ECE445

SENIOR DESIGN LABORATORY

DESIGN DOCUMENT

Machine Vision-Based Intelligent Fruit and Vegetable Picking & Sorting Robotic Arm

Team #17

Simeng Yan [simengy2]

Wenye Zhang [wenyez2]

Fengyi Jin [fengyij2]

Shengyu Xu [sx27]

Advisor: Timothy Lee

April 1, 2026

Contents

1. Introduction

    1.1 Problem statement

    1.2 Solution Overview & Visual Aid

    1.3 High-Level Requirements List

2. System Architecture

    2.1 Perception (Vision)

    2.2 Actuation: STS3215 Serial Bus Servos Joint

    2.3 Control: ESP32-S3 DeviKitC and URT-1 Serial Bus Servo Controller

    2.4 Power: 12V 10A Switching Power Supply and 12V-to-5V Converter

3. Design Decisions & Trade-offs

    3.1 Motor Selection

    3.2 Controller Choice

    3.3 Communication Protocol

    3.4 Vision Strategy

4. Mechanical & Software Logic

    4.1 Joint Layout

    4.2 Control Logic

    4.3 Vision-to-Motion Pipeline

5. Safety & Power Analysis

6. Bill of Materials

1. Project. Overview

1.1 Problem Statement

In the contemporary agricultural value chain, the post-harvest sorting of produce—such as fruits and vegetables—remains a critical bottleneck that dictates marketability and shelf-life. Currently, small-to-medium scale facilities rely heavily on manual labor to categorize products based on size, ripeness, and quality. However, this human-centric approach suffers from three fundamental flaws: high operational overhead, subjective inconsistency, and physical fatigue, leading to an estimated 10-15% error rate in sorting during peak seasons.

Furthermore, traditional fixed-automation systems are often too rigid to handle the inherent geometric variability of agricultural products. There is a technological "missing link" for a cost-effective, versatile robotic platform that can adapt to unstructured environments. To bridge this gap, a system must be developed that replaces the fallible human eye with high-speed machine vision and the manual hand with a precise, servo-actuated manipulator, ensuring 24/7 operational stability and data-driven quality control.

1.2 Solution Overview & Visual Aid

The proposed solution is an Intelligent Vision-Guided Sorting Robot, a cyber-physical system designed to transform visual entropy into structured logistics. The architecture follows a Sense-Think-Act paradigm:

Sensing: An overhead vision module captures high-resolution frames. Using a deep-learning-based perception pipeline, the system extracts the centroid coordinates, bounding box dimensions, and HSV/RGB chromatic histograms of each item in real-time.

Thinking: The central control unit performs Hand-Eye Calibration to map 2D image-space coordinates into the 3D physical workspace. It then calculates the optimal joint angles via Inverse Kinematics to position the end-effector.

Acting: A 6 DOF robotic arm, optimized for reach and payload, executes the trajectory. The servo-driven gripper receives serial bus protocol commands to apply a calibrated gripping force, securing the object and depositing it into categorized bins based on the decision-tree logic.

1.3 High-Level Requirements List

To ensure the project transitions from a laboratory prototype to a functional engineering solution, the system must adhere to the following rigorous requirements:

1 Perception Accuracy

Using the YOLO framework at 640×480 resolution, the Mean Average Precision (mAP@0.5) must be $\geq 92\%$.

Object sizing error must be ≤ ±3 mm.

Ensures the system accurately distinguishes produce grades and provides reliable coordinates for grasping.

2 Kinematic Repeatability

The robotic arm's Tool Center Point (TCP) must maintain a repeatability error of ≤ ±2mm within a workspace radius of R ≥400mm. Guarantees that the gripper aligns with the target center consistently over multiple cycles to prevent grasp failure.

3 Actuation control

The servo controllers must support a 50 Hz PWM frequency with a pulse-width resolution of 1 μs (approx. 0.09° angular precision).

High-resolution control allows for granular adjustment of the gripper's opening angle to prevent bruising the produce skin.

4 System Latency

The end-to-end "Vision-to-Motion" latency must be ≤250 ms, with the deep learning inference time specifically ≤80ms.

Low latency is critical for real-time sorting and handling targets that may be moving on a conveyor belt.

