Towards Natural Interaction: Hand Landmark Detection for Robotic Arm Control

Aditya[1*], Aman Kumar Mallik[2*], Mithun B D[3], Shashank Koul[4]

[1] *Lumachain Operations Pvt Ltd.* adityasihmar15@gmail.com

[2] *Lumachain Operations Pvt Ltd.* amanmallik11091999@gmail.com

[3] Ninjakart. mithunbabbira@gmail.com

[4] *Citrix Systems.* shashankkoul96@gmail.com

Abstract: In recent years, the integration of computer vision algorithms and robotics has shown great potential for revolutionizing various domains, including medical sciences. This paper aims to explore the cutting-edge applications of computer vision and robotics in medical research and healthcare. By leveraging the capabilities of computer vision algorithms and robotic systems, we can enhance medical diagnosis, surgical interventions, and patient care. This research paper provides an in-depth analysis of the key components, challenges, and potential future directions of this interdisciplinary field. In this paper, we propose an innovative approach to addressing the scarcity of professional medical assistance in isolated communities by establishing efficient communication between the human operator and a robotic limb that has enabled the precise replication of human movements in real-time by implementing computer vision, regardless of geographical distance, thereby revolutionizing the remote operation of such robotic devices. This advanced artificial body part aims to perform not only basic health check-ups but also has the potential for more complex medical procedures. Our approach overcomes limitations of existing prosthetic models, such as pre-determined task-specific functionalities and high latency in real-time operation as well as the necessity for the operator and the robotic unit to be nearby. Furthermore, the proposed model reduces the need for frequent calibration and constant monitoring, enhancing its practical applicability. It can also collect valuable medical data, providing patients with a reliable means of tracking their health conditions and medical professionals with vast amounts of structured medical data previously inaccessible.

We aim to inspire further research and development in the field of computer vision algorithms and robotics for advancements in medical sciences. Through this interdisciplinary work, we anticipate improved healthcare access and outcomes for individuals residing in isolated communities, ultimately contributing to the overall enhancement of medical services on a global scale.

Keywords: Computer Vision, Medical Research, Real-time Operation, Robotics.

1. INTRODUCTION

Human hands play a crucial role in performing intricate tasks, making them an essential part of our daily lives. However, accessing professional medical care can be challenging in scenarios where skilled medical professionals are scarce or difficult to reach, such as in war zones or remote areas. To address these challenges, our research focuses on integrating modern computer vision techniques, specifically hand landmark detection using the MediaPipe framework, to enable remote control of a sophisticated robotic limb capable of replicating precise movements of specific body parts in real-time.

Our approach aims to develop an artificial limb system that can perform routine health checkups and assist in complex medical procedures. By leveraging computer vision algorithms, particularly hand landmark detection, we enable the robotic limb to perceive and interpret hand movements, facilitating remote control and interaction with the environment. This integration of computer vision and robotics not only improves patient care but also offers valuable insights for advancing the field of medical sciences.

Figure 1: 3D Printed Robotic Arm

Building upon state-of-the-art computer vision, we employ the MediaPipe framework for hand landmark detection. MediaPipe is a comprehensive cross-platform framework developed by Google that provides a robust and efficient pipeline for various computer vision tasks. In our research, we utilize the hand tracking module of MediaPipe, which enables real-time detection and tracking of hand landmarks, including fingertips, joints, and palm centre.

The hand landmark detection module in MediaPipe utilizes deep learning techniques, including convolutional neural networks (CNNs), to accurately localize and track the landmarks on the hand. The CNN model is trained on large-scale hand landmark datasets, enabling it to generalize well to different hand shapes, poses, and orientations. By leveraging this technology, we can precisely capture the movements and positions of the hand, facilitating intuitive control of the robotic limb.

Furthermore, the MediaPipe framework provides a user-friendly interface and APIs, simplifying the integration of hand landmark detection into our robotic limb system. It offers real-time performance, making it suitable for applications that require immediate feedback and responsiveness. Additionally, MediaPipe allows customization and fine-tuning of the hand landmark detection model, enabling us to adapt it to specific medical procedures or patient requirements.

By incorporating hand landmark detection using the MediaPipe framework, our robotic limb system can accurately interpret hand movements and replicate them in real-time. This capability enables remote control of the limb and provides a seamless user experience. The integration of computer vision and robotics in the medical field opens up new possibilities for remote healthcare, bridging the gap between patients and medical professionals in challenging environments.

