Song Hung Nguyen
Faculty of Mechanical Engineering
Ho Chi Minh City University of Technology (HCMUT), 268 Ly Thuong Kiet Street
District 10, HCMC, Vietnam
Vietnam National University Ho Chi Minh City (VNU-HCM), Linh Trung Ward, Thu Duc District, Ho Chi Minh City, Vietnam

Duc Minh Phan
Faculty of Mechanical Engineering
Ho Chi Minh City University of Technology (HCMUT), 268 Ly Thuong Kiet Street
District 10, HCMC, Vietnam
Vietnam National University Ho Chi Minh City (VNU-HCM), Linh Trung Ward, Thu Duc District, Ho Chi Minh City, Vietnam

Ha Quang Thinh Ngo
Faculty of Mechanical Engineering
Ho Chi Minh City University of Technology (HCMUT), 268 Ly Thuong Kiet Street
District 10, HCMC, Vietnam
Vietnam National University Ho Chi Minh City (VNU-HCM), Linh Trung Ward, Thu Duc District, Ho Chi Minh City, Vietnam

Deep Learning-based Approach for Gesture Recognition with Static Hand Representation

In the era of Artificial Intelligence (AI), our science and technology have reached to a lot of milestones, especially in the field of human-robot interaction (HRI). By fusing with the image processing techniques, AI-based strategy to enhance mutual recognition of HRI via hand signs is proposed in this investigation. Primarily, robotic hardware, theoretical computation of gripper design and vision-based techniques are introduced to establish the working environment. Then, the proposed framework including controller design and interactive platform is demonstrated. Several hand signs from human operator are collected and trained. Our approach is experimented in two cases for validating the effectiveness and properness of the proposed method with varying light condition. From these results, it can be seen obviously that this scheme is applicable in different fields such as human-aware collaboration, cognitive robot or sign language translation system.

Keywords: *Deep learning, Hand gesture recognition, Human-robot interaction, Computer vision, Motion control.*

1. INTRODUCTION

Presently, robots appear everywhere in our lives, i.e. mass production [1], assembly [2], autonomous tasks [3, 4], welding [5], re-configurable body [6], bartender [7] or musical player [8]. Instead of working alone in the past time, there is a need to launch reciprocated communication between human and robot. To proceed this idea, robot must capture any motion of an operator and analyse them. With the best knowledge of robot, it can deliver more proper actions and adaptive responses due to many necessities of human.

In general, human contact is performed via the communicative information in different manners such as voice [9], image [10] or gesture [11]. Gestures are an inevitable action of daily human life. Gesture recognition via image processing is a method that integrates the complicated perception with computer vision. It is used in various fields, including engineering and research, and is indispensable for improving human– machine interaction. Owing to the continuous changes of natural gestures, the existing gesture detection technologies are unable to completely attain socially human–machine communication.

There are two types of gesture recognitions, for exa–mple static and dynamic gesture recognitions [12]. Basically, human hand is fixed for recognition and that aspects such as hand sign, shape, and location do not alter. Dynamic gestures encompass sequential frames of static gestures, inferring that it could be stored in video format or movie. However, techniques which are dep–loyed in static recognition successfully, are also utilized in dynamic one. Hence, most of researchers focus on investigating those developments for static gesture recognition.

The organization of this research is constructed as following. Section 2 summarizes the cutting-edge tech–nologies in such domains. They are investigated and categorized in sub-items, for instance challenges, the proposed solutions, hardware requirements and limita–tions. Later, in section 3, the structure of robot, mecha–nical computation of gripper as well as electrical components. Section 4 demonstrates our approach com–prising the framework of controller design, interactive platform, and training procedure. In section 5, the pro–posed method is verified in several experiments. Also, related discussions from these results of validations are noted. The content of section 6 consists of some conc–lusions and potential developments for practitioners.

2. PREVIOUS WORKS

Hand gesture recognition is designated to enable the information exchange between human and machine or robot. To capture the communicative information, sen–sing measurement [13] or visual data via digital camera [14] could be observed. In fact, the formers used many sensing devices in one robotic system. It causes some troubles because of the complex fusion techniques, difficulties in sensor management as well as high main–tenance cost. Recently, most of investigators concern on computer vision and several image processing methods. It is reasonable for them to fuse the advanced algorithms such as filtering process, machine learning techniques and adaptive control strategies.

In Table 1, summary of recent researches in these domains are listed. They are classified into two strategies such that sensor-aware method uses only non-visual data to record the working conditions and orders from human while vision-based approach is examined

Received: May 2025, Accepted: July 2025
Correspondence to: Ha Quang Thinh Ngo
Faculty of Mechanical Engineering, Ho Chi Minh City University of Technology, VNU-HCM, Vietnam
E-mail: nhqthinh@hcmut.edu.vn
doi: 10.5937/fme2503458H

by image or video. Although there has been significant progress in hand gesture recognition and human-robot interaction, most existing solutions are limited to make real-world deployment feasible. Sensor-based appro–aches such as data gloves or sEMG systems obtain pre–cise movements at the expense of being cumbersome, needing calibration. They are also user-specific due to the variability of hand sizes, muscle fatigue, or sensor placement. Wearable systems in dynamic environments are less scalable and more difficult to generalize.

Table 1. List of the cutting-edge techniques in related fields.

Approach	Type of strategy	Author(s)	Key challenge	Proposed method	Hardware device	Restriction(s)
Sensor-based method	Data glove	Derek W. Orbaugh Antillon et al [15]	Difficulty in underwater hand gesture communication due to factors such as murky water, environmental obstacles, or loss of buddy attention	Researchers developed a smart dive glove to capture finger move–ments, and machine learning algorithms to classify 13 common diving hand gestures	Five waterproofed sensors, StretchSense 10-channel capacitance measurement board	Unintentional mo–vements, such as swimming stro–kes, could mista–kenly be recog–nized as inten–tional gestures
		Haiming H. et al [16]	Traditional drones are operated via remote controllers, resulting in weak human-machine interaction and limited operator feedback	They investigated a data glove to recognize hand gestures by detecting finger bending and palm flipping	10 MPU6050 IMU sensors, STM32F103 microcontroller	During dynamic gestures, noise like hand jitter introduces additional errors
	Electromyography (EMG)	Wu, Y.,et al [17]	Most existing issues focus on whole-hand motion recognition, while subtle finger movements and pinch forces were not studied	An sEMG-based pattern recognition system was invented to predict pinch force strength and classify finger movement types	3-channel sEMG from forearm muscles and pinch force using a force transducer	Performance may vary with muscle fatigue or sensor placement variation
		Cao, L., Zhang et al [18]	Gesture recognition from low sampling frequency sEMG signals faces accuracy loss compared to traditional high-frequency systems	A novel AMPSO-SVM algorithm was innovated to comprise feature extraction, feature selection, classifier optimization and adaptive mutation probability	8 channels of sEMG sensors placed around the forearm	Classification accuracy still depends on signal quality, electrode placement, and muscle fatigue
	Wifi and radar	Rahaman, H. et al [19]	The main technical challenge for RF-sensing is the short duration of human walking events, which limits the amount of collected signal data for analysis	A reverse RF-sensing system consists of single transmitter and multiple receivers which significantly increase RSS sample size and capture short-lived walking events by CNN	IEEE 802.15.4 CC2420 transceivers, 1 node acts as a transmitter and 5 nodes as receivers	Accuracy varies depending on the layout complexity (higher in simple environments, lower in complex corridors)
Vision-based method	Image grayscaling	Yakkati, R. R. et al [20]	Efficiently classifying hand gestures from thermal grayscale images while ensuring low inference time and small model size suitable for edge devices like Raspberry Pi is a central challenge	A custom deep (CNN) architecture was studied to optimize for grayscale thermal hand gesture images using the Adam optimizer for training and cross-entropy loss for classification	FLIR Lepton 3.5 thermal camera (160×120 resolution) with PureThermal 2 breakout board connected to Raspberry Pi 4	Thermal imaging may not capture fine hand details compared to RGB or depth sensors, potentially limiting scalability to more diverse gestures
	Image smoothing	Le, H. P. et al [21]	Real-time and precise 3D object detection in chang–ing industrial environ–ments and robot and ca–mera coordinate synch–ronization, especially under occlusions, varying lighting conditions, and clutter are two key challenges	A deep learning object detection model based on single-shot that was trained to specifically detect three classes (bottle, can, cup) in real time	Intel RealSense D435, 4-DOF Robotic Arm	Precision of depth estimation, especially along the Z-axis, is the primary constraint (average error is in the range of 4.5–5.3 mm)
	Edge detection	Saez, B., Mendez et al [22]	The main challenge addressed is how to design a gesture recognition system that is privacy-preserving, low-power, low-latency, and entirely self-contained on edge devices, using only ultrasound transceivers	They implemented a gesture recognition system using ultrasonic active sonar principles with two transceivers to detect 2D arm/hand gestures	Two ultrasound transceivers, two microcontroller modules (XMC4700), Bluetooth HC-05 module, neuroshield AI board (option)	Gestures must occur in a fixed 2D plane in front of the sensors due to transducer directionality—out-of-plane gestures are not detected

