Smart Presentation Control by Hand Gestures Using Computer Vision and Google's Mediapipe

Hajeera Khanum[1], Dr. Pramod H B[2]

[1]M.Tech Student, Dept. of Computer Science Engineering, Rajeev Institute of Technology, Hassan, Karnataka, India
[2]Associate Professor, Dept. of Computer Science Engineering, Rajeev Institute of Technology, Hassan, India

---

ABSTRACT - *Using hand gestures as the system's input to control presentation, we are constructing a presentation controller in this paper. The OpenCV module is mostly utilised in this implementation to control the gestures. MediaPipe is a machine learning framework with a hand gesture detection technology that is available today. This system primarily employs a web camera to record or capture photos and videos, and this application regulates the system's presentation based on the input. The primary purpose of the system is to change its presentation slides, I also had access to a pointer that allowed me draw on slides, in addition to that erase. To operate a computer's fundamental functions, such as presentation control, we may utilise hand gestures. People won't have to acquire the often burdensome machine-like abilities as a result. These hand gesture systems offer a modern, inventive, and natural means of nonverbal communication. These systems are used widely in human-computer interaction. This project's purpose is to discuss a presentation control system based on hand gesture detection and hand gesture recognition. A high resolution camera is used in this system to recognise the user's gestures as input. The main objective of hand gesture recognition is to develop a system that can recognise human hand gestures and use that information to control a presentation. With real-time gesture recognition, a specific user can control a computer by making hand gestures in front of a system camera that is connected to a computer. With the aid of OpenCV Python and MediaPipe, we are creating a hand gesture presentation control system in this project. Without using a keyboard or mouse, this system can be operated with hand gestures.*

Keywords: OpenCV, MediaPipe, Hand Gesture Recognition, Machine Learning, presentation controller, Human Computer Interaction

1. INTRODUCTION

Industry that is 4.0, also known as the Fourth Industrial revolution, calls for automation and computerization, which are realised through the convergence of various physical and digital technologies such as sensors, embedded systems, Artificial Intelligence (AI), Cloud Computing, Big Data, Adaptive Robotics, Augmented Reality, Additive Manufacturing (AM), and Internet of Things [2]. The increased interconnectedness of digital technology's make it essential for us to carry out daily activities like working, shopping, communicating, having fun, and even looking for information and news [3]. The use of technologies and improvements in human-machine interaction allow people to identify, communicate, and engage with one another using a wide variety of gestures.

The gesture is a type of nonverbal communication or nonvocal communication that makes use of the body's movement to express a specific message. The hand or face are the most frequently used portions of the body [4]. The research has gravitated toward the new sort of Human-Computer Interaction (HCI) known as gesture-based interaction, which Krueger launched in the middle of the 1970s. Building application interfaces with controlling each human body part to communicate organically is a major focus of study in the field of human-computer interaction (HCI), with hands serving as the most practical alternative to other interaction tools given their capabilities [5].

Recognizing hand movements using Human-Computer-Interaction (HCI) might aid in achieving the necessary ease and naturalness [6]. Hand gestures serve the purpose of communicating information when engaging with other individuals. encompassing both basic and complicated hand motions. For instance, we can point with our hands towards an item or at individuals, or we can convey basic hand shapes or motions using manual articulations in conjunction with sign languages' well-known syntax and lexicon. Therefore, employing hand gestures as a tool and integrating them with computers might enable more intuitive communication between individuals [6]. To simplify things to anybody thus create Artificial Intelligence (AI) based apps, various frameworks or libraries have been developed for hand gesture detection. MediaPipe is one of them. For the purpose of employing machine learning techniques like Face Detection, Iris, Pose, Hands, , Hair segmentation Holistic, Box tracking, Object detection, Instant Motion Tracking, Face Mesh, KIFT, and Objection, Google has created the mediapipe framework. Few benefits for employing mediapipe framework's showcases include helping programmer concentrate on model and algorithm creation for application and supporting application's environment via result repeatable allover multiple architecture and gadgets. To conduct different activities, such as seeking ahead and backward through slides, drawing and erasing in a presentation, the project employs hand gestures, often no of raised fingers inside the region of interest.

The difficult component of this system is background movies or pictures that are recorded or captured while taking inputs, such as hand gestures from the user. Lightning may also sometimes affect the quality of the input obtained, which makes it difficult to recognise motions. Segmentation is the process of identifying a linked area of an image that has certain characteristics like colour, intensity, and a relationship between pixels, or pattern. Additionally, have utilised some significant packages, like mediapipe, tensorflow, numpy, and opencv-python.