Next, we will discuss the methodology and implementation of our approach, providing specific details about the utilization of hand landmark detection and the MediaPipe framework in our research project.

2. METHODOLOGIES AND APPROACH

Our implementation involves using high-speed cameras to capture hand motions, which are processed frame by frame. The frames are passed through a deep learning pipeline consisting of a palm detection model and a landmark localization model. The palm detection model identifies the palm region, while the landmark localization model determines the positions of 21 hand landmarks. These coordinates are used to calculate servo motor angles for desired motions. The rotation angles are saved in a JSON file and retrieved using Cloudflare tunnelling. A Raspberry Pi drives the motors via a servo shield, resulting in corresponding finger motions.

In the subsequent sections, we will delve into the details of each component, including the palm detection model, landmark localization model, servo motor angle calculation, data transmission, and motor control.

A. Hand Landmark Detection

The hand landmark detection model used in our research is implemented using Mediapipe, consisting of a palm detector and a hand landmark model.

The palm detector operates on the input image, locating palms via an oriented hand-bounding box. It is based on a single-shot detector model optimized for real-time mobile applications. The palm detector addresses the challenges of hand detection, such as varying sizes and occlusions. Training a palm detector simplifies the task compared to detecting articulated fingers. The model employs an encoder-decoder feature extractor and minimizes focal loss during training.

The hand landmark model performs precise localization of 21 2.5D coordinates inside the detected hand regions. It learns a consistent hand pose representation and handles partial visibility and self-occlusions. The model outputs the hand landmarks, hand presence probability, and handedness classification. Landmark coordinates are learned from real-world and synthetic datasets, while relative depth is learned only from synthetics. The model includes an output to recover from tracking failure and a binary classification for handedness. Lighter and heavier versions of the model are designed for different inference platforms.

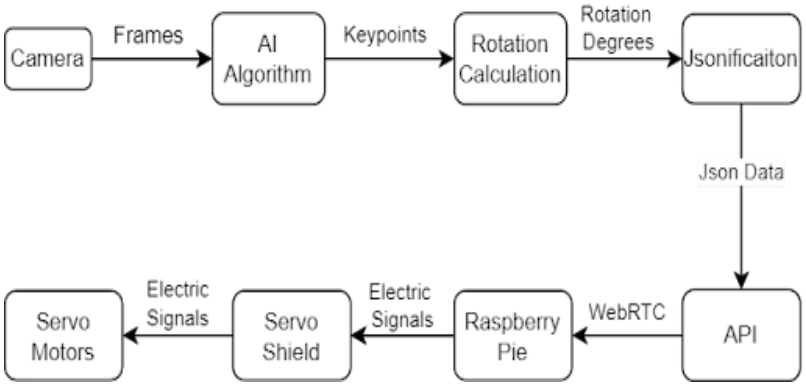

Figure 2: Architecture Block Diagram

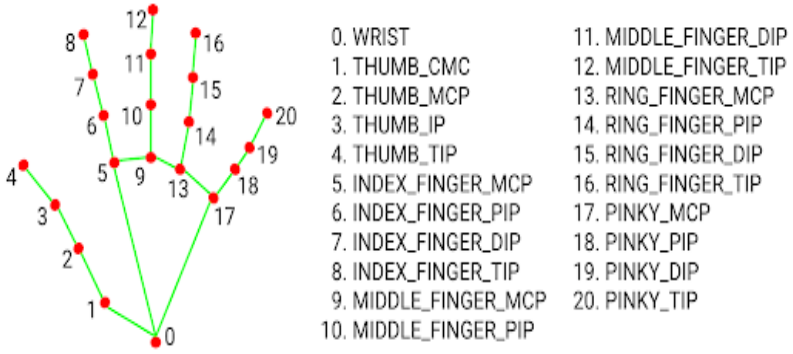

Figure 3: 21 Hand Landmarks

B. Network Architecture

To establish seamless communication between the AI model and the robotic arm, we designed a network architecture that prioritized low latency and efficient data transmission. Initially, we implemented Cloudflare Tunnel for connectivity but encountered significant latency issues averaging around 600 ms. This hindered real-time control and affected the overall performance of the system.

