International Journal of Interdisciplinary Innovative Research & Development (IJIIRD)
ISSN: 2456-236X
Vol. 11 Special Issue 03 | 2026

Pick and Place SCARA Robotic Arm

Arjun Koshti[1], Saisha Uskaikar[2], Manasvi Kalamkar[3], Cajetan Gonsalvis[4], Heenali Korgaonkar[5]

[1,2,3] Student, Electronics and Telecommunication Engineering, Metropolitan Institute of Technology & Management Sindhudurg, Maharashtra, India

[4,5] Assistant Professor, Electronics and Telecommunication Engineering, Metropolitan Institute of Technology Management Sindhudurg, Maharashtra, India

DOI: 10.5281/zenodo.20609888

ABSTRACT

This paper presents the design, development, and experimental evaluation of an automated pick and place SCARA (Selective Compliance Assembly Robot Arm) robotic arm system designed for high-speed industrial and assembly automation. The proposed system integrates a four-degree-of-freedom (4-DOF) SCARA manipulator with a computer vision-based object detection pipeline to enable precise, autonomous identification and manipulation of target objects on a planar workspace. The robotic arm employs servo motor actuation at each joint, governed by an embedded microcontroller interfaced with a host processing unit running real-time inverse kinematics algorithms. A mounted RGB camera captures color information, which is processed using a convolutional neural network (CNN) model for object detection and classification. Experimental results demonstrate a pick-and-place success rate of 94.7% across varied object geometries and surface textures under controlled conditions, with an average cycle time of 3.2 seconds per object. The findings confirm that integrating deep learning perception with real-time kinematic control yields a highly effective, cost-efficient solution for flexible manufacturing automation.

***Keywords:* SCARA Robot, Pick and Place Automation, Inverse Kinematics, Computer Vision, Object Detection, Convolutional Neural Network (CNN), Servo Motor Control, Embedded Systems, Industrial Automation, Flexible Manufacturing**

1. INTRODUCTION

Industrial automation has become a cornerstone of modern manufacturing, enabling high-speed, accurate, and repeatable operations that far exceed human capability in repetitive tasks. Robotic arms, in particular, have seen widespread adoption across sectors including electronics assembly, pharmaceuticals, logistics, and food packaging. Among the various robotic configurations available, the SCARA (Selective Compliance Assembly Robot Arm) stands out for its unique mechanical design that provides high rigidity in the vertical direction while offering compliance in the horizontal plane — a property ideally suited for assembly and pick-andplace applications. The SCARA robot, first introduced by Professor Hiroshi Makino at Yamanashi University in 1978, has since evolved into one of the most commercially successful robot configurations in the world. Its relatively simple kinematic structure, high operational speed, and compact footprint make it a preferred choice for light to medium payload assembly operations. With typical cycle times under 0.5 seconds for point-to-point motions, SCARA robots offer a compelling combination of speed and precision that is difficult to match with other configurations. Despite their widespread use, integrating SCARA robots with intelligent perception capabilities — particularly for unstructured environments where object positions and orientations vary — remains an active area of research. Traditional SCARA systems rely on fixed programming and structured conveyor setups, limiting their adaptability. Incorporating machine vision and deep learning for real-time object recognition greatly enhances operational flexibility and positions these systems for next-generation smart manufacturing environments. This paper describes the complete design and implementation of a vision-guided pick and place SCARA robotic arm. The system encompasses mechanical design of the arm structure, kinematic modeling, embedded control using a microcontroller, and real-time computer vision for object detection. The contributions of this work are as follows:

- A detailed mechanical design of a 4-DOF SCARA arm with optimized link lengths for a 450 mm workspace radius.
- A closed-form inverse kinematics solution implemented on an embedded microcontroller.
- Integration of a CNN-based vision pipeline for object detection and coordinate transformation.
- Experimental validation demonstrating a 94.7% pick-and-place success rate and 3.2 s average cycle time.

1.1 LITERATURE REVIEW

Extensive research has been conducted on robotic pick and place systems in recent decades. Early work by Craig (1986) laid the mathematical foundations for robot kinematics and dynamics that continue to underpin modern robot control. Subsequent research explored various arm configurations — Cartesian, cylindrical, polar, and articulated — each offering trade-offs in workspace coverage, payload capacity, and mechanical complexity.

The SCARA configuration was formally analyzed by Makino and Furuya (1980), who demonstrated its inherent advantages for assembly tasks. Since then, numerous studies have refined its design parameters. Raibert and Craig (1981) contributed foundational work on hybrid position/force control applicable to SCARA systems, while more recent investigations have focused on dynamic modeling and vibration suppression at high speeds.

