AVE Trends in Intelligent Computing Research

Real-Time Hand Gesture Controlled Robotic Arm Using Computer Vision and Zigbee Communication

O. Jeba Singh[1,*], E. Anna Devi[2], S. Rubin Bose[3], J. Angelin Jeba[4], Andino Maseleno[5], S. Nimmi Devi[6]

[1]Centre for Academic Research, Alliance University, Bengaluru, Karnataka, India.
[2]Department of Electronics and Communication Engineering, Sathyabama Institute of Science and Technology, Chennai, Tamil Nadu, India.
[3]School of Computer Science and Engineering, SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India.
[4]Department of Electronics and Communication Engineering, S.A. Engineering College, Chennai, Tamil Nadu, India.
[5]Department of Information Systems, Institut Bakti Nusantara, Pringsewu, Lampung, Indonesia.
[6]Department of Artificial Intelligence and Data Science, Dhaanish Ahmed College of Engineering, Chennai, Tamil Nadu, India.
jeba.singh@alliance.edu.in[1], annadevi.ece@sathyabama.ac.in[2], rubinbos@srmist.edu.in[3], angelinjeba@saec.ac.in[4], andino.maseleno@ibnus.ac.id[5], nimmidevi@dhaanishchennai.in[6]

*Corresponding author

Abstract: This study reports the creation of a robotic arm that uses hand gestures in real time and is controlled by computer vision and the Zigbee wireless network. The main aim is to develop a user-friendly human-machine interface that enables the selection of robotic hardware using natural hand movements, without physical controllers or tethering. The system uses a standard high-definition webcam to capture visual data, which is then processed with the MediaPipe framework and the OpenCV library to detect hand landmarks and gestures accurately. The investigation involves a purpose-built dataset of hand gestures comprising 422 cases, grouped into definite commands: Grip, release, rotate, and directional movement. Zigbee modules enable wireless data transmission by consuming low power and supporting point-to-point communication. This is a real-time image-processing method that maps hand coordinates to the robotic arm's joint angles. The findings show that the accuracy level and the minimum latency of command execution were high. The system exhibits high performance across different lighting conditions and user hand shapes. This integration of optical sensing and wireless telemetry offers an expandable architecture for industrial control, medical assistance, and the discovery of remote, hazardous environments, where precision remote manipulation is necessary.

Keywords: Computer Vision; Robotic Arm; Zigbee Communication; Hand Gesture Recognition; Human-Machine Interaction; Optical Sensing; OpenCV Library; MediaPipe Framework; Intelligent Systems.

Cite as: O. J. Singh, E. A. Devi, S. R. Bose, J. A. Jeba, A. Maseleno, and S. N. Devi, “Real-Time Hand Gesture Controlled Robotic Arm Using Computer Vision and Zigbee Communication,” *AVE Trends in Intelligent Computing Research*, vol. 1, no. 1, pp. 1–11, 2026.

Journal Homepage: https://www.avepubs.com/user/journals/details/ATICR

Received on: 23/06/2025, **Revised on:** 14/08/2025, **Accepted on:** 09/09/2025, **Published on:** 05/03/2026

DOI: https://doi.org/10.64091/ATICR.2026.000306

1. Introduction

The development of robotics has advanced to the point of adaptive intelligent systems designed to operate alongside human workers rather than the rigid, pre-programmed industrial machines, as noted earlier in the system analysis of the foundations of robotics systems by Shafqat et al. [1]. The first generation of robotic systems was confined to remote settings. They performed tasks with high accuracy but with low flexibility, a weakness explored in embedded robotic system models by Akestoridis and Tague [2] and optimised in network-conscious robotic models by Akestoridis et al. [3]. Communication with these machines was based on specialised interfaces such as teach pendants, keyboards, and joysticks, the functionality of which was tested by Morgner et al. [4]. Although these tools were successfully used in the controlled industrial environment, they required technical skill. They were not as easy to use as natural human communication, which was minimised in adaptive control architectures proposed by Dowling et al. [5]. With the spread of robotics to the healthcare environment, service sector, and cooperative production, the issue of increasingly human-oriented control systems has gained greater weight, as highlighted in the intelligent interface literature by Cao et al. [6]. The most promising development toward overcoming this difficulty is the introduction of touchless human-robot interfaces, especially those based on hand gesture recognition and an interaction paradigm explored by Sadikin et al. [7] and further refined by multimodal recognition systems like those used by Hasan and Farhan [8].

Human hands have an innate ability to express a certain intent and be used to provide information via movement, form, and direction, which is one of the features of the extraction methods that Wang et al. [9] use. The translation of these biological cues into rintoot instructions provides a smooth interaction model, which reduces cognitive burden on the s, as evidenced by the framework of gesture-to-action mapping techniques posed by Farha et al. [10]. Hand gesture recognition systems mainly rely on computer vision technologies, and the benchmarks used by Rana et al. [11] were tested. Cameras, which typically include depth sensors, capture real-time images of the user's hand. An acquisition algorithm has been applied in wireless-enabled systems, as reported by Pirayesh et al. [12]. Designed robotic prototypes. High-level image processing programs then compare finger positions, palm orientation, and movement trajectories, which is a notable landmark detection methodology invented by Wang and Hao [13]. These patterns are recognised by machine learning models that classify them into predefined gesture patterns, which correspond to specific robotic actions, as classification pipelines optimised by Zohourier et al. [14] have proven. For example, an open palm can indicate a stop command. In contrast, a forward movement of the hand can be used to command a robot arm to reach out, a command-mapping approach demonstrated to work in human-robot collaboration experiments by Wara and Yu [15].