	Morphological image processing	Duan, K., & Zou, Z. [23]	Most construction robots are either pre-programmed or controlled with low DoF joysticks, which are insufficient for high DoF robotic systems	A four-module framework for VR-based gesture-controlled teleoperation was designated to operate independently of the camera coordinate system	Oculus Quest 2, UR3 Universal robotic arm, OAK-D-S2 RGB camera	Size mismatch and kinematic changes between human and robotic hands can still lead to minor pose inaccuracies
	Skin color segmentation	Espejel-Cabrera, J. et al [24]	Traditional image segmentation methods perform poorly in uncontrolled environments due to lighting sensitivity, low contrast, and visual noise	A 7-step process for gesture recognition from video, with key innovations was presented	Consumer-grade cameras (e.g., smartphone or standard webcams)	Variations in skin tones or occlusions may affect performance

Vision-based techniques, on the other hand, have been well-liked due to their non-obtrusiveness. They are, however, constrained in terms of illumination sen–sitivity, background noise, and occlusion. The majority of recent research is based on pre-trained models or con–trolled laboratory datasets that generalize poorly to real-world applications. Some work in video-based dynamic gestures, which are computationally expensive and have delay, unsuitable for real-time control applications.

Besides, most systems fail to include the gesture recognition system with closed-loop control and feed–back loops within robotic systems. The absence of real-time or bi-directional communication reduces the func–tionality of HRI in dynamic settings. With these limi–tations, our study introduces a deep learning-based static gesture recognition system trained with user-specific data collected under varying lighting and background conditions. Our contributions in this works are (i) to denote a novel concept of human-robot interaction via the recognition of hand sign, (ii) to build a platform for both hardware device and algorithm, and (iii) to prove successfully the efficiency and suitability of the proposed techniques in the real-world system.

3. PRELIMINARIES

In the first part of this study, hardware platform and theoretical computation are depicted to clarify internal components and architecture of mechanism for our approach.

3.1 Structure of robot

Our robotic system totally has four DoFs which provide rotational movements. To estimate both location and direction of this end-effector, it is necessary to utilize the kinematic equations of our robot from the specified values of joint parameters.
In each link, local coordinate $X_iY_iZ_i$ is attached and transformation matrix ${}^{i-1}_{i}T$ is derived according to represent the pose of this joint to the others. The forward kinematics of its platform are modeled using Denavit–Hartenberg (D–H) parameters as Table 2 to systematically describe joint transformations. Fig. 1 exemplifies theoretical diagram of our robot with coordinate system and where

α_i: Link twist — the angle between z_{i-1} and z_i measured about the x_i axis

a_i: Link length — the distance along the x_i axis from the intersection with z_{i-1} to the origin of frame i

d_i: Link offset — the translation along the z_{i-1} axis from the origin of frame i - 1 to the intersection with the x_i axis

θ_i: Joint angle — the rotation about the z_{i-1} axis from x_{i-1} to x_i

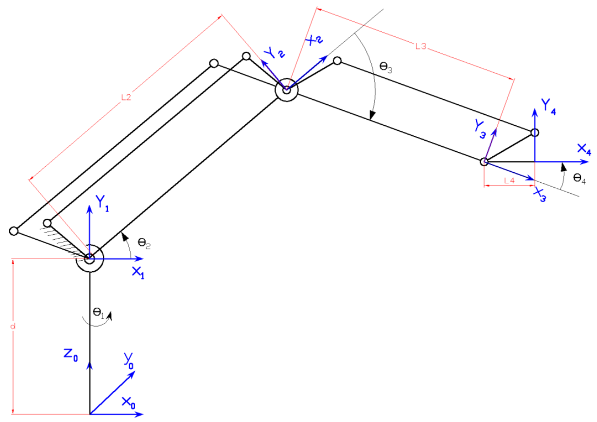


Figure 1. Theoretical diagram of our robot.

Table 2. List of D-H parameters in our robot.

Link	αi (rad)	a (m)	d(m)	θ
1	$\pi/2$	0	d_1	θ_1
2	0	a_2	0	θ_2
3	0	a_3	0	θ_3
4	0	a_4	0	$-(\theta_2+\theta_3)$

As following, A series of homogeneous transfor–mation matrices ${}^{i-1}_{i}T$ are identified for all four links.

$$
{}^{i-1}_{i}T = \begin{bmatrix} \cos\theta_i & -\sin\theta_i\cos\alpha_i & \sin\theta_i\sin\alpha_i & a_i\cos\theta_i \\ \sin\theta_i & \cos\theta_i\cos\alpha_i & -\cos\theta_i\sin\alpha_i & a_i\sin\theta_i \\ 0 & \sin\alpha_i & \cos\alpha_i & d_i \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (1)
$$

$\cos\theta_i$, $\sin\theta_i$: Trigonometric functions representing the rotation about the z axis

$\cos\alpha_i$, $\sin\alpha_i$: Trigonometric functions representing the rotation about the x axis

$a_i\cos\theta_i$, $a_i\sin\theta_i$: Determine the translational displa–cement along the x axis due to the link length

d_i: Translation along the z axis due to the joint offset

Totally, the final transformation matrix ${}^{0}_{4}T$ encapsulates the entire robotic configuration in Cartesian workspace. It permits to compute both position and orientation analysis of the end-effector.

$$ {}^{0}_{1}T = \begin{bmatrix} c_1 & 0 & s_1 & 0 \\ s_1 & 0 & -c_1 & 0 \\ 0 & 1 & 0 & d_1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2} $$

$$ {}^{1}_{2}T = \begin{bmatrix} c_2 & -s_2 & 0 & a_2c_2 \\ s_2 & c_2 & 0 & a_2c_2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3} $$

$$ {}^{2}_{3}T = \begin{bmatrix} c_3 & -s_3 & 0 & a_3c_3 \\ s_3 & c_3 & 0 & a_3c_3 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{4} $$

$$ {}^{3}_{4}T = \begin{bmatrix} c_{23} & s_{23} & 0 & a_4c_{23} \\ -s_{23} & c_{23} & 0 & -a_4c_{23} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{5} $$

where

${}^{0}_{1}T$: homogeneous transformation 4 × 4 matrix related frame 1 to frame 0

${}^{1}_{2}T$: homogeneous transformation 4 × 4 matrix related frame 2 to frame 1

${}^{2}_{3}T$: homogeneous transformation 4 × 4 matrix related frame 3 to frame 2

${}^{3}_{4}T$: homogeneous transformation 4 × 4 matrix related frame 4 to frame 3

Hence,

$$ {}^{0}_{2}T = {}^{0}_{1}T.{}^{1}_{2}T = \begin{bmatrix} c_1c_2 & -c_1s_2 & s_1 & a_2c_1c_2 \\ s_1c_2 & -s_1s_2 & -c_1 & a_2s_1c_2 \\ s_2 & c_2 & 0 & a_2s_2+d_1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{6} $$