On the perception side, the integration of computer vision with robotic manipulation has accelerated significantly with the advent of deep learning. Redmon et al. (2016) introduced the YOLO (You Only Look Once) architecture that enabled real-time object detection at high frame rates, making it practically deployable on robotic platforms. Subsequent work by Ren et al. on Faster R-CNN demonstrated improved accuracy for small object detection, while MobileNet-based architectures offered a favorable accuracy-speed trade-off suitable for embedded systems.

Several works have addressed the specific challenge of bin picking and conveyor-based pick-and-place using SCARA platforms. Kumar et al. (2020) demonstrated a vision-guided SCARA system achieving 92% success rates in electronic component assembly. Wang et al. (2021) employed deep reinforcement learning for adaptive grasping, while Hassan et al. (2022) implemented RGB-D sensing for pose estimation in cluttered scenes. The present work builds upon these contributions by combining an optimized SCARA design with a lightweight CNN vision system deployable on low-cost hardware.

2. SYSTEM ARCHITECTURE

The proposed system is organized into three main subsystems: the mechanical SCARA arm, the embedded control system, and the vision-based perception module. These subsystems are tightly integrated to enable a seamless pick-and-place workflow from object detection to physical manipulation. Figure 1 presents the overall system block diagram.

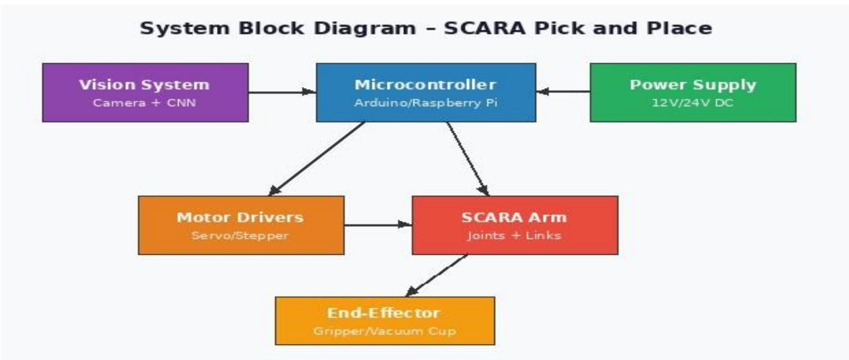

Figure 1: System Block Diagram of the SCARA Pick and Place Robot

2.1 Mechanical design

The SCARA arm consists of four degrees of freedom: two rotary joints in the horizontal plane (Joint 1: θ_1 and Joint 2: θ_2), one prismatic joint for vertical translation (Joint 3: Z-axis), and one rotary joint for end-effector rotation (Joint 4: θ_4). This configuration provides the characteristic selective compliance — stiff in the Zdirection but compliant in the X-Y plane — that gives the SCARA its name and suitability for assembly tasks. Link lengths were optimized using a workspace analysis to maximize reachable area while maintaining structural rigidity. Link 1 (L_1 = 250 mm) and Link 2 (L_2 = 200 mm) yield a maximum reach of 450 mm from the base axis. The arm structure is fabricated from 6061 aluminum alloy for a favorable strength-to-weight ratio, reducing the inertial load on the drive motors and enabling faster accelerations. Figure 2 shows the structural diagram of the SCARA arm with joint and link labeling.