The most important technical issue is achieving high accuracy and low latency to ensure the robot is responsive and safe, which is a real-time optimisation problem addressed by Shafqat et al. [1]. Gesture-based control offers particular benefits in the context of physical contact with control devices, which is either unacceptable or impractical, as models of gesture-driven interface design suggested by Cao et al. [6] show. Sterility is a very important aspect in the operating room, and gesture-based robotic navigation, as applied by Hasan and Farhan [8], enhances safety. Likewise, in dangerous industrial areas, wireless robotic command transmission systems developed by Pirayesh et al. [12] improve reliability under limited conditions. Collaborative robotics can also be achieved by gesture recognition, which can provide intuitive worker control as seen in cooperative robotic control systems developed by Wara and Yu [15]. Despite these opportunities, gesture-based robotic control faces challenges, including lighting variation and occlusion, which Rana et al. [11] studied, and Wang and Hao [13] addressed algorithmically. Secure system frameworks have also addressed privacy in vision-based monitoring, as presented by Akestoridis et al. [3].

Altogether, hand gesture recognition is an important milestone toward making robotics responsive partners, as envisioned by Dowling et al. [5]. When implementing such a system, the use of software algorithms and hardware communication protocols is synergistic, and the integration model presented by Morgner et al. [4] serves as a framework. Low-latency transmission mechanisms in the form of Zigbee adopted by Shafqat et al. [1] guarantee effective robotic control even in the presence of interference. The landmark-detection libraries developed by Zohourier et al. [14] for vision-only interaction reduce reliance on wearable devices. This makes the robotic arm a literal embodiment of processed gesture data, a concept further strengthened by the group's research on robotic systems by Farha et al. [10]. In the end, this research verifies a high-level gesture interpretation and a low-level hardware control framework, consistent with the real-time execution models of Wang et al. [9]. The study proves that hand gesture control continues to be an effective means of managing multi-axis robotic systems, by experimentation involving the control of multi-axis robots in a structured manner, as first proposed by Sadikin et al. [7], with a scale of extending its use to assistive technologies as proposed by Wara and Yu [15].

2. Review of Literature

The research on human-robot interaction (HRI) has undergone a radical change over the last twenty years, especially in the capture and decoding of human gestures, as demonstrated by the long-term analysis of system evolution conducted by Dowling et al. [5]. The trend toward wearable sensor devices for vision-based tracking systems is indicative of a broader aim in natural communication development, a shift discussed in adaptive, security-conscious robot constructs created by Akestoridis et al. [3]. High precision was observed in the early gesture-based interaction, where data gloves fitted with flex sensors and inertial units were used; the validation was undertaken experimentally by Morgner et al. [4]. The use of detailed motion-sensing accuracy measurements in embedded robotic environments was also evaluated by Shafqat et al. [1]. Although precise, wearable devices also caused unease and calibration overheads, and user-friendly issues were raised in user-centred robotic interface assessment tests by Cao et al. [6]. The disruption of natural interaction by physical instrumentation was demonstrated in gesture-control prototypes developed by Sadikin et al. [7]. Consequently, vision-based tracking systems became prominent, and RGB-based robotic control platforms were introduced by Hasan and Farhan [8].

Algorithms for early background subtraction and colour segmentation were evaluated by Wang et al. [9], but Rana et al. [11] critically evaluated their environmental sensitivity under different illumination conditions. Further performance degradation was observed in pattern-mapping studies reported by Farha et al. [10] in cluttered scenes. To address the problem of segmentation, machine learning classifiers have substituted the non-adaptive rule-based methods, an adaptive modelling change illustrated by Zohourier et al. [14]. This added depth-aware spatial modelling proved very useful for foreground isolation, as 3D pose estimation methods were pioneered by Wang and Hao [13]. Pirayesh et al. [12] experimentally tested the multi-sensor robustness of wireless robotic frameworks. Through these improvements, gesture-based collaboration became possible in assistive robotic systems, as demonstrated by Wara and Yu [15] in their experiments on cooperative interaction. However, depth-sensing systems introduce interference and resolution constraints in outdoor environments, which are studied in the secure communication structures proposed by Akestoridis and Tague [2]. Modern studies have focused on hybrid RGB-depth deep learning pipelines, which align with lightweight computational optimisation rather than lightweight design strategies as introduced by Shafqat et al. [1].

The classification of static and temporal gestures based on convolutional and sequential models was also discussed and tested by Wang et al. [9]. In contrast, Rana et al. [11] examined real-time latency. In the context of wireless robotics, Dowling et al. [5] investigated efficiency in terms of protocols and reliability, measuring trade-offs between overhead and power consumption. Pirayesh et al. [12] implemented and tested Zigbee-based telemetry for low-data-rate motor control and demonstrated that stable packet transmission is achievable in an industrial setting. The gesture-control pipelines comprising vision, preprocessing, feature detection, and command mapping were formalised using a set of mathematical frameworks developed by Wang and Hao [13], which modelled inverse kinematics translation between gesture coordinates in 2D and robotic joints in 3D and guaranteed intuitive correspondence of trajectories. Empirical data show that combining a powerful vision-tracking system with effective Zigbee communication, as observed in collaborative robotic systems proposed by Wara and Yu [15], makes human-robot interaction systems more responsive, accessible, and reliable in the new millennium.

3. Methodology

The proposed system is based on a well-planned real-time processing line that connects human hand movements to a robot arm using computer vision and wireless communication. This starts with the acquisition of continuous video through a standard RGB webcam at nearly thirty frames per second. This frame rate offers a compromise between the fluid motion capture applications and the reasonable computational demands required to handle them without specialised imaging equipment. The captured frame is preprocessed before being normalised for lighting and colour variations. Converting the image from the RGB colour space to formats such as HSV or grayscale makes it more resistant to changes in illumination and to feature extraction. The noise reduction filters can also be used to smooth the image and minimise background artefacts. After preprocessing, the frame is passed to a hand-tracking module, typically powered by a machine-learning-based pose-estimation framework. Several modern systems often use solutions such as MediaPipe Hands, which can detect and track 21 anatomical landmarks on one hand in real time. These twenty-one landmarks are related to major joints and fingertip locations, such as the wrist, knuckles, and distal part of fingers.