We denote the following symbols as

$$ \begin{aligned} & d_1 = d \quad a_2 = L_2 \quad a_3 = L_3 \\ & a_4 = L_4 \ \mathrm{s}_1 = \sin\theta_1 \quad c_1 = \cos\theta_1 \\ & s_2 = \sin\theta_2 \quad c_2 = \cos\theta_2 \\ & s_{23} = \sin(\theta_2+\theta_3) \quad c_{23} = \cos(\theta_2+\theta_3) \\ & {}^{0}_{3}T = {}^{0}_{1}T.{}^{1}_{2}T.{}^{2}_{3}T \end{aligned} \tag{7} $$

$$ = \begin{bmatrix} c_1c_{23} & -c_1s_{23} & s_1 & a_2c_1c_2+a_3c_1c_{23} \\ s_1c_{23} & -s_1s_{23} & -c_1 & a_2s_1c_2+a_3s_1c_{23} \\ s_{23} & c_{23} & 0 & a_2s_2+a_3s_{23}+d_1 \\ 0 & 0 & 0 & 1 \end{bmatrix} $$

And,

$$ {}^{0}_{4}T = {}^{0}_{1}T.{}^{1}_{2}T.{}^{2}_{3}T.{}^{3}_{4}T = \begin{bmatrix} c_1 & 0 & s_1 & a_2c_1c_2+a_3c_1c_{23}+a_4 \\ s_1 & 0 & -c_1 & a_2s_1c_2+a_3s_1c_{23}+a_4 \\ 0 & 1 & 0 & a_2s_2+a_3s_{23}+d_1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{8} $$

where

Then, the Jacobian matrix J is used to convert joint velocities into end-effector velocities, an important requirement for control and trajectory planning.

$$ \begin{bmatrix} v_0^4 \\ \omega_{0,4}^0 \end{bmatrix} = \begin{bmatrix} J_v \\ J_\omega \end{bmatrix}\dot{q} = \begin{bmatrix} J_{v1} & J_{v2} & J_{v3} & J_{v4} \\ J_{\omega 1} & J_{\omega 2} & J_{\omega 3} & J_{\omega 4} \end{bmatrix}\dot{q} \tag{9} $$

With

$$ J_{vi} = z_{i-1}^0 \times \left(O_n^0 - O_{i-1}^0\right) \tag{10} $$

$$ J_{\omega i} = z_{i-1}^0 \tag{11} $$

And

$$ z_1^0 = \begin{bmatrix} s_1 \\ -c_1 \\ 0 \end{bmatrix} \tag{12} $$

$$ z_2^0 = \begin{bmatrix} s_1 \\ -c_1 \\ 0 \end{bmatrix} \tag{13} $$

$$ z_3^0 = \begin{bmatrix} s_1 \\ -c_1 \\ 0 \end{bmatrix} \tag{14} $$

$$ O_1^0 = \begin{bmatrix} 0 \\ 0 \\ d_1 \end{bmatrix} \tag{15} $$

$$ O_2^0 = \begin{bmatrix} a_2c_1c_2 \\ a_2s_1c_2 \\ a_2s_2+d_1 \end{bmatrix} \tag{16} $$

$$ O_3^0 = \begin{bmatrix} a_2c_1c_2+a_3c_1c_{23} \\ a_2s_1c_2+a_3s_1c_{23} \\ a_2s_2+a_3s_{23}+d_1 \end{bmatrix} \tag{17} $$

$$ O_4^0 = \begin{bmatrix} a_2c_1c_2+a_3c_1c_{23}+a_4 \\ a_2s_1c_2+a_3s_1c_{23}+a_4 \\ a_2s_2+a_3s_{23}+d_1 \end{bmatrix} \tag{18} $$

Then,

$$ J_{v1} = \begin{bmatrix} -s_1\left(a_2c_2+a_3c_{23}+a_4\right) \\ c_1\left(a_2c_2+a_3c_{23}+a_4\right) \\ 0 \end{bmatrix} \tag{19} $$

$$J_{v2} = \begin{bmatrix} -c_1\left(a_2 s_2 + a_3 s_{23}\right) \\ -s_1\left(a_2 c_2 + a_3 c_{23}\right) \\ a_2 c_2 + a_3 c_{23} + a_4 \end{bmatrix} \tag{20}$$

$$J_{v3} = \begin{bmatrix} -a_3 c_1 s_{23} \\ -a_3 s_1 s_{23} \\ a_3 c_{23} + a_4 \end{bmatrix} \tag{21}$$

And

$$J_{v4} = \begin{bmatrix} 0 \\ 0 \\ a_4 \end{bmatrix} \tag{22}$$

$$J_{\omega 1} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{23}$$

$$J_{\omega 2} = \begin{bmatrix} s_1 \\ -c_1 \\ 0 \end{bmatrix} \tag{24}$$

$$J_{\omega 3} = \begin{bmatrix} s_1 \\ -c_1 \\ 0 \end{bmatrix} \tag{25}$$

$$J_{\omega 4} = \begin{bmatrix} s_1 \\ -c_1 \\ 0 \end{bmatrix} \tag{26}$$

The Jacobian is essential for inverse kinematics, control, and trajectory planning, allowing the controller to translate desired end-effector motions into appropriate joint commands.

3.2 Design of mechanical gripper

In our design, a parallel-linkage-based gripper is selected due to its mechanical strength, compactness, and finger orientation under working conditions. Symmetric linkages have four major pivot points, enabling synchronized motion of the fingers. The mechanical parameters like link lengths ($L_1 = L_2 = 40$ mm) and base distances ($d_1 = 54$ mm, $d_2 = 20$ mm, $d_3 = 43$ mm) were calculated to ensure the gripper is able to grasp objects with maximum width 80 mm.

Important design limits such as collision avoidance between links, admissible angular range, and mecha–nical clearances are addressed. The motion profile of the gripper is optimized in a way that facilitates large opening span with tight closing force. Mechanical linkage converts rotary motion of a servo motor into synchronized lateral motion of the gripping jaws and is optimally suited for gesture-commanded gestures such as pick-and-place or interactive object manipulation.

Subsequently, diagram of a gripper with parallel linkage is shown in Fig. 2, where d_1, d_2 are the distance between fixed points A – B and C – D; and d_3, d_4 are the distance between fixed points A – D and B – C. L_1, L_2, $L_3$are lengths of the corresponding links.

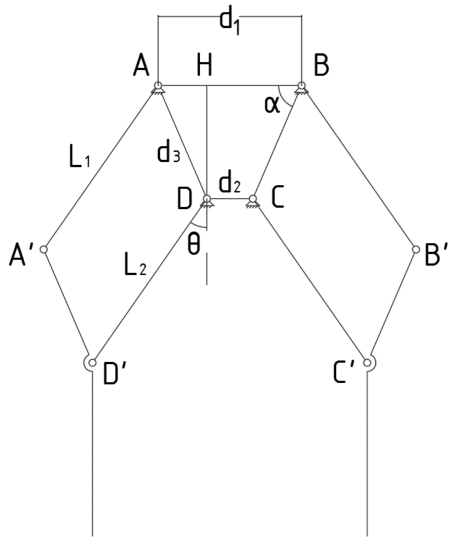


Figure 2. Theoretical diagram of our robotic gripper.

