JTE HCMUTE
ISSN: 2615-9740

JOURNAL OF TECHNICAL EDUCATION SCIENCE
Ho Chi Minh City University of Technology and Education
Website: https://jte.edu.vn
Email: jte@hcmute.edu.vn

A Stereo Vision-Based Method for Reconstructing 3-D Hand Motion in Real-Time

Trong-Pham Nguyen-Huu[1], Ngoc-Bich Le[1], Ngoc-Viet Tran[1], Tien-Tuan Dao[2], Tan-Nhu Nguyen[1*]

[1]International University, Vietnam National University – Ho Chi Minh City, Vietnam.
[2]Univ. Lille, CNRS, Centrale Lille, (LaMcube) UMR 9013, Lille, F-59000, France

[*]Corresponding author. *Email: ntnhu@hcmiu.edu.vn*

ARTICLE INFO

Received: 29/11/2024
Revised: 17/01/2025
Accepted: 04/02/2025
Published: 28/08/2025

KEYWORDS

Hand Motion Paralysis;
Hand Motion Tracking;
Stereo Vision-based Hand Motion Analysis;
Hand Motion Diagnosis;
Clinical Decision Support System for Hand Paralysis.

ABSTRACT

Hand motion paralysis negatively affects the lives of the involved patients. To recover the hand motions into their normal condition, these patients need to be taken into complex and long-term rehabilitation treatments. During rehabilitation, hand motions with full finger features need to be tracked accurately in 3-D dimension in real-time for analyzing and diagnosing hand motion paralysis. However, most studies tried to track hand motions based on contact sensors. These methods are not user-friendly. Even using contactless sensors, most of them could only detect the hand motions in 2-D image spaces. Consequently, in this study, we developed a stereo vision-based method for detecting and tracking 3-D hand features in real-time. In particular, we employed a convolutional deep neural network (C-DNN) for tracking by detecting hand-finger features. The features were tracked on left and right images captured by a stereo camera system before being reconstructed into 3-D spaces. A meta-validation procedure was conducted to compute the accuracy of the method with various light conditions, skin colors, and hand shapes. As a result, we could successfully track hand motions in real-time with acceptable accuracy. In perspective, we will apply the method for analyzing and diagnosing hand paralysis inside a clinical decision-support system.

DOI: https://doi.org/10.54644/jte.2025.1735

1. Introduction

During the life of a person, hands are used for most acts, from simple tasks such as doing daily activities to advanced tasks such as playing musical instruments [1]. By flexing and extending the fingers, people can do many jobs and create a lot of things [2]. Therefore, hand motion paralysis can lead to inconvenience and difficulty in the daily life of patients. Since the negative effects of hand paralysis have been recognized, there have been many scientists trying to improve the method for hand rehabilitation [3], [4], but they have not paid attention to hand motion assessment yet. Until the present, no system can provide the judgment for hand motion, although it has potential in the present and future.

The progress of hand motion recovery and hand skillfulness should be supervised and feedbacked to the subjects to self-control their behaviors in professional work, such as chef, musical instrument player, etc [5], [6]. To contribute to completing the system, this project will create a method for 3-D hand tracking based on a stereo–camera in real-time. This project will reconstruct the 3-D hand motions, and this data is provided for a system to assess the hand motions. So, this project is the first and primary step to make that system complete. Various previous methods have been proposed for tracking 3-D hand motions including contact and contactless, but each method still has advantages and disadvantages [7]. The contact method, commonly using IMU, which uses sensors attached to a glove to measure the position of the finger, requires people to wear a glove if they want to have a hand evaluation [8]. The primary benefit of using this method is high accuracy. As with most algorithms, the experiences using IMU to track the hand gesture have highly accurate results [8], [9]. However, this method has a drawback in that it does not potential to expand the application out of the medical field because wearing a glove with some electronic devices can lead to inconvenience and inability to do skillful tasks. Therefore, a

contactless method can solve that problem and provide the potential for application in not only the medical field but also other fields such as art, sport, etc. because of its convenience.

For contactless methods, this group includes the utilization of mono cameras [10], [11], infrared sensors [12], and stereo cameras [13]. To decide on an optimal method, each type of input device should be analyzed for its drawbacks. The first piece of equipment is a mono camera, which is the cheapest but less accurate for reconstruction. In light of the narrow viewpoint, mono cameras can have a lack of features, and it will lead to inaccuracies in the reconstruction process. Moreover, because of the lack of depth information when using mono cameras, 3-D coordinates of 2-D image points could not be reconstructed [14]. Even though some studies also tried to predict 3-D geometries using only a mono camera, the 3-D reconstruction was just optimized in the projection planes not in the real 3-D coordinate system [15]. The second device is an infrared (IR) sensor, it can reduce the drawback of calculating the depth of the hands' position but has other disadvantages that need to be considered. This device cannot capture the specific features of hands because its captured images lack texture and color, so in some special cases, the reconstruction will be wrong. Although an IR sensor can calculate the depth, these types of sensors cannot handle the occlusion of the target objects [15], [16]. Therefore, it is still not possible to be applied for 3-D reconstruction in the case of occlusion. On the other hand, stereo cameras can solve these issues. In other 3-D scanners, such as infrared sensors and laser scanners, the 3-D reconstruction is based on the time-of-flight of the light rays to the targets, so the reconstruction cannot be conducted if the objects are obscured. For the stereo camera-based reconstruction, the 3-D reconstruction is based on the pixel disparity of the targets on the left and right images. Consequently, if we can have the feature points on the pixel coordinates, the 3-D reconstruction could be conducted successfully. The Mediapipe framework can defect feature points in 2-D images even if they are obscured. Consequently, the combination of stereo-camera-based 3-D reconstruction and Mediapipe framework is necessary. In our previous study [15], we successfully employed a stereo camera system to reconstruct 3-D facial points from Mediapipe face points in real-time, but the stereo camera has not been employed to reconstruct 3-D hand motion with the Mediapipe framework.