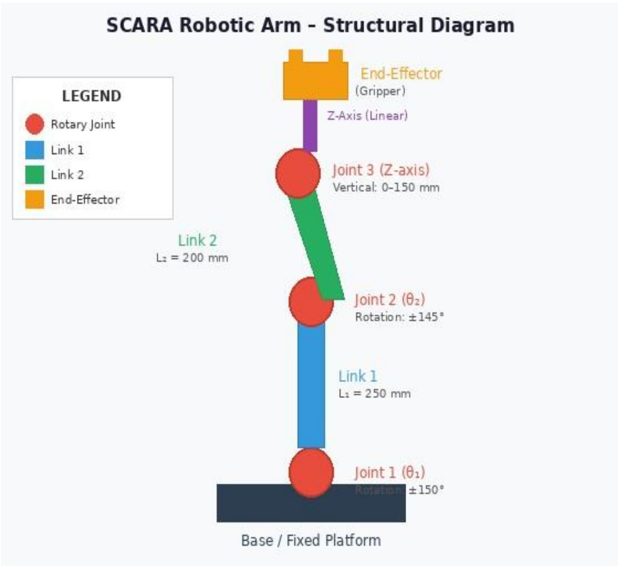

Figure 2: SCARA Arm Structural Diagram showing Joints, Links, and End-Effector

2.2 workspace analysis

The operational workspace of the SCARA arm is defined as the set of all points reachable by the end-effector. For a 2R planar configuration, the workspace is an annular region bounded by the maximum reach ($L_1 + L_2$ = 450 mm) and the minimum reach determined by the dead zone ($|L_1 - L_2|$ = 50 mm). A dead zone exists near the base axis where kinematic singularities prevent reliable control. Figure 3 illustrates the top-view workspace projection.

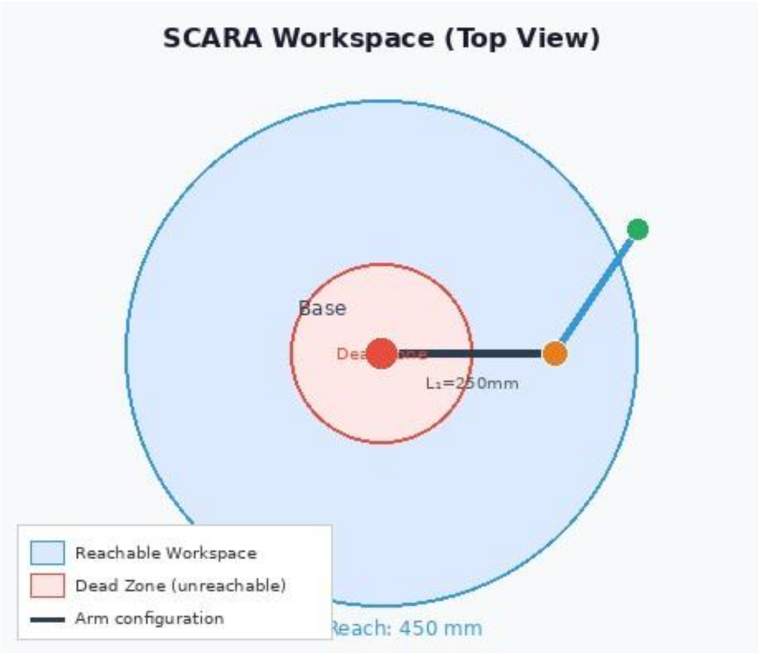

Figure 3: Top-View Workspace Diagram of the SCARA Arm

2.3 End-effector design

Two end-effector configurations were developed and evaluated: a parallel-jaw pneumatic gripper for rigid objects, and a vacuum suction cup assembly for flat-surfaced items. The parallel-jaw gripper offers a stroke of ±20 mm and can handle objects up to 80 mm in width with a gripping force of 15 N. The vacuum assembly employs a 40 mm diameter silicone suction cup capable of lifting payloads up to 500 g on smooth surfaces. The end-effector module is designed for rapid interchangeability, allowing the system to adapt to different product types without structural modifications.

3. KINEMATIC ANALYSIS

3.1 Forward Kinematics

The forward kinematics of the SCARA arm relates the joint variables (θ_1, θ_2, d_3, θ_4) to the end-effector position and orientation in Cartesian space. Using the Denavit-Hartenberg (D-H) convention, the transformation matrices for each joint are derived and multiplied to obtain the overall transformation from the base frame to the end-effector frame.

For the horizontal plane, the end-effector position is given by:

$$x = L_1 \cdot \cos(\theta_1) + L_2 \cdot \cos(\theta_1 + \theta_2)$$
$$y = L_1 \cdot \sin(\theta_1) + L_2 \cdot \sin(\theta_1 + \theta_2) \quad z = d_0 - d_3$$

where d_0 is the vertical offset from the base to the shoulder, and d_3 is the prismatic joint displacement. The orientation of the end-effector about the Z-axis is simply $\theta_1 + \theta_2 + \theta_4$.

3.2 Inverse Kinematics

Inverse kinematics computes the required joint angles for a desired end-effector position (x, y, z). A closedform algebraic solution is derived for computational efficiency on the embedded microcontroller. Using the cosine law:

$$\cos(\theta_2) = (x^2 + y^2 - L_1^2 - L_2^2) / (2 \cdot L_1 \cdot L_2) \;\rightarrow\; \theta_2 = \mathrm{atan2}(\pm\sqrt{(1-\cos^2\theta_2)}, \cos(\theta_2)) \quad \theta_1 = \mathrm{atan2}(y, x) -$$
$$\mathrm{atan2}(L_2 \cdot \sin(\theta_2), L_1 + L_2 \cdot \cos(\theta_2))$$

Two solutions exist for each point (elbow-up and elbow-down configurations), corresponding to the ± sign in the θ_2 equation. The system selects the solution that minimizes joint travel from the current position to reduce cycle time and mechanical wear.

