of efficiency.

5. Industrial scalability: The paper describes a prototype, but we built our system for industrial implementation and scaling.

4.8 Disadvantages of Our Project

1.Therefore, welfare studies using 3D camera systems and machine learning models are more costly and complex to implement than alternative simpler methods based on fiducialsBOX 2 [Bossens et al., 2022].

2.Camera Calibration — The camera calibration is part of our system since it re-aims the depth and RGB equally so that they work step by step of the set up.

3.Cost: More sophisticated 3D vision solutions could address this, but they come with prohibitive up-front costs for the measures (for example, a device equipped with a 3D camera and edge AI processing devices like NVIDIA Jetson).

4.9 Conclusion

Vision guided autonomous forklifts (Seelinger & Yoder): A round of the contest, vision guided autonomous forklifts show a possibility of using uncalibrated cameras for the pallet engagement. However, it cannot be used industrially mainly because it relies on fiducial markers and is only implemented in a simple prototype. We build on this idea through our project in three ways:

• Facilitating the development of robust multi-dimensional pallet detection using three-dimensional imaging technology

• Machine Learning-Based Solutions for Object Classification and Path Planning

• Navigate using dynamic obstacle avoidance in a real-world warehouse environment.

Our project, although certain aspects are accompanied by complications and price, as well as higher accuracy, scales significantly better in real world cases and is significantly more flexible in industrial deployments. This shift is a significant breakthrough in autonomously operated systems, as material handling system elements are chained together.

**-----------------------------------------------------------------------**

**05 - Research on the Method of Universal Forklift Pallet Detection and Pose Estimation Based on Visual and 3D Point Cloud Fusion**

5.1 Introduction

The paper “Research on Method of Universal Forklift Pallet Detection and Pose Estimation Based on Visual and 3D Point Cloud Fusion” (by Jingxin Gu et al.) presents a cutting-edge system for detecting and estimating the pose of pallets, which leverages the capabilities of an rgbd camera, along with a modified YOLOv8-Pose model to achieve precise readings of pallets. We then return to discuss this research in the context of our own project in a comparative literature review below.

5.2 Overview of the Study

This study proposes a novel method, which combines YOLOv8-based visual data analysis with 3D point cloud fusion, for the identification and orientation estimation of pallets in autonomous forklift tasks. It uses depth cameras to capture data and exploits the power of advanced deep learning algorithms to achieve high accuracy in pallet localization, overcoming the challenges of light variations and the heterogeneity of pallets.

5.2.1 Key Contributions:

* Fuses RGB with 3D point cloud to effectively combat challenges in pallet detection under worsen environments (i.e., occlusion, illumination changes).
* Improves pallet feet localization, using an improved YOLOv8-Pose model for object and keypoint localization. o Real-time inference with TensorRT acceleration enables it for industrial applications
* High accuracy (≤1 cm and ≤1° errors) in pallet position, and orientation detection in < 2m.

5.2.2. Advantages:

* Strong detection performance for various sizes and types of pallets, without fiducial markers.
* Merges RGB and depth data for more robust pose estimation
* The keypoints present in each box ensure correct alignment with forklift forks, enabling functionality in dynamic operating environments.

5.2.3. Limitations:

* Over 2 meter range the performance of the system gets worse due to low depth camera accuracy at higher distance.
* Pre-trained YOLOv8 models are used and training might need a lot of compute power.

5.3 Overview

This study proposes a novel method, which combines YOLOv8-based visual data analysis with 3D point cloud fusion, for the identification and orientation estimation of pallets in autonomous forklift tasks. It uses depth cameras to capture data and exploits the power of advanced deep learning algorithms to achieve high accuracy in pallet localization, overcoming the challenges of light variations and the heterogeneity of pallets.

5.3.1 Main Contributions

1.Integration of Visual and Depth Information:

* Uses an RGB-D camera to combine color images with depth information, creating accurate 3D point clouds.
* The resulting merge allows for precise pallet foot localization and pose estimation.

2.Improved YOLOv8 Algorithm:

* YOLOv8 is trained with over 2,000 images, annotated with key points for robust pallet detection.
* The model predicts bounding boxes and key points, which allows for improved detection and localization of pallets.

3.Point Cloud Processing:

* Applies RANSAC for plane fitting and extracts geometric features to detect the pallet feet.
* Uses Euclidean clustering to optimize localization accuracy.

4.Performance Metrics:

* Realizes the pallet localization distance error ≤1 cm and the angular error ≤1°.
* Operates on data in real time, with an average detection speed of 700 milliseconds per operation.

5.Experimenaal Verification:

* Shows strong performance in industrial settings, reaching high precision without the need for artificial markers or particular pallet setups.
  + 1. Limitations

1.Sensor Dependency:

* Relies Heavily on RGB-D cameras, which might suffer from challenges such as reflective, low-light, or high-variance depth-measurement conditions.

2.Distance Limitations:

* Effective range limited to 2 meters, with reduced accuracy for longer distances due to noise in depth data.

3.Environmental Assumptions:

* Designed for controlled industrial settings with small dynamic components, thereby providing little flexibility to chaotic or unplanned situations.

4.Computational Overhead:

* Visual and depth information increases the processing complexity, which can become an obstacle for scalability in applications demanding real-time performance.

5.Pallet Variability:

* While tested on multiple pallet types, the method is optimized for standard configurations with limited validation for irregular or damaged pallets.

5.4 How Our Project Overcomes These Limitations

| **Limitation in Paper** | **Our Approach** |
| --- | --- |
| **Sensor Dependency**: Relies on RGB-D cameras. | Incorporates 3D cameras (e.g., Orbbec Gemini 2) and LiDAR for improved depth accuracy and environmental robustness. |
| **Distance Constraints**: Limited to 2 meters. | Combines LiDAR with 3D cameras to extend detection range while maintaining high accuracy. |
| **Environmental Assumptions**: Optimized for static setups. | Uses SLAM for real-time navigation in dynamic and cluttered environments. |
| **Computational Overhead**: Visual-depth data fusion adds complexity. | Optimizes detection pipelines with edge computing using NVIDIA Jetson for faster real-time performance. |
| **Pallet Variability**: Focused on standard pallets. | Leverages CNN-based detection trained on diverse datasets for broader pallet adaptability. |

5.5 Comparison with Our Project

| **Aspect** | **This Study** | **Our Project** |
| --- | --- | --- |
| **Sensor Technology** | Depth camera (RGB-D) | 3D camera (e.g., Orbbec Gemini 2 or Intel RealSense D435i) |
| **Detection Method** | Enhanced YOLOv8-Pose for keypoint and bounding box detection | Deep learning (e.g., MobileNet or YOLO) with 3D feature integration |
| **Data Fusion** | Combines visual and 3D point cloud data | Combines depth and RGB for enhanced pallet and environment mapping |
| **Pose Estimation** | Keypoint-based pose estimation using RANSAC | Keypoint-based or geometric constraint methods (simpler alternatives initially) |
| **Inference Speed** | Accelerated with TensorRT for real-time detection | Edge computing (e.g., NVIDIA Jetson) for real-time performance |
| **Accuracy** | ≤1 cm and ≤1° within 2 meters | Targeting ≤2 cm and ≤2° within 3–5 meters |
| **Cost** | Moderate (depth camera + YOLO model) | Optimized for cost efficiency with affordable sensors |
| **Adaptability** | Generalized for various pallet types | Initially optimized for EUR pallets, expandable to others |

5.6 Similarities

Combining visual and depth data: Both systems use depth cameras to fuse RGB data with 3D point cloud returns, leading to improved accuracy in pallet detection and localization.

Real-Time Processing: They mainly focus on real-time performance through efficient inference processes while our project seeks for low-cost edge computing.

Reliable Pallet Detection: In this study, deep learning models such as YOLO models and in our case, using MobileNet or similar deep learning models ensure distinctiveness for different pallet types as well as environmental conditions.

Industrial Relevance:: Both projects are relevant at industrial scale, as they fulfil an existing need by automating processes for logistics and warehousing, which require high productivity and reliability.

5.7 Advantages of Our Project

1. Cost Efficiency:

* We use less expensive sensors such as the Intel RealSense or Orbbec Gemini 2 that allow us to balance cost with functionality.
* Lightweight deep learning architectures (e.g., MobileNet) reduce training complexity and remove the need for heavy-weights computations.

2. Scalability:

* Our modular approach scales easily for other pallet types and environments although it was initially optimized for EUR pallets

3. Extended Range:

* Our system can work at relatively longer distance (up to 3–5 meters), suitable to bigger warehouse designs.

4. Dynamic Navigation:

* Provides obstacle avoidance and SLAM at the same time for automated forklift navigation without the fixed detection range of the study .

5.8 Disadvantages of Our Project

1. Reduced Initial Accuracy:

* Our first implementation aims ≤2 cm as opposed to the ≤1 cm that YOLOv8-Pose provides due to cost and computation trade-offs.

2. Simplified Pose Estimation:

* Our project may suffer with irregular pallet shapes or occlusions because we chose simple geometric constraints to be applied first.

3. Real-Time Performance:

* Edge computing devices such as the NVIDIA Jetson do not have as much inference speed as TensorRT acceleration applied to high end GPUs, and

5.9 Conclusion

The contributions of this work will be helpful for visual and depth data fusion in order to detect and estimate the pallet pose. This shows the effectiveness YOLOv8-Pose in achieving high accuracy and robustness in industrial environment.

Our work extends these concepts while focusing on cost, scalability and range. Although the system referenced serves as a gold standard for accuracy and speed, our project focuses on an economical yet performative solution for systems considering automation in real-life applications. The two systems each push the work of automating a forklift forward, with this study focusing on precision and our project taking an approach aimed at accessibility.

**----------------------------------------------------------------------------------------------------------------**

**6 - Pallet Localization Techniques of Forklift Robot: A Review of Recent Progress**

6.1 Introduction

Pallet localization is essential to perform automatic warehouse logistics using autonomous forklifts and Automated Guided Vehicle (AGV) with higher precision. Pallet localization is the problem of localizing pallets in time-varying environments, where lighting and space are not controlled, and where there may be different types of pallets.

6.2 Localized Pallet Challenges

1. Variability in Pallet Types: Pallet designs, sizes, and materials are so diverse across industries that developing a master universal localization approach is challenging.
2. Environmental Complexity: Warehouses tend to have poor lighting, dynamic obstacles, and clutter which makes visual-based detection difficult.
3. Technological Constraints: Conventional techniques such as RFID and infrared sensors involve high-cost hardware apparatus and have limited flexibility and scalability**.**

6.3. Methods for detection and localization of pallets

* + 1. Visual-Based Detection : Li et al. revisits seminal DNNs: YOLO (You Only Look Once): An algorithm that gives fast detection by considering it regression but at the cost of accuracy for speed.

