REAL-TIME RECOGNITION OF ROTATIONAL HAND MOVEMENTS USING MEDIAPIPE AND OPENCV

Yenni Astuti[1*], Maria Asumpta Deny Kusumaningrum[2], Anggraini Kusumaningrum[3], Riani Nurdin[4], Mayang Sativa[5], Nur Priyanto Poupon[6], Tutiaditama[7]

[1,2,5]Department of Electrical Engineering, Institut Teknologi Dirgantara Adisutjipto, [3]Department of Informatics,[6]Institut Teknologi Dirgantara Adisutjipto, [4]Department of Industrial Engineering, [7]Indonesian National Military Air Force;

**corresponding author: yenniastuti@itda.ac.id*

ABSTRAK *This study proposes a real-time gesture recognition system for detecting rotational hand movements (pitch, yaw, and roll) using MediaPipe and OpenCV for extracting features and classifying gestures. The method extracts angular features from hand landmarks and classifies gestures via a lightweight rule-based approach. The system was tested under twelve environmental scenarios, combining four distances (50 – 450 cm) and three lighting conditions, with 1,500 gesture samples from 10 participants. Results show 100% accuracy at 50 cm and 200 cm across all lighting conditions, with up to 93.3% accuracy at 300 cm, surpassing similar rule-based systems in performance. These findings demonstrate a low-cost, contactless solution for interactive systems.*

Keywords *real-time gesture recognition, rotational hand movements, MediaPipe and OpenCV.*

1. INTRODUCTION

Hand gesture recognition has become a key enabler in enhancing humancomputer interaction, offering intuitive, natural, and contactless control across domains such as robotics, virtual reality, and smart control systems. The recent emergence of open-source computer vision tool sparticularly Media Pipe and OpenCV has significantly accelerated progress in real-time gesture recognition. MediaPipe, developed by Google, offers robust realtime palm landmark detection detection, while OpenCV provides versatile image processing capa bilities, enabling effective visual feedback and seamless system integration. Numerous recent studies have explored the integration of MediaPipe and OpenCV in real-time gesture recognition applications. For instance, an AI-based fitness pose counter demonstrated accurate human pose estimation at varying distances using both tool. Another study proposed a deep learning approach that combined MediaPipe land marks with OpenCV for recognizing ten distinct gestures, showing promise for immersive applications such as AR/VR, sign language translation, and interactive control systems. Media Pipe has also been used to support virtual interaction in educational settings, allowing children to engage in learning activities via hand gestures captured by a standard camera an especially valuable approach during pandemic-related restric tions. Further applications include gesture-based control of an electric wheelchair, where the system enabled intuitive and contactless navigation under varying lighting and background conditions, highlighting its potential in assistive technologies. In another

case, MediaPipe was integrated into a Rehabilitation Internet of Things (RIoT) framework for stroke therapy, allowing accurate gesture detection under diverse environmental conditions. Adaptive calibra tion, deep learning enhancements, and real-time feedback improved recognition perfor mance and patient engagement. Other studies have explored the combination of 3D hand pose estimation and deep learning to improve gesture recognition in dynamic environ ments, including hand shape estimation from RGB images, recognition of both static and dynamic gestures, and precise 3D tracking to enhance human–robot interaction. Despite these advances, most existing systems focus on static gestures or the classification of discrete symbols, often ignoring spatial orientation and continuous rotational movements. There is a clear gap in recognizing three-dimensional orientation changes specifi cally pitch, yaw, and roll which are essential for applications such as robotic control and immersive user interfaces. Moreover, environmental factors such as user distance from the camera and lighting conditions are frequently overlooked, even though they critically affect detection accuracy in real-world use. This study addresses these limitations by proposing a novel system to recognize pitch, yaw, and roll gestures. The system computes angular features between landmark vectors detected by MediaPipe and processed in OpenCV. In contrast to prior methods, this system is tested under four distances and three lighting levels to assess its robustness and real-world applicability. This study aims to develop and validate a lightweight gesture recognition system using a standard RGB webcam, MediaPipe, and OpenCV to detect rotational hand movements. Key contribu tions include: (1) a novel method for real-time angular gesture detection, (2) an analysis of environmental robustness across lighting and distance scenarios, and (3) validation of its potential for non-contact control in robotics and smart systems.