2. EXISTING SYSTEM

The author has developed an ANN application used for classification and gesture recognition, Gesture Recognition Utilizing Accelerometer. The Wii remote, which rotates in the X, Y, and Z directions, is essentially employed in this system. The author has utilised two tiers to construct the system in order to reduce the cost and memory requirements. The user is verified for gesture recognition at the first level. Author's preferred approach for gesture recognition is accelerometer-based.

Following that, system signals are analysed at the second level utilising automata to recognise gestures (Fuzzy). The Fast Fourier technique and k means are then used to normalise the data. The accuracy of recognition has now increased to 95%.

Recognition of Hand Gestures Using Hidden Markov Models - The author of this work has developed a system that uses dynamic hand movements to detect the digits 0 through 9. In this work, the author employed two stages. Preprocessing is done in the first phase, while categorization is done in the second. There are essentially two categories of gestures. both Link gestures and Key motions. The key gesture and the link gestures are employed in continuous gestures for the goal of spotting. Discrete Hidden Markov Model (DHMM) is employed for classification in this work. The Baum-Welch algorithm is used to train this DHMM. HMM has an average recognition rate range of 93.84 to 97.34%.

The author has employed inexpensive cameras to keep costs down for the consumers. Robust Part-Based Hand Gesture Recognition Using Kinect Sensor. Although a kinect sensor's resolutions lower than that of other cameras, it is nevertheless capable of detecting and capturing large pictures and objects. Only the fingers, not the entire hand, are paired with FEMD to deal with the loud hand movements. This technology performs flawlessly and effectively in uncontrolled settings. The experimental result yields an accuracy of 93.2%.

3. RELATED WORK

3.1 Hand Gesture Recognition

Computer science's key field of gesture recognition develops technology that tries to understand human motions so that anybody may use basic gestures to communicate with a device without touching it directly. Gesture recognition is the process of tracking gestures, representing them, and translating them into a specific instruction[8]. The goal of hand gesture recognition is to identify from explicit hand movements as input, then process these motions representation for devices by mapping as output. The software sub-tree includes an image recognition tool that accepts video feeds as input. Using third-party tools like OpenCV, it is possible to locate MediaPipe objects that are visible in the camera's field of view.

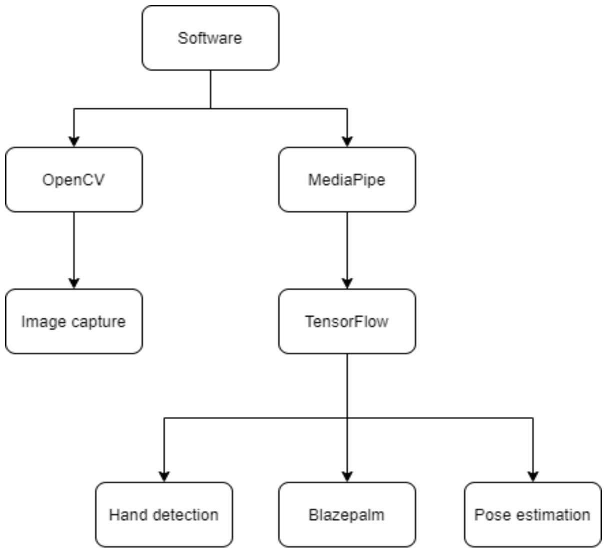

Fig -1: Software Sub-Tree

There are three hand gesture recognition techniques that may be found in various types of literature:

Machine Learning Techniques: The condensation method, PCA, HMM [9][10][11][12], sophisticated particle filtering, and stochastic processes and approaches based on statistical models for dynamical gestures produced the resultant output.

Algorithm perspectives: collection of manually defined, encoded constraints and requirements for defining gestures in dynamic gestures. Galveia [13] used a 3rd-degree numerical solution (construct a 3rd-degree polynomial equation, detection, decreased complexity for equations, and comparative handled in gestures libraries) to ascertain the dynamic element of the hand movements.

Rule-based methods: Appropriate for dynamic movement alternatively fixed gestures, that are inputs with a pre-encoded body of norms [5]. The characteristics of input movements are retrieved, and they are evaluated to the encoded principles that regulate the flow of the motions that are recognised Advances in Engineering Research, volume 207 102. synchronization between rules-based gestures and input that is accepted as recognised gestures upon output [13].

3.2 MediaPipe Framework

For the recognition of hand gestures, there are several machine learning frameworks and tools available today. MediaPipe is among them. The MediaPipe is just a framework created to provide manufacturing-ready machine learning, which requires build infrastructure to execute inference above any sort of sensory information and has released code to go along with scientific work [7]. The function of the media processor model, inference model and the data manipulation are all drawn from a perceptual pipeline in MediaPipe [14]. Other machine learning systems like OpenCV 4.0, Tensor flow, PyTorch, MXNet, CNTK, also employ graphs of computations.