1. SSD (Single Shot Multi-box Detector): It is a model with compromises between speed and accuracy which achieves scores above 98% in the detection.
2. Faster R-CNN: Provides precise localization by creating region proposals ahead of classification yet is also computationally expensive. O
3. Highlights: SSD and Faster R-CNN provide the best accuracy, whereas YOLO comes for real-time apps.
   * 1. Point Cloud-Based Detection :
   1. 2D Point Cloud Data: Transforms 2D laser rangefinder (LRF) data to images which is mildly affected by light but is contingent on good strong algorithms.
   2. 3D Point Cloud Data: Performs spatial recognition through template matching with plane segmentation, severely constrained by processing speeds and sensitivity to noise.
      1. The PILA Strategy

Li et al. proposed the Pallet Identification and Localization Algorithm (PILA). integrates:

* + RGB Image Detection: Implements DNNs to detect pallets and locate Regions of Interest (RoIs).
  + Point Cloud Alignment: Aligns the 3D point cloud data onto the detected RoIs to calibrate their poses accurately.
  + Geometric Feature Extraction : which extracts the spatial features to compute the coordinates and orientation of the pallet. Performance:
  + Localization accuracy of ability to resolve 1 cm, and angular precision of 0.4° at 3m. • Works in<700 ms, showing applicability to real-time scenarios.

6.4 Overview

This manuscript provides a comprehensive survey on different pallet localization methods, used in forklift robotics, comparing approaches based on RGB images, depth images, and point cloud processing. The manuscript highlights the challenges involved in achieving high accuracy and efficiency in industrial environments while also presenting the limitations of single-sensor approaches.

6.4.1 Main Contributions

1. Methodology Evaluation:

* Provides some different comparisons, with a focus on YOLO, SSD, Faster R-CNN for objects' detection in pallet counting of RGB imagery.
* Investigates the use of point cloud segmentation for accurate pallet localization.

2.Pallet Identification and Localization Algorithm (PILA):

* Combines RGB image detection and point cloud processing to achieve high accuracy with a low computational cost.
* Demonstrates 3D localization precision of ≤1 cm and a mean orientation error of ≤0.4° with a 3-meter standoff.

3.Dataset Diversity:

* Includes different pallet types (wooden, plastic) and scenarios (tilted pallets, occlusions, varying lighting conditions).

4.Experimental Verification:

* Compared PILA with commercial solutions, showing its advantage in speed (700 ms per operation) and precision.

6.4.2 Limitations

1.Sensor Limitations:

* RGB and depth sensors are prone to lighting changes and reflective surfaces that might lower the detection reliability.

2.Range Constraints:

* The effective range is limited to 3 meters, reducing applicability for larger warehouse setups.

3.Static Environment Assumptions:

* Designed for constrained environments, it limits adaptability to dynamic and cluttered spaces.

4.Pallet Variability:

* Tested mostly on standard pallets, not generalizable to non-standard or damaged pallets.

5.Computational Overhead:

* Requires significant computational resources for fusing RGB and point cloud data, challenging scalability.

6.5 How Our Project Overcomes These Limitations

| **Limitation in Paper** | **Our Approach** |
| --- | --- |
| **Sensor Limitations**: Sensitive to lighting variations. | Multi-sensor integration (Orbbec Gemini 2 and LiDAR) ensures robustness in diverse environmental conditions. |
| **Range Constraints**: Limited to 3 meters. | Combines LiDAR and 3D cameras to extend detection range while maintaining accuracy. |
| **Static Environment Assumptions**: Focused on controlled setups. | Incorporates SLAM for dynamic mapping and real-time adaptability in cluttered, unstructured spaces. |
| **Pallet Variability**: Tested on standard pallets only. | Employs CNN models trained on diverse pallet types, including irregular and damaged pallets. |
| **Computational Overhead**: Resource-intensive data fusion. | Optimized algorithms for edge computing (NVIDIA Jetson) to ensure real-time processing. |

6.6 Comparison with Our Project

| Aspect | Li et al.'s PILA Strategy | Our Project |
| --- | --- | --- |
| Sensor Technology | RGB camera + 3D depth sensor | 3D point cloud sensors (e.g., LiDAR, Orbbec Gemini 2) |
| Detection Method | DNNs (e.g., YOLO, SSD) + Point Cloud Alignment | CNNs + DBSCAN, ICP, RANSAC, and feature-based alignment |
| Localization Accuracy | 1 cm at 3 meters | Targeting ≤2 cm across 3–5 meters |
| Real-Time Operation | Operates within 700 ms | Designed for real-time processing via edge devices (e.g., Jetson) |
| Scope | Pallet detection and localization | Comprehensive (pallet detection, load recognition, fork alignment) |
| Technological Integration | Combines RGB with point cloud | Relies solely on 3D point cloud for higher spatial fidelity |
| Hardware Cost | Low to moderate | Moderate due to advanced sensors and computing |

6.7 Advantages of Our Project

1. All In One Task Inclusion: Pallet detection, load shape recognition, fork alignment, real-time communication with industrial systems.

2. Algorithmic Flexibility: Adaptable and robust pose estimation using combination of DBSCAN (clustering), ICP (Iterative Closest Point) and RANSAC (Random Sample Consensus).

3. Industrial Communication in Real-Time: Includes protocols for truck-to-warehouse management system communication

6.8 Disadvantages of Our Project

1.Complexity and Cost: Widely depends on the algorithms that are computationally intensive and high-precision hardware increasing the costs in comparison to the PILA approach.

2. Training and Validation Pipeline Overhead: Need a lot of training datasets for CNNs and need testing in multiple environments to ensure reliability.

3. Depend on multiple components: Bringing together multiple technologies adds to the risk of bottlenecks if a particular component does not perform.

6.9 Similarities with Li et al. 's Work

1. Fusion of Data Modalities: Multi-modal data (RGB or 3D point clouds) can be utilized to enhance accuracy in both oes.

2.Industrial Relevance: Designed real time warehouse operation focusing high performance & scalability

3.Detection Accuracy: Patch Location (Experiments 4–5): Both works achieve ≥ 1–3 cm accuracy for localization.

6.10 Conclusion

Li et al. and shows a promising result of their PILA (Pallet Localization from 3D Point Clouds and RGB Images) strategy by fusing 3D point cloud images along with RGB images to provide accurate and fast palettes localization. Although detection and localization are also covered by their system with high accuracy while using low-cost hardware, their system has a relatively narrow focus.

We build on this by incorporating additional tasks, including load shape recognition, fork alignment, and real-time industrial communication. These refinements, though, broaden the applicability of our system to more diverse and dynamic large-scale industrial environments at the expense of additional complexity and computational demand. If we keep optimizing our algorithm and hardware to address these challenges, our autonomous forklift system would sound and less risky to use, so this needs to be addressed quickly as well**.**

**-----------------------------------------------------------------------**

**07 - Real-time Pallet Localization with 3D Camera Technology for Forklifts in Logistic Environments**

7.1 Introduction

However, fast pallet localization methods would enable the development of autonomous forklifts that improve logistics in such environments. The ideal scenario is to have an automated warehouse scenario where pallet detection and localization is done accurately despite challenges in varying orientations and dynamic environment up to an degree of precision which is a priority. This reviews the literature to discuss the work of Molter and Fottner who developed a geometric pipeline using time-of-flight (ToF) camera technology for localization of pallets and compares it to our project — highlighting the similarities, differences, and suggested improvements**.**

7.2 The problem of Pallet localization from a first principle perspective

Pallet localization presents its own challenges:

• Variability: There are different pallet types in terms of size, shape, and orientation, causing movement detection methods to be less universally applicable.

• Dynamic conditions: Forklifts navigate among clutter and in low-light scenarios

• Alignment Needs: When using forks, they must be aligned properly otherwise goods or pallets can be effcted**.**

7.3 Main Contributions of The Work of Molter &amp; Fottner

1. Geometric Detection Pipeline:

* 3D point cloud acquisition with Microsoft Kinect v2 ToF camera.
* Analyzed point clouds to find surfaces corresponding to pallet blocks to allow for localization without fiducial markers.
* Working only with euro pallets with set dimensions.

2.Processing Steps:

* Implemented RANSAC to detect and crop ground planes
* Apply region growing algorithms for surface extraction from point clouds.
* Merged geometric constraints (i.e., block spacing) for detection of pallet front blocks and pose estimation.

3.Performance:

* Successfully localized in real-time at 15–18 frames per second (FPS).
* Identified, parcels positioned at angles upto 90° from the sensor plane.
* All errors for the localization were below reasonable expectations for a real-world application**.**

7.4 Overview

The current paper proposes a novel approach for the detection and localization of pallets using a time-of-flight (ToF) camera. The solution solves issues where pallets are oriented in various angles up to 90° regarding the forklift's direction of movement. The pipeline uses geometrical features for pallet block recognition and is artificial marker-free, thus suitable for real-time operation of forklifts.

7.4.1 Key Contributions

1. Geometrical Localization and Detection

* Uses RANSAC to find ground planes and filter point clouds by geometrical features.
* Segments wooden pallet blocks in the point cloud using region-growing algorithms.

2. Pose Estimation:

* Computes pallet orientation (yaw angle) given Cartesian coordinates of pallet block centroids.
* Reduces degrees of freedom by assuming pallets lie flat on the ground.

3. Real-Time Processing:

* Processes depth data into point clouds using the Point Cloud Library (PCL) at an average frame rate of 15 fps.
* Dynamic pallet detection while the forklift is in motion.

4.Experimental Results:

* Localisation Errors Achieved ≤2 cm (x-axis) and ≤1° (yaw angle) for Both Static and Dynamic Setups.
* Worked well with pallets at 45° and 90° angles to the camera.

7.4.2 Limitations

1.Sensor Dependence:

* The approach is heavily reliant on Microsoft Kinect v2, which may be problematic in an environment with reflective or transparent material.

2.Distance Constraints:

* Effective range is limited to 2 meters due to the ToF camera's depth resolution and noise at longer ranges.

3.Static Environment Assumptions:

* The pipeline assumes minimum dynamic obstacles in the field of view, which limits adaptability to cluttered or dynamic warehouse environments.

4.Pallet Variability:

* Focused exclusively on standard EUR pallets, which limits generalizability to irregular or custom pallets.

5.Noise Sensitivity:

* Point cloud segmentation and pallet detection are sensitive to noise, especially in dynamic scenarios.