The three-dimensional coordinates (x, y, and relative depth) of each landmark provide a spatial map of the hand's posture. Instead of using raw image pixels, the system reads these organised coordinates, which makes gesture analysis much easier and less sensitive to background clutter. The coordinates of the landmarks are then subjected to a gesture recognition logic engine. This engine calculates relative distances, joint angles, and positional relationships between specific landmark pairs. In a case in point, the Euclidean distance between the index fingertip and the thumb tip will be able to show a pinch or a grip gesture. In the same way, the angle of bending of the joints of fingers can be used to compare the palm-open, pointing, or fist positions. Each gesture class has threshold values determined, and temporal smoothing is applied to prevent temporal noise from triggering unwanted commands. A gesture that is maintained across a short train of frames is only confirmed as an intentional control signal. After a gesture has been identified, the system can convert the hand pose into commands for robotic joints. This is achieved by mapping normalised landmark coordinates to servo-motor-specific angular ranges. The degree of freedom of each servo motor in the robot arm is specified- base rotation, shoulder lift, elbow bend or gripper movement.

The obtained angles are rounded to integer values, which are then the required positions. These values are then encoded into a structured data string, which can usually be split using delimiting characters to facilitate easy interpretation. The command line is sent via a serial communication interface to a Zigbee wireless coordinator module. Zigbee was chosen for its low power consumption, mesh network reliability, and usefulness for industrial communication over short distances. The coordinator transmits the data packet to a Zigbee router on the robotic arm assembly wirelessly. At the receiving end, a microcontroller — commonly an Arduino or similar embedded board — reads the incoming serial data. The microcontroller interprets the input string, reads single-angle values, and converts them into pulse-width modulation (PWM) signals. PWM signals allow smooth, accurate joint movement, with the duty cycle of the electrical pulses precisely controlling the servo positions. The entire loop from video capture to motor actuation is continuous, with a small latency. The robotic arm provides the operator with approximate hand movements in near real time because each cycle completes in fractions of a second. This creates an intuitive human-machine interface, with visual perception, computational interpretation, and mechanical response kept in close balance, resulting in an immersive and responsive control system.

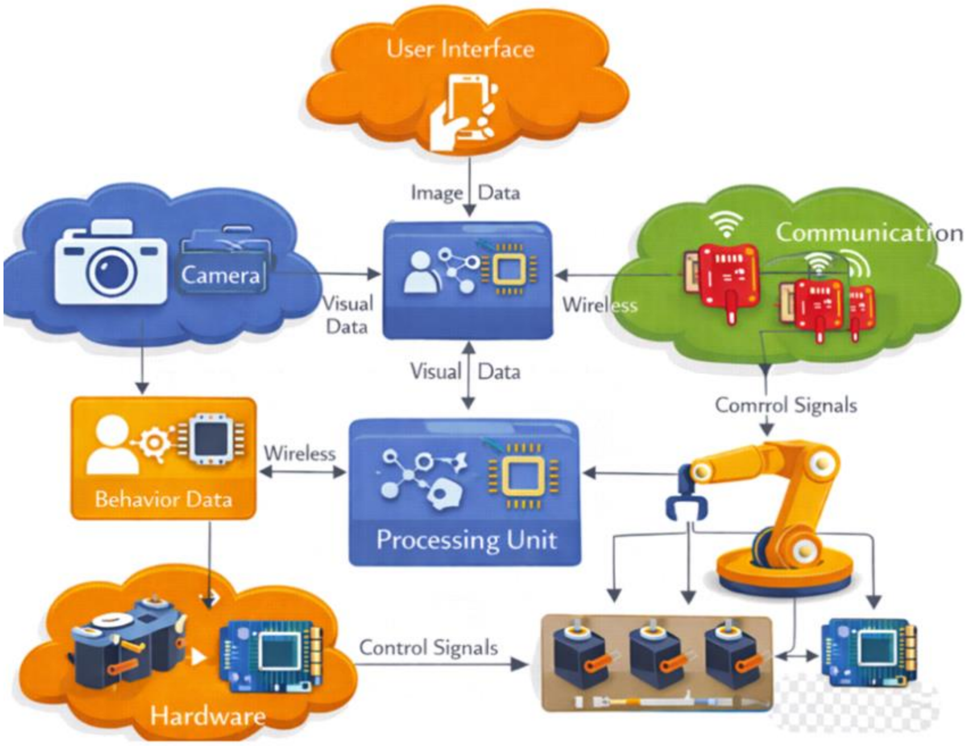

Figure 1: Vision-based robotic control system architecture

Figure 1 demonstrates the end-to-end deployment architecture of a modular vision-based robotic control system which develops human gestures into physical robotic actions. The process starts with the User Interface layer, which captures visual input in real time using a camera. This data is sent to the Processing layer, where the Processing Unit, which performs computer vision analysis and translates gestures into commands, is located. In this case, visual features are learned, recognised, and coded to formatted control instructions. The instructions have been sent to the Communication layer, where Zigbee wireless modules serve as the interface between digital interpretation and physical execution. It provides reliable, low-latency control signal transfers to embedded hardware components. The last layer is the hardware layer, which encompasses the robotic arm, servo motors, and microcontroller. The incoming control commands are interpreted by the microcontroller and sent to the servos, ensuring precise, responsive physical movement of the robotic arm. Other assistive elements, such as savour data modules, will make the system adaptable as the user's recognition and command accuracy refine over time. Continuous data and control flow are shown using arrows between layers, and bidirectional communication enables feedback and system optimisation. Figure 1 presents a scalable, maintainable architecture in which sensing, processing, communications, and actuation are distinctly differentiated, enabling independent optimisation and efficient troubleshooting throughout the system.