2. System Architecture (Subsystems)

2.1 Perception (Vision)

The Perception subsystem utilizes cost-effective optical hardware coupled with a lightweight deep-learning model to achieve industrial-grade sorting precision under budget constraints. Its primary objective is to overcome the sensor noise inherent in consumer-grade hardware to extract the spatial pose and attribute data of agricultural produce.

2.1.1 Hardware Selection and Environmental Compensation

The system employs the Logitech C270i as the primary visual sensor. While designed for consumer applications, it is optimized for this project through specific software and environmental configurations:

Technical Specifications: The camera supports a 720p (1280x720) resolution at 30fps with a diagonal Field of View (FoV) of 60°.

Mounting Strategy: An "Eye-to-Hand" configuration is adopted, with the camera fixed vertically approximately mm above the sorting plane. This height ensures the 60° FoV covers the entire workspace while maintaining sufficient pixel density to minimize localization errors.

Illumination Optimization: Due to the relatively small CMOS sensor in the C270i, significant image noise occurs in low-light conditions. To compensate, a 3000K-5000K adjustable LED ring light is integrated to maintain a constant luminance threshold, mitigating detail loss and stabilizing the camera's auto-exposure behavior.

1.1.2 Vision Processing Framework: Lightweight Perception via YOLOv8

The software architecture is centered around YOLOv8-nano (YOLOv8n). This model utilizes advanced pruning and quantization techniques to minimize computational overhead while maintaining high detection accuracy on embedded platforms.

Chromatic Robustness: Captured RGB streams from the C270i are first converted to the HSV color space. This approach effectively filters out chromatic deviations caused by the C270i's automatic white balance fluctuations, ensuring stable color recognition for items such as ripening fruits.

Object Detection and Multi-task Branching:

YOLOv8n Inference: The model outputs bounding boxes (x,y,w,h) and class probabilities. The 720p resolution provides sufficient detail for the model to accurately define the edges of the produce.

Dimension Estimation: Based on the pixel width of the bounding box, the system applies a calibrated linear mapping factor K to estimate the physical diameter: $D = K \cdot W_{pixel}$

Dynamic Filtering: To counter image jitter produced by the consumer-grade sensor, a Kalman Filter is implemented to smooth the target centroid coordinates, ensuring the stability of the grasping coordinates sent to the robotic arm.

1.1.3 Coordinate System Mapping and Hand-Eye Calibration

Since the Logitech C270i lens exhibits noticeable radial distortion, the system utilizes Zhang's Calibration Method to obtain the camera's intrinsic parameters and undistort the raw image.

The pixel coordinates (u,v) are mapped to the robotic arm's base coordinates ($X_{base,}Y_{base}$) using an affine transformation matrix M_{affine}, calculated via the least-squares method:

$$\begin{bmatrix} X_{base} \\ Y_{base} \\ 1 \end{bmatrix} = M_{affine} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}$$

By employing sub-pixel corner detection during the calibration phase on the 720p image, the system achieves a high degree of mapping accuracy, compensating for the hardware's resolution limitations.

2.2 Actuation: STS3215 Serial Bus Servos Joint

The actuation of the 6-DOF robotic arm is powered by six STS3215 Serial Bus Servos. This subsystem is responsible for translating logical control commands into precise physical motion.

2.2.1 STS3215 Servo Specifications

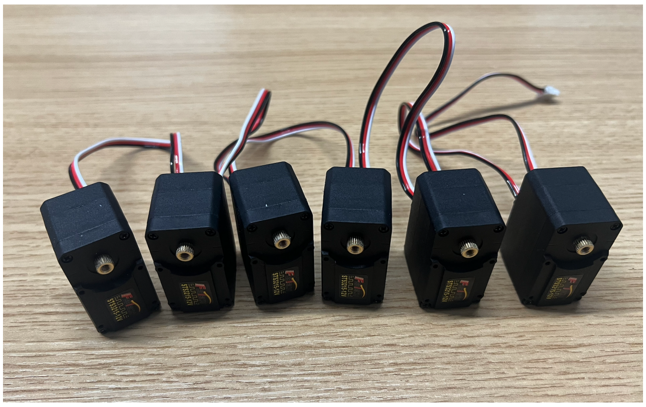
The STS3215 is a high-torque, industrial-grade smart servo designed specifically for multi-joint robotic applications. Unlike traditional PWM servos, it integrates several core components into a single compact housing.