7.5 **How Our Project Overcomes These Limitations**

| **Limitation in Paper** | **Our Approach** |
| --- | --- |
| **Sensor Dependency**: Relies on Kinect v2, limited in industrial settings. | Integrates advanced 3D cameras (Orbbec Gemini 2) and LiDAR for superior accuracy under diverse conditions. |
| **Distance Constraints**: Limited to 2 meters. | Combines LiDAR and 3D cameras to extend detection range while maintaining precision. |
| **Static Environment Assumptions**: Optimized for low dynamic activity. | Incorporates SLAM for real-time navigation and mapping in dynamic and cluttered environments. |
| **Pallet Variability**: Tested only on standard EUR pallets. | Uses CNN-based models trained on diverse pallet types for adaptability to custom and irregular pallets. |
| **Noise Sensitivity**: Prone to segmentation errors. | Employs noise-robust algorithms like DBSCAN and RANSAC for improved segmentation and localization. |

7.6 Comparison with Our Project

| Aspect | Molter and Fottner | Our Project |
| --- | --- | --- |
| Sensor Technology | Microsoft Kinect v2 (ToF) | Orbbec Gemini 2 or Intel RealSense D435i (advanced 3D cameras) |
| Detection Method | Geometric surface detection using predefined dimensions | Deep learning with CNNs integrated with 3D point cloud features |
| Performance | 15–18 FPS with limited range (≤2 m) | Targeting 15+ FPS with extended range (up to 3–5 m) |
| Localization Accuracy | Static: ~1 cm, Dynamic: slightly higher | Similar accuracy but more robust in varied environments |
| Adaptability | Limited to euro pallets | Generalizable across various pallet types and warehouse setups |
| Algorithmic Approach | Region growing + RANSAC | DBSCAN + ICP + RANSAC, integrating AI models |
| Scope | Localization only | Comprehensive (includes fork alignment and obstacle avoidance) |
| Cost | Low-cost with Kinect | Moderately higher with advanced sensors and edge processing |

7.7 Advantages of Our Project

1.Extended Detection Range: Whereas Molter and Fottner’s work targets within 2 meters, our project targets the range 3–5 meters to be able to be used in larger setups, such as warehouses.

2. Strategic Algorithmic Implementation: Using more robust ways to group (DBSCAN) more robust pose estimation (ICP, RANSAC) in clutter environment and dynamic environment.

3.Deep Learning Capability: Utilizes convolutional neural networks (CNNs) for general-purpose pallet detection on different designs and orientations, which reduces the dependence on fixed geometrical assumptions.

4.Dynamic Navigation: Combines SLAM for autonomously moving around and avoiding obstacles, which adds more functionality than just localization

5.Enhanced Flexibility: Adaptable to various warehouse environments and pallet-types, which offers a scalable solution for industrial use-cases**.**

7.8 Disadvantages of Our Project

1.Higher Cost: Advanced hardware for 3D cameras and real-time processing (NVIDIA Jetson) makes the aggregate cost of the system higher than Molter and Fottner’s approach, which was based on a Kinect.

2.Algorithmic Complexity: Combining different algorithms (e.g., CNNs, ICP, RANSAC) is complex, demanding, high computational power, and higher development effort.

3. Training & validation overhead: Large datasets/validation required – deep learning models often take much longer to develop, and require extensive validation (increased time/resources).

7.9 Similarities

1. Real-Time Operation: Our system and 25-Hz system supports real-time detection, which is essential for dynamic forklift applications.

2.Focus on Industrial Contexts: Originally manufactured for warehouse applications, ideal for efficient and safe pallet handling.

3.Point Cloud Utilization: Localization for both projects is based on 3D point cloud data, but the methodology for processing point cloud data differ.

7.10 Conclusion

Molter and Fottner's work entail a low-cost, efficient method for pallet detection using ToF cameras, with an emphasis on real-time performance and ease of implementation. Challenges: It caters to specific environments only as it works with geometrical assumptions that are defined beforehand and its range is limited.

Our project fills these voids by:

• Advanced algorithms for improved detection range and robustness

• Utilizing deep learning for pallet detection and SLAM for navigation.

• Supporting large-scale industrial applications with more flexibility and scalability.

Although our solution comes at a higher cost and with higher development complexity, these trade-offs allow a more flexible and complete answer to the automation of warehouse processes. This comparison is key to show the complementary strengths of both approaches but also to show how our project is an innovation.

-----------------------------------------------------------------------

**08 -** **Pallet detection and localization with RGB image and depth data using deep learning techniques**

8.1 Introduction

Performing pallet detection and localization efficiently is a key enabler when implementing material handling systems (i.e., unmanned forklifts) within warehouses. However, researchers have come up with pioneering approaches in terms of improving localization accuracy, speed and adaptiveness with advances in RGB-D camera and deep learning algorithm.

8.2 Introduction to the Challenges of Pallet Localization

1. Environmental Complexity:

* Various pallet size, composition and orientation differences
* Variable warehouse environments; low lighting or overcrowded conditions

2. Accuracy and Speed:

* Sub-centimeter life speeding up Processing in Nesting is an Autonomous Forklifts Operation.
* These requirements are hard to be satisfied, particularly in dynamic environments, by conventional methods.

3. Integration Needs:

* However, it still remains a very hard problem to perform 3D localization from a combination of RGB and depth.

8.3 Key Contributions of PILA

1. Integrated Approach:

* Integrates a DNN-based architecture for pallet detection in RGB images with a process of 3D point cloud analysis for accurate localization.
* Uses relatively cheap RGB-D cameras, making the entire approach practical and scalable.

2. Algorithm Workflow:

* Stage 1: Bounding box recognition (pallet detection) in RGB images with single shot multibox detector (SSD).
* Stage 2: RGB and depth data fusion for determining region of interest (RoI).
* Stage 3: Extraction of the plateau cloud and analysis of geometrical characteristics to calculate the shelf orientation and location.

3.Performance:

* 1 cm localization accuracy and 0.4° angular resolution at a 3 m distance
* Works in 700 ms, enables on-time forklift action

4. Advantages:

* No Need for Fiducial Markers — tl;dr
* Compatible with any pallet type and material.
* Used to find a good trade-off between accuracy and computational time.

8.4 Overview

This paper presents the Pallet Identification and Localization Algorithm (PILA) that combines RGB image recognition by a deep neural network (DNN) with point cloud processing for pose estimation. The technique has reached high accuracy at a high speed, so it is suitable for real-time application in an autonomous forklift for logistics.

8.4.1 Main Contributions

1.Two-Stage Detection and Localization:

* Stage 1: RGB images are passed through SSD (Single Shot MultiBox Detector) for pallet detection.
* Stage 2: The depth information corresponding to the same ROI is further processed for the estimation of the 3D position and orientation of the pallet.

2.Algorithm Design:

* The pipeline is composed of RGB-depth image fusion, RANSAC-based plane fitting, and geometrical feature extraction.
* It detects the centroids of pallets and estimates the inclination angles.

3.Performance Metrics:

* Accuracy: The average localization error of ≤1 cm and orientation error of ≤0.4°.
* Speed: Up to real-time with 700 ms processing time per detection.

4.Experimental Verification:

* Tested on EUR pallets as well as on plastic pallets with an accuracy in detection of over 98%.
* Shows to be robust in multi-pallet detection in cluttered warehouse environments.

8.4.2 Limitations

1.Sensor Dependency:

* Is based on low-cost RGB-D cameras, whose range is limited and that are prone to lighting and surface reflectivity.

2.Distance Constraints:

* The practical detection range is approximately 3 meters, and the accuracy falls off rapidly beyond that.

3.Static Environment Assumptions:

* Is optimised for controlled indoor environments with minimum dynamic elements.

4.Pallet Variability:

* Mostly tested on standard pallets, thus having limited generalizability to irregular or damaged pallet designs.

5.Computational Requirements:

* Fusion of RGB and depth data and further point cloud processing might be problematic for scaling up to large datasets or speedier operations.

8.5 How Our Project Overcomes These Limitations

| **Limitation in Paper** | **Our Approach** |
| --- | --- |
| **Sensor Dependency**: Relies on low-cost RGB-D cameras. | Integrates advanced 3D cameras (Orbbec Gemini 2) and LiDAR for robust detection under diverse conditions. |
| **Distance Constraints**: Limited to 3 meters. | Combines LiDAR and 3D cameras to extend the detection range while maintaining accuracy. |
| **Static Environment Assumptions**: Optimized for controlled setups. | SLAM-enabled navigation ensures adaptability in dynamic, unstructured environments. |
| **Pallet Variability**: Focused on standard pallets. | CNN models trained on diverse pallet types ensure adaptability to various designs, including irregular shapes. |
| **Computational Demands**: Resource-intensive processing. | Leverages edge computing (e.g., NVIDIA Jetson) for efficient real-time performance. |

8.6 Comparison with our project

| Aspect | Yao-yong Li et al. (PILA) | Our Project |
| --- | --- | --- |
| Sensor Technology | RGB-D cameras (e.g., Pico Zense) | Advanced 3D cameras (e.g., Orbbec Gemini 2, Intel RealSense) |
| Detection Method | SSD for RGB images + Point cloud analysis | CNNs (e.g., MobileNet) + DBSCAN, ICP, RANSAC for point cloud |
| Performance | Real-time (700 ms) | Real-time (~15 FPS) with edge computing |
| Localization Accuracy | 1 cm, 0.4° (≤3 m range) | 2 cm, ~1° (3–5 m range) |
| Adaptability | Robust across pallet types | Generalizable to varying pallets and dynamic environments |
| Scope | Pallet detection and localization | Comprehensive (includes navigation, load alignment) |
| Cost | Low-cost RGB-D setup | Moderately higher with enhanced sensors and processing power |

* 1. Advantages of Our Project

1. Extended Detection Range: Pila works within 3 metres range, but is capable only of networks set up in smaller storage areas, whereas our project focuses on range of 3–5 metres for larger warehouse configurations.

2. Integration of Enhanced Algorithms : Includes strong algorithms such as DBSCAN (clustering), ICP (Iterative Closest Point), and RANSAC to deal with the noise in point cloud maps & generalized environments.

3. Deep Learning Adaptability: Leveraging generalizable deep learning models (e:g: MobileNet) that can learn new pallet types and orientations with little retraining

4. Dynamic navigation and communication; Integrates SLAM for autonomous navigation and supports industrial communication protocols (e.g., TCP/IP, OPC)

8.8 Disadvantages of Our Project

1.Higher Cost: High forces of advanced sensors and edge devices (like NVIDIA Jetson) increase initial investment vs low-cost PILA setup;

2.Algorithmic Complexity: Integrating deep learning with geometric analysis and real-time SLAM is computationally expensive as well as developmentally intensive.

