AI-Powered Robotic Systems for Intelligent Retail Shelf Management

Hira Ajmal
ajmalhira.ajmal@gmail.com

***Abstract*—Efficient retail shelf organization is crucial for maintaining product visibility and optimizing customer experience. This study presents an AI-driven robotic system designed to autonomously detect, rearrange, and replenish items in dynamic retail environments. By integrating advanced perception models with knowledge-based reasoning, the system enables precise object recognition and strategic placement, even in cluttered spaces. A multi-layered planning framework ensures seamless manipulation and adaptation to real-time shelf conditions. Experimental evaluations demonstrate the robot's ability to execute intelligent restocking and organization strategies, offering a scalable solution for automated retail management.**

I. Introduction

The emergence of robotic automation has significantly transformed various domains, with autonomous mobile manipulators gaining increased accessibility. A fundamental research challenge is enabling these robotic platforms to interact with and rearrange their surroundings efficiently, addressing complex real-world manipulation objectives in variable and unstructured settings.

Retail and warehouse environments present suitable initial application areas for such technology, particularly in automated product handling. Typical responsibilities include replenishing shelves, organizing misplaced goods, and adjusting item configurations. Unlike factory settings, where operations follow predefined steps, these environments exhibit semi-structured but unpredictable characteristics. Items may be missing, placement rules evolve, and storage areas may be partially concealed, demanding adaptive solutions.

A retail scenario provides structured cues that robotic agents can exploit. Items are arranged to maximize visibility, their dimensions and shapes align with human handling capabilities, and packaging often includes useful details such as weight and branding. However, these factors also present obstacles; identical items are frequently adjacent, complicating segmentation and manipulation without disturbing surrounding objects.

The targeted robotic functionalities include initializing an empty display, restocking depleted goods, decluttering disordered shelves, and adjusting inventory layouts. Additionally, tasks extend to warehouse operations such as item retrieval, packing, and preparation for dispatch. The necessity for prolonged autonomous operation fosters opportunities for continual learning and adaptive improvement.

Recent advancements have demonstrated interest in autonomous item retrieval and placement. The Amazon Picking Challenge of 2015 emphasized the development of integrated perception and manipulation solutions within warehouse environments. While promising, many entries simplified challenges related to environmental clutter and high-level decision-making.

This work introduces an autonomous robotic system capable of executing structured reordering within shelving units. The methodology utilizes qualitative spatial relationships, such as positioning cereals adjacent to coffee products. A dynamic grid-based approach segments the target area, allocating compartments for different product categories, ensuring optimal orientation and accessibility.

The approach surpasses previous benchmarks by recognizing and handling objects despite substantial occlusion, identical adjacent items, and cluttered arrangements. The system extends beyond simple object transfer by integrating semantic reasoning, essential for achieving structured organization in real-world applications. Additionally, the avoidance of unnecessary shelf clearance is prioritized, ensuring precision in manipulation.

Summarizing the specific constraints within this framework, the primary contributions of this research include:

- **Perception:** Operating within an environment characterized by dense clutter, no predefined object positioning, extensive occlusion, recurring textures, and complex arrangements.
- **Knowledge-Based Reasoning:** Resolving occlusion through implicit secondary adjustments, implementing an autonomous solver for shelf organization, and interpreting abstract task instructions.
- **Manipulation:** Knowledge-driven object handling with integrated failure mitigation to ensure robustness and efficiency.

II. Related Work

A. Autonomous Assistants in Retail Spaces

The demanding task of organizing shelves in retail spaces has led to the emergence of robotic assistants. These robotic systems have been designed to offer diverse customer support but often lack direct interaction with store merchandise. Generally, such robots function in a partially automated manner, where they navigate users to a specified product within a designated section, yet the retrieval remains manual. Fully independent systems have recently surfaced, primarily optimized for warehouse logistics, while retail settings remain an area for further exploration.

Fig. 1: Retail environment featuring the robotic agent and identified objects within a sample scene.

Current methodologies for self-positioning have been refined to accommodate large-scale indoor environments such as supermarkets. However, merely determining the shelf location does not suffice for seamless object interaction. To efficiently manage items within a storage unit, a robotic system must acquire contextual awareness regarding individual products and their precise orientation.

A novel shelf-monitoring robot has been introduced, offering real-time inventory tracking while autonomously traversing aisles, notifying personnel about restocking needs. Rather than assigning analytical and handling responsibilities to human workers, the approach focuses on a fully automated solution.