High Torque and Reduction: It features a full-metal gearbox with a significant reduction ratio, delivering a peak torque of 19.3kg.cm (at 12V). This is sufficient to handle the structural weight of a mid-sized arm and the payload of various fruits.

Integrated Sensing: Each unit contains a 12-bit high-precision magnetic encoder, allowing for 360-degree absolute position feedback and continuous rotation.

Feedback Telemetry: The servo can communicate real-time data back to the controller, including position, speed, load, voltage, and internal temperature, enabling closed-loop control and safety monitoring.

2.2.2 Serial Bus Topology



The system utilizes a Serial Bus Topology to connect all six actuators, offering a significant advantage over parallel wiring.

Daisy Chain Connection: All six servos are connected in series using a 3-wire cable (Signal, VCC, GND). The output of the first servo connects to the second input, and so on. This "daisy chain" minimizes wire clutter within the 3D-printed arm structure.

Unique ID Addressing: Each actuator is assigned a unique digital ID (1 through 6). The ESP32 controller broadcasts data frames onto the bus, and only the servo with the matching ID executes the command.

2.2.3 Operational Logic

The actuators operate on a 12V DC power supply distributed via the URT-1 Relay Board. The URT-1 acts as the hardware bridge, converting the ESP32’s full-duplex UART signals into the half-duplex signals required by the STS3215 bus. This setup ensures that high-current power for the motors is isolated from the sensitive logic circuitry of the microcontroller, preventing electrical interference during high-load sorting tasks.

2.3 Control: ESP32-S3 DeviKitC and URT-1 Serial Bus Servo Controller

The control subsystem serves as the "nervous system" of the robotic arm, bridging the gap between high-level vision processing and low-level mechanical execution. It consists of two primary components: the ESP32-S3 DevKitC microcontroller and the URT-1 Serial Bus Servo Controller.

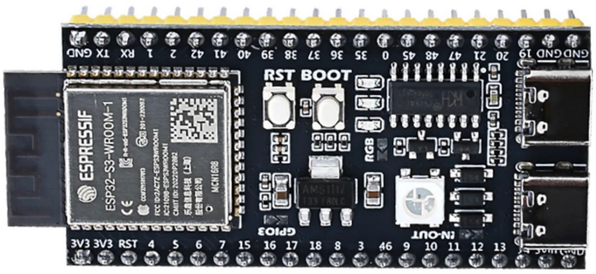
2.3.1 ESP32-S3 DevkitC:The Core Processor

The ESP32-S3 is a powerful, dual-core 32-bit LX7 microcontroller that manages the logic and synchronization of the 6-DOF system.

High-Speed Computation: With a clock speed of up to 240MHz and integrated hardware acceleration for AI instructions, the ESP32-S3 efficiently processes the Inverse Kinematics algorithms required to convert target coordinates into 6-joint angles in real-time.

Dual-Core Task Management: The dual-core architecture allows for a "Parallel Processing" strategy: one core handles high-speed UART communication with the servos, while the other core manages Wi-Fi/Serial communication with the external vision-processing computer.

Peripheral Connectivity: It provides multiple hardware UART ports, which are essential for stable, high-speed (1Mbps) data exchange with the servo bus via the URT-1 board.



2.3.2 URT-1 Serial Bus Servo Controller: The Signal Bridge

The URT-1 acts as a specialized hardware interface between the standard microcontroller and the serial bus servos. In this project, the URT-1 performs following functions:

Asynchronous Half-Duplex Conversion: The ESP32-S3 uses standard full-duplex UART (TX/RX). The URT-1 integrates high-speed switching circuitry to convert these signals into the single-wire half-duplex protocol required by the STS3215 servos, ensuring error-free data broadcasting and feedback retrieval.

Power Distribution Hub: The 12V 10A power from the main supply is injected directly into the URT-1. It features reinforced copper traces capable of handling the high-current demands of six simultaneous actuators, acting as a safe distribution point that protects the ESP32 from electrical noise and voltage spikes.

2.4 Power: 12V 10A Switching Power Supply and 12V-to-5V Converter

The power subsystem is designed to provide a stable and high-capacity energy supply to both the high-torque actuators and the sensitive control electronics. Given the 6-DOF configuration, the system must handle significant transient currents during rapid acceleration and heavy-load sorting tasks.