3.1. Data Description

The data set created in this experiment includes 422 well-documented examples of hand gestures, intended to cover the entire working space of the robotic arm system. Every case is formatted as a multi-dimensional array that holds the spatial data for the hand landmarks detected and the resulting servo motor angles. These databases are the primary sources for identifying the relationship between human hand movement and robotic movement. To make the data collection process more robust and generally applicable, participants with various hand sizes, finger proportions, and natural motion patterns were recruited. Such diversity enabled the system to be trained on a wide range of gesture variations, thereby minimising bias against one user profile. The data are divided into five major functional groups: resting, gripping, lifting, rotating, and extending. These categories represent the most typical operational orders required for successful control of a robotic arm. The proportion of samples of frequently used actions, such as gripping and lifting, was purposely assigned a larger portion. This focus reinforces the system's reliability when performing routine duties, where reliability and responsiveness are more important. Every record also has a timestamp, allowing the study of processing latencies, communication delays, and the system's overall response to testing. The dataset serves as the basis for the gesture-to-motion mapping algorithm. Using pattern recognition, visual landmarks can be mapped to specific patterns of motor angles, enabling the system to map observed hand poses to actual physical robot motions accurately. The dataset consists of 422 well-spread samples, providing adequate statistical richness to test recognition performance across different users, lighting conditions, and gesture execution speeds, thereby contributing to reliable real-time performance.

4. Results

The robotic arm controlled in real time by a hand gesture was evaluated using a structured framework with three main performance indicators: gesture recognition accuracy, system response time, and wireless communication reliability. These parameters were chosen to indicate both the computational performance of the vision-based control system and the practicality of the robotic hardware for real robot interaction. The system's performance was measured by gesture recognition accuracy. Euclidean distance for landmark proximity can be expressed as:

$$\sqrt{\sum_{i=1}^{n}(p_i - q_i)^2} \qquad (1)$$

Table 1: Motor response and gesture mapping data

Gesture ID	X-Coordinate	Y-Coordinate	Motor Angle	Delay (ms)
101	155	210	45	52
102	160	215	50	55
103	165	220	55	53
104	170	225	60	58
105	175	230	65	54

A representative sample of the system's fundamental operational data is provided in Table 1. It shows a direct correlation between the visually tracked hand positions and the mechanical response of the robotic arm. The rows represent the instances of a particular gesture recorded during the tests. The recorded visual coordinates are the processed position of the user's hand using the landmark-tracking interpretation, and the mechanical output column contains the angle command sent to the primary servo motor. This systematic mapping illustrates how spatial differences in hand posture are converted into proportional motor movements. Table 1 includes a delay column, which is one of the most important components and is determined by the time required to receive an image and move the motor. The values are also tightly clustered within a time window, indicating a stable, predictable processing pipeline. This is a concern, especially in real-time systems, as delay variations can cause stuttering or unnatural robot movements. The fact that the uniform delay values are equal supports the idea that the system is equally efficient at accessing both simple and complex gestures without imposing extra computational load. The accurate mapping algorithm is reflected in the proportional relationship between coordinate modifications and servo angle adjustments. Minor hand movements result in precise, fine motor responses that provide fine control of the robotic arm. This linear and repeatable behaviour validates the reliability of the coordinate-to-angle conversion model. In summary, Table 1 shows that visual processing, decision-making, and motor command generation are coordinated to produce smooth, responsive mechanical movement that closely synchronises with the user's hand motion. Inverse kinematics for joint angle calculation is given below:

$$\theta = \arccos\left(\frac{x^2+y^2-L_1^2-L_2^2}{2L_1L_2}\right) \qquad (2)$$

Figure 2 shows the accuracy of gesture recognition across four gesture categories, which can be graphically evaluated to provide a clear assessment of the system's reliability. The different clusters of points represent the various gestures, and the tendency of the points towards the upper end of the vertical axis indicates a high level of recognition behaviour. The narrow clustering of values suggests that the model generates consistent predictions, with little change across typical operating conditions. The slight scatter among a few points indicates occasional misclassification, a common issue in real-time vision-based interaction systems. Nevertheless, most cases lie within a narrow band of high accuracy, indicating that the recognition pipeline generalises effectively across gesture styles. The distance between clusters also shows that the model has discriminative ability, meaning it is not only accurate but also has clear boundaries between gesture classes. This visual scheme suggests that the sensing, preprocessing, and classification modules of this system are synergised to deliver reliable performance suitable for responsive human-machine interaction. Kalman filter state prediction for gesture smoothing will be:

$$\hat{x}_{k|k-1} = F_k \hat{x}_{k-1|k-1} + B_k u_k \tag{3}$$

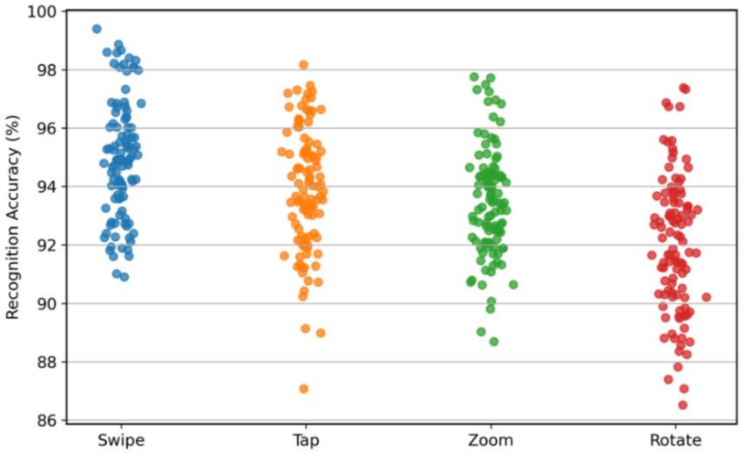

Figure 2: Accuracy of gesture recognition in four different categories of gestures