B. Visual Processing in Densely Packed Retail Settings

Detecting products within congested retail environments presents numerous challenges, particularly as product arrangements shift due to customer activity. Furthermore, visibility may be hindered due to obstructions, and items can be methodically stacked while exhibiting varied orientations and appearances. Given the vast array of goods available in retail outlets, a reliable recognition system must accommodate differences in structural design, physical size, and surface textures.

To address these complex conditions, products can be equipped with RFID tags, facilitating effortless identification and spatial positioning. Alternatively, intrusive methods aim to recognize objects in disordered settings by altering their placement and observing scene modifications. Nevertheless, An image-based detection strategy is adopted, eliminating the need for object alterations or physical interaction. A single perception model may face challenges under these conditions; therefore, integrating multiple object recognition techniques enhances robustness and adaptability, ensuring compatibility with the intended deployment scenario.

C. Cognitive Reasoning and Adaptive Manipulation

For objects situated beyond immediate reach due to obstructing elements, prior research has utilized sampling-driven planning strategies to initially displace hindering objects, though implementations have largely remained in simulated environments.

An alternative approach implements similar concepts but deploys humanoid robots capable of handling barriers such as doors while progressing towards an objective. The ultimate goal aligns with ours: empowering robotic systems to instinctively perform prerequisite handling actions, even when these actions were not explicitly incorporated into their predefined tasks. Enabling such adaptive capabilities allows robots to respond dynamically to present conditions without direct human intervention.

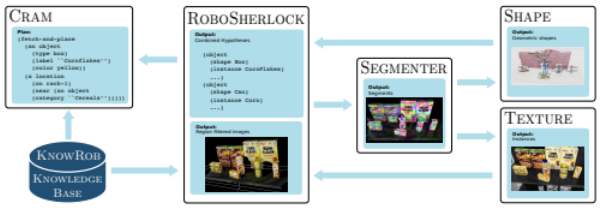

Fig. 2: Overview of the system framework, depicting cognitive inference (left) and perception (right) components.

III. System Overview

Solving the challenges outlined in this study necessitates a seamless integration of essential modules required by an independent robotic system. Figure 2 illustrates the comprehensive framework, showcasing the specific functionalities assigned to each component.

The architecture comprises three core elements, all of which have been recently introduced. These were chosen primarily due to their open-source nature, accessibility, and their capability to provide a modular and extensible structure:

- CRAM – advanced robotic planning and logical decision-making
- KNOWROB – centralized framework for processing and inferring knowledge
- ROBOSHERLOCK – knowledge-assisted robotic perception system

In the subsequent sections, it elaborate on the two fundamental perception modules incorporated in ROBOSHERLOCK, which play a crucial role in the recognition pipeline: a texture-based and a shape-based detection mechanism.

The system initiates with a generalized blueprint for shelf organization, formulated by CRAM. Initially, ROBOSHERLOCK is tasked with identifying the objects present. Utilizing spatial awareness data stored in the semantic repository of KNOWROB, the preliminary data is refined to retain only those areas pertinent to the current operation, such as the designated shelf.

Following this, the refined visual data undergoes instance and shape classification to infer potential products and their respective placements, subsequently relaying this information back to CRAM via ROBOSHERLOCK. Internally, the task planning mechanism evaluates multiple execution strategies

for reordering the shelf, prioritizing them based on cost efficiency computed through an A*-based pathfinding algorithm implemented in CRAM.

The most optimal approach is then selected for execution. Once the best course of action is determined, a corresponding manipulation sequence is formulated and carried out by the robotic agent. If any complications arise during execution (e.g., failure to grasp an object, accidental dropping, or inability to generate a feasible manipulation trajectory), control is reverted to CRAM, where a fallback mechanism is activated to devise an alternative execution plan.

Further elaboration on these components is provided in the following sections, with Figure 2 depicting additional sample outputs from the various perception and knowledge-processing subsystems.

IV. Perception in Complex Retail Settings

Object detection in retail environments presents substantial difficulties due to the diverse range of object textures, geometric forms, and spatial arrangements in confined areas. However, these settings provide contextual information that can be leveraged to improve perception tasks.

To address this challenge, an integrated approach is employed, combining prior knowledge with real-time perception. Specifically, predefined knowledge of the spatial structure of shelves and the expected inventory is utilized. This structured knowledge enables the dynamic optimization of perception tasks by narrowing the search space and concentrating on relevant regions.

Simultaneously, the system necessitates the ability to identify and distinguish individual objects on shelves, even under dynamic conditions such as varying orientations, occlusions, or novel object instances. Since training a perception model on all possible product variations is impractical, the system must be capable of generalizing and adapting to unseen scenarios.