Owing to the width of target object as 54 mm, we choose the initially opening range of our gripper from 0 to 80mm. Firstly, it is essential to estimate the optimal distance of d_2 such that two bars CC’ and DD’ are not touched when the gripper is fully closed. Based on their dimensions, we select:

$$d_2 = 20\,\text{mm} \tag{27}$$

Similarly, the distances d_1, d_3, are evaluated so that when the gripper is fully closed, two bars AA’ and DD’ are not reached. According to our experiences, minimum lateral distance $Ah_{\min} = 21$ mm is obtained before they contact. We have:

$$d_{1\min} = 2AH_{\min} + d_2 = 2\times 12 + 20 = 44\,\text{mm}$$

Then,

$$d_1 = 54\,\text{mm} \tag{28}$$

The distance d_3 is computed such that the spur gears do not contact with the links CC’ and DD’. Hence,

$$d_3 = d_4 = 43\,\text{mm} \tag{29}$$

Later, the angle α is:

$$\begin{aligned} \alpha &= 90° + \arcsin\frac{AH}{d_3} \\ &= 90° + \arcsin\frac{\left(54-20\right)/2}{43} = 113.28° \end{aligned} \tag{30}$$

Therefore, the link lengths L_1, L_2 can be calculated. As our requirements, maximum opening range of our gripper can reach to 80 mm. We elect $L_1 = L_2$ to form a parallelogram linkage. Minimum length of this gripper is

$$\begin{aligned} &L_{1\min} = L_{2\min} \\ &= \frac{80-d_2}{2} = \frac{80-20}{2} = 30\,\text{mm} \end{aligned} \tag{31}$$

From above computation, we decide to pick as

$$L_1 = L_2 = 40\text{ mm} \tag{32}$$

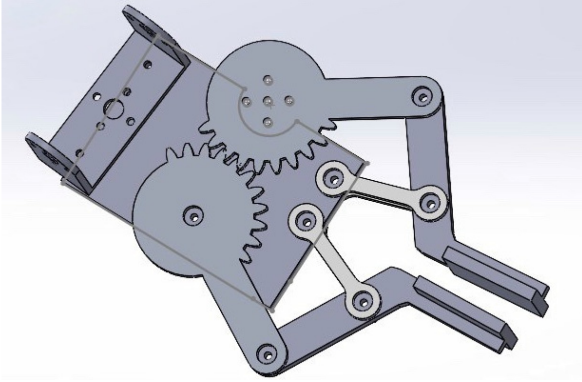


Figure 3. 3D design of the proposed gripper.

3.3 Design of electric components

In Fig. 4, the electrical arrangement includes sensors, actuators, a microcontroller, and communication modu–les for robot control and real-time gesture recognition. The TIVA C TM4C123G microcontroller is employed as the central controller owing to its processing power, GPIO (General Purpose Input Output) flexibility, and PWM functionality. It regulates three DC motors (each of which is interfaced with an encoder) for joint actuation, and an RC servo for gripper control.

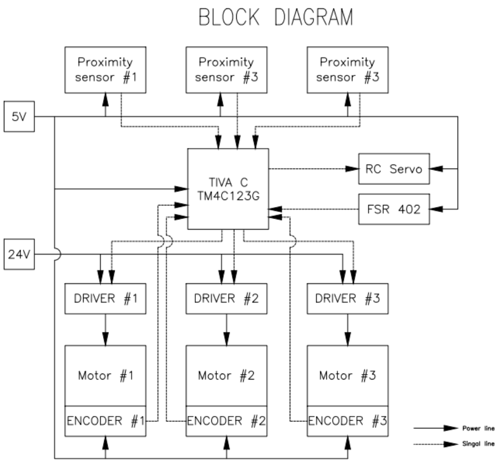


Figure 4. Block diagram of electrical components.

It gives feedback via three proximity sensors for environmental interaction and an FSR-402 force sensor for grip contact detection. Motor drivers are used for power supply control to the DC motors for accurate speed and position control. The system operates from a 24V power rail with regulated supply lines for logic-level devices. The signal path between sensors, drivers, and the controller is shielded to reduce electromagnetic interference, which is critical to maintaining control fidelity in gesture recognition applications.

4. THE PROPOSED METHOD

In this section, our concept introduces the control archi–tecture and gesture-based interaction framework deve–loped in the study. In the controller design, mechanism selections of driver and signal control are exemplified to handle properly. Far ahead, the proposed framework to detect and classify the hand sign from human is portrayed.

4.1 Controller design

Controller structure is derived from either modular control logic, MPC (Model Predictive Control) [25] or PID (Proportional-Integral-Derivative) feedback loops [26] with gesture-driven decision modules. Each motor is controlled by a PID controller with empirically adjusted individual gains for minimum overshoot, fast settling time, and minimal steady-state error. Velocity of each motor is regulated to 30 RPM (Round Per Minute) to balance responsiveness and control stability.

Selecting the proper driver for high-precision speed control involves one that would respond quickly during high-gain operation and be able to stabilize far-away pole positions in its dynamic response. We check this by comparing the linearity of the motor speed against the supply voltage input modulated by the PWM (Pulse Width Modulation) duty cycle.

The PWM signal, generated by the microcontroller and fed to the BTS7960 driver, is step-wise increased from a duty cycle representing a pulse width value of 10 to 250. The encoder is utilized to measure the resulting motor speeds, and the relationship between PWM input and motor output speed is given as Table 3.

Table 3. Relation between duty and speed of three main driving motors.

Duty	Motor 1	Motor 2	Motor 3
10	0	0	0
30	6	6	6
50	13	13	13
70	19	19	19
90	26	23	23
110	29	26	29
130	36	33	33
150	39	39	39
170	46	43	46
190	53	49	49
210	59	56	53
230	63	59	59
250	66	66	66
255	66	66	66

The driver is powered at 24 V supply voltage. A 50% duty cycle PWM signal is employed to transfer an average of 12 VDC to the motor. From the linear relation of PWM duty cycle and motor speed previously derived, we get the unit of transfer function for the integrated driver–motor system. The PWM signal is modulated at a frequency of approximately 5 kHz to suppress high-frequency harmonics involved in square wave modulation to provide stable and smooth motor operation.

To estimate the transfer function of a motor using the system identification toolbox as Fig. 6, the measu–rement data must be typically imported into MATLAB and apply pre-processing operations like filtering or detrending. The model order, poles, and zeros can be specified and evaluated. The resulting identified model can be validated using independent data sets or grap–hical comparison between simulated and actual outputs. This transfer function can serve as the main source to develop control systems for the dynamic behavior of the motor.

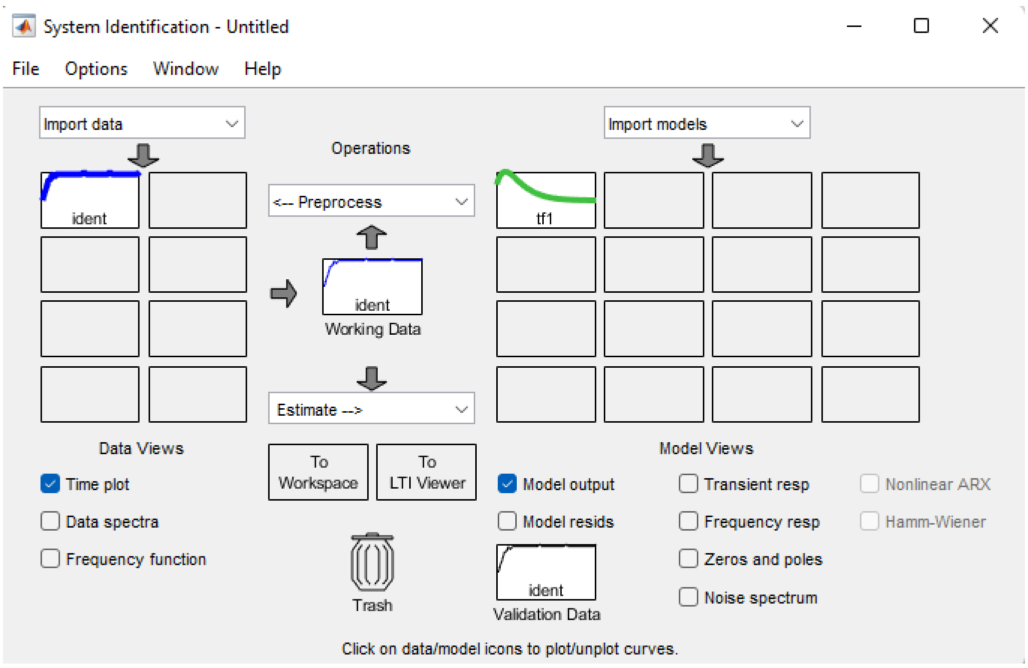
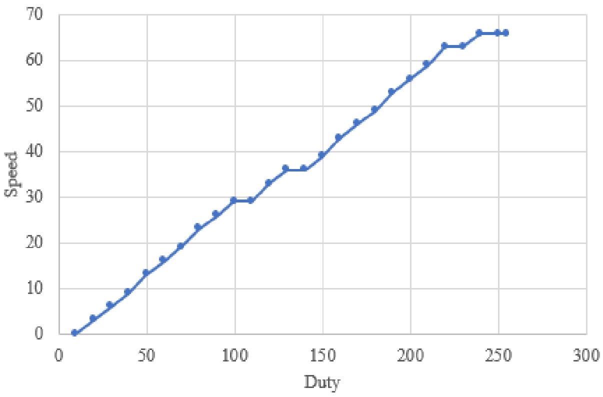