After analyzing and choosing the most optimal device for 3-D hand tracking, this project will use stereo cameras and combine them with MediaPipe and OpenCV to reconstruct the 3-D hand of humans. Even though the proposed procedure of 3-D reconstruction of 2-D hand feature points using a stereo camera is classical, our procedure is conventional and necessary for reconstructing 3-D hand features in real-time from the pairs of 2-D hand features. Moreover, these 3-D reconstruction results are necessary for biomechanical simulation of the hand motion in real-time.

In detail, this project will create a system including two mono cameras in a fixed position, then use OpenCV to do camera calibration to get the camera parameters. Based on those parameters, the system can easily scale the objects in a 2-D image into the real values in 3-D space, this is the reconstruction step. Moreover, MediaPipe is also necessary in this project to do hand detection. It is an open–source library that helps developers with detecting human poses. In this project, it is used to detect hand landmarks including twenty-one points of left and right hand. So, the detailed procedure of this project is constructing a system of cameras firstly, doing camera calibration, then detecting hand captured by the stereo camera and reconstructing it. After completing the reconstruction step, it is crucial to compute the accuracy of results to ensure the efficiency of this method. The accuracy will be determined by comparing the real values of hand models with the same reconstructed ones and then calculating the error between those values.

2. Materials and Methods

2.1. Stereo Vision System Design

This project does not require a complex hardware architecture, it needs only two mono cameras and a fixed frame to make the camera parameters constant. This is a crucial stage to reduce error in further experiments [17]. However, there are some essential requirements for the camera system in the experiment process. Firstly, it must be fixed in a frame with a constant distance between two cameras because a system has one set of parameters that are used for reconstruction. If there is any change in the distance between 2 cameras, it will lead to an error when reconstructing. In this project, the system is

constructed as the Figure 1. In particular, the two Rapoo C260 webcams were configured so that they were aligned horizontally. The distance between the two cameras was 5 cm. The employed webcam was configured with the resolution of full high-definition (HD) (1920x1080). The two cameras have a frame rate of 24 frames per second. With this camera configuration, we recommend the user put the hand from 45 to 55 cm apart from the two cameras.

2.2. 2-D Hand Motion Tracking

To track the hand, two mono cameras must be done stereo tracking first. In this process, each camera must be calibrated to get the camera parameters, this step will compute camera intrinsic parameters [18] including the principal points, distortion coefficient, and focus length; and camera extrinsic parameters [19] including rotation matrix and translation vector, then the system will rectify images from both left and right camera. After that, the key points of both rectified images are extracted and matched together to get the points in 3-D coordinates [20]. After matching the stereo camera, the next step is hand detection and tracking in real-time. In this project, instead of creating a new AI model to detect hand landmarks, MediaPipe, especially MediaPipe's Hand Landmark will be applied to do that task. This model utilizes C-DNNs to achieve real-time and accurate hand and finger tracking [21]. A lightweight C-DNN model initially detects the palm region, followed by a more complex C-DNN model that predicts 21 3D hand landmarks [21]. C-DNNs are optimized for efficiency, enabling low-latency, low-power inference while maintaining high accuracy. By applying it to the camera, the system can detect and track the hand gestures. More specifically, each mono camera detects the hand gestures which are 21 points on a hand, as shown in Figs. 1,2. The result of this process is a set of hand landmarks in 2-D coordinates.

Figure 1. *The hardware parameters of the project.*

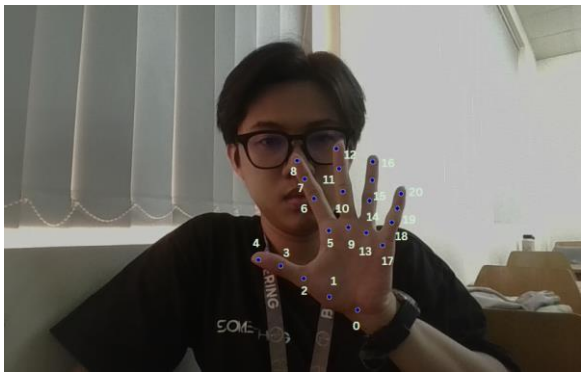

Figure 2. *Twenty-one hand landmarks were detected by MediaPipe.*

2.3. 3-D Hand Motion Reconstruction

The reconstruction process was done by following the block diagram (Fig. 3). More specifically, this project will again use the camera parameters taken from the calibration stage. These parameters are crucial for 3-D reconstruction because they help to correct distortion accurately, align corresponding points, scale properly, and position the reconstructed 3-D model in space. The next step of the reconstructing process is converting 2-D points of the hand into 3-D points. To do this, the 2-D points should be rectified by applying the OpenCV library to strengthen the distortions and make those points match the corresponding points in 3-D space, therefore, this act will enhance the quality of 3-D reconstruction. After that, the set of hand landmarks will be triangulated by using OpenCV as well. This step is essential because it helps to connect data points into a mesh, create structures, and indicate the relationship between points. Then, a set of hand landmarks in 3-D space is released and saved in an off-file (Fig. 4).