2.4.1 Dual-Voltage Rail Architecture

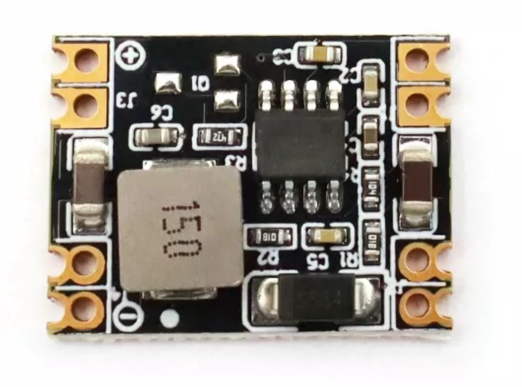
To ensure operational stability and prevent logic interference, the system implements a split-rail architecture. The high power 12V rail is dedicated to the 6 * STS3215 Serial Bus Servos. It draws directly from the main switching power supply to provide the high voltage and current required for peak torque (up to 19.3kg.cm per joint). A secondary, regulated rail provides clean power to the ESP32-S3.

2.4.2 Power Components & Specifications

产品机型	S-120-12	输出功率	120W
输出电压	12V	保护功能	过载/过流/短路
输入电压	200-240VAC	产品尺寸	199*98*38mm

12V 10A Switching Power Supply:

This AC-to-DC converter serves as the main energy source. A 120W (12V 10A) rating was selected to provide a 30% safety margin over the estimated maximum simultaneous draw (approx. 7A-8A) of the 6-joint assembly during high-speed fruit sorting.

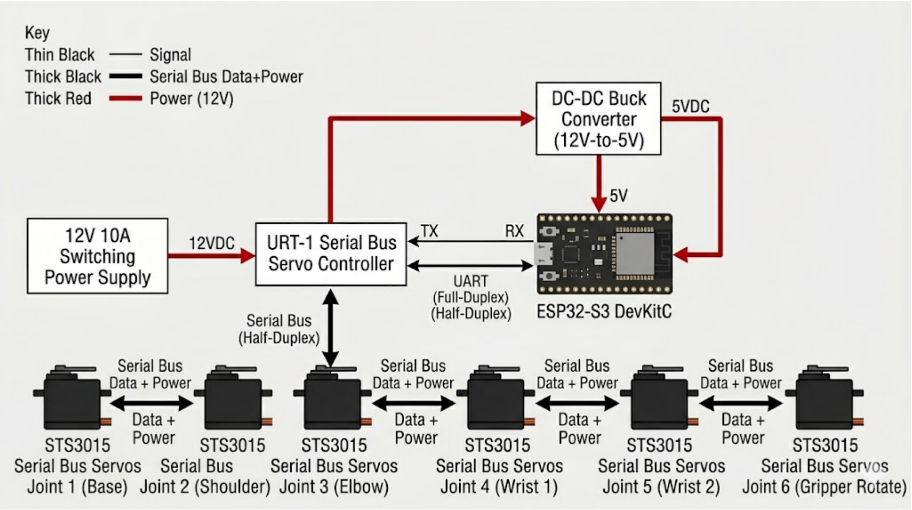


DC-DC Buck Converter:

A high-efficiency buck module scales the 12V input down to a steady 5V (min. 2A). This regulated output powers the ESP32-S3 DevKitC and ensures consistent performance of the peripheral sensors.

URT-1 Power Distribution:

The URT-1 board acts as the physical junction for the 12V rail. It features heavy-duty PCB traces and electrolytic capacitors that act as local reservoirs, smoothing out current ripples generated by the servos' PWM drivers.



3. Design Decisions & Trade-offs

3.1 Motor Selection

The selection of actuators is a critical decision that balances mechanical performance, budget constraints, and development efficiency. During the initial design phase, the DMJ4310 brushless modular motor was considered due to its superior torque density and high-speed capabilities. However, the DMJ4310 was ultimately bypassed in favor of the STS3215 Serial Bus Servo for several strategic reasons.

Firstly, the DMJ4310 utilizes the CAN protocol, which significantly increases the complexity of the control hardware and firmware, requiring specialized transceivers and a more sophisticated interrupt-driven software architecture. In contrast, the STS3215 employs a streamlined asynchronous serial bus protocol that is natively supported by the ESP32-S3's UART peripherals, dramatically reducing the time required for protocol implementation.