Handling occlusions, deformable structures, and reflective surfaces adds complexity to object recognition. Therefore, the framework integrates multiple features, including color distribution, texture patterns, and geometric structure, to achieve accurate identification.

A. Perception Framework

The proposed system incorporates an advanced perception framework designed to efficiently analyze unstructured visual data. The framework interprets perception tasks as query-based processes, progressing through three distinct stages: (1) proposing candidate regions of interest within the captured images, (2) enriching these regions with semantic annotations, and (3) ranking and refining the hypotheses to maximize accuracy.

A key feature of this framework is the utilization of prior knowledge about the retail setting, which enhances perception capabilities. This stored knowledge enables efficient filtering of visual input, ensuring that only semantically significant regions undergo further analysis. Additionally, awareness of environmental context helps predict probable object categories, facilitating anomaly detection, such as misplaced items.

Furthermore, the system dynamically selects the most suitable recognition algorithms based on the task at hand. It also fuses outputs from multiple perceptual modules, resolving inconsistencies and improving reliability. To maintain coherence in object recognition across frames, the framework implements a memory module that preserves a record of detected entities.

The architecture supports seamless integration of diverse perception models. For this implementation, two specialized recognition modules—texture-based and shape-based—are incorporated as dedicated processing units. The framework ensures consistency by merging outputs from these modules and interfacing them with the reasoning component.

B. Object Recognition Modules

To address the complexities of retail perception, two recognition models were implemented within the framework, each catering to different object identification challenges. The texture-based model (Section IV-B1) facilitates instance-level recognition using prior knowledge, whereas the shape-based model (Section IV-B2) enables the classification of novel objects, ensuring adaptability in unfamiliar scenarios.

Both recognition strategies have been extensively evaluated in automated inventory management and product arrangement tasks. While perception in retail differs from logistics-based scenarios, common challenges—such as occlusion, clutter, and ambiguous object boundaries—necessitate similar solutions. Below, provided an overview of the recognition process.

Initially, raw RGBD data undergoes optional pre-processing steps, such as median-based noise reduction or depth-enhanced virtual scans. Subsequently, two segmentation techniques are applied:

- **Type I: Uninformed segmentation**—This method does not rely on prior knowledge of known objects. It over-segments the scene into primitive regions, forming the foundation for subsequent recognition tasks.
- **Type II: Knowledge-driven segmentation**—This approach refines segmentation by merging adjacent regions based on specific heuristics, such as surface continuity and convexity.

The over-segmentation step is particularly crucial, as it preserves object boundaries while generating granular segments for downstream processing. These refined segments provide a structured representation of the scene, enabling robust object analysis.

After segmentation, the system applies geometric filtering to remove irrelevant regions before passing the remaining data to the recognition modules.

1) Texture-based Object Recognition: The texture-based module employs a dual-phase recognition strategy, combining visual pattern extraction with depth-based geometric constraints. This method constructs a database of 3D object models augmented with visual descriptors.

To build this database, real-world items are digitally reconstructed using either autonomous scanning methods or

manual 3D modeling tools. Once stored, these models serve as reference points for object matching during perception tasks.

During recognition, segmented image patches are analyzed to extract visual features. A hypothesis generation step then employs a RANSAC-based algorithm to estimate object placement, ensuring geometric consistency between detected keypoints. The resulting pose estimates are validated against depth information, refining the hypotheses and eliminating false positives.

Final object identifications are determined based on hypothesis confidence scores. If multiple instances of an item are present, recognition is iteratively applied to detect each occurrence independently.

2) Shape-based Object Recognition: The shape-based recognition module focuses on unsupervised classification, utilizing hierarchical learning techniques to identify object structures. It employs a multi-level symbolic representation of surface geometry, which facilitates recognition without reliance on pre-existing models.

The recognition process begins with feature extraction, where geometric components are categorized into hierarchical symbols representing different levels of detail. A graph-based structure then encodes the relationships between these symbols, forming a model of shape constellations.

During inference, an energy-minimization algorithm matches observed object structures to stored shape categories, determining the most probable classification. This hierarchical approach enhances robustness by incorporating structural details across multiple scales, leading to higher recognition accuracy.

By integrating these modules, the system ensures comprehensive perception capabilities, enabling effective object detection in diverse retail scenarios.