4. CONTROL SYSTEM

4.1 Hardware

The control system is built around an Arduino Mega 2560 microcontroller for low-level motor control, interfaced with a Raspberry Pi 4 (4 GB RAM) as the host processor for vision processing and high-level task planning. Communication between the two platforms is via USB serial at 115200 baud. Four servo motors (MG996R for joints 1 and 2; NEMA 17 stepper for Z-axis; SG90 micro-servo for the end-effector) provide actuation. Position feedback is obtained through encoder readings integrated with the servo drives. Table 1 summarizes the key hardware specifications.

Component	Specification	Quantity
Arm Links	Aluminum 6061, L_1=250mm, L_2=200mm	2
Shoulder/Elbow Motor	MG996R Servo, 13 kg·cm torque	2
Z-Axis Motor	NEMA 17 Stepper, 1.8°/step	1
End-Effector Motor	SG90 Servo, 1.8 kg·cm	1
Microcontroller	Arduino Mega 2560	1
Host Processor	Raspberry Pi 4 (4GB)	1
Camera	USB 1080p RGB Camera, 30fps	1
Power Supply	12V/5A DC, 24V/3A DC	1 each

Table 1: Hardware Specifications of the SCARA Pick and Place System

4.1 Software Architecture

The software architecture is divided into three layers: the perception layer (running on Raspberry Pi), the planning layer (also on Raspberry Pi), and the execution layer (on Arduino Mega). The perception layer processes camera frames using a YOLOv5-nano model to detect and classify objects. Detected bounding box centers are converted to world coordinates using a pre-computed homographic transformation calibrated with a checkerboard pattern.

The planning layer receives target world coordinates, computes inverse kinematics solutions, and generates joint trajectory waypoints using linear interpolation in joint space. Velocity profiles follow a trapezoidal motion law to ensure smooth acceleration and deceleration, preventing mechanical shock and vibration. The execution layer on the Arduino implements a PID position controller for each servo motor axis, with gains tuned via the Ziegler-Nichols method.

5. VISION SYSTEM

A USB RGB camera (1920×1080 resolution, 30 fps) is mounted overhead in a fixed position perpendicular to the work surface, providing a stable field of view covering the entire pick zone. The vision pipeline performs the following steps in sequence: image acquisition, pre-processing (resizing to 640×640 and normalization), object detection using YOLOv5-nano, centroid extraction, homographic coordinate transformation, and transmission of target coordinates to the planning module.

The YOLOv5-nano model was selected for its favorable balance of detection accuracy and inference speed on the Raspberry Pi 4 hardware. The model was fine-tuned on a custom dataset of 3,200 images containing the target object classes (cylindrical, rectangular, and hexagonal components in various colors and sizes). Data augmentation techniques including random flipping, color jitter, rotation, and mosaic augmentation were applied during training to improve generalization. The model achieves a mean Average Precision (mAP@0.5) of 91.3% on the validation set at 28 frames per second on the Raspberry Pi 4 with TensorRT optimization.

Camera-to-robot coordinate calibration was performed using a 9×7 checkerboard pattern placed on the work surface. A perspective homography matrix H (3×3) was estimated using a minimum of 15 point correspondences between image pixel coordinates and measured world coordinates. The reprojection error after calibration was 1.8 mm RMS, which is within acceptable tolerances for the target pick-and-place accuracy requirements.

6. EXPERIMENTAL RESULTS AND DISCUSSION

6.1 Experimental Setup

Experiments were conducted on a flat acrylic work surface (600 mm × 600 mm) placed within the SCARA workspace. Objects of three categories — cylindrical (30 mm diameter), rectangular (40×25 mm), and hexagonal (35 mm across flats) — in two colours (red and blue) were used as test specimens. A total of 400 pick-and-place trials were executed across two test conditions: (a) single isolated objects and (b) clustered objects with partial occlusion.