Secondly, from a cost-benefit perspective, the STS3215 combines the motor, reduction gearbox, 12-bit magnetic encoder, and control circuitry into a single affordable housing, while the DMJ4310 nearly triples the per-joint cost. Considering the dimension and application of our robotic arm, we don’t have to use such professional motors and their high load capability and operation accuracy.

Therefore, the STS3215 was selected as the optimal choice for this 6-DOF fruit sorting arm, providing a highly cost-effective and "easy-to-implement" platform without sacrificing the precision and feedback telemetry necessary for delicate sorting tasks.

3.2 Controller Choice

The selection of the ESP32-S3 DevKitC as the primary microcontroller was a strategic decision driven by the computational demands of a 6-DOF vision-guided system.

Unlike traditional 8-bit or simple 32-bit microcontrollers, the ESP32-S3 features a Dual-Core Xtensa® LX7 processor running at up to 240MHz. This dual-core architecture allows for a dedicated task-partitioning strategy: one core is assigned to high-frequency Inverse Kinematics calculations and constant telemetry monitoring from the 6 servos, while the second core handles the high-speed communication with the external system. This prevents "task-blocking," ensuring that the mechanical arm maintains smoothly.

Furthermore, the ESP32-S3 offers superior Serial Communication performance, featuring multiple hardware UARTs with large FIFO buffers that comfortably sustain the 1Mbps baud rate required by the STS3215 bus. In contrast, standard microcontrollers often struggle with such high-speed serial data, leading to buffer overflows or timing jitters that could cause unpredictable motor behavior. Additionally, the integrated 2.4GHz Wi-Fi and Bluetooth capabilities provide a future-proof interface for wireless debugging or remote data logging.

Consequently, the ESP32-S3 provides the necessary computational "overhead" and peripheral flexibility to manage a sophisticated fruit-sorting pipeline within a single, cost-effective package.

3.3 Communication Protocol

The selection of the Single-Wire Asynchronous Serial Bus Protocol for the 6-DOF robotic arm was primarily driven by the need for mechanical simplicity and system scalability. Traditional PWM-controlled servos require a dedicated signal wire for every single motor, which would result in a bulky and inflexible "wiring loom" of at least 18 individual cables (Power, Ground, and Signal for each of the 6 joints) passing through the arm's rotating pivots.

By adopting the serial bus protocol, the system architecture is collapsed into a Daisy Chain configuration, where a single 3-wire harness carries both 12V power and bi-directional data to all six actuators in series.

This significantly reduces mechanical friction at the joints and minimizes the risk of wire fatigue or snagging during high-speed sorting maneuvers. Furthermore, unlike the simple one-way pulses of PWM, this asynchronous protocol allows for Half-Duplex bi-directional telemetry. This means the ESP32-S3 can not only send positioning commands at a high-speed 1Mbps baud rate but also query each STS3215 servo for real-time status updates, such as internal temperature and torque load.

This feedback loop is essential for the "Soft-Touch" requirement in fruit sorting, as it allows the controller to detect if a fruit is being squeezed too hard or if a joint has reached a physical limit, a level of sophistication that traditional parallel wiring schemes cannot achieve without a massive increase in cable complexity.

3.4 Vision Strategy

A critical architectural decision in robotic integration is the spatial relationship between the imaging sensor and the manipulator. For this project, an Eye-to-Hand (Fixed-Camera) configuration was selected over the Eye-in-Hand (On-Arm) approach. This choice was dictated by the operational requirements of agricultural sorting and the physical constraints of the Logitech C270i hardware.

3.4.1Global Perspective and Planning Efficiency

The Eye-to-Hand setup provides a comprehensive, static Field of View (FoV) of the entire workspace. In agricultural applications, produce often arrives in randomized, unstructured clusters. A fixed overhead camera allows the YOLOv8n model to perform "one-shot" detection across the entire sorting area. This enables the system to calculate an optimized sorting sequence for multiple items simultaneously, whereas an Eye-in-Hand system would require the arm to perform time-consuming scanning motions to "search" for targets, significantly reducing the overall throughput.

