Design of Ingredient Recognition and Automatic Feeding System for Frying Robot Based on Machine Vision

Zhanqiang Liu[1], Liang Dong[2], Jianling Zhao[1,*], Jianguo Wang[3]
[1]*Shandong Hehe Information Technology Co., Ltd., Jinan, China*
[2]*Jinan BOSAI Network Technology Co., Ltd., Jinan, China*
[3]*Shandong Tongyun Network Technology Co., Ltd., Jinan, China*
**Corresponding Author*

Abstract: In order to improve the accuracy of food ingredient recognition and the synergy of feeding control in stir-frying automation equipment, a machine vision-based intelligent stir-frying system is designed to construct a hierarchical architecture consisting of visual recognition, control decision-making, and multi-axis actuator units. A high-resolution industrial camera combined with MobileNetV2 network is used for image feature extraction and classification, and control labels and scale parameters are output for the execution system to schedule the feeding path. The path planning and closed-loop control is realized by integrating the STM32F407 master chip and PD control strategy. Experiments show that the system has a recognition accuracy of 93.6% under the conditions of multiple types of ingredients, an average feeding path error of ±1.8 mm, and the response delay is controlled within 842 ms, which verifies the validity and practicability of the synergistic mechanism of visual recognition and automatic control.

Keywords: Machine Vision; Intelligent Stir-Fry Machine; Ingredient Recognition; Automatic Feeding

1. Introduction

With the advancement of smart manufacturing and kitchen automation, intelligent cooking systems are emerging as a key trend in modern food processing equipment. Among them, accurate ingredient recognition and responsive feeding mechanisms are critical to achieving stable, repeatable dish quality. Traditional cooking robots often suffer from poor adaptability in complex kitchen environments, where diverse ingredient shapes and lighting variations present significant challenges to real-time visual analysis. Recent studies have highlighted the effectiveness of integrating machine vision with embedded control systems to enhance perception and actuation synergy. For example, Bao and Luo [1] designed a fully automatic cooking robot incorporating multi-sensor feedback and execution coordination. Zhang et al. [2] applied STM32F407 chips in a cooking system to ensure low-latency control, while Cui et al. [3] proposed a modular design for frying automation that combines visual classification and feeding path optimization. These works lay the groundwork for further improving automation intelligence in cooking scenarios. Building upon these foundations, this paper presents a machine vision-based stir-frying robot that achieves high-accuracy ingredient classification and efficient automatic feeding under multi-task coordination.

2. Intelligent Frying Machine System Design Scheme

2.1 System Overall Architecture Design

To ensure accurate ingredient recognition and responsive feeding control, the system adopts a layered, modular architecture integrating machine vision with automation, as shown in Figure 1. It consists of a perception layer , a decision-making laye, and an execution layer. The perception layer integrates CMOS industrial cameras and load cells for multi-dimensional data collection, connected to the edge computing unit via USB3.0. The decision layer uses an embedded AI module (e.g., NVIDIA Jetson Xavier NX) running a lightweight CNN to classify images and send control commands to the STM32F407 chip [1]. The execution layer features a three-axis stepper drive and a multi-bin sorting mechanism to perform quantitative feeding based on path planning. A hybrid UART + I²C communication protocol ensures low-latency coordination across modules. All components use a unified data interface, minimizing signal bottlenecks and ensuring smooth interaction between recognition and control modules.

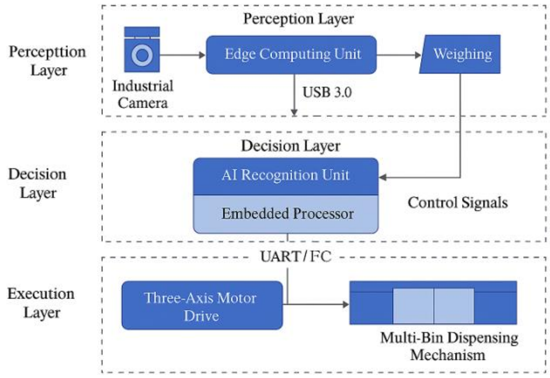

Figure 1. Overall Architecture of Intelligent Stir-Fry Machine System

2.2 Design of Ingredient Visual Recognition Module

Positioned between the perception and decision-making layers, the visual recognition module identifies multi-category ingredients and provides classification labels and feature parameters for control. To address challenges such as complex lighting, occlusion, and ingredient diversity, the module follows a three-stage design: multi-channel perception, lightweight feature extraction, and integrated classification. A 12MP CMOS industrial camera is paired with a high-CRI LED ring light to maintain gray balance and edge clarity [2]. During preprocessing, histogram equalization and YUV color space conversion reduce background complexity. Features are then extracted using an optimized MobileNetV2 CNN, which incorporates a linear bottleneck and depthwise separable convolutions to lower computational cost.The model output layer accesses a softmax classifier whose multiclass output Pi represents the predicted probability of the ith class of ingredients, satisfying:

$$P_i = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}}, \quad i = 1,2,\ldots,n \tag{1}$$

Where zi is the neuron output corresponding to the ith class, n is the total number of ingredient classes, and Pi is used for the design of the feeding mapping logic in the decision layer. Key parameters, such as lens focal length, sensor sensitivity, and processing speed, are detailed in Table 1. The module outputs two data channels: one for ingredient type labels (e.g., "Ginger", "Garlic") and another for size estimation, enabling precise control of multi-bin dispensing units.