V. Intelligent Decision-Making and Mobile Manipulation in Retail Scenarios

Effectively organizing items on a store shelf demands a diverse set of cognitive skills. A robotic system must autonomously devise a structured approach to achieve its objective, determining the optimal arrangement, sequence, and positioning of products. Executing this plan necessitates performing a sequence of grasping and placement actions in an environment that may be constrained and cluttered. A pre-established model of the shelf, inclusive of dimensions and object grasp configurations, is utilized to dynamically construct a spatial representation and compute feasible approaches for handling items.

These operations rely on dynamic decision-making processes influenced by highly variable data, including object orientations, robot positioning, and current versus target shelf configurations. A robust memory management framework is employed to store episodic interactions, facilitating the evaluation and refinement of devised strategies.

A. Intelligence-Driven Planning Framework

For efficient rearrangement of shelf products, an autonomous agent requires a structured methodology to determine an optimal sequence of actions, transforming an initial state S_i into a target configuration S_f. To achieve this, an enhanced heuristic-based search algorithm has been implemented with the following capabilities:

1) extbfAction Sequence Generation: The system generates parameterized sequences of fundamental robotic actions, ensuring $A(S_i) = S_f$. These include:
   - Grasp and release operations
   - Object transfer between robotic arms
   - Vertical torso adjustments for reaching varied shelf heights
   - Positional realignment for improved accessibility
2) extbfMulti-Solution Synthesis: Upon computing an initial solution, the algorithm assigns an artificially elevated cost to it, prompting the derivation of alternative solutions in order of decreasing efficiency. This approach ensures a repertoire of viable strategies for real-time decision-making.
3) **Categorical Goal Matching:** The planner is capable of satisfying category-based constraints instead of precise object placements, facilitating broader solution spaces and faster convergence.

Traditional heuristic search methods lack these enhancements. The modifications introduced are particularly beneficial for shelf-stocking tasks but are equally applicable in diverse robotic manipulation domains.

To execute a heuristic search, a state discrepancy measure and an estimated cost function are essential. If a state S_i deviates from S_j, implying a non-zero divergence metric, it is classified as entropic. The heuristic function H is redefined as:

$$H(S_i, S_j) = \frac{\text{Number of misaligned items in } S_i \text{ relative to } S_j}{\text{Total items in } S_i} \tag{1}$$

Generally, $H(S_i, S_j) \neq H(S_j, S_i)$ due to the influence of item count on the denominator. This formulation accounts for scenarios where available items are fewer than the total possible placements, a condition never reversed in practical applications.

The transition metric D, which determines the cost of transitioning between two states, is given by:

$$D(S_i, S_j) = \sum_{k=1}^{n} c(A_k) + C_r \tag{2}$$

where $c(A_k)$ represents the weighted cost of each atomic action A_k, and C_r accounts for deviations in the robot's positioning and grasp states. The symmetry condition holds as:

$$D(S_i, S_j) = D(S_j, S_i) \tag{3}$$

Ensuring that H never exceeds D upholds the admissibility criterion for heuristic search:

$$D(S_i, S_j) \geq H(S_i, S_j) \tag{4}$$

The generated sequence of operations is executed incrementally. If localized failures occur, such as invalid trajectories, replanning mechanisms intervene to rectify deviations. However, since the planner is based on symbolic representations, it does not inherently account for real-world uncertainties. Should the failure frequency exceed a predefined threshold, an entirely new action sequence is generated based on the latest environmental conditions.

B. Motion Strategy

The advanced motion strategies are devised and executed using the robotic planning framework CRAM. Within this framework, trajectory formulation is essential whenever the viability of a manipulation operation is assessed or when such an action is carried out. To facilitate path generation and actuator control, The MoveIt! framework is integrated, which excels in unconstrained path planning but demonstrates enhanced effectiveness when its planning domain is enriched with obstruction data. To ensure structured environmental knowledge is accessible to intelligent robotic systems governed by CRAM, the KNOWROB knowledge inference module is incorporated.

In the implementation, obstructions include elements such as shelving units, surrounding structural barriers, and manipulable objects. Figure 3 illustrates an example configuration of a retail environment, displaying supermarket furnishings. When objects are positioned on the shelf, the robotic system is informed about the shelving framework and actively avoids unnecessary contact during handling.

The planning domain must be managed externally to MoveIt! to be viable for manipulation tasks. While static elements like shelving and walls are retrieved from a predefined dataset, objects in the environment are detected dynamically. As modifications occur in real-world object arrangements, the system updates its planning domain accordingly, ensuring an accurate representation of restricted zones during robotic arm navigation.