JTE HCMUTE
ISSN: 2615-9740
JOURNAL OF TECHNICAL EDUCATION SCIENCE
Ho Chi Minh City University of Technology and Education
Website: https://jte.edu.vn
Email: jte@hcmute.edu.vn

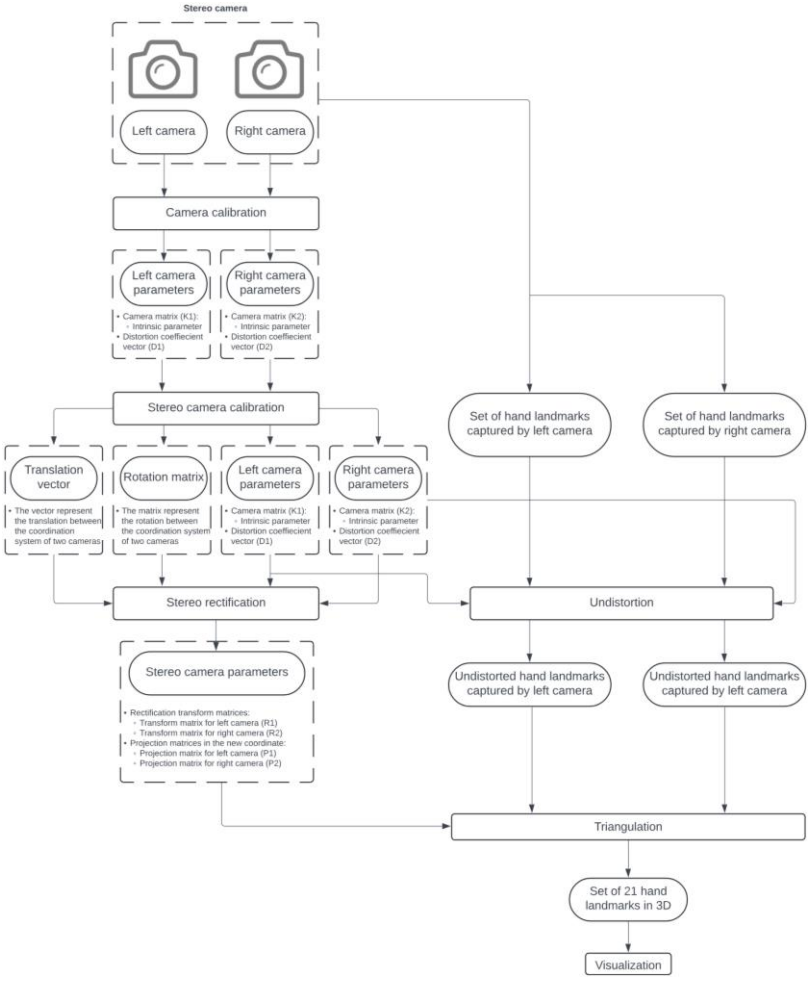

Figure 3. *The process of hand reconstruction.*

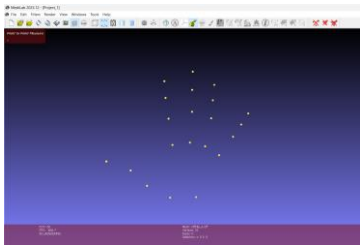

Figure 4. *The result after reconstructing hand landmarks.*

2.4. Metadata Validation

After getting the set of 3-D hand gesture points from the reconstruction step, those parameters will be validated to ensure the accuracy of the project. This stage includes stereo camera validation, and hand motion reconstruction validation in many different conditions. Firstly, stereo camera validation is done by getting the root mean square value of stereo camera calibration and checking the error between the standard chessboard with a fixed square size, 20mm in this experiment, and the reconstructed one. Computing root mean square (RMS), is calculated by Eq. 1:

$$RMS = \sqrt{\frac{1}{N}\sum_{1}^{N} d_i^2} \tag{1}$$

where: $N = 7 \times 9$ is the total number is chessboard points. d_i is the Euclidean distance between the corresponding points P_s^i and P_r^i, defined as in Eq. 2. P_s^i is the ith standard chessboard point, P_r^i is the ith reconstructed chessboard point.

$$d_i = \sqrt{\left(x_s^i - x_r^i\right)^2 + \left(y_s^i - y_r^i\right)^2 + \left(z_s^i - z_r^i\right)^2} \tag{2}$$

However, the OpenCV library [22] also has a function to calculate it, so the value is just calculated by using the function stereoCalibrate() of that library. This value shows the accuracy of camera parameters computed in the above step. After getting the RMS as expected, the reconstructed chessboard is validated. In more detail, a mesh including the same corner matrix as the testing chessboard and each point far from the next one is exactly 20mm will be the ground truth value for validating the stereo

camera. Then, the chessboard captured from the camera will be reconstructed, and the corners matrix into an off file (*.off), and that data will be used to compute the error value between testing and ground truth value. Note that the ground truth chessboard coordinates were formed by the edge lengths of each square with the z-coordinates of zero. Before computing the point-to-point distances between the reconstructed chessboard corners and the ground truth ones, the computed corners were rigidly registered to the ground truth corners using the Singular Value Decomposition registration algorithm [23]. This registration procedure minimized the rigid differences between the two point sets, so only non-rigid differences were computed.