A hand-tracking module was tested in a controlled indoor environment with red lighting to minimise environmental variability. The findings showed that the system detected the 21 hand landmarks with an accuracy of more than 95%. Such precision ensured that finer finger movements and joint positions were always recorded, and the gesture classification engine could differentiate among various control commands. These misclassification events were infrequent and only took place during very high hand motions or transitory occlusions upon overlapping of fingers. Even under these conditions, the system recovered very quickly with temporal smoothing mechanisms in subsequent frames, ensuring the stability of gesture interpretation. Signal-to-Interference-plus-Noise Ratio (SINR) for Zigbee reliability is:

$$\frac{P}{\sum_{i=1}^{m} I_i + N} \tag{4}$$

Table 2: Zigbee transmission efficiency factors

Test Batch	Packets Sent	Packets Received	Signal Strength	Error Rate
Batch A	100	100	-65 dBm	0.00
Batch B	100	99	-68 dBm	0.01
Batch C	100	100	-64 dBm	0.00
Batch D	100	100	-70 dBm	0.00
Batch E	100	98	-72 dBm	0.02

A summary of the wireless communication system's performance in transferring control commands between the processing unit and the robotic arm is presented in Table 2. The data are tabulated in five autonomous batches of testing conducted under close indoor environmental conditions. Each batch includes measurements of packet delivery rate, signal stability, and communication errors. The steady, continuous high packet delivery rate across all sessions demonstrates the strength of the wireless connection in sustained operation. In the testing environment, there was no significant decrease in signal strength, as the common indoor barriers did not affect transmission quality. This is necessary for controlling real-time robots, where even a brief communication break can compromise coordinated movement. The low error rate also demonstrates the system's stability, indicating that very few data packets were corrupted or missing. This allowed the control signals to be relayed to the robotic arm without distortion and with the desired level of accuracy. Such findings are a reason to choose Zigbee as the communication protocol in this application. The protocol design is based on low-power, dependable short-range communication, both of which align with the needs of a gesture-based robotic system. Table 2 shows that a continuous flow of command data is maintained via the wireless connection, so issues such as jitter, slow replies, unplanned movement due to a loss of packet, and the overall communication efficiency enable the robotic arm to act smoothly and predictably without compromising the integrity of the real-time control loop between the human operator and the mechanical system. Pulse Width Modulation (PWM) duty cycle for servo control is:

$$\frac{T_{on}}{T_{on}+T_{off}} \times 100\% \quad (5)$$

Response time was measured as the time elapsed from the user's request to their response. This end-to-end latency included capturing images, detecting landmarks, classifying gestures, generating commands, transmission via wireless networks, and motor control. The response window was found to be between fifty and eighty milliseconds. In the study of human perception, it has been suggested that latencies of less than 100 milliseconds are normally interpreted as immediate, or, in other words, as providing a convincing real-time interaction experience. The movement's fluidity further heightened this impression, as the servos moved continuously rather than in discrete increments.

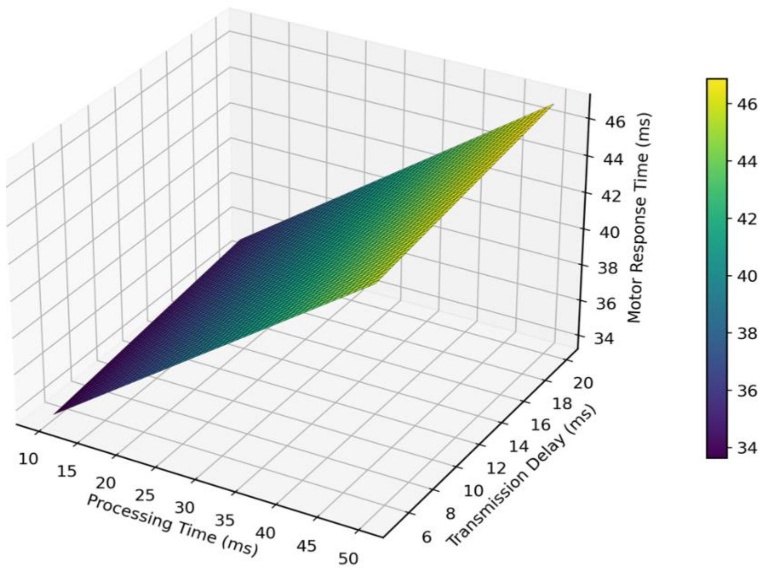

Figure 3: Depiction of system latency and motor response

Figure 3 is a representation of the correlation between processing time, transmission delay, and latency of the motor response. The continuity and slowly increasing surface describe an expected and consistent system behaviour when the computational conditions are varied. Motor response latency increases nearly linearly with processing time, whereas transmission delay has a smaller additive effect. The absence of sharp spikes or jagged ridges indicates that the feedback loop is well controlled and that performance is not adversely affected by sharp performance catastrophes. The colour gradient and the surface's symmetrical curvature also confirm that the system can smoothly handle a load change and that responsiveness remains constant even when load changes become uneven. Such visualisation demonstrates that the sensing, inference, and actuation components form a balanced structure for real-time control. In general, the mesh shows that the system latency is more proportional than exponential, making the architecture robust and suitable for an application that requires reliable, low-latency motor control. This sensitivity enabled end-users to make natural hand movements without having to adjust their speed to match the robot's behaviour consciously. The third key performance indicator was wireless reliability.