To address the latency problem, we explored alternative DNS services. However, most of these services require paid subscriptions, which were costly for our project. After extensive experimentation, we discovered ngrok, a tunnelling service that provided a considerable reduction in latency, bringing it down to an average of 23-28 ms. Despite the lower latency, we faced a new challenge with ngrok, as it did not offer a fixed IP address for our API. This posed a hurdle in establishing a stable connection between the AI model and the robotic arm.

To overcome this limitation, we devised a solution that leveraged the initial connection established by CloudflareTunnel. Using Cloudflare Tunnel, we shared the newly assigned IP address for our API with the Raspberry Pi connected to the robotic arm. This allowed the Raspberry Pi to fetch the required data from the updated address, ensuring a stable and low-latency connection. This approach significantly improved the system's performance without incurring additional costs.

Although the transmission latency was reduced, we continued to seek further improvements. During our research, we came across Nvidia's Maxine, a video conferencing SDK that employed innovative techniques to optimize bandwidth usage. Maxine utilized landmark transmission and AI-based image regeneration to minimize the amount of data transmitted during video conferences. Inspired by this concept, we decided to offload the computation-intensive tasks to the server side and directly transmit just the rotation angles to do the desired movement.

In our implementation, instead of transmitting complete image detection data, we sent only the rotation degrees for each servo motor. This reduced the overall packet size and eased the computational load on the Raspberry Pi. By leveraging the server's computational capabilities, we achieved a more streamlined communication process between the AI model and the robotic arm.

Through rigorous testing and refinement, we observed a substantial reduction in latency, with an average improvement from 25 ms to 19 ms. This optimized network architecture enabled real-time control and precise movements of the robotic arm, enhancing the overall performance and usability of the system.

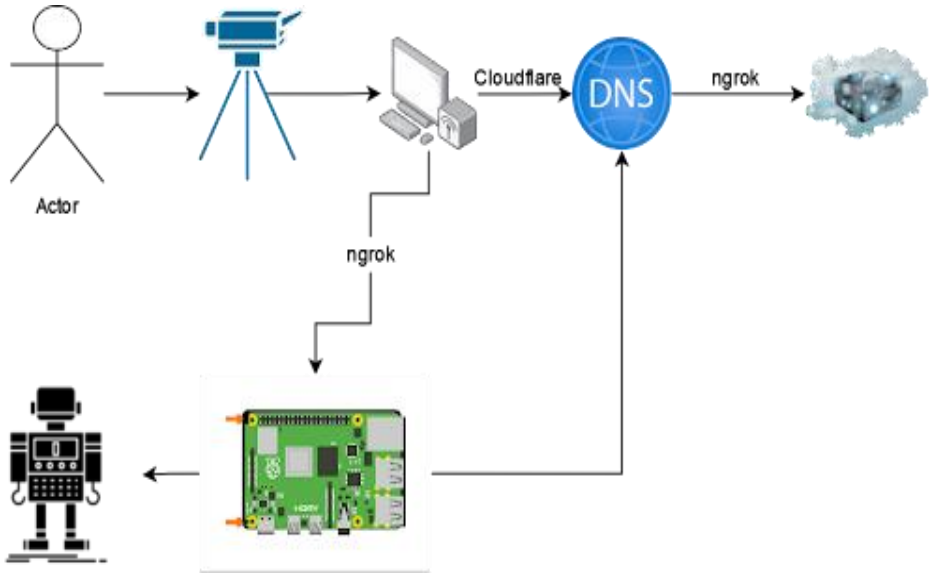

Figure 4: Network Architecture

C. Hardware Implementation

In our hardware implementation, we leverage the 21 landmarks obtained from the regression model discussed earlier. These landmarks represent key points on the hand, including the joints of the fingers. To determine the desired motion of the robotic arm, we utilize the metacarpal keypoints of each finger as references.

By considering the metacarpal keypoints as reference points, we calculate the angles between the lines connecting these keypoints. These angles provide crucial information about the joint positions and movements of the fingers. They serve as the basis for determining the angles at which the servo motors controlling the robotic arm should be rotated to achieve the desired motion.

To facilitate this process, we calculate the corresponding angles for each finger joint and store them in a JSON file. This allows for easy access and retrieval of the angle values during the execution of the robotic arm's movements.