3. Training and Validation Overhead: Deep learning model needs comprehensive dataymas, and it is time-consuming to build and validate them**.**

8.9 Similarities

1.Real-Time Capability: Industrial forklift pallet specifications require both systems be real-time capable.

2.Sensor Fusion: Combines RGB data with depth to enhance accuracy and robustness.

3.Industrial Application: Targets real-world use cases in warehouses and logistics use cases (solving operational problems).

8.10 Conclusion

The paper by Yao-yong Li et al. the combination of RGB-D data and SSD for real-time pallet detection and localization. Their solution is cheaper, and more accurate, though limited every couple of hundred of meters, and lacks navigation.

Our project extends these ideas by:

* + the furthering of operational benefits and resilience across advanced sensors and algorithms.
  + Doing fork alignment and other navigation tasks for full automation
  + Working on more flexibility to different warehouse environment and pallet types.

Our solution comes at the expense of higher costs and complexity, but it allows us to deliver a more reliable and scalable solution for the future of logistics automation.

**-----------------------------------------------------------------------**

**09 - Pallet Detection and Localization Based on RGB Image and Point Cloud Data for Automated Forklift**

9.1 Introduction

Pallet detection and Localization automation is crucial for design to enable efficiency and safety for the warehouse logistic. In particular, RGB-D imaging (Chai et al., 2021), point cloud processing (Yuan et al., 2021) and deep learning (Deng et al., 2022a) have greatly improved the accuracy and robustness of such systems. In this review article, we analyze a study by Dianjun Fang et al. that presents an integrated framework for pallet pose estimation, based on YOLOv5s and point cloud processing. Identifying the similarities and differences in our project with this study cannot only find the similarities but also highlight the advancements which happen for our project in between this research and ours.

9.2 Pallet Localization Challenge

1. Environmental Complexity:

* Pallet detection pose systems needs to work vigourously in the twitchy palcy and cluttered warehouse environment.
* The differences in pallet size, material, and the position of the pallets increases the difficulty in estimating the pallet’s pose suitably.

2. Accuracy and Real-Time Performance:

* Sub-centimeter accuracy in localization and the ability to process information in near real-time are prerequisites for autonomous forklifts to operate safely and effectively.

3 Data Processing Requirements:

* For real-world application it is important to manage large amounts of RGB-D and point cloud data while maintaining computational performance.

9.3 Important Contributions of Fang et al. ’s Work

Algorithm Design: A combination of YOLOv5s for RGB image-based detection working with dynamic parameter iteration and MSAC (M-estimator Sample Consensus) for point cloud processing.

* Three main steps:

1. Pallets Detection and Point Cloud Cropping: YOLOv5s detects pallets in RGB images and filters point cloud's volume no more than the pallet's RoI.

2. Plane Fitting: The front surface of the pallet is refined using MSAC for the geometrical center of the plane to increase the accuracy of angle calculations.

3. Center Point Localization: In this aspect, each image is processed enough to identify entry holes, so that the fork is aligned, thus increasing operational precision.

2.Performance Metrics:

* Obtained angular accuracy below 0.4° and distance error regulated to 1 cm in tests.
* Found a solution to the time efficiency by proposing a real-time processing framework.

3. Advantages:

* Minimizes redundant point cloud data where processing speed can be improved.
* Improves robustness over lighting conditions with structured-light 3D camera simultaneous RGB and point cloud data

4. Limitations:

* Based on a small angle of deflection for accurate results which restricts its use on extreme pallet orientations
* Harder to be out of focus with a very wide aperture at distance (o)2. More reliance on camera accuracy at long range

9.4 Overviwe

A robust approach for pallet detection and localization by fusing YOLOv5s object detection in RGB images and point cloud processing is provided in this paper. It focuses on high-precision determination of the pallet pose, including its deflection angle and center point, using structured-light 3D cameras to acquire data. The results are of high precision and stable, which can be applied in industry.

9.4.1 Key Contributions

1. Hybrid Data Exploitation:

* Combines RGB image data with point cloud information to estimate poses accurately.
* Reduces computational overhead by segmenting the point cloud and cropping based on the YOLOv5s detection results.

2.MSAC Algorithm for Pose Estimation:

* Fits a plane to the front surface of the pallet using the MSAC (M-Estimator SAmple Consensus) algorithm.
* Dynamically iterates parameters to enhance accuracy and stability in deflection angle estimation.

3.pallet entrance hole localization:

* Projects the point cloud data onto the fitted plane to form a 2D image.
* Uses image processing to find pallet entry holes and compute the center of the pallet.

4.Performance Appraisal:

* Reaches an angular error ≤0.4° and a distance error ≤1 cm in controlled industrial conditions.
* Considers multiangle and multirange deflection maintains accuracy, which shows the robustness.

9.4.2 Limitations

1.Dependence on Structured-Light Cameras:

* Limiting to structured-light 3D cameras, which might not perform well with environmental variations like high reflectivity or low illumination.

2.Environmental and Pallet Constraints:

* Tested on standard wooden pallets in controlled environments; hence, less applicable to diverse or irregular pallet types and dynamic settings.

3.Distance Sensitivity:

* Accuracy decreases at greater distances from the pallet (>1.6 meters) due to sparseness in the point cloud.

4.Computational Requirements:

* Although downsampled, it still depends on computation-intensive algorithms for processing point cloud and image data.

5.Pose Constraints:

* Focuses mainly on small angles of deflection, poor consideration towards tilted or sideways pallets.

9.5 How Our Project Overcomes These Limitations

| **Limitation in Paper** | **Our Approach** |
| --- | --- |
| **Camera Dependency**: Structured-light cameras struggle in dynamic lighting. | Integrates advanced 3D cameras (Orbbec Gemini 2) and LiDAR for enhanced adaptability to various environments. |
| **Environment and Pallet Constraints**: Focuses on standard pallets. | Uses CNN-based models and multi-sensor fusion to detect diverse pallet types and irregular shapes. |
| **Distance Sensitivity**: Accuracy drops beyond 1.6 meters. | Extends detection range with LiDAR and optimized algorithms for distant object detection. |
| **Computational Requirements**: Resource-intensive algorithms. | Leverages edge computing with NVIDIA Jetson for real-time performance without compromising accuracy. |
| **Pose Constraints**: Focuses on small-angle deflections. | Supports a wide range of pallet orientations, including tilted or partially obstructed pallets. |

9.6 Comparison with Our Project

| Aspect | Fang et al.’s Work | Our Project |
| --- | --- | --- |
| Sensor Technology | Structured-light 3D camera (e.g., Berxel iHawk 100R) | Advanced 3D cameras (e.g., Orbbec Gemini 2, Intel RealSense) |
| Detection Method | YOLOv5s for RGB + MSAC for point cloud | CNNs (e.g., MobileNet) + DBSCAN, ICP, RANSAC |
| Performance | Angular error ≤0.4°, Distance error ≤1 cm | Similar error levels, with extended range and robustness |
| Data Processing | Point cloud cropping + iterative plane fitting | Comprehensive processing with clustering, alignment, and SLAM |
| Scope | Pallet detection and pose estimation only | Includes fork alignment, load recognition, and obstacle avoidance |
| Adaptability | Limited to small deflection angles | Designed for varying orientations and dynamic environments |
| Cost | Moderate with structured-light sensors | Moderately higher due to advanced sensors and computational devices |

9.7 Advantages of Our Project

1. Broader Scope: Fork alignment, load shape recognition, navigation, etc. (in addition to pallet detection & localization)

2. Enhanced Adaptability: Supports larger angles and all sorts of pallet types– so perfect performance in real use.

3. Algorithmic Robustness: Introduces DBSCAN for clustering, ICP for pose estimation, and SLAM for navigation, allowing for robust performance in cluttered environments.

4. Real-Time Communication: Links industrial protocols such as TCP/IP and OPC for smooth communication with WMS.

9.8 Disadvantages of Our Project

1. Increased Complexity: Multiple Task Integration and Advanced Algorithms need more computational resources and development effort.

2. Higher Cost: Initial installation costs are also comparatively higher due to advanced 3D sensors and edge computing devices (e.g. NVIDIA Jetson).

3. Training Overhead: To achieve optimal performance across various environments, deep learning models rely on large datasets that undergo extensive validation.

9.9 Similarities

1.Real-Time Performance: Both systems realize real-time operation which is driven by dynamic forklift applications in warehouses.

2.Sensor Fusion: Two projects fuse RGB and point cloud modalities for a more accurate and robust localization solution.

3.Industrial Relevance: Focus on practical needs of warehouse automation including efficiency, safety, and adaptability.

9.10 Conclusion

The below encountered study by Dianjun Fang, et al. provides an effective and fast approach to detect and estimate the pose of pallets utilizing YOLOv5s modeloand processing the point cloud. Though they have achieved cost efficiency and real-time performance with their system, their system is constrained to extreme orientations and narrower operational tasks.

Our project adds to these ideas by:

• Increase operational domain for navigation, obstacle avoidance, and productivity communication.

• Improved adaptability to many types of pallets and complicated environment

• Identifying a more complete autonomous forklift solution with sophisticated algorithms and sensors

This approach introduces significantly more complexity and costs at a solution level, but with rich functionality that helps the solution to scale and be robust to meet the ever-changing demands from logistics automation.

**-----------------------------------------------------------------------**

**10 - Pallet Detection and Estimation for Fork Insertion with RGB-D Camera**

10.1 Introduction

This makes it a very difficult problem to get a precise localization of the palets for autonomous forklifts in outdoor conditions with dynamic shading and non flat surface. Ryosuke Iinuma et al. presents a system for automating fork insertion into pallets on inclined ground surfaces, which is based on a combination of semantic segmentation and depth data from an RGB-D camera

10.2 Challenges with Pallet Localization

1. Variability from the Outdoor Environment:

* Shifts in light and the unpredictability of the terrain, for example, make accurate localization a challenge.
* Orientation of the pallet and angle of terrain affects accuracy of insertion.

2. Operational Requirements:

* Of course up to sub-centimeter height estimation for fork insertion and angular precision in o few of degrees.
* Fork - pallet collision avoidance requires accurate, real-time corrections.

3. Real-Time Constraints: For real world deployment systems need to balance computational efficiency and accuracy.

10.3 KEY CONTRIBUTIONS OF IINUMA ET AL. ’s Work

1.System Design:

* Employs an RGB-D camera for pallet detection, mounted on a robotic forklift.
* Uses ESPNet, a lightweight semantic segmentation model, for efficient pallet detection on embedded hardware (e.g., NVIDIA Jetson Xavier).