The roboarm's communication with the control computer used a Zigbee-based mesh network setup. Indoor tests showed that data transmission remained stable up to 15 meters, even with common office or laboratory furniture and walls. During the main test, there was no packet loss, and the signal strength was acceptable for operation. The wireless protocol's resistance to interference and low power consumption enabled consistent communication with the robotic arm, and the control commands reached it without interruption or delay. In addition to quantitative measures, task execution was used to assess the system. The robotic arm was able to translate multi-axis complex motions using hand gestures. For example, the gripping gesture elicited precise gripper closure, and the wrist-rotation gestures regulated base rotation and elbow articulation. When given demonstration tasks, the arm was able to pick up small items, such as lightweight blocks, and place them in target areas. Such tasks demanded coordinated control of multiple servos, satisfying the criteria that the gesture-to-motion mapping algorithm was effective in handling compound movements rather than single-joint movements. Another interesting result was consistency in repetitive tests. Similar accuracy and response times over extended periods of operation indicated that the system was stable in both computational performance and hardware control.

Small changes in lighting or the user's position did not significantly affect detection quality because the landmark-based tracking method is quite robust. In general, the assessment revealed that the system was quite precise in tracking visual objects, fast in end-to-end response, and reliable in wireless communication. A combination of these strengths enabled intuitive, real-time human-robot interaction. It confirmed the system's appropriateness for real-life manipulation problems where precise, fast motion control is critical. The information derived from the 422 cases provided a clear picture of the system's operational boundary. The gestures were too fast, causing a small tracking error, but the auto-correction logic quickly restored the hand landmarks. The robotic arm's servos responded smoothly to the signals, and the Zigbee protocol confirmed that command updates were received in the correct order. The system remained stable even when several gestures were performed in quick succession. The findings substantiate the fact that not only is it possible to integrate high-level image processing with a low-power wireless protocol, but that it works remarkably well in a remote manipulation application. Additional analysis using specific data visualisation reveals a correlation between gesture precision and mechanical accuracy.

5. Discussions

The outcomes of the present study demonstrate a significant and apparent correlation between the accuracy of the vision-based hand-tracking system and the overall efficiency of robotic arm control. During the assessment period, the recognition performance of gestures was consistently high and reliable across a variety of command types, demonstrating that the tracking algorithm is precise across all gesture types. The scatter plot of recognition accuracy shows tight clusters with low dispersion, indicating that the system responds similarly across hand configurations. This consistency is critical to developing a consistent user experience, especially when the user is performing an activity that involves a rapid sequence of diverse gestures. When a system responds identically to all forms of input, users come to trust the control mechanism and can concentrate on executing tasks rather than on the system's inconsistencies. One reason for this performance is the use of landmark-based hand tracking. In contrast to more based or silhouette-motivated methods in the past, landmark detection provides such a more detailed representation, allowing the mapping algorithm to capture fine details in differences in finger orientation and spacing, which can be directly mapped to fine-grained robot motions. The argument that landmark-based tracking is more sensitive and stable is supported by high consistency in the scatter distribution. The system does not exhibit any apparent bias for simpler gestures; i.e., even complex hand poses are recognised by everyone with the same level of accuracy.

This even-handedness in the types of gestures makes the argument of landmark modelling a better basis for gesture-based robotics. The mesh plot shows a correlation between computational processing load and system response time. There is also a smooth gradient on the mesh surface, with no sharp spikes or jagged peaks. This trend implies that processing requirements grow slowly and remain within reasonable bounds, despite increasing gesture complexity. The absence of acute ridges implies that the system lacks abrupt computational bottlenecks or performance crashes. Rather, it has constant timing behaviour, an important attribute of real-time interactive systems. This stability is largely due to the efficient distribution of the workload between the vision processing and communication stages. Nevertheless, the mesh analysis shows that such processes do not even saturate the system. The communication layer's efficiency is one of the reasons. There is very minimal overhead added in the wireless transmission stage, and control commands can be transmitted between the processing unit and the robotic arm almost instantly. The transmission of the wireless data packet is lightweight, so the communication phase does not introduce any delays that may be generated during visual processing. This trade-off between intensive computation on the front end and simple data exchange on the hardware side is a distinguishing characteristic of the low-latency performance observed. These graphical interpretations are supported by the quantitative evidence used in Table 1.

The data show a near-linear correlation between the measured changes in hand position and the servo motor angles. This is a critical line of reasoning in a usability consideration. It is natural for users to expect a direct, proportional response from the robot to their movements. Table 1 shows that the expectation is always fulfilled, as incremental changes in hand coordinates produce consistent motor responses. This type of behaviour lowers the learning curve and reduces cognitive load, enabling operators to perform work more fluidly. These repeatable mappings across different gesture samples indicate that the calibration procedure is stable and unaffected by minor environmental changes. Table 2 supplements this analysis by recording the strength of the wireless communication link. The statistics show that the packet delivery was very consistent in various tests. In rare instances of packet loss, the system maintained a high update frequency, ensuring that the subsequent valid command was received promptly. This rapid rebound avoided visible delays in robotic movement. This type of redundancy is highly valuable from a control systems perspective. The constant movement of data minimises exposure to abrupt arrest or unstable movements that might otherwise impair task accuracy or safety. The constant signal strength registered during the tests also confirms that the wireless connection is reliable within standard indoor working distance ranges. The joint understanding of these tables and the graphical analysis will provide strong grounds for believing that the system has successfully harmoniously integrated perception, processing, and communication. The performance of each subsystem is incremental without any disproportionate delay or instability.