Next, the hand motion tracking reconstruction is validated in a variety of conditions. The data used in these experiments are from the MANO dataset [24], this dataset provides many 3-D hand poses from numerous different people with both left and right hands. To ensure the system can track all hand activities, the models chosen for the validation stage must be different and have different motions. But this stage will be done more advanced in future research, in the study of hand motion tracking, now this project just focuses on how well the system can reconstruct the hand of distinguished people, so there are only two types of simple hand shapes including one hand with clenching fingers and one hand with extending fingers. After choosing two different hand poses, their 3-D files then are printed. Firstly, two hand models were captured by the system in three different light conditions. This experiment is used to find out which light range the system has the best performance. On the other hand, the hand models are dyed with three different colors which represent the color of Asian, European, and African skin. This project is not only forwarded to the Vietnamese or Asians but also forward to people all over the world. Therefore, validating the system in many different color conditions is essential.

The configuration of our experiment is shown in Fig. 5. In this experiment, a meta hand was printed with our 3-D printer. The employed 3-D printer was the 3-D Ender-3 S1 with a layer resolution of 0.4 mm. Our stereo camera included two Full HD Rappo Rapoo C260 webcams. A light source was also employed for controlling the light conditions of the capturing environment. We also used a personal laptop with an average hardware configuration for this experiment of core i5-12450H CPU and 16GB RAM. While capturing the printed meta hand, the hand was put around 38 – 45 cm away from the left camera. These distances were the optimal capturing position of our further hand motion rehabilitation system.

Figure 5. *The experimental configuration for evaluating the accuracy of hand feature point reconstruction: (a) The experiments included a printed meta hand, a stereo camera, a light source, and a computer; (b) the printed meta hand was put 38 – 45 cm away from the left camera.*

Note that to evaluate the accuracy of the reconstructed 3-D hand feature points. We also employed the Statistical Shape Model of the hand to generate the hand shape in 3-D spaces. The generated 3-D hand shapes were printed using our 3-D printing machine. The printed hand shapes were manually analyzed to determine the ground truth hand feature points based on the hand anatomical structure [25]–[27]. The ground truth values of the phalanges were computed from the ground truth feature points. The ground truth lengths of the phalanges were compared with the reconstructed ones using the Euclidean distance metric.

3. Results

3.1. Stereo Camera Accuracy

3.1.1. Single camera calibration

The target root mean square value of camera calibration for every single camera is under 0.5 because the hand motion tracking project requires high sensitivity with the change of hand's state for rehabilitation judgment, it is necessary to recognize the small movement of each finger of the patient. In this experiment, the root means the square value of the left and right single camera is approximately 0.308 and 0.307 respectively. These values are acceptable in this project because they meet the requirements.

3.1.2. Single camera calibration

As same as single camera calibration, the RMS value for stereo camera calibration also must be less than 0.5 pixels to ensure the accuracy and sensitivity of the system. In this experiment with a fixed frame of stereo camera, the rms value of stereo camera calibration is approximately 0.41. This value is under the maximum value, so it is acceptable, and the parameters can be used for 3-D reconstruction. Note that the calibration errors were computed as the pixel differences between the projected chessboard corners and the ground-truth chessboard corners. The projected chessboard corners were the projection of the real 3-D chessboard coordinates onto the image plane of the camera using the estimated camera's intrinsic parameter. The ground truth chessboard corners were detected using the chessboard corner detection algorithm from the OpenCV library [28].

3.1.3. Chessboard reconstruction

The chessboard reconstruction validation is used to measure the difference between the reconstructed chessboard and the standard chessboard (Fig. 6). After reconstructing 32 chessboards captured from many views, each chessboard has 7x9 points, which means there are 2016 points were used to calculate the accuracy of the system. The error of the system is approximately 1.714 mm, this value is moderate enough to ensure the system can be used for the next step. However, there is an error when reconstructing, it may be caused by the frame because it is not sustainable enough to prevent the error while calibrating and then it led to a small deviation when reconstructing.