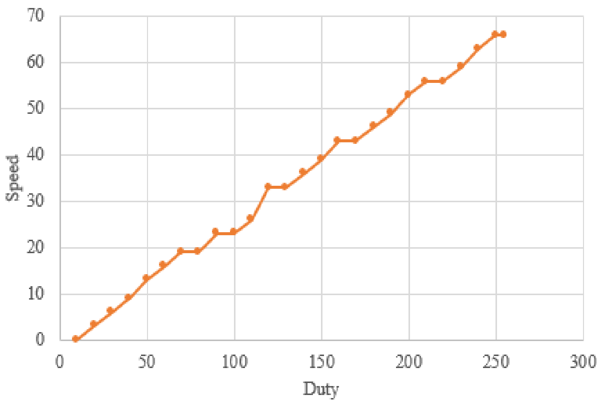
Figure 6. Illustration of the computational method for transfer function using system identification tool.

4.2 Interactive platform

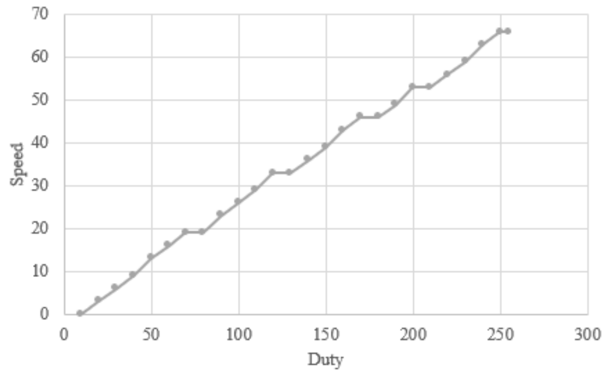
Gesture inputs are classified based on a deep learning-based model that has been trained with image data collected in controlled environments. Upon recognition of a gesture, corresponding control commands are converted into motor actuation signals. Smooth state transition and fast response for human-robot interaction are ensured with this approach.



(a)



(b)



(c)

Figure 5. Experimental graph of the driving response in motor 1 (a), motor 2 (b) and motor 3 (c).

The interactive system is used to connect the robotic hardware and the visual recognition system. It comprises the camera system for gesture input, the processing unit (GPU-enabled PC), and the robotic controller. Hand gestures are captured and pre-processed and transmitted to the trained deep neural network for classification.

The identified label is subsequently transmitted to the robot controller via UART or USB serial com–munication. Bi-directional communication is also offe–red by the platform for feedback, e.g., confirmation of gesture completion or grip pressure alerts. The modu–larity of our platform offers scalability for future tasks of higher complexity and integration of additional sensors or end-effectors. In Table 4, a set of hand gestures which are commonly met in daily activities, becomes nine signs to train our robotic system. These actions are defined by users and could be extended due to each specific application.

In training process, we consider that current average loss is less than 0.01 and the mAP (mean Average Precision) nearly reach to 100% as Fig. 7. Nevertheless, file weights should be checked to avoid overfit con–

ditions. If our training file is overfit, execution soft–ware just recognizes images that are the same as trained data and do not recognize gestures if different lighting conditions, background… and reduce the various input data.

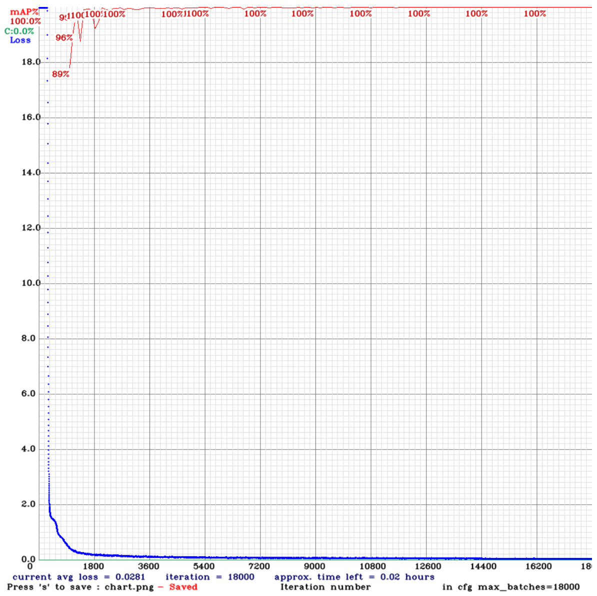


Figure 7. Chart of the training result after 18000 iterations.

4.3 Novelties and motivation

In this research work, the novelty of our design is the development of a tightly integrated system combining deep learning-based gesture recognition with real-time robot control in changing, unconstrained settings. In contrast to several recent studies that use controlled data or pre-trained models with poor flexibility, our tech–nique consists of training a proprietary deep neural net–work from scratch on static hand gesture images attained specifically from our experiments. This information was captured with various lighting condi–tions and with different backgrounds, resolving one of the main challenges of gesture-based human–robot collaboration: environmental variance robustness.

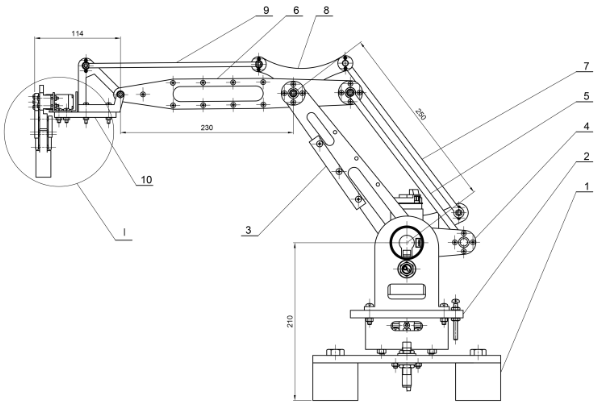
The deep learning architecture is trained to recog–nize nine basic hand gestures, which are then assigned to specific robot commands such as movement direc–tions, gripper operation, or stop actions. For ensuring real-time performance, we train the model for rapid inference on a GPU-based processing unit offering high responsiveness suitable for interactive applications. These results validate the capability of the model to generalize in diverse test environments, additionally validated in the experimental phase.

Another relevant innovation is the direct incor–poration of this AI-based recognition module within a modular robotic control system. Our system features closed-loop motor control with PID algorithms and bi–directional communication between perception unit and robot controller. This allows the robot not only to respond to gesture inputs but also provide status feedback, promoting usability and safety in collabo–rative work.

Furthermore, we heavily encourage user-defined gesture mapping and on-line adaptability to allow non-expert users to tailor the system to their own operational needs. Unlike several earlier works that focus on software or algorithmic efficiency alone, our work is the whole gamut - perception to physical actuation—so that the system is very valuable in real-world applications such as industrial HRC, assistive robotics, and sign-language interfaces. These combined traits emphasize the innovation and utility of our approach in enabling intelligent, vision-based human–robot interaction.

5. RESULT AND DISCUSSION

To verify the efficiency and properness of the proposed system comprising robotic execution and gesture recognition, the real-world platform as Fig. 8 has been built according to our design. In this system, host computer is highly computational ability with Core i9 14900K, 64GB DDR5, RTX 4070 Super VGA, 1TB SSD, OS Windows 10 Professional. The working envi–ronments of whole system are both indoor and outdoor where daily light might change continuously. Human stands in front of digital camera and perform numerous gestures by hand to control robot arm.