In CRAM, the representation of the surroundings synchronizes with the robot's internal belief framework upon detecting critical state changes. These changes encompass but are not confined to, adjustments in navigation, object manipulation, and perception.

C. Grasp Strategy and Implementation

Once objects are recognized and their intended placement positions are determined, a loosely defined retrieval-and-positioning task must be refined for precise execution. Keeping this task definition flexible allows an autonomous system to optimize key parameters for execution:

```
(retrieve-and-position
    (an item
        (type container)
        (label ``CerealBox'')
        (shade golden))
    (a target
        (on shelf-1)
        (adjacent-to (an item
            (category ``BreakfastFoods'')))))
)
```

Fig. 3: Robot in a simulated retail setting (left); digital reconstruction of the environment for decision-making (right).

Assuming the system has detected a golden container labeled "CerealBox," it must determine:

- The optimal method to secure the object.
- The precise three-dimensional placement coordinates.

CRAM subsequently updates its representation with an object mirroring the size and estimated pose of the cereal box. The MoveIt! planning framework then generates a controlled trajectory where the gripper moves into proximity, initiates secure contact with the item, and performs the grasp. A motion sequence is then planned to transport the box safely away from the shelf.

For placement, a similar trajectory computation is executed; however, the final coordinates require precise determination. The shelving structure contains dimensional details alongside metadata defining surfaces capable of supporting objects. Within CRAM, the loosely defined target location is refined using contextual information and modeled data to generate a specific six-degree-of-freedom placement coordinate:

$$\text{Placement Accuracy} = \frac{\text{Correctly Positioned Items}}{\text{Total Objects Handled}} \tag{5}$$

This transformation ensures that the robotic system autonomously selects a feasible and efficient arrangement within the shelving unit, optimizing both spatial constraints and execution feasibility.

D. Memory Retention and Data Evaluation

For an intelligent system to make autonomous choices, it must integrate pre-existing structured data with real-time contextual attributes. While foundational knowledge is systematically formulated, transient situational data is often complex to capture and analyze when interpreting a robot's decision-making process.

Within the framework, a robust data archival system is implemented to log all dynamic influences affecting decision-making while the robotic system executes its predefined plans.

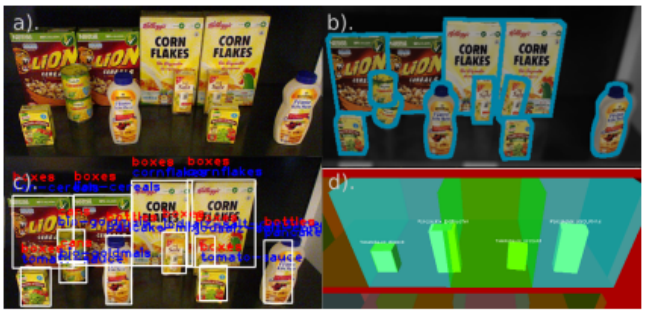

Fig. 4: Processing of perception data for a sample scenario: (a) Initial image (b) Instance recognition output (c) Processed results within RoboSherlock (d) Objects designated for manipulation as interpreted by the robot's cognitive model.

This accumulated repository of episodic interactions provides a structured dataset for retrospective assessment of robotic conduct. Such post-execution analysis aids in identifying latent discrepancies and refines future decision-making protocols.

The heuristic evaluation of recorded sequences assists in deducing patterns that remain undetectable during real-time execution. By leveraging this historical dataset, alternative courses of action can be simulated to improve strategy formation and enhance overall performance efficiency.

VI. Experimental Assessment

The performance of the framework is assessed by analyzing the effectiveness of the generated execution plans derived from actual sensory input. Additionally,Two operational scenarios involving a PR2 robot are presented, illustrating the reorganization of items and the decision-making process within cluttered environments.

A. Configuration

To validate the practicality of the holistic perception and knowledge-driven manipulation paradigm, The implementation is carried out on a PR2 robot operating within a simulated retail environment. As shown in Figure 5, this setting includes a storage unit resembling those in supermarkets, stocked with a variety of products differing in size, shape, texture, and weight, sourced from multiple grocery categories.

The core advantage of the methodology is the endowment of an autonomous system with intelligence-supported methodologies for restructuring misplaced products within such storage arrangements. Given an initial configuration and a target layout, the robot must formulate decisions to accomplish the rearrangement optimally. Leveraging CRAM, the robot executes a structured sequence to fulfill its objective:

- Analyzing the existing shelf occupancy and extracting symbolic representations of detected objects and their spatial orientations.
- Formulating an optimized sequence of actions to transition the rack into a pre-defined arrangement.
- Carrying out these tasks through base movement, torso adjustments, and executing diverse grasping and placement operations.
- Effectively handling execution failures by reassessing the present conditions, recalculating the motion strategy, and proceeding with task execution.