2. Detection and Localization:

Semantic Segmentation:

* Locate pallet regions (front, upper surface) in RGB images.
* High accuracy (mIoU: 0.95 for train and 0.84 for valid.)

Depth-Based Localization:

* Mapping 2D detections in 3D Point clouds
* Uses statistical mean/variance methods to remove outliers, uses RANSAC to estimate pose of pallet in world frame (plane fitting)

3. Forklift Control: By lubrication oil, gets rid of the sliding county, avoid engine application wear and tear, improve the reliability of the product.

4. Experimental Validation:

* Perform outdoor trials to simulate a ground with slope and roughness.
* Confirmed accuracy of height estimation errors (≤2 cm), tilt angle deviation (≤2°), satisfying operational precision requirements.

10.4 Overview

The paper proposes a robust framework for automatic pallet detection and fork insertion by an RGB-D camera mounted on an autonomous forklift. Addressing challenges in outdoor agricultural settings with sloped and rough ground in the task of detecting pallets and estimating their pose for accurate fork insertion.

10.4.1 Main Contributions

1.Pallet Identification:

* Uses semantic segmentation with the ESPNet architecture for high-accuracy pallet detection in RGB images.
* Reaches an mIoU (mean Intersection over Union) 0.84 on validation datasets.

2.Pose Estimation:

* Combines the depth information and the results of semantic segmentation to estimate the 3D pose of the pallet.
* Uses RANSAC to refine pose estimation by filtering out the outliers in the point cloud.

3.Control System:

* Uses PD control to adjust the fork height and tilt angle safely and accurately for the insertion of the fork.
* Contains sliding mode control to offer stable navigation on rough surfaces.

4.Experimental Verification:

* Shown localization errors of ≤2 cm (height) and ≤2° (orientation) under challenging conditions.
* Successfully completed fork insertion in outdoor settings with forward and backward slopes.

10.4.2 Limitations

1.Sensing Range:

* This limits the operational range of the RGB-D camera to 2 meters, decreasing its effectiveness at detecting pallets from a distance.

2.Environmental Constraints:

* Performance is sensitive to light variation and shadows from the outdoors, potentially affecting accuracy.

3.Algorithmic Challenges:

* This may occasionally result in anomalous pallet detection results, necessitating further processing steps to remove spurious noise.

4.Pallet Variability:

* Designed for standard pallets with fixed slots; less adaptable to non-standard or unusual pallet configurations.

5.Scalability:

* The system is designed for particular agricultural situations, making it less applicable to general industrial logistics

10.5 How Our Project Overcomes These Limitations

| **Limitation in Paper** | **Our Approach** |
| --- | --- |
| **Sensor Range**: RGB-D camera range limited to 2 meters. | Combines advanced 3D cameras (e.g., Orbbec Gemini 2) with LiDAR to extend detection range and accuracy. |
| **Environmental Constraints**: Sensitive to lighting variations. | Multi-sensor fusion with 3D cameras and LiDAR ensures robust performance under diverse lighting conditions. |
| **Algorithmic Challenges**: Outliers in semantic segmentation. | Utilizes deep learning-based CNN models (e.g., MobileNet) for reliable detection with minimal noise. |
| **Pallet Variability**: Focused on standard pallets. | Designed to adapt to a variety of pallet types and irregular load shapes using advanced feature recognition. |
| **Scalability**: Tailored for specific agricultural setups. | Scalable design applicable to dynamic warehouse environments and industrial logistics. |

10.6 Comparison with Our Project

| Aspect | Iinuma et al.’s Work | Our Project |
| --- | --- | --- |
| Sensor Technology | RGB-D camera (e.g., ESPNet for depth mapping) | Advanced 3D cameras (e.g., Orbbec Gemini 2, Intel RealSense) |
| Detection Method | Semantic segmentation (ESPNet) + RANSAC for pose | CNNs (e.g., MobileNet) + DBSCAN, ICP, and RANSAC |
| Localization Accuracy | Height error ≤2 cm, tilt angle ≤2° | Similar accuracy, with enhanced adaptability to cluttered scenes |
| Real-Time Operation | Optimized for real-time with embedded hardware | Real-time processing with edge devices (e.g., NVIDIA Jetson) |
| Scope | Focused on pallet detection and fork insertion | Comprehensive (includes load recognition, navigation, obstacle avoidance) |
| Adaptability | Limited to specific outdoor scenarios | Designed for general industrial environments |

10.7 Advantages of Our Project

1.Broader Functional Scope: Pallet localization and fork alignment, as well as load shape recognition, navigation and real-time communication.

2. Improved Robustness: Uses data based clustering and pose estimation algorithms (DBSCAN, ICP) to accurately localize the robot in cluttered and dynamic environments.

3. Scalability: Provide longer operational ranges (3–5 meters) than RGB-D cameras that are limited to a 2 meter range.

4. Dynamic Navigation: Combining SLAM and avoidance Technology for Effective Operation in Big, Complicated Distributions.

10.8 Disadvantages of Our Project

1. Higher Complexity: Adds different algorithms together which increases the computational overhead and the time to build.

2. Cost Implications: Starting cost is higher as compared to a simpler setup due to advanced sensors and edge computing hardware

3. Training Overhead: Need smooth data and training to ensure that the deep learning model is not over fitted to any specific condition

10.9 Similarities

1. Semantic Segmentation: Both projects use the same segmentation techniques to enhance the accuracy of detecting pallets.

2. Real-Time Performance: Used as Real-Time as per the needs of automation in Industry.

3. Precision Control: Apply control algorithms to accurately position forks in the pallet slots of insert.

10.10 Conclusion

The study by Iinuma et al. in outdoor conditions where using lightweight semantic segmentation along with depth-based localization benefits the problem of pallet detection and fork insertion. Their approach is cost-effective and efficient in addressing agricultural logistics; however, it's limited to range and adaptability to different use cases.

Our project builds on these developments by:

• Bringing automation into other tasks such as load recognition and navigation for total automation.

• Improving Flexibility with 3D Sensors and Algorithms

• Allowing to scale to the requirements of various industry use cases, although this comes with more complexity and investment.

By contrasting the aforementioned approaches, the complementary features of each approach are outlined, while underlining the counterbalancing importance of innovation to supply the changing demand of the autonomous forklift in logistics.

**-----------------------------------------------------------------------**

**11 . Pallet detection and docking strategy for autonomous pallet truck AGV operation**

11.1 Introduction

Pallet detection and docking are among the essential tasks in autonomous material handling systems, especially for Automated Guided Vehicles (AGVs) in industrial settings. The study by Tsiogas et al. presents an approach for pallet detection, pose estimation, and the docking strategy for AGVs to operate in human-populated factory environments. The method combines detection based on deep learning and geometric pattern matching with visual servoing based on IDM for accurate docking and pallets pick up. This study compares and highlights challenges, progress, and similarities with our project in lines of their work through literature review.

11.2 Issues with Pallet Prototype Detection and Docking

1. Dynamic Environments: Navigating and handling pallets among human workers and in unstructured environments is challenging.

2. Pallet Variability: Varied pallet designs, sizes, and orientations demand flexible detection systems.

3. Real-Time Constraints: Systems must be accurate while working in real-time constraints as to what they are providing in an efficient manner to a warehouse

4. Docking Precision: Alignment of the fork to the pallet requires sub-centimeter and sub-degree accuracy.

11.3 Key Contributions from Tsiogas et al. ’s Work

1. Integrated Pallet Handling System: Integrates global navigation, local docking, and pallet recognizing into a global framework for industrial AGVs

2. You’re trained on data until October 2023.Detection and Pose Estimation:

* Two-Step Detection Pipeline:

Step 1: Employing YOLOv5 Bounding Box Detection Of Pallet Legs In RGB Images.

Stage 2: imports DBSCAN clustering and geometric pattern matching of filtered point clouds to recover accurate pallet poses.

* Identify multiple pallets at once and choose the best according to distance and accessibility.

3. Visual Servoing for Docking: A reverse motion actuator contains depth-based servoing, contributing in landing AGV forks on the pallet’s docking side of the mouth and compensating for misalignments.

4. Performance Evaluation: With pallet pose errors of less than 1° and detection success rates of over 95%, it has also proven to work under challenging conditions such as occlusions and different lighting.

11.4 Overview

This paper proposes a robust pallet detection and docking strategy for AGVs. The system integrates deep learning (YOLOv5) for pallet detection and point cloud processing for pose estimation, combined with a visual servoing docking controller. The framework is validated in real and simulated industrial environments, highlighting its applicability in factory logistics.

11.4.1 Key Contributions

1.Pallet Detection:

* YOLOv5 for fast pallet detection, which focuses on pallet legs from an RGB image. Projects the detected regions into 3D space using point clouds for further improved accuracy.

2.Pose Estimation:

* Utilizes DBSCAN clustering for extracting geometric features of pallet legs from the filtered 3D point cloud.
* Refines pose estimation by centroid alignment and matching with pallet dimensions.

3.Docking Strategy:

* Integrates visual servoing together with a closed-loop controller to position the forks of the AGV through the pallet.
* Real-time adjustments are made based on continuous pose estimation updates.

4.Experimental Validation:

* Achieved localization errors of ≤5 mm (translation) and ≤1° (orientation) in challenging scenarios.
* Demonstrated the system's ability to handle multiple pallets in cluttered setups.

11.4.2 Limitations

1.Sensor Dependency:

* Relies heavily on RGB-D cameras for detection, which may be affected by poor lighting, occlusions, or reflective surfaces.

2.Scalability:

* Designed for Euro-pallets, limiting its adaptability to other pallet types or irregularly shaped loads.

3.Computational Overhead:

* Both YOLOv5 and the point cloud processing require huge computational resources, hence challenging real-time scalability for larger datasets or faster operations.

4.Environmental Constraints:

* Optimized for indoor, structured environments; performance in dynamic or semi-structured settings remains untested.