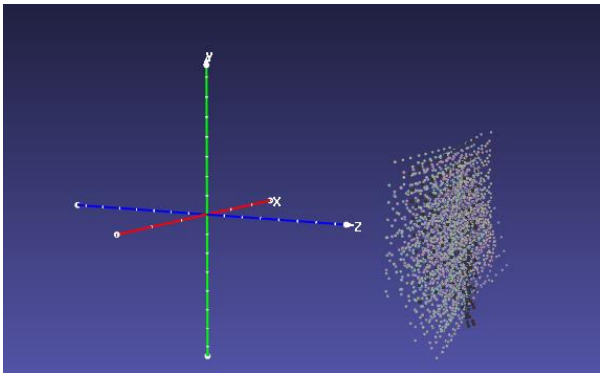

Figure 6. *Chessboard reconstruction.*

3.2. Hand Motion Tracking in Various Light Conditions

Fig. 7 shows the experiment conducted by reconstructing a yellow extending hand in three different light conditions. In particular, with normal light conditions, the features on the hand can be optimally recognized by the MediaPipe framework, so the mean error when reconstructing the hand feature points was the smallest of 3.81 mm. In the high light condition, because of the high contrast in the surface of the hand, features on the hand surfaces were not displayed clearly. Consequently, the Mediapipe framework could hardly detect feature points, so the reconstruction errors were higher (mean = 4.34 mm) than those in the normal light condition (mean = 3.81 mm). In the low-light condition, because of the very low light contrast between the hand surface and the background, the hand features could not be detected effectively using the MediaPipe framework. Experimental results show that the mean errors in

the low-light condition were highest at 4.69 mm. Overall, for all light conditions, the mean reconstruction errors were from 3.81 mm to 4.69 mm. This range was in acceptable values for hand motion rehabilitation applications [7], [29].

3.3. Hand Motion Tracking in Various Skin Colors

Fig. 8 shows the reconstruction results of hand feature points when the skin colors varied among yellow, white, and black. In particular, the proposed framework worked most effectively with yellow skin colors with a mean reconstruction error of 3.81 mm. In the white and black skin colors, the reconstruction errors were higher with the means of 4.63 mm and 4.72 mm, respectively. All skin colors were analyzed in normal light conditions. These reconstruction errors showed that when the hands were in the yellow skin colors, all features of the hand were visualized the most clearly. In the white and black colors, the features were not visualized effectively because of the low light contrast of the skin with the environment.

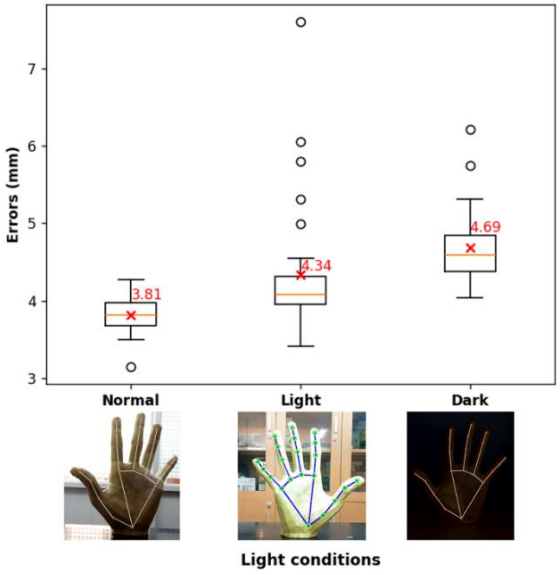

Figure 7. *The reconstruction of 3-D hand feature points in the normal, light, and dark light conditions.*

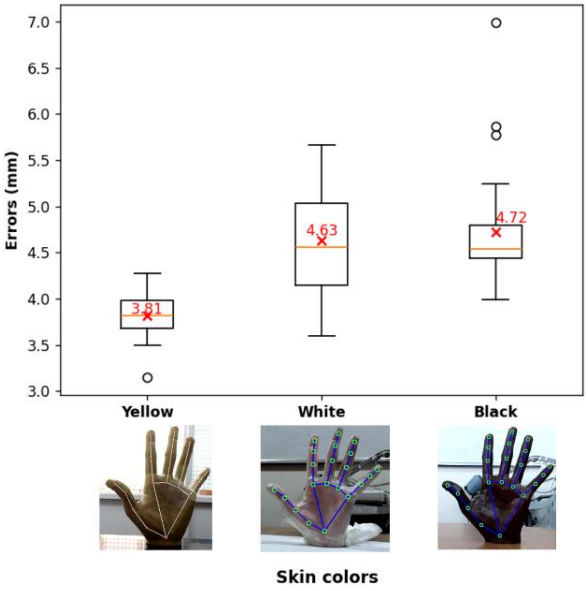

Figure 8. *The reconstruction of 3-D hand feature points in the yellow, white, and black skin colors.*

3.4. Hand Motion Tracking in Various Hand Gestures

The experiment is conducted by reconstructing yellow hands with extending and flexing state in the normally light conditions. For the extending hand, the system reconstructs well all points of hand gesture and has an error of approximately 3.81 mm, as shown in Figs. 7 and 8. For flexing hands, the result is approximately 4.04 mm, this result needs to be enhanced by widening the view of the stereo camera because the construct file shows that there are some hidden points and the estimated points for them to reconstruct are still not good enough.

4. Discussion

Hand motion tracking in 3-D real-time is crucial in not only the medical field but also in daily life. In the medical field, it helps doctors track the rehabilitation of hands, it will provide a more accurate judgment than people using bare eyes to observe. In daily life, this project can be applied to train someone who needs a skillful hand to do their jobs such as playing guitar, cooking, sculpture, etc. This project will be a helpful tool for tracking the accuracy and precision of each finger and then providing advice to people. The similarity between these two purposes is that the system must be able to recognize the small change in each hand gesture in real-time. To complete this project, it is required to be able to reconstruct the hand motion in 3-D space, because the high accuracy of the system will lead to high sensitivity of the project.