2. METHODOLOGY

This study developed a real-time hand gesture recognition system designed to classify rotational hand movements specifically pitch, yaw, and rollusing the MediaPipe framework and OpenCV library. The system was built to operate with a standard RGB webcam and was implemen ted using Python programming.

2.1. System Overview

The proposed system consists of four primary stages: image acquisition, hand landmark detection, feature extraction, and gesture classification. During runtime, an RGB camera captures the user's hand motion. MediaPipe's pre-trained hand-tracking model processes the video stream to detect and track 21 key hand landmarks (see Figure 1) in real time. These land marks represent anatomically significant points, such as the wrist, fingertips, and joints. Figure 2 illustrates the overall system process, comprising five sequential stages: image acquisition, hand-landmark detection, feature extraction, gesture decision, and output display.

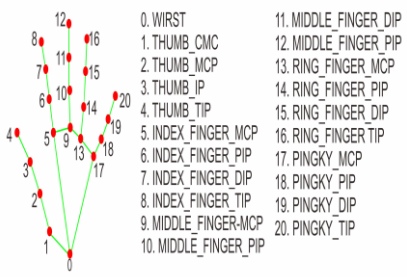


Figure 1. Hand landmark used for feature extraction [13]

To detect and differentiate rotational hand gestures, the system computes directional vectors between specific hand landmarks detected by MediaPipe. These vectors are used to calculate angular movements that represent the orientation of the hand in three-dimensional space. The gesture classification focuses on three types of rotation: pitch, yaw, and roll. Pitch refers to vertical movement of the hand that reflects rotation along the X-axis, such as nodding motions. Yaw corresponds to horizontal rotation along the Y-axis, typically observed as left or right turns of the hand. Roll represents rotational movement around the Z-axis, often caused

by wrist twisting. To quantify these motions, the system tracks the spatial relationship among landmark coordinates especially those involving the wrist, index fingertip, and thumb tip and calculates angular displace ment using trigonometric functions. The temporal variation of these angles over successive video frames provides a basis for determining the type of gesture being performed. Instead of complex neural networks, this system uses empirically derived rule-based thresholds to classify the gestures. The classification determines whether angular changes fall within empirically defined thresholds to determine whether a pitch, yaw, or roll motion has occurred. This approach ensures the system remains lightweight and responsive. The angular thresholds for classifying pitch, yaw, and roll gestures were determined through an initial trial involving five participants. Each participant performed each gesture ten times at each distance under each lighting mode. The resulting angular values were analyzed to define empirical threshold ranges for each gesture class.

2.2. Experimental Setup

To evaluate system performance, testing was conducted under control of variations in environmental conditions. Four distances were tested: 50 cm, 200 cm, 300 cm, and 450 cm between the user and the camera. Additionally, lighting conditions were adjusted to represent three categories: bright, moderate, and dim. The evaluation involved multiple repetitions of each gesture class, recorded and processed to determine detection accuracy. All experiments were conducted using an Acer Aspire 3 A314-22 laptop, equipped with an AMD Ryzen 3 3250U processor, 4 GB RAM, and a built-in HD webcam. The built-in webcam captures 1280×720 resolution at 30 fps. No external sensors, webcams, or depth cameras were used in this study. The experimental scenario is shown in Figure 3. The dataset comprised of three gesture types under three lighting conditions, with each gesture repeated five times by ten participants. The code for this experiment is built using Python 3.11.3 version.

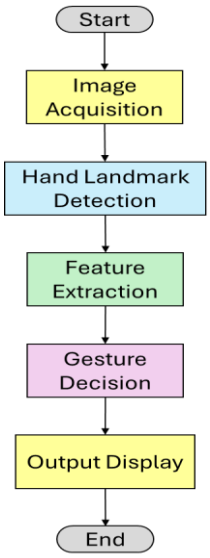


Figure 2. Flowchart of the proposed system

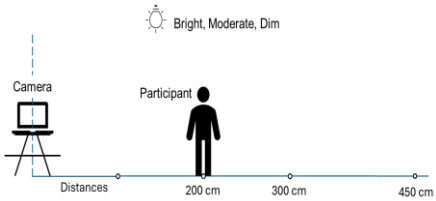


Figure 3. Experimental scenario

Accuracy was calculated as correct predictions divided by total attempts (Equation 1). A total of 1,500 gesture samples were recorded, with 10 participants each performing three gesture types—pitch, yaw, and roll—under four distances (50 cm, 200 cm, 300 cm, 450 cm) and three lighting conditions (bright, moderate, dim). Each gesture was repeated five times per condition by each participant.