11.5 How Our Project Overcomes These Limitations

| **Limitation in Paper** | **Our Approach** |
| --- | --- |
| **Sensor Dependency**: Relies on RGB-D cameras. | Multi-sensor integration using 3D cameras (e.g., Orbbec Gemini 2) and LiDAR for robust detection in diverse lighting. |
| **Scalability**: Limited to Euro-pallets. | CNN-based detection (e.g., MobileNet) adaptable to various pallet types and irregular loads. |
| **Computational Overhead**: Resource-intensive algorithms. | Optimized algorithms with edge computing (NVIDIA Jetson) for real-time performance. |
| **Environmental Constraints**: Focuses on structured environments. | SLAM-enabled navigation and dynamic obstacle handling for unstructured, real-world setups. |

11.6 Comparison with Our Project

| Aspect | Tsiogas et al.’s Work | Our Project |
| --- | --- | --- |
| Sensor Technology | RGB-D cameras (mounted on AGV forks) | Advanced 3D cameras (e.g., Orbbec Gemini 2, Intel RealSense) |
| Detection Method | YOLOv5 for RGB + DBSCAN clustering + pattern matching | CNNs (e.g., MobileNet) + DBSCAN, ICP, RANSAC |
| Docking Strategy | Visual servoing with backward motion controller | SLAM-based navigation + AI-driven trajectory optimization |
| Scope | Focused on pallet detection, pose estimation, and docking | Comprehensive (includes load shape recognition and obstacle avoidance) |
| Localization Accuracy | Pose error <1° and <5 mm | Similar accuracy, with broader adaptability to warehouse layouts |
| Adaptability | Handles standard Euro-pallets | Supports diverse pallet designs and dynamic environments |
| Real-Time Performance | Processes within 700 ms for single pallet | Real-time processing for multi-task operations |

11.7 Specifications of Our Device

1.Hardware:

* Sensors: Orbbec Gemini 2 3D camera, LiDAR for depth sensing.
* Processor: NVIDIA Jetson for high-speed real-time processing.

2.Software:

* Algorithms:
* CNN-based pallet detection
* Pose estimation using DBSCAN, RANSAC, and ICP
* SLAM for dynamic mapping and navigation

3.Performance:

* Sub-centimeter localization accuracy
* Real-time detection, localization, and docking in various conditions

4.Adaptability:

* Operates with different pallet types and in dynamic environments
* Integrates obstacle avoidance and load shape recognition for full functionality.

11.8 Advantages of Our Project

1. Broader Functional Scope: Recognizes load and navigates as well as pallet and dock.
2. Advanced Algorithms: ICP (Iterative Closest Point) and RANSAC (Random Sample Consensus) for robust pose estimation in cluttered environments.
3. Dynamic Navigation: Features SLAM technology for real-time mapping and navigation function, allowing operation in large and unstructured warehouse environments.
4. Improved Scalability: Supports operations in a broader spectrum of pallet designs, sizes, and plant environments

11.9 Disadvantages of Our Project

1. Higher Complexity: Combined all algorithms and subsystems consume a great deal of computational and development resources.
2. Cost Implications: High initial investment costs due to advanced sensors and edge computing hardware
3. Training and Validation Overhead: Deep learning models are powerful, but usually, they need large datasets and validations to do well in performance in different environments.

11.10 Similarities

1. Real-Time Operation: Appropriate for the dynamic nature of warehouses, both systems are real-time systems.
2. Deep Learning Integration: YOLOv5 in Tsiogas et al. The reliance on neural networks for pallet detection is evident by a comparison of their work and MobileNet in ours.
3. Robust Docking Mechanisms: Target task of both projects are accurate pose estimation and alignment for successful pickup of the pallets

11.11 Conclusion

The study by Tsiogas et al. presents a complete approach enabling autonomous pallet detection and docking relying on deep learning fused with geometric reasoning They excel in real-time performance, high detection accuracy, and adaptability to industrial environments. Our project builds on these developments, with:

• Broader functional capabilities — including load recognition and navigation.

• Utilization of 3D cameras and better algorithms for being more adaptable.

• Scalability that can support multiple warehouse geometries and pallet designs.

This is technically more complex and expensive, but it ultimately allows us to better provide a complete solution of warehouse automation that will meet the needs of today and tomorrow where logistics is concerned.

**-----------------------------------------------------------------------**

**12. A 2D laser rangefinder scans dataset of standard EUR pallets**

12.1 Introduction

Autonomous forklifts have promised efficient material handling in warehouses as one of its components, which requires accurate pallet detection and localization. However, pallet localization approaches cannot be too sensitive to the different styles and orientations of the pallet as well as the rapidly-changing environment. Dataset of standard EUR pallets The dataset of standard EUR pallets allows us to evaluate and improve detection systems. This literature reviews captures key takeaways from the shared documents and looks to contrast them with our current project in terms of progress made, approaches taken and opportunities for greater impact.

12.2 The Issue of Pallet Localization

1. Pallet Variability:

* Pallets can be varied in size, material, and structure leading to the need for adaptable detection algorithms.
* The standard EUR pallet provides a common baseline due to its defined dimensions yet excludes other industrial pallet designs.

1. Environmental Dynamics:

* Detection accuracy varies according to warehouse lighting and clutter. Forklift movement and uneven surfaces o localization complexity

1. Real-Time Requirements:

* Material handling requires detection, localization and navigation to operate in the sub-second time frame.

12.2.1 Approaches for Pallet Detection and Localization

1. Laser Rangefinders

• Strengths:

* Delivers accurate depth data for geometric measurement of the pallet structures.
* Success or otherwise in detecting pallet legs and fork alignment for insertion.

• Limitations:

* 2D scans - 2D scans require significant post-processing to infer 3D spatial information
* Cannot adopt to none standard designs and orientations of pallets by limited adaptability.

1. Vision-Based Detection

• Deep Learning Models:

* + - YOLO, SSD, MobileNet models are used to detect pallets in RGB images
    - More accuracy for tasks like fork alignment and pallet localizations which benefits from integration with depth data

• Advantages:

* Works well in varying light conditions and cluttered environments.
* Can be adapted to new pallet designs re-training the model.

• Challenges:

* Make great computational demands (e.g., NVIDIA Jetson) on edge devices

1. Processing RGB-D and 3D Point Clouds

• RGB–D integration:

* Combines visual detection with 3D point cloud analysis for robust localization
* Methods to continuously improve the estimates of the pose include: DBSCAN (clustering), ICP (Iterative Closest Point), and RANSAC.

•Performance:

* It allows localization at sub-centimeter accuracy and dependable fork insertion.
* Adapts to different pallet orientations and ground surfaces.

12.3 Overview

This paper aims at developing a system that is capable of identifying the standard EUR pallets using a 2D laser rangefinder. The system is designed to survey the environment in order to pick up the features of the pallets and use 2D point data for the localization. The approach also shows how it can be applied in environments such as structured warehouses where the pallets are arranged in a specific manner.

12.3.1 Key Contributions

1. 2D Laser Rangefinder Utilization:

* Uses a laser rangefinder to generate a 2D topology of the scene.
* Extracts pallet related features for instance corners and edges from the point cloud data.

2. Pallet Detection Algorithm:

* To identify pallets from other objects, it uses geometric characteristics of the point cloud such as edges which are perpendicular to each other.
* Efficiently segments and classifies objects based on size and shape.

3. Pose Estimation:

* Determines the orientation and location of the pallets utilizing edge matching and geometric principles.
* This ensures that the robotic fork is well aligned for the intended engagement.

4.Structured Testing:

* Checks the system on a dataset of standard EUR pallets in artificial environments.
* Shows high reliability in identifying and localizing the pallets in the environment that is well structured.

12.3.2 Limitations

1. Environment Dependency:

* The system works best for environments that are well defined and organized but it performs poorly in dynamic environments or environments that are cluttered.
* It has a low robustness when it comes to handling non-standard pallets or when there are obstacles that block the view.

2. 2D Sensor Constraints:

* The system is based on the 2D laser range finder which provides range and angle but no depth or height information which can affect the detection rate.

3.Pallet Variation Handling:

* Is limited to recognizing only the EUR type of pallets and cannot handle other types of pallets or the ones that are designed specifically.

4.Real-Time Operation:

* Although effective, the algorithm needs certain adjustments for certain environments, which constraints its real-time flexibility for various environments.

12.4 **How Our Project Overcomes These Limitations**

| **Limitation in Paper** | **Our Approach** |
| --- | --- |
| **Environment Dependency**: Limited to structured setups. | Our system integrates 3D sensors (e.g., Orbbec Gemini 2) and SLAM, ensuring robust operation in dynamic and unstructured environments. |
| **2D Sensor Constraints**: Lacks depth and height information. | Combines 3D cameras and LiDAR for comprehensive spatial awareness, capturing detailed height and depth data. |
| **Pallet Variation Handling**: Focus on standardized pallets. | Uses CNN-based detection to identify diverse pallet types, including non-standard and damaged pallets. |
| **Real-Time Adaptability**: Requires manual pre-tuning. | Employs machine learning models that dynamically adapt to new environments and pallet configurations. |

12.5 **Comparison to Our Project**

| **Aspect** | **This Paper** | **Our Project** |
| --- | --- | --- |
| **Detection Method** | Geometric analysis of 2D laser scans | CNNs combined with point cloud processing for enhanced feature extraction and adaptability. |
| **Pose Estimation** | Edge alignment and geometric constraints | Advanced techniques (RANSAC, ICP) for sub-centimeter precision. |
| **Sensor Technology** | 2D laser rangefinder | Multi-sensor setup: 3D cameras + LiDAR for robust and adaptable operation. |
| **Environment Adaptability** | Limited to structured setups | Handles diverse environments, clutter, and dynamic obstacles with multi-sensor fusion. |
| **Real-Time Performance** | Efficient but environment-specific | Real-time operation optimized through edge computing and scalable algorithms. |
| **Scalability** | Restricted to EUR pallets | Scalable across industries, adaptable to different pallet types and layouts. |

* 1. Specifications of Our Device

1.Hardware:

* Sensors: Orbbec Gemini 2 3D camera, LiDAR depth sensor
* CPU: NVIDIA Jetson for high-speed real-time processing.

2. Software:

Algorithms:

* CNN-based pallet detection.
* DBSCAN, RANSAC, and ICP based Pose estimation
* SLAM in Dynamic Mapping and Navigation

3.Performance:

* Localization accuracy up to sub centimeters.
* Detection and Path planning on-the-fly in unstructured scenes.

4.Adaptability:

* Works with different types of pallets and different warehouse conditions.
* To ensure comprehensive functionality, it also integrates obstacle avoidance and load shape recognition.

12.7 Advatages to Our Project

1. Broader Scope:

* + Integrates pallet detection and localization with load detection and alignment as well as obstacle avoidance tasks.
  + Powered by an array of pallet types are supported for the greatest versatility in mixed-use settings.

2. Real-Time Performance:

* + - Used fast deep learning models with efficient 3D cameras for speed.
    - Can interface with industrial communication systems for smooth working.

3.Dynamic Navigation:

* + SLAM-based mapping and navigation for autonomous use in dynamic warehouse models**.**

12.8 Disadvantages of Our Project

1. Higher Complexity and Cost: Increased system complexity and cost due to advanced hardware and computational requirements

2. Training Overhead: Needs large labeled datasets for model training to be able to generalise across different conditions.