Table 1. Parameters of Key Components of the Ingredient Recognition Module

Module Components	Parameter item	Value/ Specification
Industrial Camera	Resolution	12MP (4032 × 3024)
Lens	Focal length	8mm fixed focal length
Fill light system	Color temperature	5500K
AI Processing Module	Algorithmic Platform	Jetson Xavier NX
Recognition Models	Network Architecture	MobileNetV2
Image frame processing rate	Real-time processing capability	⩾30FPS

2.3 Automatic Feeding Execution Module Design

Using the label and size data from the recognition module, the feeding execution module performs precise, position-controlled, and quantitative feeding. It adopts an integrated X-Y-Z three-axis planar drive system coupled with a multi-compartment feeding unit, forming a "three-dimensional positioning + multi-channel feeding" execution architecture [3]. As shown in Figure 2, the actuator platform is supported by a two-stage synchronous belt drive and linear guide rails to realize X-Y axis planar positioning, while the Z-axis adopts an electric screw lifting mechanism to control the feeding height, taking into account the speed and load stability. The end of the actuator is installed with multi-station suction cups and flip feeding structure, which can realize fast switching and stable release between multiple ingredient bins. Corresponding to the ingredient category label c ∈ C output from the recognition module and the hopper center position (xd,yd), a fifth degree polynomial trajectory interpolation function is used to generate a continuous smooth path:

$$x(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3 + a_4 t^4 + a_5 t^5 \tag{2}$$

$$y(t) = b_0 + b_1 t + b_2 t^2 + b_3 t^3 + b_4 t^4 + b_5 t^5 \tag{3}$$

where t ∈ [0,T] is the planning time period and the coefficients ai,bi are determined by the boundary conditions (position, velocity, acceleration) to ensure the balance between the end trajectory continuity and the mechanical stiffness response [4]. In order to support the instantaneous feeding control under the high-frequency recognition response, the system adopts a closed-loop stepping motor with a high-resolution photoelectric encoder to realize real-time position feedback, and the relevant parameter configurations are shown in Table 2.The actuator module also reserves a CAN bus communication interface to ensure a low-latency command transmission between it and the main control chip and the decision-making system, and real-time transmission of the current position and feeding status back to the decision-making layer, which provides a feedback interface to the subsequent control The integrated system provides a feedback interface.

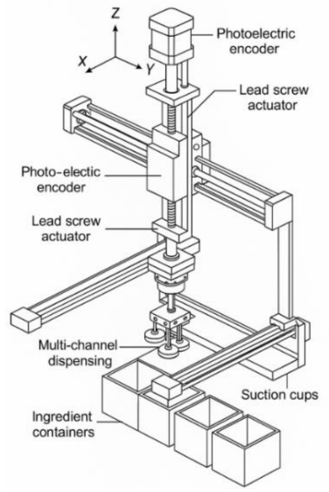

Figure 2. Structure of the Automatic Feeding Actuator Module

Table 2. Key Parameters of the Automatic Feeding Module Actuator.

Module Component	Parameter	Numerical value/ specification
Stepping motor	Step angle	1.8° /step
Optical encoder	Resolution	1000 PPR
Screw lifting mechanism	Maximum stroke	150 mm
Suction cup end structure	Number of feed bin connections	4 pcs
Control cycle	Minimum command cycle	5 ms
Communication protocols	Control and feedback interface	CAN 2.0B

2.4 Decision-Making and Control Integration System

In order to realize efficient closed-loop control from image recognition to action execution, this system designs an integrated control system based on state management logic and instruction mapping rules, and builds a "recognition-decision-execution" full-process scheduling architecture. The control core is based on STM32F407 main control chip, adopts embedded real-time operating system (RTOS) architecture, and integrates task priority scheduling mechanism and state transfer control model [5]. The control process is divided into six states in the form of a state machine: standby state, recognition startup, execution preparation, path computing, action control and reset reset, and the control logic is driven by the output of the visual recognition module of the ingredient labels c, the size valuation m and the image confidence Pc to transfer the state. The path calculation module executes the trajectory function by scheduling thread calls, combining the end feedback position (xf,yf,zf) with the desired coordinates (xt,yt,zt) to perform PD control, and the error controller is calculated as follows:

$$u(t) = K_p \cdot e(t) + K_d \cdot \frac{de(t)}{dt}$$

Where e(t)=xt-xf is the instantaneous position error, Kp and Kd are the proportional and differential coefficients, respectively. The control commands are distributed to the multi-axis motors and the multi-bin flip module through the CAN 2.0B bus to ensure the synchronized response of each execution node. All modules in the system follow the unified communication protocol, and the main command field structure is listed in Table 3, including frame header identification, command category, data valid bits and CRC check fields, and supports the data link abnormal self-recovery mechanism. In order to be compatible with future module expansion, the system reserves the interrupt interface and I² C sub-node expansion channel, so that the new identification module or feeding mechanism can be embedded into the scheduling system without reconfiguring the main logic.