$$Accuracy\ (\%) = \frac{Number\ of\ correct\ predictions}{Total\ gesture\ attempts} \times 100\% \qquad (1)$$

3. RESULTS AND DISCUSSION

The proposed system was evaluated based on its ability to recognize three types of rotational hand gestures pitch, yaw, and rollunder different lighting conditions and camera distances. Testing was conducted by ten participants, each performing the gestures multiple times at four distances (50 cm, 200 cm, 300 cm, 450 cm) and under three lighting intensities (bright, moderate, dim).

3.1. Gesture Detection Accuracy

Table 1 summarizes the average recogni tion accuracy of each gesture across all lighting and distance variations. The highest average accuracy (100%) was consistently achieved at 200 cm, regar dless of lighting conditions. At 300 cm, accuracy remained high, especially for yaw and roll gestures. In contrast, 450 cm showed the most degradation in recognition, particularly in bright lighting, where all gestures failed to be detected. The reported accuracy values represent the aggregate performance across all repetitions and participants for each condition. While standard deviation values were not computed explicitly, consistent results were observed, particularly for yaw and roll gestures at close and medium distances. Although standard deviation values were not computed, future studies should include statistical dispersion to strengthen reproducibility and confidence in the results.

Distance	Lighting	Pitch (%)	Yaw (%)	Roll (%)	Average (%)
50 cm	Bright	100	100	100	100
200 cm	Bright	100	100	100	100
300 cm	Bright	80	100	100	93.3
450 cm	Bright	0	0	0	0
50 cm	Moderate	100	100	100	100
200 cm	Moderate	100	100	100	100
300 cm	Moderate	100	100	100	100
450 cm	Moderate	60	100	100	86.7
50 cm	Dim	100	100	100	100
200 cm	Dim	100	100	100	100
300 cm	Dim	0	80	100	60
450 cm	Dim	0	60	60	40

Table 1. Average accuracy of gesture recognition by distance and lighting

3.2. Influence of Distance and Lighting

The system achieved optimal performance at 50 cm and 200 cm under all lighting conditions. As distance increased, perfor mance dropped especially in the pitch gesture, which showed the most sensitivity to poor lighting and greater distance. The accuracy decline at 450 cm suggests that hand landmarks became too small or unclear for MediaPipe to reliably detect, particularly under bright or dim lighting where overex posure or insufficient contrast can occur. At 300 cm, performance remained strong, notably for yaw and roll. However, pitch recognition under dim lighting completely failed (0%), while yaw and roll still maintained usable accuracy.

3.3. Error Patterns and Gesture Robustness

Among all the test results, the yaw gesture demonstrated the most consistent perfor mance, achieving high accuracy across nearly all environmental conditions except at 450 cm under bright lighting, where all gesture recognitions failed. The roll gesture also exhibited strong reliability, with only a moderate decline in accuracy under dim lighting at the farthest distance. In contrast, the pitch gesture characterized by vertical hand rotation proved to be the most error-prone. The system showed a marked decline in detection accuracy at 300 cm under dim lighting, with complete failure observed across all lighting conditions at 450 cm. The reduced performance likely stems from vertical movements causing occlusion or reduced angular contrast between land marks. Yaw and roll, on the other hand, produce more pronounced lateral motion that is easier for 2-D landmark-based systems like MediaPipe to detect, even at greater distances. These results indicate that detec tion accuracy is strongly influenced by gesture type, with more visually prominent motions being more reliably recognized. Compared to deep learning-based gesture recognition methods, the proposed system achieves comparable accuracy for yaw and roll gestures under controlled conditions while using significantly fewer computational resources. For instance, reported over 95% accuracy in indoor settings with CNN models, whereas our system achieved 100% accuracy at 200 cm across all lighting levels using only rule-based classification. Howe ver, unlike the deep learning-based methods which may generalize better across gesture types, our system showed perfor mance limitations in detecting pitch gestures at longer distances and dim lighting.