This project provides a useful method to reconstruct 3-D hand motion in real-time, which is a primary stage for 3-D hand motion tracking in real-time. The stereo camera is proven to have higher accuracy than the mono camera when reconstructing, the error for reconstruction by using the mono camera is from 10 – 20mm, while the stereo camera is only 1 – 4 mm [6]. It has better vision than infrared sensors because IR sensors cannot detect the hand feature, and it lacks an algorithm for 3-D reconstruction. This paper shows that the project can work well in many different conditions, so it is suitable for almost of people all over the world and in many different light conditions. Moreover, this method also proves that it can detect some distinguished hand shapes and hand motions in real-time. It has a crucial meaning for the big project because it initially shows that the big goal is feasible and the stereo–vision–based method can contribute to the final project. Regarding the use of infrared or laser scanners for reconstructing the 3-D hand feature points, the accuracy of the IR sensor was about 1 cm to 10 cm with ranges from 1 to 5 meters. The accuracy of the laser scanner can even be smaller about from 1 mm to 10 mm. However, because the infrared sensors and laser scanners cannot handle the occlusion while capturing, so only stereo camera systems were suitable for 3-D hand motion reconstruction.

It is important to note that, the accuracy of the 3-D scanning devices is highly dependent on the geometrical and/or visual characteristics of the target surface [30]. This concept is only true with infrared-based sensors because these sensors measure distances by the time-of-flight of the light from the sensor to the target object and reflecting from the target object [31]. However, in this study, we employed a stereo camera to triangulate the 2-D hand features. These features were detected with a deep neural framework called Mediapipe. This framework is stable with various light conditions and skin colors [15]. Moreover, 3-D reconstruction could not be conducted by the infrared or laser scanners if the target objects were obscured. The Mediapipe can infer the hand feature points even if they were obscured [15]. Consequently, our proposed method can handle abnormal light conditions and object obscurity better than other methods based on infrared or laser scanning devices.

We agreed that the infrared scanning sensors have a wider plausible range of light conditions in comparison with the cameras. This is because these infrared-light-based sensors emit light instead of observing light from environments like optical cameras [32]. Although the IR scanning sensors can work in various light conditions than optical cameras, they cannot handle occlusion [33]. In a stereo camera, if the pixel coordinates of the target object were in the two cameras, the 3-D coordinates of them could be reconstructed [34].

In the literature, although we could find some repositories in GitHub (https://github.com/) conducting a reconstruction of 3-D hand features in real-time using a stereo camera, we did not use these available codes. Moreover, these repositories have not evaluated the calibration and reconstruction accuracy in various light conditions and skin colors. Additionally, we could not find any publications that implemented the same procedure in the literature. Although the contribution of this study is relatively small, the results of this study will be valuable for real-time biomechanical simulation of hand motion in our clinical decision-support system for hand-motion rehabilitation.

Although this method has proved to be useful for the final aims, it still has three main drawbacks that should be fixed to enhance the quality of the project. Even though, in this study, we just developed a procedure for combining the advantages of the stereo camera system and hand feature detection from Mediapipe. This procedure employed conventional processes: (1) stereo-vision-based triangulation and (2) DNN-based Mediapipe framework, but this combination is necessary for further real-time biomechanical simulation. Because of its target application in biomedical engineering, we need various validation and testing procedures in various light conditions and skin colors. In perspective, we will employ the real-time-reconstructed 3-D hand feature points for analyzing and diagnosing hand-motion paralysis. Moreover, these could be used for 3-D hand reconstruction and bone-skeleton prediction. 3-D hand feature point reconstruction in real-time with various light conditions and skin colors is important to hand gesture recognition for Augmented Reality applications. We can also find promising applications of the proposed methods in robotic hand controls for controlling the motion of the robot hand, entertainment and gaming for controlling the game objects, education and training for surgery and music training, and other industrial applications for analyzing and evaluating manual assembly processes. Moreover, the system has an error for reconstruction, this error is still acceptable but after this project, the hardware will change to another frame which is more reliable and sustainable to reduce

the error from the hardware. The other issue is the experiments were only done with 2 different types of hand shapes, the result is just used to prove the feasibility of the next step, it has no meaning in showing the accuracy of the system. Therefore, the next research process will focus on doing experiments with various hand shapes, sizes, and motions. Another problem is the width of the camera view, it is not wide enough to capture some hidden points, and it leads to high accuracy when reconstructing a flexing hand. We also have drawbacks relating to the analysis of the flexing motion of the hand. In perspective, we will employ the two stereo camera systems for capturing these motions and compare the results with the marker-based reconstruction method.

5. Conclusions

In conclusion, hand motion tracking in 3-D real-time is crucial in both medical and daily life. In the medical field, it helps doctors track hand rehabilitation and provides more accurate judgment. In daily life, it can be applied to train individuals for tasks. This paper aims to reconstruct 3-D hand motions in real-time, ensuring high accuracy and sensitivity. This method works well in various conditions, making it suitable for people worldwide and in various light conditions. It can detect distinguished hand shapes and motion in real-time, demonstrating the feasibility of the project. Although there are some drawbacks the solution for them is proposed and will be addressed in the next process of research.