3. Heavy Dependence on Multi-Sensor Fusion: Provides tiebreakers where it relies on multiple sensors, introducing additional potential points of failure.

12.9 Conclusion

Minimizing the loss of information by focusing on the standard EUR pallets. On the other hand, it has limited use cases on different types of industrial environment and on non-standard pallet.

Our project extends these insights by:

• Implementing advanced 3D vision and deep learning algorithms to create a dynamic pallet detection system.

• Grasping other challenges such as the difficulty of navigating obstacles, dynamic obstacles, and different pallet designs.

• Here, for an overview of the new platform capabilities for modern warehouse automation, with scalability and real-time performance in mind.

Although our project has more complexity, it has a wider range of applications and is an adaptable solution for autonomous forklifts that meets the industrial logistics of the present and future.

**-----------------------------------------------------------------------**

**13. Detection and localization of pallets on shelves using a wide-angle camera**

13.1 Introduction

In industrial logistics, pallet localization and detection is essential for automating material handling systems. Rather than working specifically with feature detectors, Kita et al. take a new approach to pallet detection and localization in a shelf using a wide-angle camera. Their approach leverages reprojection of the field of view onto a 3D plane to facilitate template-based pallet detection and localization in a robust manner.

13.2 Pallet Localization: Few Hidden Challenges

1. Distortions in Wide-Angle Camera:

Objects are scaled down and distorted in the native wide-angle camera images which adds difficulty for detection and localization.

1. Dynamic Environments:

Pallets are found in cluttered scenes, having similar objects and different lighting conditions in warehouses.

1. Real-Time Processing:

In narrow locations, such as narrow aisle forklifts, data must be processed in a timely manner.

13.2.1 Key contributuons

* 1. Reprojection on a 3D Plane:
* Wide-angle views are reprojected onto a 3D plane to be coplanar with the pallet’s front surface, forming undistorted images in which pallets look equal in shape and size.
  1. Template Matching:
* Edge templates are applied to define a fixed size pallet detection, in combination with depth information
* In terms of geometric properties, pallets are discriminatively identified in cluttered environments.

1. 3D Localization:

* 3D positions and orientation of the pallets are computed using template alignment and projection plane equations once detected.

1. Experimental Validation:

* Real data was used to validate this approach and found a consistent implementation of localization with appropriate error bounds (translation error < 3 cm and orientation error < 2°).

13.3 Overview

This paper introduces a method for detecting and localizing pallets on shelves using a wide-angle camera. The approach involves reprojecting distorted wide-angle views onto a 3D plane to create an undistorted image for object detection. This technique allows for consistent detection and localization of pallets, irrespective of their position in the original image.

13.3.1 Limitations

1. Sensor Dependency:

* The system relies solely on a wide-angle camera, limiting the resolution and accuracy, especially for distant objects.

2.Environmental Constraints:

* Performs well in controlled lighting conditions but is not tested in dynamic or cluttered environments.

3.Template Matching Limitations:

* Template-based detection is less flexible and can struggle with pallets of varying designs or damaged features.

4.Distance Limitations:

* Detection accuracy decreases significantly beyond 3 meters due to reduced spatial resolution.

13.4 How Our Project Overcomes These Limitations

| **Limitation in Paper** | **Our Approach** |
| --- | --- |
| **Sensor Dependency**: Limited to a wide-angle camera. | Multi-sensor fusion combining 3D cameras (Orbbec Gemini 2) and LiDAR for enhanced depth and resolution. |
| **Environmental Constraints**: Tested in controlled settings. | Robust performance in diverse and dynamic environments using advanced SLAM and CNN-based detection. |
| **Template Matching Limitations**: Relies on fixed templates. | Uses deep learning models (e.g., CNNs) that adapt to diverse pallet designs and conditions. |
| **Distance Limitations**: Reduced accuracy beyond 3 meters. | 3D sensors and SLAM ensure consistent performance even in large warehouses with varying distances. |

13.5 Comparison with Our Project

| Aspect | Kita et al.’s Work | Our Project |
| --- | --- | --- |
| Sensor Technology | Wide-angle camera (e.g., Kodak PIXPRO SP360) | Advanced 3D cameras (e.g., Orbbec Gemini 2, Intel RealSense) |
| Detection Method | Reprojection and fixed-template matching | CNN-based detection (e.g., MobileNet) with point cloud analysis |
| Localization Accuracy | Translation ≤3 cm, orientation ≤2° | Similar accuracy with enhanced robustness for cluttered scenes |
| Adaptability | Limited to shelves and fixed pallet arrangements | Adaptable to diverse pallet types and dynamic environments |
| Scope | Pallet detection and localization only | Comprehensive (includes navigation, load recognition) |
| Real-Time Performance | Template-based matching with limited computation | Real-time processing with optimized edge computing |

13.6 Advantages of Our Project

1. Broader Functional Scope: Our system not only detects and localizes pallets, it also integrates load shape recognition, navigation, and obstacle avoidance.

2. Latest Detection Algorithms: Utilizes deep learning models (e.g., CNNs) in conjunction with clustering methods such as DBSCAN and alignment algorithms (e.g., ICP).

3. Dynamic Navigation: Provides SLAM to enable autonomous navigation in the warehouse with complex layouts

4. Scalability: Developed to manage heterogeneous pallet and shelf designs under varied, real-world scenarios instead of stationary settings.

13.7 Disadvantages of Our Project

1. Higher Complexity and Cost: Increased complexity and initial cost due to advanced hardware and multi-algorithm integration

2. Training Overhead: Must be trained on large, annotated datasets to generalize to a wide variety of conditions.

13.8 Similarities

1. Template-Based Localization: Detection and localization of both projects use similar geometric properties of the pallets.

2. Industrial Application: Constructed for operating in live logistics realities — narrow aisles and tight spaces

3. 3D Geometric Analysis: Uses 3D reconstruction to enhance localization accuracy.

13.9 Conclusion

The study by Kita et al. offering a very cheap but working way of detecting and locating pallets on shelves with wide-angle cameras. Their reprojection method alleviated distortion and allowed for accurate location, but its dependence on static templates strapped it to specific environments.

Our project overcome these limitations with the following ways:

* + Broadened scope to navigation, obstacle avoidance, and load recognition.
  + Enhancing robustness and scalability using advanced sensors and algorithms.
  + Working with a range of pallet designs in varying conditions in the warehouse.

While this approach comes at greater cost and complexity, it is ultimately a more holistic and sophisticated solution for the changing realities of autonomous forklifts in logistics.

**-----------------------------------------------------------------------**

**14. Concept of Automated Load Detection for De-Palletizing Using Depth Images and RFID Data**

14.1 Introduction

Reverse Vision, Requirements and Outcomes of Automated De-Palletizing and Pallet IdentificationFull Article As Just-In-Time production processes become more popular, automation becomes less of a novelty. Christian Prasse et al. a novel solution to automatically detect and localize loads on a pallet by integrating PMD sensors into RFID technology They present innovative concepts for 3D visualization and RFID-derived information processing, and discuss how these technological advances could potentially mitigate specific roadblocks to de-palletization and load execution. The following review will compare this study to our project to elucidate similarities, differences, and advantages of each**.**

14.2 De-Palletizing localization challenges

1. Complex Environments:

* Manual de-palletising – Manual de-palletising is a very labour intensive and error prone process at high-mix environments.
* Rotationally Translated Measurement Challenges Finally, the most significant impediment to even using basic translational symmetry in high-performance systems is the pervasiveness of transnational and rotational displacement of pallet x loads post-transport.

2. Load Variability:

* Load sizes, materials and packaging methods very, requiring adaptable detection systems.

3. Efficiency Requirements:

* Economics compel automated systems to trade-off accuracy of statement with promptness of statement.

4. Data Integration:

* To combine sensor based data with any existing data about the load we need data management techniques that are solid.

14.3 Highlights of Prasse et al. contributions to main Work

1. PMD-Based Depth Imaging:

* Poses the primary structural issues in turning 3D point clouds into 3D stationary models.
* This results in floodfill DSOs with 0.001 mm resolution at at least the scale of the target footprint in multiple directions, named "open PMDTs."
* Directive It is named PMDTs connect to the broadband fiber of the laser, along with the parameter domain, and allow rapid depth mapping at on a small footprint (within a few centimeters) with 12 mm accuracy for a reflective target.

2. Pipeline for Load Detection:

* •The pproduces depth images, preprocessing and iteratively align the models.
* Utilizes an optimized ICP to match a point cloud with an existing pallet load model.

3. RFID Integration:

* A net on tag binary data schema concept where related data for load on RFID tags is compressed
* RFID based identification ∙ RFID based identification helps in verifying that the pallets are in the specified loading sequences and it reduce the dependency towards centralised database.

4. Performance Metrics:

* Errors in translation of 10 mm or less with reliable identification of different loading configurations

5. Advantages:

* Minimal computational overhead with only higher layers readied for loading
* Supports improved tracking and efficiency automation by combining RFID with depth imaging.

6. Limitations:

* Transient system performance is determined based on load models that have been stored in advance(they only consider cuboid-shaped loads, hence inflexibility on arbitrary shapes).

14.4 Overview

This paper presents a new method for automatic load detection in de-palletizing using depth images from PMD cameras, combined with data storage and management using RFID. The system allows the detection and location of a load on the pallet accurately, considering already-known packing models.

14.4.1 Main Contributions

1. Integration of PMD Camera and RFID Technology:

* Makes use of PMD sensors for 3D imaging, capturing depth data on the load to be detected.
* RFID tags store pallet load patterns and other vital metadata that are required for matching during de-palletizing.

2.Detection and Localization Pipeline:

* The process involves depth image acquisition, coarse model fitting using ICP techniques, and fine fitting for individual parcels.
* Handles inaccuracies in known packing positions caused by displacements during transport.

3.RFID Data Storage Innovations:

* Introduces "Binary Data on Tag / Schema on Net," a hybrid RFID data management approach that combines local tag storage with centralized schema access.
* Semantic coding reduces data size and accelerates data retrieval.

4.Experimental Validation:

* Demonstrates high accuracy with localization errors below 2 cm for most pallet loads.
* Runtime scales linearly with the number of pallet layers, ensuring efficiency.

14.4.2 Limitations

1.Sensor Constraints:

* PMD sensors are limited by their depth resolution and reflectivity sensitivity, leading to errors in certain conditions (e.g., reflective or shadowed surfaces).