(a)

(b)

Figure 8. Computer-based 3D design and practical manipulator in our platform.

Two experiments are conducted under different li–ghting conditions such that test case 1 (TC1) has high light intensity with minimal background noise and test case 2 (TC2) contains moderate lighting with same controlled background. Nine gestures are utilized to command actions (up, down, left, right, forward, back, open, close and stop). To evaluate the output performance of our system, it is required to construct the confusion matrix as Fig. 9.

where:

True Positives (TP): The model predicted a label and is correctly matched with ground truth.

True Negatives (TN): The model has not made the prediction for the label and is not covered in the ground truth.

False Positives (FP): The model has predicted a label but is not in the ground truth (Type I Error).

False Negatives (FN): The model does not make a prediction, but it is part of the ground truth. (Type II Error).

Precision: The accuracy of our model told us in general how often we were correct when we predicted a hand gesture.

Recall: How often our model correctly predicted in general and how good our model is at predicting a specific hand gesture.

Accuracy: How often our model accurately predicted all classed.

		Ground truth		
		+	-	
Predicted	+	True positive (TP)	False positive (FP)	Precision = TP / (TP+FP)
	-	False negative (FN)	True negative (TN)	
		Recall = TP / (TP+FN)		Accuracy = (TP+TN) / (TP+FP+TN+FN)

Figure 9. Properties of confusion matrix.

Predicted \ Ground truth	Up	Down	Right	Left	Forward	Back	Open	Closed	Stop	
Up	200 11.11%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	100.00% 0.00%
Down	0 0.00%	200 11.11%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	100.00% 0.00%
Right	0 0.00%	0 0.00%	183 10.17%	4 0.22%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	97.86% 2.14%
Left	0 0.00%	0 0.00%	17 0.94%	194 10.78%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	91.94% 8.06%
Forward	0 0.00%	0 0.00%	0 0.00%	2 0.11%	200 11.11%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	99.00% 1.00%
Back	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	200 11.11%	0 0.00%	0 0.00%	0 0.00%	100.00% 0.00%
Open	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	200 11.11%	0 0.00%	0 0.00%	100.00% 0.00%
Closed	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	200 11.11%	0 0.00%	100.00% 0.00%
Stop	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	200 11.11%	100.00% 0.00%
	100.00% 0.00%	100.00% 0.00%	91.50% 8.50%	97.00% 3.00%	100.00% 0.00%	100.00% 0.00%	100.00% 0.00%	100.00% 0.00%	100.00% 0.00%	98.72% 1.28%

Figure 10. Confusion matrix of test case 1

Both test condition confusion matrices indicate high classification accuracy at an average of more than 97%. Most gestures were correctly identified, while there was minimal confusion between term Left and Right due to their visual similarity in bounding box features.

Predicted \ Ground truth	Up	Down	Right	Left	Foward	Back	Open	Closed	Stop	
Up	195 10.83%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	100.00% 0.00%
Down	0 0.00%	200 11.11%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	100.00% 0.00%
Right	0 0.00%	0 0.00%	165 9.16%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	97.86% 2.14%
Left	3 0.17%	0 0.00%	35 1.94%	200 11.11%	7 0.39%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	81.63% 18.37%
Forward	0 0.00%	0 0.00%	0 0.00%	0 0.00%	193 10.72%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	100.00% 0.00%
Back	2 0.11%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	200 11.11%	0 0.00%	0 0.00%	0 0.00%	99.00% 1.00%
Open	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	200 11.11%	0 0.00%	0 0.00%	100.00% 0.00%
Closed	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	200 11.11%	0 0.00%	100.00% 0.00%
Stop	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	200 11.11%	100.00% 0.00%
	97.50% 2.50%	100.00% 0.00%	92.50% 7.50%	100.00% 0.00%	96.50% 3.50%	100.00% 0.00%	100.00% 0.00%	100.00% 0.00%	100.00% 0.00%	97.39% 2.61%

Figure 11. Confusion matrix of test case 2

Due to two confusion matrices in both Fig. 10 and Fig. 11 for different environments, it could be observed that:

+ Accuracy in both working conditions is greater than our initial prospect (>90%).

+ The proposed algorithm does not obscure hand gestures with bounding box that differs significantly from each other.

+ However, right, and left gestures are regularly jumbled with each other because their bounding boxes are quite similar.

In practice, the setting parameters of above PID controller is not suitable because of the external impact (device errors, noise and so on). Thus, we regulate PID scheme again with the initially setting values which take in three sets of the proposed controller in previous section. In our design, we deliver 30 rpm to all motors because at this speed, the driving motor would move in our control and would not be very slow. It is expected that tuning PID control in the real-world application is approximately in the numerical simulation:

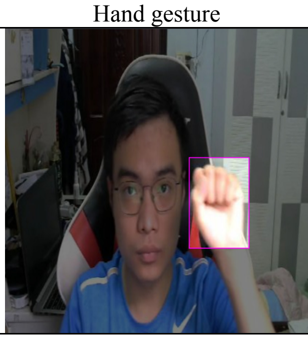
+ Settling time: $T_s < 0.333$ s

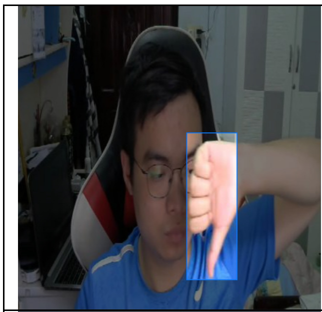
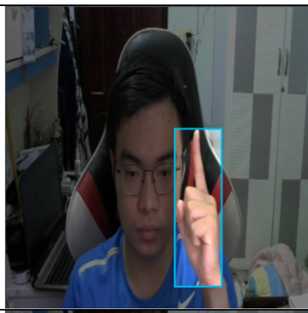
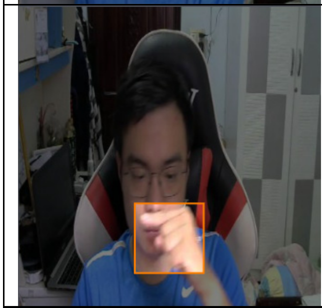
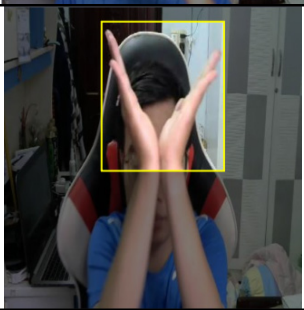
+ Overshoot: $OS < 5\%$

+ Steady – state error: ESS < 5%

For instance, we obtain the updated set of PID controller for the driving motor 1 is $K_{1P} = 2.5$; $K_{1I} = 35.2$; $K_{1D} = 0$. According to this adjustment, experimental result to validate the driving performance of motor 1 is exemplified as Fig. 12.

Table 4. A collection of nine hand gestures used for our robotic arms

Hand gesture	Label	Action	Hand gesture	Label	Action
	Front	Motor 1 rotates clockwise		Back	Motor 1 rotates counter – clockwise

	Down	Motor 2 rotates clockwise		Up	Motor 2 rotates counter – clockwise
	Left	Motor 3 rotates clockwise		Right	Motor 3 rotates counter – clockwise
	Open	RC servo rotates to $0°$ position		Close	RC servo rotates to $120°$ position
	Stop	Stop all system			

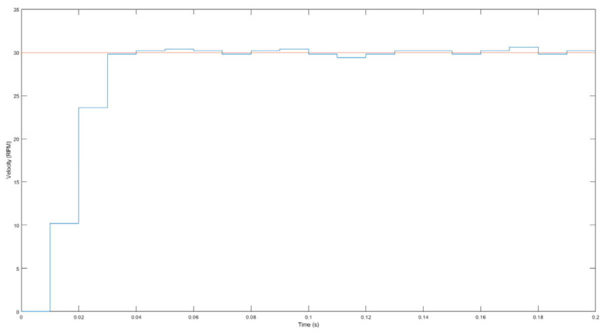


Figure 12. Experimental response of the tracking velocity for the driving motor 1.