3.4. System Implications

The experimental results confirm that the proposed rule-based system is capable of accurately recognizing three dimensional

rotational hand gestures using only an RGB webcam and angular analysis of hand landmarks. Its simplicity and efficiency make it particularly suitable for low-cost, real-time applications such as touchless human computer interaction in educa tional, assistive, or smart control environments. The system performed optimally at close to medium distances, demonstrating reliable accuracy under various lighting conditions, especially for yaw and roll gestures. How ever, the system's performance is highly dependent on environmental factors. Gesture recognition declined significantly over longer distances, especially under extreme lighting conditions both overly bright and dim. These findings highlight the limitations of relying solely on MediaPipe's 2D landmark detec tion, which struggles with small or poorly lit features at greater ranges. In addition, gesture-specific challenges were observed. Pitch gestures proved to be less robust compared to yaw and roll, due to factors such as partial hand occlusion and reduced angular differen tiation in the camera's field of view. To improve the system's robustness and generalizability, future work could incorporate adaptive thresholding, real-time landmark confidence scoring, or lightweight machine learning techniques to complement the rule-based approach. These enhance ments could help compensate for environ mental variability and user-specific differen ces, making the system more reliable for broader applications.

4. CONCLUSION AND RECOMMENDATIONS

4.1. Conclusions

- *This study has introduced and evaluated a real-time hand gesture recognition system designed to identify rotational hand movements namely pitch, yaw, and rollusing only a standard RGB webcam, MediaPipe, and OpenCV.*

- *The system employs a rule-based approach by extracting angular features from hand landmarks, eliminating the need for additional sensors or deep learning models. This lightweight and efficient method enables responsive gesture detection suitable for low-cost, non-contact applications. Experimental results demonstrated that the system achieved perfect recognition accuracy (100%) for all gesture types at 50 cm and 200 cm across all lighting conditions. At 300 cm, the system maintained high performance, particularly under bright and moderate lighting, although pitch gestures showed reduced accuracy under dim conditions.*

- *Performance declined significantly at 450 cm, with pitch gestures being the most affected, especially under bright lighting. In contrast, yaw and roll gestures remained consistently robust, likely due to their more prominent lateral motion and better landmark visibility in the camera's view. These findings validate the effecti veness of the proposed rule-based system for recognizing 3D rotational gestures in constrained environments.*

- *The study contributes a simple yet reliable solution for gesture-controlled interfaces that can be deployed in education, assistive technology, and smart systems without requiring computationally expen sive models.*

- *Future development will aim to improve the system's adaptability across users and environmental conditions. Enhan cements such as dynamic thresholding, landmark confidence weighting, and the integration of lightweight machine learning techniques are proposed to strengthen gesture classification under uncertainty. Moreover, expanding the gesture set and supporting continuous gesture sequences will further enhance the system's relevance for real-world interactive applications.*

4.2. Recommendations

- *Uses MediaPipe and OpenCV to extract features and classify gestures. This method extracts angular features from key hand points and classifies gestures*

using a lightweight, rule-based approach.

- *This method is particularly noteworthy because it is inexpensive and even relatively cost-effective, making it highly effective.*

- *Acknowledgments. The authors would like to express their sincere gratitude to the institute for research and community service (lppm) of institut teknologi dirgan tara adisutjipto for funding and supporting this research through the 2025 internal research grant program (no.011/k2.1/ lppm/vi/2025).*

5. REFERENCES

[1] Amanda Muchsin Chalik, Bilal Abdul Qowy, Faiz Hanafi, & Ahlijati Nuraminah. (2021). Mouse Tracking Tangan dengan Klasifikasi Gestur Menggunakan OpenCV dan Mediapipe. Jurnal Ilmiah Teknik Informatika Dan Komunikasi, 1(2), 10–18. https://doi.org/10.55606/juitik.v1i2.323

[2] Amprimo, G., Ferraris, C., Masi, G., Pettiti, G.,&Priano, L.(2022). GMH-D: Combining Google MediaPipe and RGB-Depth Came ras for Hand Motor Skills Remote Assess ment.2022 IEEE International Conferen ce on Digital Health (ICDH), 132–141. https://doi.org/10.1109/ICDH55609.2022. 00029