During experimentation, a PR2 robot is presented with a shopping rack containing disorganized items. Following the computation of a suitable manipulation sequence, the robot systematically relocates all items to their designated positions.

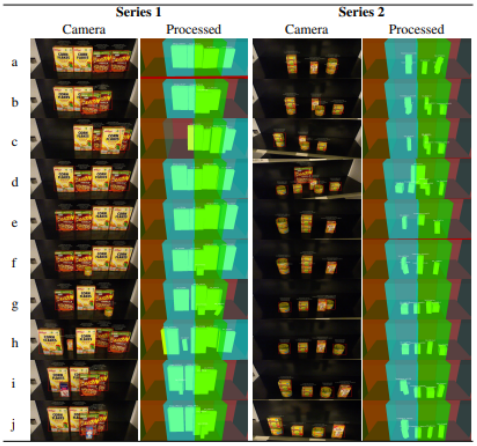

Fig. 5: Various experimental setups displaying different object placements within a shopping rack. The left side contains the raw visual input, while the right side represents the system-processed object identification outputs.

Action $A_{i,k}$	**Grasp**	**Position**	**Torso Adjustment**	**Base Navigation**
Weight w	1.2	1.2	2.0	1.0

TABLE I: Adjusted heuristic weight parameters employed in the modified A* pathfinding algorithm for computing the total cost of an execution sequence.

B. Experimental Assessment

The framework is evaluated by examining the effectiveness of task sequences generated from real-world sensory data. Furthermore, two execution sequences on a PR2 robot are presented, highlighting object reorganization and strategic decision-making in scenarios with occluded items.

C. Experimental Configuration

To validate the effectiveness of the integrated perception and cognition-driven manipulation approach, The implementation was carried out on a PR2 robot functioning within a commercial retail environment. As illustrated in Figure 3, this setup features a shelving unit similar to those in standard supermarkets, containing products of varying sizes, textures, and weights across multiple categories.

The advantage of the methodology lies in endowing an autonomous robot with knowledge-informed tactics for reorganizing products in these storage units. Given an initial

layout and a target configuration, the robot must determine the appropriate manipulative actions. Utilizing CRAM, the robot is provided with a repertoire of methods to fulfill its objectives:

- Identifying the current state of shelf occupancy, extracting symbolic data regarding the objects and their positions.
- Formulating an action strategy to transition the rack from its current state to the intended arrangement.
- Implementing this plan by adjusting the robot's base, torso, and executing sequential pick-and-place maneuvers.
- Responding dynamically to unexpected failures by backtracking, revising the plan, and continuing execution.

In the trials, the robot is presented with a disordered retail shelf containing misplaced products. After computing an optimal action strategy, the robot proceeds to reposition all items to their designated locations.

D. Planning Trials

To highlight the significance of synchronizing system components in handling intricate retail shelving scenarios, examined a collection of representative cases. These scenarios illustrate how slight modifications in sensory data influence the planner's output and subsequent object manipulation.

Figure 5 displays twenty shelf arrangements, categorized into two distinct groups. Each scenario includes raw camera input and its corresponding processed data. Variations in object count and positioning affect perception, resulting in instances of occlusion (cases 1.b-c, f-j), irregularly shaped objects (cases 1.f, h-j), and stacked formations (cases 2.a, e-f, b-c, i-j). In some instances, objects were either misidentified or completely undetected (cases 2.c, e). Table 2 presents the specifics of planned task sequences based on these 20 cases.

The planner's objective was to cluster available objects and evenly allocate them across two shelves. Initially, none of the objects were positioned correctly.

The cost in Table 2 represents the cumulative expense of all individual operations, as computed using a modified A* search algorithm. The associated action costs are detailed in Table 1. Grasping and placing are relatively straightforward operations, whereas adjusting the torso requires additional time. Since execution duration is a key metric for optimizing shelf organization, torso movements are weighted more heavily. Conversely, moving the robot's base is swift, while transferring an item between hands is more time-consuming.

As indicated in Table 2, task complexity does not significantly inflate the action cost, but rather extends the computational time required for planning. For example, scenario 2.d was particularly demanding, necessitating 50 seconds for computation and yielding a total cost of 32.4.