At this moment, the performance of this PID set is measured as:

+ Settling time: $T_s = 0.03$ s
+ Overshoot: $OS = 0\%$
+ Steady – state error: $E_{SS} = 2.5\%$

This set of PID scheme satisfies our design criteria. In the same way, we have sets of PID control for motor 2 and 3 as following $K_{2P} = 2.9$; $K_{2I} = 34.6$; $K_{2D} = 0$; $K_{3P} = 1.8$; $K_{3I} = 26.4$; $K_{3D} = 0$.

In such module of computer vision, one graphical user interface as Fig. 13 which utilizes both OpenCV (library for image processing) and convolutional neural network (a deep net combining convolutional layers and connected layers), is implemented by Visual C# lan–guage programming. In the left side, there are several functional buttons for an operator to manipulate. Screen which is displayed in the right-centre side, is captured from our digital camera.

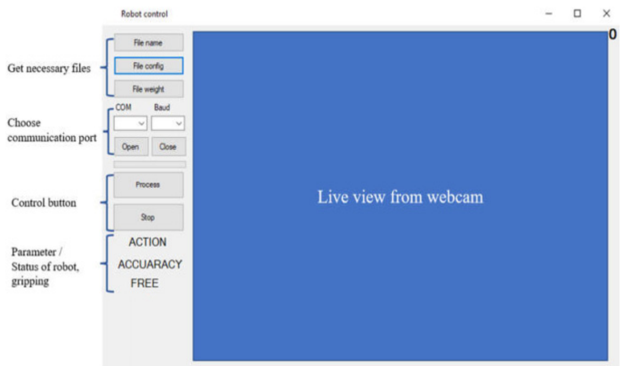
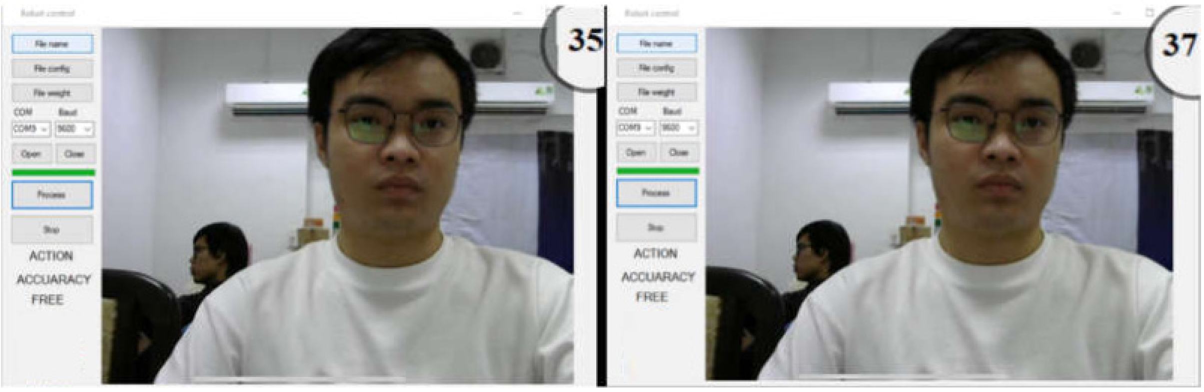


Figure 13. Graphical user interface in our design.



(a) (b)

Figure 14. FPS range display in our program, (a) 35 FPS and (b) 37 FPS.

Whenever an operator clicks on Start button, this program would create a separate thread for image processing tasks, the main thread still works to control our interface. To proceed the image processing technique, we use the Otsu threshold - multi level algorithm to extract the error prone area with a suitable small size. To associate with peripherals, the open-source YOLO is mainly deployed. By editing this source code to alter the input parameters and increase or decrease the selective thresholds for the prediction probability of each class, NET framework is also embedded this program.

To estimate the response time of our system, the processing sequence could be divided into two routines such that the consuming time comprises the image processing routine and motion control routine. In the module of image processing, from when the digital camera takes these pictures, recognizes the hand gesture, and transmits the signal to the CPU. Hence, the processing time in this stage is computed and shown on the right corner of the screen. In this experiment, our time ranges from 0.027 – 0.028 second (from 35 to 37 FPS approximately). In the module of motion control, from receiving signal, our program evaluates the PID scheme and transfers signal to the specific motor. This time is nearly equal 0.005 second. Therefore, our system archives the real time condition since the total response time is about 0.032 – 0.033 second which is less than 0.03333 second (maximum sampling time of our camera).

6. NOTES TO PRACTITIONERS

The gesture-controlled robot platform introduced in this investigation has a huge prospect in numerous real-life scenarios of intuitive, touchless, and adaptive man-machine interaction. First, in manufacturing environ–ments, the system can be used for human-robot colla–boration in assembly or material handling where the operator issues motion instructions through simple hand gestures without the need for physical interfaces, redu–cing downtime and enhancing safety. Second, in assis–tive robotics, the platform assists individuals with phy–sical disabilities in terms of enabling control of robotic devices using gesture commands, making them capable of independent living in their everyday activities such as object grasping or navigation in our home.

In addition, the ability of the system to detect user-declared static hand gestures accurately under varied lighting conditions categorizes it as a candidate for sign language recognition, which is a fundamental aspect of real-time hearing-impaired translation tools. The structure can also be implemented in distant and adversarial situations, such as search-and-rescue missions or servicing of high-risk zones, where gesture-based teleoperation enables secure and efficient robot teleoperation at a distance.

Due to its modularity and ability to respond in real-time, the platform is perfectly suitable for educational and research purposes, and it serves as a tangible example of our integration of AI with mechatronic systems. It serves as a testbed for researching a variety of problems with regard to control systems, computer vision, and human-centered robotics design. On the whole, the flexibility, non-invasiveness, and scalability of the system ensure it can be modified to meet the requirements of various fields that require human-in-the-loop control of robots.

7. CONCLUSION

In this investigation, an innovative concept to mani–pulate the robotic platform via gesture recognition scheme was developed. Firstly, the real-world system including industrial manipulator, software program–ming, digital camera and design of mechanical gripper are shown. A set of hand signs from an operator is then defined to represent their orders. Later, host computer would drive the robotic hardware owing to those signs.

Further developments are encouraged. In the comp–licated scenario such as multiple persons or crowded, the advanced detection method should be deployed to enhance. To deal with the series of hand signs, more precise and faster processing techniques could be integrated in the video analysis. Additional vision-based filtering procedure could be considered to achieve high quality of data although the working conditions are either inside or outside environment.

ACKNOWLEDGMENT

We acknowledge Ho Chi Minh City University of Technology (HCMUT), VNU-HCM for supporting this study.

REFERENCES

[1] Stecke, K. E., Mokhtarzadeh, M. (2022). Balancing collaborative human–robot assembly lines to optimise cycle time and ergonomic risk. *Inter–national Journal of Production Research*, *60*(1), 25-47.

[2] Nourmohammadi, A., Fathi, M., Ng, A. H. (2022). Balancing and scheduling assembly lines with human-robot collaboration tasks. *Computers & Operations Research*, *140*, 105674.

[3] Nguyen, T. P., Nguyen, H., Ngo, H. Q. T. (2024). Developing and evaluating the context-aware per–formance of synchronization control in the real-time network protocol for the connected ve–hicle. *Mobile Networks and Applications*, *29*(4), 1273-1294.

[4] Wang, Z., Yan, H., Wang, Z., Xu, Z., Wu, Z., Wang, Y. (2024, July). Research on autonomous robots navigation based on reinforcement learning. In *2024 3rd International Conference on Robotics, Artificial Intelligence and Intelligent Control (RAIIC)* (pp. 78-81). IEEE.

[5] Ngo, H. Q. T., Nguyen, T. H., Nguyen, H. A., Mo, V. N. Q., Le, Q. L., Vu, T. A., ... Thai, T. T. (2023). Model and simulation of tracking motion for accurate welding using a universal Cobot. *Inter–national Journal of Computer Applications in Technology*, *71*(1), 43-51.