[3] Anand, A., Pandey, A., Mehndiratta, S., & Rathore, H. G. P. K. R. (2023). Smart AI based Volume Control System: Gesture Recognition with OpenCV & Mediapipe Integration. 023 International Conference on Smart Devices (ICSD), 1–6. https://doi. org/ 10.1109/ICSD60021.2024.10751362

[4] Avola, D., Cinque, L., Fagioli, A., Foresti, G. L., Fragomeni, A., & Pannone, D. (2022). 3D hand pose and shape estimation from RGB images for keypoint-based hand gesture recognition. Pattern Recognition,129.https://doi.org/10.1016/j. patcog.2022.108762

[5] Arif, M., Haryono, G. S., Arsyad, N. F., Ramadhani, R., Sahid, A., & Rosyani, P. (2024). Hand Detection System Based on MediaPipe and OpenCV for Gesture Recognition.

[6] Biswas, S., Saw, R., Nandy, A., & Naskar, A. K. (2025). Attention-enabled hybrid convolutional neural network for enhan cing human–robot collaboration through hand gesture recognition. Computers and ElectricalEngineering,123(PartA),110020 https://doi.org/10.1016/j.compeleceng.20 24.110020

[7] Budiman, S., et al. (2022). Finger Gesture Recognition for Volume Control on Computers Using OpenCV and Media pipe. Antivirus Journal, 16(1), 1-10.

[8] Dhuzuki, N.H.M., Zainuddin, A.A., Zaman, N.A.S.K., Razmi, A.N.M.A., Kaitane, W.S., Puzi, A.A., Johar, M.N., Yazid, M., Nordin, N.A.M., Sidek, S.N., & Zaki, H.F.M. (2025). Design and Implementation of a Deep Learning Based Hand Gesture Recog nition System for Rehabilitation Internet-of-Things (Riot) Environments Using Mediapipe. IIUM Engineering Journal, 26(1),353–372.https://doi.org/10.31436/ IIUMEJ.V26I1.3455

[9] Gao, Q., Chen,Y., Ju,Z.,&Liang,Y. (2022). Dynamic Hand Gesture Recog nition Based on 3D Hand Pose Estimation for Human-Robot Interaction. IEEE Sen sors Journal,22(18),17421–17430. https://doi. org/10.1109/JSEN.2021.3059685

[10] Gil-Martín, M., Marini, M. R., Martín-Fernández, I., Esteban-Romero, S., & Cinque, L. (2025). Hand Gesture Recog nition Using MediaPipe Landmarks and Deep Learning Networks. Interna tional Conference on Agents and Artificial Intelli gence, 3(Icaart), 24–30.https://doi.org/ 10.5220/0013053500003890

[11] Han, J. S., Lee, C. I., Youn, Y. H., & Kim, S. J. (2022). A Study on Real-time Hand Gesture Recognition Technology by Machine Learning-based MediaPipe.

Journal of System and Management Sciences, 12(2), 468–482.https://doi.org/ 10.33168/JSMS.2022.0225

[12] Huda.,M.R.,Ali.,M.L.,&Sadi, M.S. (2025). Developing a real-time hand-gesture recognition technique for wheel chair control. Plo SONE,20(4 April). https://doi. org/10. 1371/journal.pone.0319996

[13] Irfan, I., Abdul muthalib, M., Kartika, K., & Meliala, S. (2023). Pengiraan Pose Model Manusia Pada Repetisi Kebugaran Ai Pemograman Python Berbasis Kompu terisasi [Human Pose Estimation for Fitness Repetitions Using AI-Based Python Programming]. Infotech Journal, 9(1),11–19. https://doi.org/10.31949/info tech. v9i1.4233

[14] Kavitha, M. N. Saranya, S. S. Pragat heeswari, E. Kaviyarasu, S. Ragunath, N., & Rahul, P. (2024). A real-time hand-gesture recognition using deep learning techniques. Proc. Int. Conf. Artif. Intell. Smart Energy (ICAIS), 489–500. https:// doi.org/10.1007/978-3-031-61471-2_37