Perceptual inaccuracies do not directly reflect in the planner's results. In scenario 2.c, two stacked items were perceived as a single object, leading to a simpler yet incorrect plan. Similarly, in case 2.e, lower stacked items were entirely undetected, resulting in an erroneous final placement and an artificially low-cost, quickly computed task sequence. Such discrepancies are mitigated by the failure detection and correction mechanisms in CRAM. After each manipulation step, the robot reassesses the scene and recalibrates its approach if inconsistencies arise.

E. Robot Execution Trials

To validate the practical feasibility of the proposed framework,Multiple trials were performed on a PR2 robot, with the experimental results depicted in Figure 6. These trials were structured to highlight the robot's cognitive capabilities in organizing objects effectively.

The first experiment (Figure 6, upper sequence) demonstrates the process of rearranging objects based on similarity. The initial object placements and the perception results are depicted in Figure 4. Upon recognizing the objects, the robotic agent devises an optimized strategy to position similar objects adjacent to each other. In this instance, it determines that the pancake mix and tomato sauce should be swapped.

In the second trial, the system showcases its inference-driven decision-making capabilities when faced with occluded items requiring manipulation. This is accomplished through qualitative spatial reasoning based on the robot's position relative to the shelf. The robot utilizes its knowledge base to infer object relationships and deduces that the Lion cornflakes box is obstructed by a salt container. To retrieve the cornflakes without interference, the system prioritizes removing the salt container first.

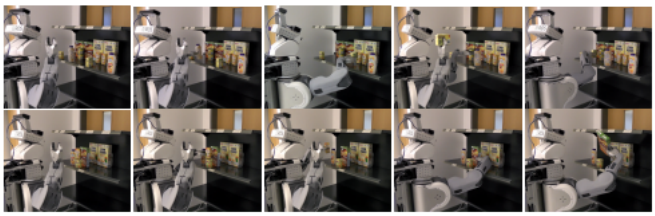

Fig. 6: Top row: Reorganizing objects based on target arrangement. Left to right: initial object placement, grasping the pancake mix, holding the tomato sauce, repositioning the pancake mix, and placing the tomato sauce. Bottom row: Handling occlusion. Left to right: original object placement, reaching for the salt, relocating the salt, grasping the cereal, and moving the cereal.

VII. Conclusion

This research introduces an innovative implementation of intelligence-driven manipulation within a routine setting that necessitates advanced perception and logical inference abilities.

A fully operational and cohesively integrated system has been demonstrated, incorporating a cognition-augmented perception module, an optimized planner for reorganization strategies, and the essential manipulation proficiencies within a robotic architecture. The key challenges addressed in this scenario encompass various facets of robotics, such as object

recognition in complex settings and the development of goal-driven methodologies for efficient robotic operations.

The practicality of the approach has been validated using a robotic agent operating in a setting that mimics real-world retail scenarios. Presently, assessing the proficiency of comprehensive robotic frameworks remains a challenge, as conventional methods focus primarily on evaluating discrete components. the experimental outcomes highlight the correlation between the outputs of an actual perception system and the formulation of manipulation strategies for robotic execution. Establishing a measurable link between environmental understanding and the effectiveness of robotic plans facilitates further exploration of alternative methodologies to enhance robotic functionality.

To summarize, this study underscores the significance of both sophisticated perceptual frameworks and cognition-driven strategic planning in achieving autonomous manipulation within a dynamic and complex real-world setting.

REFERENCES

[1] M. Beetz, F. Balint-Benczedi, N. Blodow, D. Nyga, T. Wiedemeyer, and Z.-C. Marton, "RoboSherlock: Unstructured Information Processing for Robot Perception," in *International Conference on Robotics and Automation*, 2015.

[2] M. Beetz, L. Moesenlechner, and M. Tenorth, "CRAM – A Cognitive Robot Abstract Machine for Everyday Manipulation in Human Environments," in *International Conference on Intelligent Robots and Systems*, 2010.

[3] M. Beetz, M. Tenorth, and J. Winkler, "Open-EASE – A Knowledge Processing Service for Robots and Robotics/AI Researchers," in *International Conference on Robotics and Automation*, 2015.

[4] D. Canelhas, T. Stoyanov, and A. Lilienthal, "Improved local shape feature stability through dense model tracking," in *International Conference on Intelligent Robots and Systems*, 2013.

[5] S. Chitta, I. Sucan, and S. Cousins, "MoveIt! [ROS topics]," *Robotics & Automation Magazine*, vol. 1, no. 19, pp. 18–19, 2012.