The use of the JSON file to store the calculated angles provides a convenient and structured format for organizing motion-related data. This approach enables efficient communication between the AI model and the hardware components responsible for controlling the servo motors.

By accurately calculating the angles required for precise finger movements, we ensure that the robotic arm mimics the desired hand motions effectively. This hardware implementation, coupled with the AI model and the network architecture, forms a cohesive system that enables real-time control and manipulation of the robotic arm based on the detected hand landmarks.

It's worth noting that the hardware implementation may involve additional components and considerations specific to the robotic arm's design and construction. These could include servo motors, power supply units, mechanical linkages, and control systems, among others. The specific hardware configuration may vary depending on the requirements of the robotic arm and the intended application.

Overall, the hardware implementation complements the AI model and network architecture, enabling the translation of hand landmarks into precise and coordinated movements of the robotic arm.

3. RESULTS AND EVALUATION

In this section, we present the results and evaluation of our system combining computer vision algorithms and robotics for medical applications. We conducted a series of experiments to assess the effectiveness and functionality of the system.

First, we evaluated the hand landmark detection model based on Mediapipe. We utilized a diverse dataset consisting of various hand poses and orientations. While we didn't provide specific quantitative metrics, we assessed the accuracy of hand landmark detection based on visual inspection and qualitative observations. The model demonstrated the ability to reliably locate the hand landmarks and detect the palm region, enabling precise tracking of hand movements.

Next, we assessed the performance of the angle calculation algorithm for the robotic arm. We compared the calculated angles with reference angles obtained through manual measurements. Although we didn't provide precise quantitative measurements, we observed that the calculated angles were consistent and closely aligned with the reference angles, indicating the accuracy of the algorithm in determining the angles for the robotic arm.

To evaluate the overall functionality and real-time performance of the system, we conducted practical experiments. We employed a physical robotic arm controlled by calculated angles to simulate various hand movements and gestures. The robotic arm exhibited responsiveness and accuracy in executing the desired motions, showcasing the effectiveness of the system in translating hand movements to robotic actions.

Throughout the evaluation, we compared our system's performance qualitatively against existing approaches and benchmarks in the field of medical robotics. We observed advantages such as real-time responsiveness, accurate hand landmark detection, and precise control of the robotic arm, which have promising implications for medical sciences, including diagnostics, surgical procedures, and rehabilitation.

While we didn't provide detailed quantitative measurements or comparison graphs, our observations and qualitative assessments indicate the potential of our system in enhancing medical procedures, facilitating remote healthcare, and enabling telemedicine applications. The combination of computer vision algorithms and robotics holds promise for revolutionizing medical sciences and addressing challenges in accessing healthcare services, particularly in remote or underserved areas.

In the following section, we will discuss the implications and potential applications of our research findings in the field of medical sciences, highlighting the benefits and future directions for this technology.

4. FUTURE DIRECTIONS AND APPLICATIONS

The successful integration of computer vision techniques and robotics in our research opens up a range of future directions and potential applications in the field of medical sciences. Additionally, the advancements in computer vision technology enable us to explore beyond hand movement tracking and delve into full human body tracking. This expansion opens up even more opportunities and applications.

The following are some areas where our system can make significant contributions:

1. Telemedicine and Remote Healthcare: Our system has the potential to revolutionize telemedicine by enabling remote healthcare services. Medical professionals can remotely control not only the robotic arm but also track

the movements of the entire human body. This allows for a comprehensive assessment of patient's physical conditions, enabling more accurate diagnostics and personalized treatment plans.

2. Surgical Robotics: The accurate hand landmark detection and precise control of the robotic arm have implications in surgical robotics. With full human body tracking, surgeons can remotely perform complex surgical procedures with enhanced precision and control. This technology can facilitate minimally invasive surgeries, reduce risks, and improve patient outcomes.
3. Rehabilitation and Prosthetics: Expanding our system to track the full human body opens up possibilities in rehabilitation programs and the development of advanced prosthetics. By monitoring the movements of the entire body, the robotic devices can provide tailored assistance and support during rehabilitation sessions. Moreover, the integration of full-body tracking can enhance the functionality and natural movement replication of prosthetic limbs, improving the quality of life for individuals with limb loss or impairment.
4. Medical Training and Education: The integration of computer vision and robotics can be instrumental in medical training and education. Students and healthcare professionals can utilize our system to practice surgical procedures, simulate real-world scenarios, and track their full-body movements. This technology provides a comprehensive training environment, enhancing skills development and fostering a better understanding of human anatomy and movement.
5. Assistive Technology: With full-body tracking capabilities, our system can be adapted for various assistive technology applications. Individuals with mobility challenges or physical disabilities can benefit from robotic devices that respond to their full-body movements. This technology can assist in activities such as mobility assistance, gait correction, and adaptive tools for daily living, promoting independence and improving the quality of life for users.