[15] Mazhar, O., Ramdani, S., & Cherubini, A. (2021). A Deep Learning Framework for Recognizing Both Static and Dynamic Gestures. Sensors, 21(6), 1–20. https://doi.org/10.3390/s21062227

[16] Odeh, A., & Odeh, N. (2024). OpenCV and its Applications in Artificial Intelligent Systems. 2024 International Conference on Intelligent Computing, Communication, Networking and Services (ICCNS), 242–249.https://doi.org/10.1109/ICCNS62192. 2024.10776047

[17] Oktaviyanti,D., Nugroho, A., & Suni, A.F. (2022). Pemanfaatan Hand Tracking untuk Membuat Program Virtual Painter sebagai Alternatif Menggambar Digital [Utilizing Hand Tracking to Develop a Virtual Painter Program as a Digital Drawing Alternative].Petir,15(2),287–294. https://doi.org/10.33322/petir.v15i2.1523

[18] Oudah, M., Al-Naji, A.,& Chahl, J.(2020). Hand Gesture Recognition Based on Computer Vision:A Review of Techni ques. In Journal of Imaging (Vol.6,Issue, 8).https://doi.org/10.3390/jimaging60800 73

[19] Owda, N. A. A. Y., & Owda, M. (2025). Real-Time Detection and Analysis of Facial Movement Disorders Using MediaPipe and OpenCV: A Streamlit-Based Framework. 2025 12th Interna tional Conference on Information Techno logy (ICIT),13–18.https://doi.org/10.1109/ ICIT64950.2025.11049234.

[20] Palani, P., Panigrahi, S., Jammi, S. A., & Thondiyath, A. (2022). Real-time Joint Angle Estimation using Mediapipe Frame work and Inertial Sensors. 2022 IEEE 22nd International Conference on Bioin formatics and Bioeng ineering (BIBE), 128–133.https://doi.org/10.1109/BIBE55 377.2022. 00035.

[21] Prabowo, R., et al. (2024). Implemen tation of a Virtual Mouse Based on Hand Gesture Recognition Using MediaPipe and OpenCV. Scientific Journal of Informatics and Communi cation Engineering (JUITIK), 1(2), 10-18.

[22] Quiñonez, Y., Lizarraga, C., & Aguayo, R. (2022). Machine Learning solutions with MediaPipe. 2022 11th International Conference On Software Process Impro vement (CIMPS), 212–215. https://doi.org/10.1109/CIMPS57786.202 2.10035706

[23] Santoso, B. (2023). Hand Gesture Recog nition for Human-Computer Interaction (Case Study: Rotation and Position Recognition). National Seminar on Infor mation Technology Applications (SNATI)

[24] Suhartono, & Ramadhan, F. (2025). Hand Gesture Recognition for Presenta tion Slide Control Using Mediapipe Framework, OpenCV and LSTM Model. JUPTI: Indonesian Journal of Education

and Technology, 4(1), 100-110.

[25] Srinivasan, R., Kavita, R., Kavitha, M., Mallikarjuna, B., Bhatia, S., Agarwal, B., Ahlawat, V., & Goel, A. (2023). Python And Opencv For Sign Language Recognition. 2023 International Confe rence on Device Intelligence, Computing and Communication Techno logies, (DICCT),1–5.https://doi.org/10.1109/DIC CT56244.2023.10110225

[26] Tao, X., & Ko, S. Y. (2024). Tilt-X: Development of a Pitch-Axis Tiltrotor Quadcopter for Maximizing Horizontal Pulling Force and Yaw Moment. Applied Sciences (Switzerland), 14(14). https://doi.org/10.3390/app14146181

[27] Tomar, Y., Himanshu, Devi, S., & Kaur, H. (2023). Human Motion Tracker using OpenCv and Mediapipe. 2023 3rd Inter national Conference on Innovative Mecha nisms for Industry Applications (ICIMIA), 1199–1204.https://doi.org/10.1109/ICIMI A60377.2023.10425865

[28] Vijaya, J., Singh, A., Singh, K., & Kumar, A. (2024). AI-Enabled Real-Time Exercise Monitoring with MediaPipe and OpenCV. 2024 15th International Conference on Computing Communication and Networ king Technologies (ICCCNT), 1–5. https://doi.org/10.1109/ICCCNT61001.20 24.10724252