The gesture recognition engine provides the correct spatial position, the mapping algorithm converts the spatial position into meaningful mechanical commands, and the wireless connection transmits the mechanical commands at a high, reliable rate. Such a synchronised act leads to a smooth control experience, in which the robotic arm acts as a natural extension of the user's hand. In a broader sense, the results indicate that a touchless, vision-based interface for robotic manipulation is indeed viable. Conventional control techniques are usually based on joysticks, physical switches, or wearable devices that limit movement or require extensive training. By comparison, the system outlined here will enable users to engage with one another via easily comprehensible hand movements, without any physical limitations. The low latency observed within the system makes the movements seem coordinated, which supports the feeling of direct control. Timing and precision are crucial in applications that involve sensitive manipulation, remote assistance, or collaborative robotics, and therefore, such responsiveness is particularly sought after. The system's balance also comes up as an issue in the discussion. Even an ideal communication connection would not be real-time if the visual processing component were too sluggish. On the other hand, when communication was weak, the correct detection of gestures did not result in a uniform movement. The existing framework is successful since the two components are optimised together. Visual computation provides high accuracy at a reasonable processing time, whereas wireless transmission provides close-to-instantaneous delivery with minimal data loss.

The system's overall effectiveness stems from this synergy. The other significant implication of the findings is the scalability of the framework. The trends in the smoothness of the mesh plot indicate that increasing gesture complexity can be done in moderation without a significant drop in system performance. This leaves the chance of adding to the vocabulary of gestures or adding other robotic joints. With a proportional increase in computational resources, the current communication infrastructure seems able to accommodate a larger set of commands. The reliability measures indicate sufficient headroom to execute more complex control scenarios without compromising stability. During the demonstration, the robotic arm performed coordinated tasks, including catching, picking up, and putting down items. This work required timed multi-axis movement, which could be achieved only through precise interpretation of gestures and prompt motor feedback. The fact that the system performs these functions continuously highlights the suitability of the mapping strategy and the sensitivity of the control loop. The viewers remarked that the motion was fluid and continuous, and that there was no apparent delay between the human hand movement and the robot's response. The general analysis proves that three key performance pillars, such as accuracy, latency, and reliability of communication, are all at a high level of maintenance. These three strengths make the system stand out from older methods that usually succeed in only a single aspect, to the detriment of others. The balance among all three gives the framework a credible foundation for real-world gesture-based robotics.

6. Conclusion

This paper presents the design and testing of a real-time robotic control system that can be entirely controlled by human hand gestures, supported by a wireless communication structure. The study demonstrated, using a dataset of 422 gesture instances, that vision-based landmark tracking could achieve high recognition accuracy and be applied to controlling mechanical hardware with high precision. The system demonstrated the ability to read natural hand gestures and translate them into a set of meaningful commands for robots to act on, without the need for wearable sensors or physical controllers. The system's stability and responsiveness were identified during performance evaluations. The timing analysis based on visual inspection revealed that response latency was consistently low, with little fluctuation across gestures and levels of motion complexity. This was necessary so the robotic arm would respond almost immediately to user input, resulting in a smooth, intuitive interactive experience. This responsiveness is necessary in applications that require real-time feedback and accurate motion control. The system's success was also due to wireless performance. Communication tests established that the wireless connection was stable and reliable, with very low packet loss during the testing period. This reliable data transfer ensured constant synchronisation between the processing unit and the robotic hardware, eliminating delays and movement-related side effects. The translation between the visual hand coordinates and the servo motor angles was also linear and predictable, enabling the robotic arm to move in a manner comparable to the user's movements. All in all, with strong vision tracking, low-latency processing, and effective wireless communication, this system can be considered an efficient and easy-to-use solution for touchless robotic manipulation.

6.1. Limitations

Despite the system having good real-time performance and reliability, many constraints were realised during experimentation. The greatest limitation is associated with the environmental lighting. Since the hand-tracking module relies on visual data from a regular camera, drastic changes in lighting may reduce detection accuracy. Bright backlighting can make the hand appear as a silhouette, and severe shadows can lead to inaccurate landmark estimates. These elements introduce noise into the tracking process and may temporarily misclassify gestures. Another challenge is the occlusion. The two-dimensional camera image cannot construct appropriate depth relations between joints when the fingers overlap or when foreign bodies partially cover the hand. This weakness is more pronounced in complex gestures that involve closed fists or crossed fingers. The system can fail to accurately track some landmarks in the absence of reliable depth information, leading it to lose track of them temporarily.

The range of wireless communication is also a constraint on deployment flexibility. The Zigbee-based link performs well within a typical indoor range but falls short of the long-range performance of industrial wireless protocols. The move to warehouses or outdoor operations would demand more networking facilities. The system currently supports only one user. When more than one hand is in the camera shot, the gesture's ownership is confused, which may result in unintended instructions. This constraint constrains multi-operative or collaborative situations. Finally, the mechanical limitations of the robotic arm influence general motion fidelity. The arm has a fixed number of joints and limited degrees of freedom and cannot fully replicate the human hand's articulation. Consequently, complex gestures in the movement are reduced to large mechanical movements.