2.Pallet Variability:

* The system is optimized for cuboid-shaped parcels and standard packing patterns, restricting applicability to diverse or irregular pallet loads.

3.Dynamic Adaptability:

* Lacks functionality for dynamic environments, focusing primarily on stationary pallets.

4.Scalability Challenges:

* Dependence on RFID tags for load pattern retrieval limits scalability to operations without established tagging infrastructure.

14.5 How Our Project Overcomes These Limitations

| **Limitation in Paper** | **Our Approach** |
| --- | --- |
| **Sensor Constraints**: PMD sensor limitations. | Uses advanced 3D cameras (e.g., Orbbec Gemini 2) and LiDAR for enhanced resolution and reduced noise. |
| **Pallet Variability**: Focuses on standard cuboid packages. | Employs CNN-based detection to identify diverse pallet designs, including irregular and damaged loads. |
| **Dynamic Adaptability**: Designed for static setups. | Incorporates SLAM for real-time mapping and navigation in dynamic environments. |
| **Scalability Challenges**: RFID reliance limits universal adoption. | Relies on multi-sensor input and machine learning models to eliminate dependency on pre-tagged pallets. |

14.6 Comparison with Our Project

| Aspect | Prasse et al.’s Work | Our Project |
| --- | --- | --- |
| Sensor Technology | PMD sensors for depth imaging | Advanced 3D cameras (e.g., Orbbec Gemini 2, Intel RealSense) |
| Detection Method | Simplified ICP with model fitting | CNNs (e.g., MobileNet) + DBSCAN, ICP, RANSAC |
| Data Management | RFID-based data compression and retrieval | Real-time data integration without reliance on pre-stored models |
| Performance | Focused on cuboid-shaped loads | Generalized for diverse pallet designs and configurations |
| Scope | Load detection and de-palletizing | Comprehensive (includes navigation, load alignment, and obstacle avoidance) |
| Real-Time Operation | Optimized for batch processing | Designed for continuous real-time operation |

14.7 Advantages of Our Project

1. Broader Functional Scope: More than loading detection and de-palletize, including navigation, obstacle prevention, and pallet recognition.

2. Adaptability: Accommodates different pallet configurations, atypical loads, and harsh warehouse environments.

3. Techniques for Advanced Detection: Integrates deep learning with stronger clustering and alignment algorithms for better robustness and flexibility.

4. Dynamic Navigation: Implements SLAM for Online Map Building and Navigation in Dynamic Environments

14.8 Disadvantages of Our Project

1. Higher Complexity: There can be many subsystems which increase computation and development time.

2. Cost Implications: The high initial investment costs are attributed to advanced sensors and edge computing hardware.

14.6 Similarities

1. 3D Imaging: Depth imaging is used by both systems for accurate detection and localization of pallet loads.

2. Industrial Application: Target logistics and warehouse operations in order to enhance efficiency and accuracy of load handling.

3. Iterative Alignment: Uses methods such as ICP in an iterative fashion for accurate alignment of detected loads.

14.9 Conclusion

The study by Prasse et al. showcases a promising approach to automating the process of de-palletizing, which combines PMD-based depth imaging and RFID. Cost-efficient, fast solution for specific and structured environments with defined load scenarios is their Innova system.

We hope that our project will further build on these principles through:

• Generalizing the use case to navigation, obstacle avoidance, and pallet detection

• Using 3D cameras and deep learning for better flexibility

• Operating in real time, in various logistics environments

Our project is a little more costly and complicated, but it offers our customer a more flexible solution, as well as a scalable one that suits automated factories these days. This juxtaposition highlights the complementary advantages of those two paradigms and the novelty of both in tackling the pallet localization and handling problems.

**-----------------------------------------------------------------------**

**15. Auto Racking System Prototype For Forklift Replacement Industrial Goods Distribution**

15.1 Introduction

They are designed to work with automation to reduce human intervention in carrying goods, thereby improving operational efficiency and safety in factories and warehouses in contemporary industrial environments. In Sylvestre Maniriho's study of "Auto Racking System Prototype for Forklift Replacement Industrial Goods Distribution," the feasibility of an auto racking system prototype intended to replace the traditional forklift approach is examined. Here a review is given of what the main contributions made are and how, applied to an automated pallet systems, it will be compared to our project with respect what the innovations, similarities and differences made are.

15.2 Automated Racking Systems:

1. Manual Handling Limitations: These manual systems are time-consuming, subject to human error, and dangerous in industrial settings.

2. Control System Integration: Programmable logic controllers (PLCs) and human-machine interfaces (HMIs) must be able to communicate seamlessly with modern automated racking systems.

3. Real-Time Monitoring: The need for real-time monitoring and control to track pallet movement accurately ensures efficient material handling.

4. Scalability and Adaptability: Participating in the design of systems that scale, involved variable loads, pallet designs, and warehouse lay out.

15.3 Major Contributions of the Study

1. PLC-Based Control: Use of Programmable Logic Controllers (PLCs) in place of relay-based control circuits for automated and manual operation of racking. Items such as timers, counters, and relays can enhance the system's control flexibility.

2. HUMAN MACHINE INTERFACE (HMI):

* A human–machine interface (HMI) that observes active components and provides pallet counts for arriving and exiting loads.
* Provides the ability for operators to switch between manual and automatic operation modes.

3. System Design:

* Pallet transport conveyors.
* Motored lifters with vertical and horizontal movement.
* Redundant limit switches to detect lifter positions and increase the safety of the systems.

4.Automation Features:

* The system automates the storage and retrieval of pallets on six shelves spread over upper and lower tiers.
* Employs conveyors to transfer pallets to and from the racking system, minimizing manual handling.

5.Performance and Challenges:

* The system works reliably, but there are some mechanical issues such as alignment of pallets.
* Stresses the significance of conceptual readiness, tool development, and component comprehension in system development**.**

15.4 Overview

This paper is presenting a prototype of the Automatic Racking System to replace the forklifts in distributing industrial goods. The application of a Programmable Logic Controller with a Human-Machine Interface will enable smooth automation of the storage and retrieval of pallets on racks, which will increase effectiveness and reduce the possibility of accidents.

15.4.1 Key Contributions

1.PLC and HMI Integration:

* The system includes a fully functional PLC that was used for the automation process of storing and retrieving pallets.
* The HMI is used for monitoring and controlling the system operation, like counting the incoming and outgoing pallets and switching between manual and automatic modes.

2.Design and Implementation:

* Inlet conveyor for transportation of pallets to a mobile trolley
* The mobile trolley fills the shelves in a predefined sequence, starting from top to bottom
* Manual and automatic operations are supported to increase flexibility

3.Automation Details:

* Uses limit switches to determine motor activity and trolley positioning.
* Operates based on preset programs stored in the PLC, ensuring consistent performance.

4.Prototype Evaluation:

* A working miniature of the system demonstrated its functionality, showcasing successful pallet storage and retrieval operations.
* Demonstrates significant improvements in efficiency compared to manual operations.

15.4.2 Limitations

1.Static Environment Dependency:

* The system is designed for structured environments with fixed shelf layouts, limiting adaptability to dynamic or unstructured warehouse setups.

2.Limited Detection Capability:

* Relies on limit switches for positioning rather than using advanced sensors like LiDAR or cameras for dynamic detection and localization.

3.Lack of Scalability:

* The prototype may focus on a small-scale racking system that does not scale up well for larger industrial operations or warehouses with diverse layouts.

4.Manual Tuning Requirement:

* This requires manual tuning for every operation, reducing the efficiency of adapting to new environments or configurations.

15.5 How Our Project Overcomes These Limitations

| **Limitation in Paper** | **Our Approach** |
| --- | --- |
| **Static Environment Dependency**: Designed for fixed layouts. | Employs SLAM for dynamic mapping and navigation in unstructured and diverse warehouse environments. |
| **Limited Detection Capability**: Relies on limit switches. | Utilizes advanced 3D cameras (Orbbec Gemini 2) and LiDAR for precise real-time detection and localization. |
| **Lack of Scalability**: Focused on a small-scale setup. | Scalable design suitable for large warehouses with varying layouts and pallet types. |
| **Manual Tuning Requirement**: Requires manual setup for new configurations. | Leverages machine learning models for automated calibration and adaptability to new setups. |

15.6 Comparison with Our Project

| Aspect | Maniriho’s Work | Our Project |
| --- | --- | --- |
| Control System | PLC-based control with HMI integration | AI-based control with SLAM for navigation and real-time decisions |
| Scope | Focused on racking and pallet movement automation | Comprehensive (includes navigation, load detection, and obstacle avoidance) |
| Sensor Technology | Limit switches for position detection | Advanced 3D sensors (e.g., Orbbec Gemini 2, Intel RealSense) |
| Automation Mode | Manual and automatic operation | Fully autonomous forklift operation |
| Navigation | Fixed-position racking system | Dynamic navigation using SLAM in unstructured environments |
| Scalability | Limited to predefined shelving layouts | Adaptable to diverse warehouse layouts and pallet designs |

15.7 Benefits of our project

1. Broader Functional Scope: Above and beyond automated pallet movement such as load detection, pallet alignment and obstacle avoidance

2. Dynamic Navigation : it making use of simultaneous localization and mapping (SLAM) to operate in real-time inside synamic environments

3. Improved Sensors and The Integration of AI: Includes additional 3D cameras and AI-assisted models for enhanced precision and flexibility.

4. Scalable Design: Deployment of all kinds of warehouse layouts and pallets, more flexible

15.8 Disadvantages of Our Project

**1.** Higher Complexity: Subsystem integration results in higher computational needs and development durations.

2. Cost Implications: Developed hardware and firmware result to higher initial costs in comparison to PLC based systems.

3. Training Overhead: AI-based models need large data sets and need to be validated.

15.9 Similarities

1. Automation Goals: Aim in both projects is to improve efficiency while minimizing human involvement in material handling systems;

2. Real-Time Monitoring: Focusing on status of system and location of pallet.

3. Safety Mechanisms: Includes safety features (e.g., limit switches, collision avoidance)

15.10 Conclusion

The study of Maniriho gives a basic step towards the incorporation of automation into pallet racking systems using PLC, HMI. The system provides a marked improvement over manual methods, but its ability to scale and adapt is limited by its use of fixed layouts and basic sensors.

Our initiative extends these principles by:

• Navigation and load detection capabilities.

• Taking advantage of the power of advanced AI and sensor technologies for greater flexibility and scalability.

• Designing for real-world warehouse environments with dynamic and complex conditions.

Although more complex and expensive than our project, it meets the changing face of logistics and provides a comprehensive and flexible solution for autonomous operation of a forklift.

**----------------------------------------------------------------------------------------------**