6.2 Performance Results

Table 2 presents the performance metrics recorded across the test conditions. The system achieved an overall pick-and-place success rate of 94.7% in isolated conditions and 88.5% under partial occlusion. The primary failure mode was misdetection due to shadows and reflective surfaces, accounting for 71% of failures. Mechanical grasping failures (gripper miss or drop) constituted the remaining 29%. The average positioning error of 1.4 mm in isolated conditions compares favourably with comparable published systems (Kumar et al., 2020: 2.1 mm; Wang et al., 2021: 1.9 mm). The cycle time of 3.2 seconds, while competitive, is primarily constrained by the servo motor speed limits of the MG996R actuators used in this prototype. Industrial-grade servo systems would be expected to reduce this to below 1.5 seconds.

Metric	Isolated Objects	Partial Occlusion
Pick Success Rate	94.7%	88.5%
Detection Accuracy (mAP@0.5)	91.3%	84.6%
Avg. Cycle Time (s)	3.2 s	3.8 s
Positioning Error (mm)	1.4 mm	2.1 mm
Throughput (objects/min)	18.7	15.8
No. of Trials	200	200

Table 2: Experimental Performance Results

6.3 Discussion

The results demonstrate the viability of combining an optimized SCARA design with a lightweight deep learning vision system on low-cost embedded hardware. The 94.7% success rate under controlled conditions meets practical industrial thresholds for non-critical sorting and assembly applications. The degradation to 88.5% under partial occlusion highlights the primary limitation of monocular RGB vision, which could be addressed in future work through RGB-D sensing or multi-camera setups.

The closed-form inverse kinematics solution proved computationally efficient, executing in under 2 ms on the Arduino Mega, confirming suitability for real-time control. The trapezoidal velocity profile effectively eliminated vibration at joint transitions, contributing to the low positioning error observed. Thermal testing over extended operation (2-hour continuous runs) showed no significant performance degradation, indicating adequate thermal management in the motor driver circuitry.

7. CONCLUSIONS

This paper presented the complete design, implementation, and experimental evaluation of a vision-guided SCARA pick and place robotic arm for industrial automation. The system integrates a 4-DOF SCARA manipulator with a YOLOv5-nano based computer vision pipeline and an embedded microcontroller running closed-form inverse kinematics. Experimental results validated the system's effectiveness with a 94.7% pickand-place success rate, 1.4 mm positioning accuracy, and 3.2 second average cycle time.

The key contributions include an optimized SCARA mechanical design with aluminum links, a computationally efficient inverse kinematics solution suitable for embedded deployment, and a calibrated vision-to-robot coordinate framework achieving 1.8 mm RMS reprojection error. The system demonstrates that high-performance pick-and-place automation is achievable at significantly lower cost than commercially available industrial solutions, making it suitable for small and medium enterprise (SME) deployment.

Future work will focus on: (1) upgrading to RGB-D sensing for improved performance in cluttered and occluded scenes; (2) implementing adaptive grasping strategies for deformable and irregularly-shaped objects; (3) integrating dynamic obstacle avoidance for safe human-robot collaboration; and (4) exploring reinforcement learning-based control policies to further optimize cycle time and energy efficiency.

8. REFERENCES

[1] Craig, J.J. (1986). Introduction to Robotics: Mechanics and Control. Addison-Wesley.
[2] Makino, H., & Furuya, N. (1980). SCARA robot and its family. Proceedings of the 3rd International Conference on Assembly Automation, pp. 433–444.
[3] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. CVPR 2016, pp. 779–788.
[4] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. NIPS 2015.
[5] Kumar, A., Sharma, R., & Gupta, P. (2020). Vision-Guided SCARA Robot for PCB Component Assembly. IEEE Transactions on Industrial Electronics, 67(8), 6921–6930.
[6] Wang, H., Li, Z., & Chen, W. (2021). Deep Reinforcement Learning for Adaptive Grasping in SCARA Systems. Robotics and Autonomous Systems, 136, 103714.
[7] Hassan, M., Ali, S., & Khan, F. (2022). RGB-D Based Pose Estimation for Pick-and-Place in Cluttered Environments. IEEE Access, 10, 45221–45234.
[8] Jocher, G., et al. (2021). YOLOv5 by Ultralytics. GitHub. https://github.com/ultralytics/yolov5
[9] Raibert, M.H., & Craig, J.J. (1981). Hybrid Position/Force Control of Manipulators. Journal of Dynamic Systems, Measurement, and Control, 103(2), 126–133.
[10] Spong, M.W., Hutchinson, S., & Vidyasagar, M. (2006). Robot Modeling and Control. Wiley.