[6] Zhu, Q., Huang, S., Wang, G., Moghaddam, S. K., Lu, Y., & Yan, Y. (2022). Dynamic reconfiguration optimization of intelligent manufacturing system with human-robot collaboration based on digital twin. *Journal of Manufacturing Systems*, *65*, 330-338.

[7] Ngo, H. Q. T., Nguyen, T. P., Nguyen, H. (2018, November). Investigation on barbot to serve human in public space. In *2018 4th International Conference on Green Technology and Sustainable Development (GTSD)* (pp. 300-305). IEEE.

[8] Wang, H., Zhang, X., Iida, F. (2024). Human-robot cooperative piano playing with learning-based real-time music accompaniment. *IEEE Transactions on Robotics*.

[9] Tran, P., Jeong, S., Lyu, F., Herrin, K., Bhatia, S., Elliott, D., ... & Desai, J. P. (2022). FLEXotendon glove-III: Voice-controlled soft robotic hand exoskeleton with novel fabrication method and admittance grasping control. *IEEE/ASME Transactions on Mechatronics*, *27*(5), 3920-3931.

[10] Le, H. T. N., & Ngo, H. Q. T. (2025). Application of the vision-based deep learning technique for waste classification using the robotic manipulation system. *International Journal of Cognitive Computing in Engineering*, *6*, 391-400.

[11] Huu, P. N., & Phung Ngoc, T. (2021). Hand gesture recognition algorithm using SVM and HOG model for control of robotic system. *Journal of Robotics*, *2021*(1), 3986497.

[12] Qi, J., Ma, L., Cui, Z., Yu, Y. (2024). Computer vision-based hand gesture recognition for human-robot interaction: a review. *Complex & Intelligent Systems*, *10*(1), 1581-1606.

[13] Qi, W., Ovur, S. E., Li, Z., Marzullo, A., Song, R. (2021). Multi-sensor guided hand gesture recognition for a teleoperated robot using a recurrent neural network. *IEEE Robotics and Automation Letters*, *6*(3), 6039-6045.

[14] Ngo, H. Q. T. (2023). Using an HSV-based App–roach for Detecting and Grasping an Object by the Industrial Manipulator System. *FME Transac–tions*, *51*(4).

[15] Antillon, D.W.O. et al. (2022). Glove-based hand gesture recognition for diver communication. *IEEE Tran–sactions on Neural Networks and Learning Sys–tems*, *34*(12), 9874-9886.

[16] Huang, H., Wu, D. E., Liang, Z., Sun, F., Dong, M. (2022). Virtual interaction and manipulation control of a hexacopter through hand gesture recognition from a data glove. *Robotica*, *40*(12), 4375-4387.

[17] Wu, Y., Liang, S., Ma, Y., Li, B. (2024). Prediction and classification of sEMG-based pinch force between different fingers. *Expert Systems with Applications*, *237*, 121635.

[18] Cao, L., Zhang, W., Kan, X., Yao, W. (2021). A Novel Adaptive Mutation PSO Optimized SVM Algorithm for sEMG□Based Gesture Recognition. *Scientific Programming*, *2021*(1), 9988823.

[19] Rahaman, H., Dyo, V. (2021). Tracking human motion direction with commodity wireless net–works. *IEEE Sensors Journal*, *21*(20), 23344-23351.

[20] Yakkati, R. R., Yeduri, S. R., Cenkeramaddi, L. R. (2021, December). Hand gesture classification usi–ng grayscale thermal images and convolutional neural network. In *2021 IEEE International Sym–posium on Smart Electronic Systems (iSES)* (pp. 111-116). IEEE.

[21] Le, H. P., Phan, X. T., Phan, T. Q., Tran, T. T. M. (2025). Development of an intelligence vision for a robot system to pick and place objects. *FME Transactions*, *53*(2), 233-242.

[22] Saez, B., Mendez, J., Molina, M., Castillo, E., Pegalajar, M., Morales, D. P. (2021). Gesture re–cognition with ultrasounds and edge computing. *IEEE Access*, *9*, 38999-39008.

[23] Duan, K., Zou, Z. (2024). Morphology agnostic gesture mapping for intuitive teleoperation of construction robots. *Advanced Engineering Informatics*, *62*, 102600.

[24] Espejel-Cabrera, J., Cervantes, J., García-Lamont, F., Castilla, J.S.R., Jalili, L. D. (2021). Mexican sign language segmentation using color based neuronal networks to detect the individual skin color. *Expert Systems with Applications*, *183*, 115295.

[25] Hiep, T. T., Cong, V. D., Phuong, L. H. (2023). Model Predictive Collision-Free Path Following Control for Nonholonomic Mobile Robots. *FME Transactions*, *51*(2), 193

[26] Hiep, T. T., Thang, V. V., Thong, H. D., & Tri, D. Q. (2023). Control of the Side Brush Street Sweeper for

Various Road Surfaces Using PID and Sliding Mode Controllers. *FME Transactions, 51*(3).

NOMENCLATURE

α_i	Link twist — the angle between z_{i-1} and z_i measured about the x_i axis
a_i	Link length — the distance along the x_i axis from the intersection with z_{i-1} to the origin of frame i
d_i	Link offset — the translation along the z_{i-1} axis from the origin of frame i - 1 to the intersection with the x_i axis
θ_i	Joint angle — the rotation about the z_{i-1} axis from x_{i-1} to z_i
${}^{0}_{1}T$	Homogeneous transformation 4 × 4 matrix related frame 1 to frame 0
${}^{1}_{2}T$	Homogeneous transformation 4 × 4 matrix related frame 2 to frame 1
${}^{2}_{3}T$	Homogeneous transformation 4 × 4 matrix related frame 3 to frame 2
${}^{3}_{4}T$	Homogeneous transformation 4 × 4 matrix related frame 4 to frame 3
d_1, d_2	Distance between fixed points A – B and C – D
d_3, d_4	Distance between fixed points A – D and B – C
L_1, L_2, L_3	Lengths of the corresponding links
T_s	Settling time
OS	Overshoot
E_{SS}	Steady – state error
K_{nP}	Proportional gain of the driving motor n^{th}
K_{nI}	Integral gain of the driving motor n^{th}
K_{nD}	Derivative gain of the driving motor n^{th}

Acronyms and abbreviations

PID	Proportional-Integral-Derivative
AI	Artificial Intelligence
HRI	Human Robot Interaction
AMPSO	Adaptive Mutation Particle Swarm Optimization
RF	Radio Frequency
CNN	Convolutional Neural Network
SVN	Support Vector Machine
DoF	Degree-of-Freedom
sEMG	Surface Electromyography
YOLO	You Only Look Once
FPS	Frame Per Second
CPU	Central Processing Unit

ПРИСТУП ЗАСНОВАН НА ДУБОКОМ УЧЕЊУ ЗА ПРЕПОЗНАВАЊЕ ГЕСТОВА СА СТАТИЧКИМ ПРЕДСТАВЉАЊЕМ РУКЕ

С.Х. Нгујен, Д.М. Фан, Х.К.Т. Нго

У ери вештачке интелигенције (ВИ), наша наука и технологија су достигле многе прекретнице, посебно у области интеракције човека и робота (ИЧР). Спајањем са техникама обраде слика, у овом истраживању је предложена стратегија заснована на ВИ за побољшање међусобног препознавања ИЧР путем ручних знакова. Првенствено, уведени су роботски хардвер, теоријско прорачунавање дизајна хватаљке и технике засноване на виду како би се успоставило радно окружење. Затим је демон–стриран предложени оквир, укључујући дизајн контролера и интерактивну платформу. Прикупљено је и обучено неколико ручних знакова људског оператера. Наш приступ је експериментисан у два случаја ради валидације ефикасности и исправности предложене методе са различитим условима осветљења. Из ових резултата се јасно види да је ова шема примењива у различитим областима као што су сарадња свесна људи, когнитивни робот или систем за превођење знаковног језика.