[6] L. Chuan, A. Johari, M. Wahab, D. Nor, N. Taujuddin, and M. Ayob, "An RFID warehouse robot," in *International Conference on Intelligent and Advanced Systems*, 2007.

[7] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, May 2002.

[8] D. Ferrucci and A. Lally, "UIMA: An Architectural Approach to Unstructured Information Processing in the Corporate Research Environment," *Natural Language Engineering*, vol. 10, no. 3-4, pp. 327–348, 2004.

[9] C. Gharpure and V. Kulyukin, "Robot-assisted shopping for the blind: Issues in spatial cognition and product selection," *Intelligent Service Robotics*, vol. 1, no. 3, pp. 237–251, 2008.

[10] T. Kanda, M. Shiomi, Z. Miyashita, H. Ishiguro, and N. Hagita, "An Affective Guide Robot in a Shopping Mall," in *International Conference on Human Robot Interaction*, 2009.

[11] S. Koo, D. Lee, and D.-S. Kwon, "Unsupervised object individuation from RGB-D image sequences," in *International Conference on Intelligent Robots and Systems*, 2014.

[12] R. Krug, T. Stoyanov, M. Bonilla, V. Tincani, N. Vaskevicius, G. Fantoni, A. Birk, A. Lilienthal, and A. Bicchi, "Improving Grasp Robustness via In-Hand Manipulation with Active Surfaces," in *International Conference on Robotics and Automation – Workshop on Autonomous Grasping and Manipulation: An Open Challenge*, 2014.

[13] R.-G. Mihalyi, K. Pathak, N. Vaskevicius, T. Fromm, and A. Birk, "Robust 3D Object Modeling with a Low-Cost RGBD-Sensor and AR-Markers for Applications with Untrained End-Users," *Robotics and Autonomous Systems*, vol. 66, Apr. 2015.

[14] C. Mueller, K. Pathak, and A. Birk, "Object recognition in RGBD images of cluttered environments using graph-based categorization with unsupervised learning of shape parts," in *International Conference on Intelligent Robots and Systems*, 2013.

[15] C. Mueller, K. Pathak, and A. Birk, "Object shape categorization in RGBD images using hierarchical graph constellation models based on unsupervisedly learned shape parts described by a set of shape specificity levels," in *International Conference on Intelligent Robots and Systems*, 2014.

[16] K. Okada, A. Haneda, H. Nakai, M. Inaba, and H. Inoue, "Environment manipulation planner for humanoid robots using task graph that generates action sequence," in *International Conference on Intelligent Robots and Systems*, 2004.

[17] J. Papon, A. Abramov, M. Schoeler, and F. Wörgötter, "Voxel Cloud Connectivity Segmentation – Supervoxels for Point Clouds," in *Computer Vision and Pattern Recognition*, 2013.

[18] M. Stilman, J.-U. Schamburek, J. Kuffner, and T. Asfour, "Manipulation planning among movable obstacles," in *International Conference on Robotics and Automation*, 2007.

[19] T. Tasaki, S. Tokura, T. Sonoura, F. Ozaki, and N. Matsuhira, "Mobile robot self-localization based on tracked scale and rotation invariant feature points by using an omnidirectional camera," in *International Conference on Intelligent Robots and Systems*, 2010.

[20] M. Tenorth and M. Beetz, "KnowRob: A knowledge processing infrastructure for cognition-enabled robots," *International Journal of Robotics Research*, vol. 32, no. 5, pp. 566–590, 2013.

[21] N. Vaskevicius, K. Pathak, and A. Birk, "Fitting superquadrics in noisy, partial views from a low-cost RGBD sensor for recognition and localization of sacks in autonomous unloading of shipping containers," in *International Conference on Automation Science and Engineering*, 2014.

[22] N. Vaskevicius, K. Pathak, A. Ichim, and A. Birk, "The Jacobs robotics approach to object recognition and localization in the context of the ICRA'11 Solutions in Perception Challenge," in *International Conference on Robotics and Automation*, 2012.

[23] T. Wiedemeyer, F. Balint-Benczedi, and M. Beetz, "Pervasive 'Calm' Perception for Autonomous Robotic Agents," in *International Conference on Autonomous Agents and Multiagent Systems*, 2015.

[24] Will Knight, Technology Review, "Robot Makes Sure Stores Don't Run Out of Doritos," Available: http://www.technologyreview.com/news/543281/robot-makes-sure-stores-dont-run-out-ofdoritos, Accessed: 2015-11-16.