Acknowledgments

This research is funded by International University, VNU-HCM under grant number SV2023-13-BME.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] S. L. Gorniak, E. D. Collins, K. Goldie Staines, F. A. Brooks, and R. V Young, “The Impact of Musical Training on Hand Biomechanics in String Musicians.,” *Hand (N. Y).*, vol. 14, no. 6, pp. 823–829, Nov. 2019.
[2] S. Brorsson, A. Nilsdotter, E. Pedersen, A. Bremander, and C. Thorstensson, “Relationship between finger flexion and extension force in healthy women and women with rheumatoid arthritis,” *J. Rehabil. Med.*, vol. 44, no. 7, pp. 605–608, 2012.
[3] A. Borboni *et al.*, “Robot-Assisted Rehabilitation of Hand Paralysis After Stroke Reduces Wrist Edema and Pain: A Prospective Clinical Trial,” *J. Manipulative Physiol. Ther.*, vol. 40, no. 1, pp. 21–30, Jan. 2017.
[4] V. Nazari, M. Pouladian, Y. P. Zheng, and M. Alam, “A Compact and Lightweight Rehabilitative Exoskeleton to Restore Grasping Functions for People with Hand Paralysis,” *Sensors*, vol. 21, no. 20, p. 6900, Oct. 2021.
[5] A. Schiavio and M. Benedek, “Dimensions of Musical Creativity,” *Front. Neurosci.*, vol. 14, Nov. 2020.
[6] A. Floel and L. G. Cohen, “Translational Studies in Neurorehabilitation: From Bench to Bedside,” *Cogn. Behav. Neurol.*, vol. 19, no. 1, pp. 1–10, Mar. 2006.
[7] L. Wade, L. Needham, P. McGuigan, and J. Bilzon, “Applications and limitations of current markerless motion capture methods for clinical gait biomechanics,” *PeerJ*, vol. 10, p. e12995, Feb. 2022.
[8] C. Mummadi *et al.*, “Real-Time and Embedded Detection of Hand Gestures with an IMU-Based Glove,” *Informatics*, vol. 5, no. 2, p. 28, Jun. 2018.
[9] D. Zhang *et al.*, “Fine-Grained and Real-Time Gesture Recognition by Using IMU Sensors,” *IEEE Trans. Mob. Comput.*, vol. 22, no. 4, pp. 2177–2189, Apr. 2023.
[10] K. Kwon, H. Zhang, and F. Dornaika, “Hand pose recovery with a single video camera,” in *Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation (Cat. No.01CH37164)*, vol. 2, pp. 1194–1200.
[11] Muslikhin, J. R. Horng, S. Y. Yang, and M. S. Wang, “Object Localization and Depth Estimation for Eye-in-Hand Manipulator Using Mono Camera,” *IEEE Access*, vol. 8, pp. 121765–121779, 2020.
[12] C. S. Lee, S. Chun, and S. W. Park, “Tracking hand rotation and various grasping gestures from an IR camera using extended cylindrical manifold embedding,” *Comput. Vis. Image Underst.*, vol. 117, no. 12, pp. 1711–1723, Dec. 2013.
[13] J. Zhang, J. Jiao, M. Chen, L. Qu, X. Xu, and Q. Yang, “3D Hand Pose Tracking and Estimation Using Stereo Matching,” Oct. 2016.
[14] S. Jiang *et al.*, “In-Hand 3D Object Reconstruction from a Monocular RGB Video,” Dec. 2023.
[15] T. N. Nguyen, A. Ballit, and T. T. Dao, “A Novel Stereo Camera Fusion Scheme for Generating and Tracking Real-Time 3-D Patient-Specific Head/Face Kinematics and Facial Muscle Movements,” *IEEE Sens. J.*, vol. 23, no. 9, pp. 9889–9897, May 2023.
[16] R. M. Jans, A. S. Green, and L. J. Koerner, “Characterization of a Miniaturized IR Depth Sensor With a Programmable Region-of-Interest That Enables Hazard Mapping Applications,” *IEEE Sens. J.*, vol. 20, no. 10, pp. 5213–5220, May 2020.
[17] F. Alqahtani, J. Banks, V. Chandran, and J. Zhang, “3D Face Tracking Using Stereo Cameras: A Review,” *IEEE Access*, vol. 8, pp. 94373–94393, 2020.
[18] Y. F. Ji, “Vision-based sensing for assessing and monitoring civil infrastructures,” in *Sensor Technologies for Civil Infrastructures*, Elsevier, 2022, pp. 309–333.