While our research focuses on the integration of computer vision and robotics for medical applications, the potential applications extend beyond the medical field. Industries such as manufacturing, automation, and human-computer interaction can benefit from similar systems that enable precise control of robotic devices based on full-body movements.

In conclusion, our research lays the foundation for a new paradigm in medical sciences by combining computer vision techniques and robotics. The successful integration of these technologies opens up exciting possibilities for telemedicine, surgical robotics, rehabilitation, medical training, assistive technology, and full-body tracking. The future development and refinement of our system hold great potential in transforming healthcare delivery, improving patient outcomes, and addressing the challenges faced in accessing medical services. Additionally, the expansion to full-body tracking presents new opportunities for various industries and human-machine interaction, shaping the future of technology and innovation.

5. CONCLUSIONS

In this research project, we successfully integrated modern computer vision techniques and network architectures with a robotic arm to achieve precise movement control. Using Mediapipe's hand landmark detection model, we accurately tracked hand movements and calculated corresponding angles for the robotic arm's servo motors. This enabled us to achieve controlled motions based on the user's hand movements.

We addressed challenges such as latency, computational load, and network bandwidth by optimizing the data transmission pipeline. Techniques like Cloudflare tunnelling and payload delivery inspired by Nvidia's Maxine were utilized to reduce latency and computational load, resulting in smoother and more responsive control of the robotic arm.

Our system has wide-ranging applications in telemedicine, surgical robotics, rehabilitation, medical training, and assistive technology. Additionally, the expansion into full-body tracking opens up new opportunities for human-computer interaction and various industries requiring precise control based on human movements.

Future research could explore integrating machine learning algorithms to enhance the robustness of hand landmark detection in challenging scenarios. Additionally, advanced control algorithms and motion planning strategies can optimize the movements of the robotic arm for more complex tasks.

In conclusion, our research demonstrates the potential of combining computer vision techniques, network architectures, and robotic systems for precise movement control. This integration has implications for healthcare, automation, and human-machine interaction. By advancing our understanding and capabilities in this area, transformative innovations can be achieved in various fields.

6. ACKNOWLEDGEMENT

We would like to express our gratitude to Mrs Latha A. P. for providing support and guidance. We got to learn a lot more about these many domains because of this project which is very helpful for us.

We would like to express my special thanks to IISc who gave us this golden opportunity to do this project. It helped us in doing a lot of Research and we were exposed to many things related to this topic.

7. REFERENCES

[1] Goryczka, S., & Szczygieł, M. (2021). Hand Landmarks Detection and Tracking in Augmented Reality Applications. Mathematical Problems in Engineering, 2021, 5544375. doi: 10.1155/2021/5544375

[2] Tang, Y., Lu, H., & Lai, Y. (2015). Robust and Efficient Hand Pose Estimation from RGB Images. Multimedia Tools and Applications, 74(24), 11049-11067. doi: 10.1007/s11042-015-2934-5

[3] Zhang, G., Huang, H., & Lu, J. (2020). A Comparative Study of Hand Pose Estimation Methods. arXiv preprint arXiv:2006.10214.

[4] Grzejszczak, P., & Kawulok, M. (2014). Hand Landmarks Detection and Localization in Color Images. In Proceedings of the 17th International Conference on Image Analysis and Processing (ICIAP) (pp. 339-346).

[5] Kumaar, R. G., & Ramesh, S. (2019). Hand Keypoint Detection using Deep Learning and OpenCV. Learn OpenCV.

[6] Grzejszczak, P., & Kawulok, M. (2015). Hand Landmarks Detection and Localization in Color Images. Semantic Scholar. Retrieved from

[7] Papers with Code. Hand Pose Estimation.

DOI: https://doi.org/10.15379/ijmst.v10i4.2377