3.4.2 Mechanical Dynamics and Payload Optimization

By decoupling the camera from the manipulator, we preserve the mechanical agility of the robotic arm. Mounting the Logitech C270i and its associated cabling at the end-effector (Eye-in-Hand) would increase the rotational inertia and the cantilever load on the servos. This additional mass necessitates higher torque for acceleration and deceleration, leading to increased mechanical wear and potential oscillations. Maintaining the camera on an external rigid frame ensures that the arm's full payload capacity and dynamic response are dedicated solely to the high-speed manipulation of the produce.

3.4.3 Calibration Stability and Computational Economy

From a computational standpoint, the Eye-to-Hand configuration offers superior stability. The transformation matrix (T_{cam}^{base}) remains constant once the initial checkerboard calibration is performed. This simplifies the spatial mapping to a static linear operation, reducing the real-time processing overhead on the controller. In contrast, an Eye-in-Hand setup requires the coordinate transformation to be dynamically recomputed at every time-step using the arm's forward kinematics. Such a process is highly sensitive to mechanical backlash and joint encoder errors, which could compromise the sub-pixel localization accuracy required for precise grasping.

3.4.4 Mitigation of Visual Occlusion

The inherent trade-off of a fixed camera is the risk of the robotic arm obstructing the camera's line-of-sight during operation. To mitigate this, the system incorporates a "clearance-position" logic in its control cycle. After depositing an item into the categorized zone, the arm returns to a predefined neutral home position that lies outside the camera's primary sensing cone. This ensures that the perception layer always has an unobstructed view for the subsequent detection cycle, maintaining a high-reliability feedback loop.

4. Mechanical & Software Logic

4.1 Joint Layout

This manipulator adopts a 6-axis serial articulated joint layout with the following spatial arrangement logic:

J1 (waist): Vertical axis – provides waist rotation.

J2 (shoulder) & J3 (elbow): Horizontal axes, parallel to each other – enable pitching motions of the upper arm and forearm.

J4 (wrist 1): Axis along the forearm – provides forearm roll.

J5 (wrist 2): Axis [perpendicular to / parallel to] J4 – provides secondary roll (or lateral tilt, depending on implementation).

J6 (wrist 3 / tool): Axis oriented [please specify] – provides final fine adjustment of the end-effector orientation.

This configuration modifies the conventional 6-axis industrial robot layout (J4 roll, J5 pitch, J6 roll) by replacing the pitch joint (J5) with an additional roll joint, which can be beneficial for applications requiring continuous tool rotation or wrist singularity avoidance, such as agricultural product sorting.

4.2 Control Logic

The software control logic is responsible for translating high-level task coordinates into synchronized physical execution while ensuring the structural safety of the arm.

4.2.1 Inverse Kinematics and Synchronized Motion

To achieve precise fruit positioning, the system implements an Inverse Kinematics solver on the ESP32-S3. When the vision system provides the target Cartesian coordinates (x, y, z) and orientation (α, β, γ), the IK algorithm calculates the corresponding angles for all six STS3215 joints. To ensure fluid and natural movement, we utilize Synchronized Write instructions via the serial bus protocol. This allows the controller to package the target positions and velocities for all six servos into a single data frame. By doing so, every joint initiates and concludes its motion simultaneously, minimizing trajectory deviation and mechanical vibrations during high-speed sorting.

4.2.2 Current-Feedback Based Overload Protection

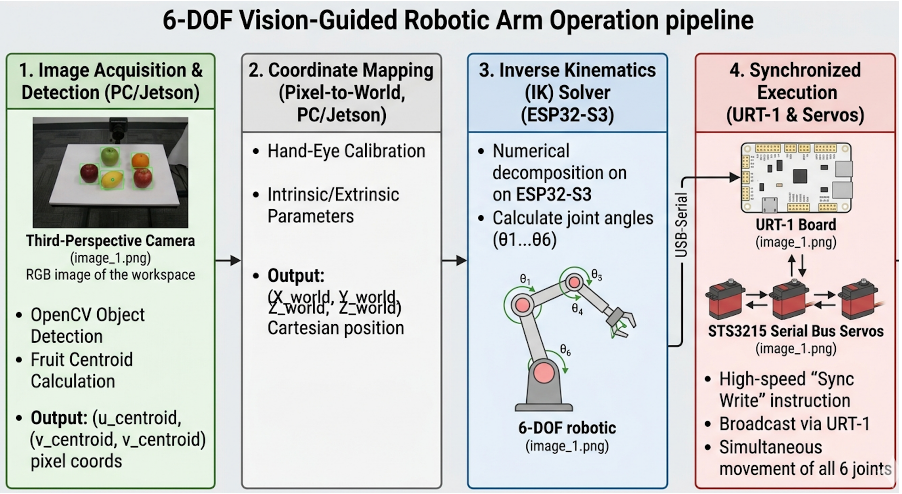
Safety and delicacy are prioritized through a closed-loop feedback system utilizing the real-time telemetry capabilities of the serial bus servos. The control firmware constantly monitors the current consumption of each actuator at a 50Hz sampling rate.