[19] A. Khan, M. M. Hossain, A. Covaci, K. Sirlantzis, and C. Xu, "Light field imaging technology for virtual reality content creation: A review," *IET Image Process.*, vol. 18, no. 11, pp. 2817–2837, Sep. 2024.
[20] E. Adil, M. Mikou, and A. Mouhsen, "A novel algorithm for distance measurement using stereo camera," *CAAI Trans. Intell. Technol.*, vol. 7, no. 2, pp. 177–186, Jun. 2022.
[21] F. Zhang *et al.*, "MediaPipe Hands: On-device Real-time Hand Tracking," Jun. 2020.
[22] A. Zelinsky, "Learning OpenCV---Computer Vision with the OpenCV Library (Bradski, G.R. et al.; 2008)[On the Shelf]," *IEEE Robot. Autom. Mag.*, vol. 16, no. 3, pp. 100–100, Sep. 2009.
[23] J. Dongarra *et al.*, "The Singular Value Decomposition: Anatomy of Optimizing an Algorithm for Extreme Scale," *SIAM Rev.*, vol. 60, no. 4, pp. 808–865, Jan. 2018.
[24] J. Romero, D. Tzionas, and M. J. Black, "Embodied hands," *ACM Trans. Graph.*, vol. 36, no. 6, pp. 1–17, Dec. 2017.
[25] S. P. Kildare and K. Malone, "Skeletal Anatomy of the Hand," *Hand Clin.*, vol. 29, no. 4, pp. 459–471, Nov. 2013.
[26] C. A. Moran, "Anatomy of the Hand," *Phys. Ther.*, vol. 69, no. 12, pp. 1007–1013, Dec. 1989.
[27] H. Chim, "Hand and Wrist Anatomy and Biomechanics: A Comprehensive Guide," *Plast. Reconstr. Surg.*, vol. 140, no. 4, pp. 865–865, Oct. 2017.
[28] K. Pulli, A. Baksheev, K. Kornyakov, and V. Eruhimov, "Real-time computer vision with OpenCV," *Commun. ACM*, vol. 55, no. 6, pp. 61–69, Jun. 2012.
[29] L. C. Shum, B. A. Valdés, and H. M. Van der Loos, "Determining the Accuracy of Oculus Touch Controllers for Motor Rehabilitation Applications Using Quantifiable Upper Limb Kinematics: Validation Study," *JMIR Biomed. Eng.*, vol. 4, no. 1, p. e12291, Jun. 2019.
[30] H. A. Mercado, A. P. Vanegas, and A. G. Marrugo, "Robust 3D surface recovery by applying a focus criterion in white light scanning interference microscopy," *Appl. Opt.*, vol. 58, no. 5, p. A101, Feb. 2019.
[31] Y. J. Park and C. Y. Yi, "Time of Flight Distance Sensor–Based Construction Equipment Activity Detection Method," *Appl. Sci.*, vol. 14, no. 7, p. 2859, Mar. 2024.
[32] A. Karim and J. Y. Andersson, "Infrared detectors: Advances, challenges and new technologies," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 51, p. 012001, Dec. 2013.
[33] K. Saleh, S. Szenasi, and Z. Vamossy, "Occlusion Handling in Generic Object Detection: A Review," in *2021 IEEE 19th World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, 2021, pp. 000477–000484.
[34] S. Dill *et al.*, "Accuracy Evaluation of 3D Pose Reconstruction Algorithms Through Stereo Camera Information Fusion for Physical Exercises with MediaPipe Pose," *Sensors*, vol. 24, no. 23, p. 7772, Dec. 2024.

Trong-Pham Nguyen-Huu. The student is a senior student at the Biomedical Engineering School in International University; Vietnam National University - Ho Chi Minh City. He has worked for the Healthcare Visionary lab which is located at A1.408, International University, Zone 6, Linh Trung Ward, Thu Duc City, Ho Chi Minh City, Vietnam.
Email: bebeiu21258@student.hcmiu.edu.vn. ORCID: https://orcid.org/0009-0000-8925-7395

Ngoc-Bich Le earned his Master's and Ph.D. degrees in mechatronics science from Southern Taiwan University of Science and Technology – Taiwan in 2007 and 2010, respectively. His current research focuses on wearable devices, biomechanics, robotics, and artificial intelligence**.** He currently a lecturer at the Biomedical Engineering School in International University; Vietnam National University - Ho Chi Minh City.
Email: lnbich@hcmiu.edu.vn. ORCID: https://orcid.org/0000-0001-7431-0157.

Ngoc-Viet Tran graduated with engineering and master's degrees in the field of biomedical engineering at the International University; Vietnam National University - Ho Chi Minh City. He has expertise in the design and manufacturing of medical instrumentation. He is currently a full-time technician in the School of Biomedical Engineering, International University, Vietnam National University Ho Chi Minh City, Vietnam.
Email: tnviet@hcmiu.edu.vn. ORCID: https://orcid.org/0009-0007-3812-7794.

Tien-Tuan Dao is a Full Professor in Biomedical Engineering and Biomechanics at Centrale Lille Institut, France since 2020. His research interests concern computational biomechanics, knowledge and system engineering, and in silico medicine. He is in the Univ. Lille, CNRS, Centrale Lille, UMR 9013-LaMcube-Laboratoire de Mécanique, Multi-physique, Multiéchelle, Lille, F-59000, France.
Email: tien-tuan.dao@centralelille.fr. ORCID: https://orcid.org/0000-0002-5088-3433.

Tan-Nhu Nguyen received a Ph.D. in Biomedical Engineering and Biomechanics at Université de Technologie de Compiègne, France, in 2020. His current research interest is muscle modeling coupled with a serious game for facial rehabilitation. He is currently a full-time lecturer at the School of Biomedical Engineering, International University; Vietnam National University - Ho Chi Minh City, Vietnam. Mobile: +84389046652.
Email: ntnhu@hcmiu.edu.vn. ORCID: https://orcid.org/0000-0003-3343-0886.