Table 3. Control Command Communication Protocol Structure

Field Name	Number of bytes	Function Description
Frame Header	1 Byte	Start flag (0xAA)
Command	1 Byte	Command Type (0x01

Type		Recognition Class, 0x02 Execution Class)
Payload Length	1 Byte	Effective data byte length
Data Payload	N Byte	Position command, feeding information

3.System Realization and Experimental Verification

3.1 Experimental Platform Construction and Experimental Design

Based on the hardware and software collaboration to build the experimental platform, the design includes task perception, instruction transmission, path execution and data feedback of the complete closed-loop verification system. The platform is built using Jetson Xavier NX as the edge recognition processing core, supporting GPU-accelerated reasoning, with a high-resolution industrial camera for image acquisition, and transmitted to the local model running environment via USB3.0 high-speed bus. The control execution part is based on STM32F407 main control chip, which completes the image label receiving and the action sending of the feeding device through UART and CAN bus respectively. The mechanical structure part selects three-axis linear module and stepping drive system. In order to match the system logic flow, the experimental design is divided into two parts: single-module function test and overall co-simulation, and the test content includes: verification of multi-class ingredient recognition accuracy, positional trajectory tracking error analysis, statistics of feeding action time delay and overall response closed-loop test. All experiments are based on the key performance parameters of the system as indicators, set standard input incentives and constraints boundaries to ensure that the results can be traced back and the process can be reproduced.

3.2 Analysis of Experimental Results

Based on the integrated experimental platform built, the visual recognition module, path execution system and feeding control logic are verified for functionality and performance. In the multi-category ingredient recognition test, the system realizes 6 categories of target classification under 30FPS image stream, with an average recognition accuracy of 93.6%, in which the recognition rate of solid categories (e.g., peanuts, garlic cloves) is generally higher than that of shredded structures. In the trajectory tracking experiment, by recording the deviation between the actual path of the end actuator and the planned trajectory, the average trajectory error is calculated to be ±1.8 mm, and the maximum deviation does not exceed 3.4 mm, which meets the demand for high-precision localization. In the control response analysis, the average response delay of the system from the recognition output to the completion of feeding is 842 ms, and the stability fluctuation is ±5.2% in 50 repeated tests. As shown in Fig. 3, the non-linear growth trend between path error and response time in multi-task concurrent scenarios suggests that task scheduling and system load management strategies need to be further optimized in the future. The whole system shows good controllability and repeatability in multi-task linkage, module coordination and data closed-loop capability, which verifies the effectiveness of the recognition control fusion mechanism.

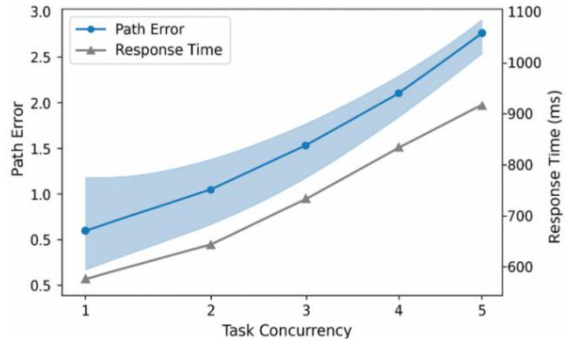

Figure 3. Distribution of Path Error and Response Delay under the Condition of Task Concurrency

4. Conclusion

The system realizes accurate classification and closed-loop automatic feeding control of multi-class ingredients based on visual recognition, and shows good stability and accuracy in multi-task response and path planning. Due to the image acquisition conditions and complex ingredient morphology, there is still room for improvement in the recognition accuracy between occlusion and similar classes. In the future, multimodal sensing and dynamic task scheduling mechanism can be introduced to further expand the versatility and intelligence level of the system.

References

[1] Sha, F., Meng, J., Zheng, X., & Jiang, Y. (2025). Sustainability Under Fire: How

China-US Tensions Impact Corporate ESG Performance?. Finance Research Letters, 107882.
[2] Hu, X. (2025). Learning to Animate: Few-Shot Neural Editors for 3D SMEs.
[3] Hu, X. (2025). Low-Cost 3D Authoring via Guided Diffusion in GUI-Driven Pipeline.