6.2. Future Scope

There are several opportunities to expand and improve the existing system. Among such upgrades, there is the addition of depth-sensing or stereoscopic vision. These methods would provide three-dimensional spatial information, enabling better landmark detection and addressing problems with occlusion and finger overlap. More accurate mapping between hand posture and robot joint angles would have also been possible with depth information. Advanced machine learning models offer the opportunity to achieve adaptive performance on the software side. With the implementation of user-specific learning, the system had the potential to reduce the need to interpret gestures into specific movement patterns and to increase comfort and recognition accuracy over time. Strategies for continuous learning might also help the system remain resistant to variations in lighting or background. Mesh networking can be used to increase communication capabilities. The distributed wireless architecture would increase the working distance and enable a single control station to coordinate multiple robotic arms simultaneously. This type of scalability would be useful in manufacturing lines, remote maintenance or collaborative robotic settings. The other interesting direction is the incorporation of haptic feedback devices. Physical contact between the robotic arm and the objects could be simulated using wearable elements that provide either vibration or pressure. Such a feedback loop would provide a more interactive and realistic experience of controlling by releasing resistance of weight or contact forces. Lastly, the flexibility of the framework, through modifications to specialised applications, creates new research opportunities. Robotics with gestures would be useful for underwater exploration, hazardous work, or space missions where direct human presence is not possible. Such applications would be facilitated by intuitive remote manipulation, good wireless communication and visual tracking.

Acknowledgement: The authors would like to express their sincere gratitude to Alliance University, Sathyabama Institute of Science and Technology, SRM Institute of Science and Technology at Ramapuram, S.A. Engineering College, Institut Bakti Nusantara, and Dhaanish Ahmed College of Engineering for their valuable academic support, research resources, and encouragement throughout the completion of this work.

Data Availability Statement: Data is available upon request from the corresponding author.

Funding Statement: The authors received no specific funding for conducting this research.

Conflicts of Interest Statement: The authors declare that there are no conflicts of interest regarding the publication of this paper. All cited sources and references have been appropriately acknowledged.

Ethics and Consent Statement: Ethical guidelines were followed throughout the study. Necessary permissions and informed consent were obtained from the relevant institutions and participants involved in the research.

References

1. N. Shafqat, D. J. Dubois, D. Choffnes, A. Schulman, D. Bharadia, and A. Ranganathan, "Zleaks: Passive inference attacks on Zigbee-based smart homes," in Proceedings of the International Conference on Applied Cryptography and Network Security, *Springer*, Cham, Switzerland, 2022.
2. D. G. Akestoridis and P. Tague, "HiveGuard: A network security monitoring architecture for Zigbee networks," *in Proceedings of the IEEE Conference on Communications and Network Security (CNS),* Tempe, Arizona, United States of America, 2021.
3. P. Morgner, S. Mattejat, Z. Benenson, C. Müller, and F. Armknecht, "Insecure to the touch: Attacking ZigBee 3.0 via Touchlink commissioning," *in Proceedings of the 10th ACM Conference on Security and Privacy in Wireless and Mobile Networks,* Boston, Massachusetts, United States of America, 2017.
4. D. G. Akestoridis, M. Harishankar, M. Weber, and P. Tague, "Zigator: Analyzing the security of Zigbee-enabled smart homes," *in Proceedings of the 13th ACM Conference on Security and Privacy in Wireless and Mobile Networks (ACM WiSec),* Linz, Austria, 2020.
5. S. Dowling, M. Schukat, and H. Melvin, "A ZigBee honeypot to assess IoT cyberattack behaviour," *in Proceedings of the 28th Irish Signals and Systems Conference (ISSC),* Killarney, Ireland, 2017.

6. X. Cao, D. M. Shila, Y. Cheng, Z. Yang, Y. Zhou, and J. Chen, "Ghost-in-ZigBee: Energy depletion attack on ZigBee-based wireless networks," *IEEE Internet of Things Journal*, vol. 3, no. 5, pp. 816–829, 2016.
7. F. Sadikin, T. van Deursen, and S. Kumar, "A ZigBee intrusion detection system for IoT using secure and efficient data collection," *Internet of Things*, vol. 12, no. 12, p. 100306, 2020.
8. N. A. Hasan and A. K. Farhan, "Security improve in ZigBee protocol based on RSA public algorithm in WSN," *Engineering and Technology Journal*, vol. 37, no. 3, pp. 67–73, 2019.
9. Y. Wang, C. Chen, and Q. Jiang, "Security algorithm of Internet of Things based on ZigBee protocol," *Cluster Computing*, vol. 22, no. 3, pp. 14759–14766, 2019.
10. F. Farha, H. Ning, S. Yang, J. Xu, W. Zhang, and K. K. R. Choo, "Timestamp scheme to mitigate replay attacks in secure ZigBee networks," *IEEE Transactions on Mobile Computing*, vol. 21, no. 1, pp. 342–351, 2022.
11. S. M. S. Rana, M. A. Halim, and M. H. Kabir, "Design and implementation of a security improvement framework of ZigBee network for intelligent monitoring in IoT platform," *Applied Sciences*, vol. 8, no. 11, pp. 1–15, 2018.
12. H. Pirayesh, P. K. Sangdeh, and H. Zeng, "Securing ZigBee communications against constant jamming attack using neural network," *IEEE Internet of Things Journal,* vol. 8, no. 6, pp. 4957–4968, 2021.
13. X. Wang and S. Hao, "Don't Kick Over the Beehive: Attacks and Security Analysis on Zigbee," *in Proceedings of the ACM SIGSAC Conference on Computer and Communications Security (CCS*), Los Angeles, California, United States of America, 2022.
14. A. Zohourian, S. Dadkhah, E. C. P. Neto, H. Mahdikhani, P. K. Danso, H. Molyneaux, and A. A. Ghorbani, "IoT Zigbee device security: A comprehensive review," *Internet of Things*, vol. 22, no. 7, p. 100791, 2023.
15. M. S. Wara and Q. Yu, "New replay attacks on Zigbee devices for Internet-of-Things (IoT) applications," *in Proceedings of the IEEE International Conference on Embedded Software and Systems (ICESS)*, Shanghai, China, 2020.