During the "Close Gripper" phase, the system detects the spike in current that occurs when the fingers contact an object. Once the current hits a predefined threshold tailored to the fruit firmware profile, the motor maintains its position without further increasing torque, achieving a "soft-touch" grip.

If a joint encounters an unexpected obstacle, the sudden rise in current triggers an Emergency Torque-Off routine. This immediate response prevents gear stripping in the STS3215 gearbox and mitigates potential damage to the 3D-printed structural components, ensuring the long-term reliability of the 6-DOF assembly.

4.3 Vision-to-Motion Pipeline

The integration of visual perception and mechanical execution is managed through a structured, multi-stage pipeline. This pipeline ensures that fruit identified in the camera's digital image is accurately translated into coordinated physical movement by the 6-DOF arm.



5. Safety & Power Analysis

5.1 Power Budget

The power budgets calculated based on the maximum potential draw of the six STS3215 serial bus servos operating at a nominal voltage of 12V. Each STS3215 unit has a rated stall current of approximately 1.5A to 1.7A when exerting peak torque.

In a worst-case scenario where all six joints are simultaneously initiated or encountering high resistance during a heavy-lift sorting task, the theoretical instantaneous current demand could reach upwards of 9A to 10.2A.

However, under typical fruit-sorting trajectories, only two or three primary joints—specifically the base, shoulder, and elbow—operate at high load, while the wrist and gripper servos consume significantly less power. By averaging these operational states, we estimate a continuous working current of approximately 4A to 6A, which translates to a power consumption range of 48W to 72W during active movement.

To ensure system reliability and longevity, we have selected a 12V 10A (120W) industrial switching power supply, providing a significant current redundancy of approximately 30% to 40% over the expected peak operational load. This overhead is critical for maintaining voltage stability.

Furthermore, the URT-1 board and the thick-gauge 12V bus wiring are designed to handle this 10A capacity without significant resistive heating. This conservative power overhead ensures that the robotic arm can perform high-frequency sorting cycles over extended periods without thermal throttling or electrical failure, providing the necessary robust foundation for the vision-to-motion pipeline.

5.2 Risk Mitigation

The safety of the 6-DOF robotic arm is ensured through a multi-layered strategy across three primary domains:

Mechanical & Spatial Safety: To prevent structural failure and self-collision, the firmware implements software-defined joint limits. Additionally, the third-perspective camera provides "Zone Monitoring," enabling the vision system to trigger an immediate interruption and halt motion if a human or foreign object enters the workspace.

Electrical & Thermal Protection: The system leverages real-time telemetry from the STS3215 servos to monitor current draw and internal temperatures. If a stall or thermal threshold (e.g., 70°C) is detected, an automatic torque-off command is executed. The 12V power supply further includes integrated short-circuit and overload protection to safeguard the logic circuitry.

Ethical & Operational Safeguards: A physical Emergency Stop (E-Stop) button is integrated to provide a hardware-level power cut independent of software. To ensure safe human-robot interaction, the arm features rounded-edge 3D designs and compliant gripper materials, justifying the project's adherence to standard robotic safety codes and

"Soft-Touch" operational ethics.

6. Bill of Materials

Components	Type	Amount	Cost (RMB)
Serial Bus Servo	Feetech STS3215	7 (one for spare)	110*7=770
Microcontroller	ESP32-S3 DevKitC	1	28.31
Servo Controller	URT-1 Serial Bus Relay Board	1	50
Vision Sensor	USB Camera (720P)	1	100
Power Supply	AC-DC 12V 10A	1	37.68
Buck Converter	DC-DC (12V to 5V 3A)	1	4.59