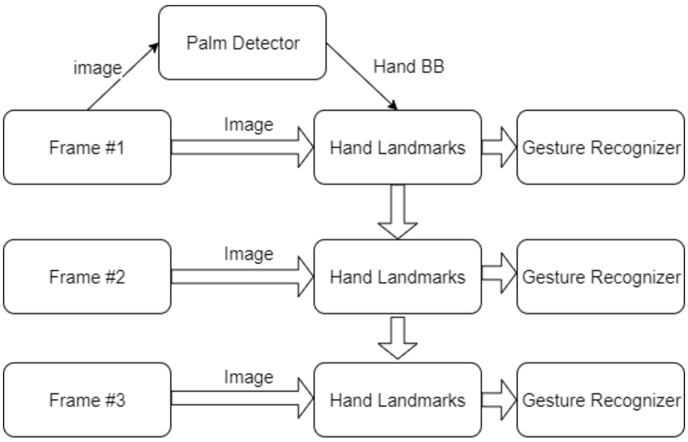

Fig -2: Overview of the Hand Perception Pathway

Two hand gesture recognition models are implemented by the MediaPipe of Figure 2 as follows [14]:

1. A palm detecting described the basic the acquired picture and transforms the images with such an aligned object of a hand.
2. The hand landmark model processes an image with a clipped bounding box and produces 3D hands key points just on hands.
3. The gesture recognition system that organises 3D hands key point into a distinct set of motions after classifying them.

3.3 Palm Detection Pattern

The BlazePalm first palm detector was implemented inside the MediaPipe framework. The detection of the hand is a difficult process. In order to model the palm using square bounding boxes to prevent other aspect ratios and lowering the amount of hooks by the ratio of 3–5, non-maximum reduction method must first be trained on the palm rather than the hand detector. Finally, limit the focus loss during training with help from a huge number of anchoring caused by the high exist in a wide using encoder-decoder of image retrieval that is employed for larger scene context-awareness even for tiny objects.

3.4 Hand Landmark

Accomplishes accurate crucial point clustering of 21 main points with only a 3D touch coordinates that is done within the identified hand areas and immediately generates the coordinates predictor that is a representation of hand landmarks within MediaPipe [15][16].

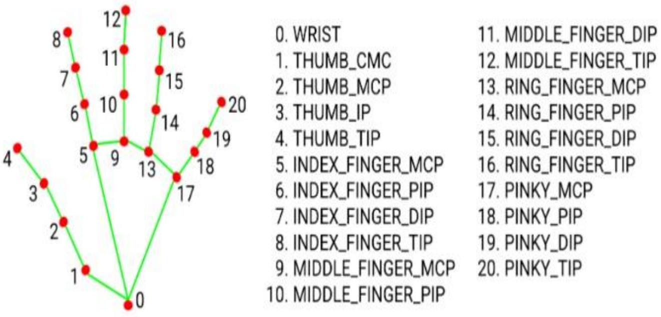

Fig -3: MediaPipe Hand Landmark [18]

Every touch of a landmark had already coordinate is constituted of x, y, and z in which x and y have been adjusted to [0.0, 1.0] besides image width and length, while z portrayal the complexity of landmark. A depth of the ancestral landmark, which is located at the wrist. The value decreases the more away the landmark is from the camera.

4. SYSTEM ARCHITECTURE AND METHODOLOGY

The code for this project was created in the python language utilising the Opensv and NumPy packages. In this work, the libraries that will be utilised for further input and output processing are initially imported. MediaPipe, OpenCV and numpy are the libraries that are utilised in this project and that need to be imported. Video inputs come from in out main camera.To recognise the video as input from our camera, mediapipe is now being utilised, and the mphand.hands module is being used to detect the gesture. Then, in order to access the presentation, we utilised pointer. The input picture must next be converted to an RGB image to finish the input processing. Then it's your chance to specify the thumb and finger points in input. Numpy is utilised to transform this process needed output. presentation is handled using the hand range in this procedure. The Python language's NumPy library is essential for computing. It includes a variety of elements:

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 07 | July 2022 www.irjet.net p-ISSN: 2395-0072

1. effective N-dimensional array
2. Object broadcast and C integration tools
3. Capability for the Fourier analysis and pseudo random

4.1 Identifying Hand Gesture by MediaPipe

Across platforms including Android, iOS, the web, edge devices, and many applicable ML pipelines, MediaPipe is a module for processing video, audio, and various sorts of related data. With the use of this module, a variety of tasks may be completed

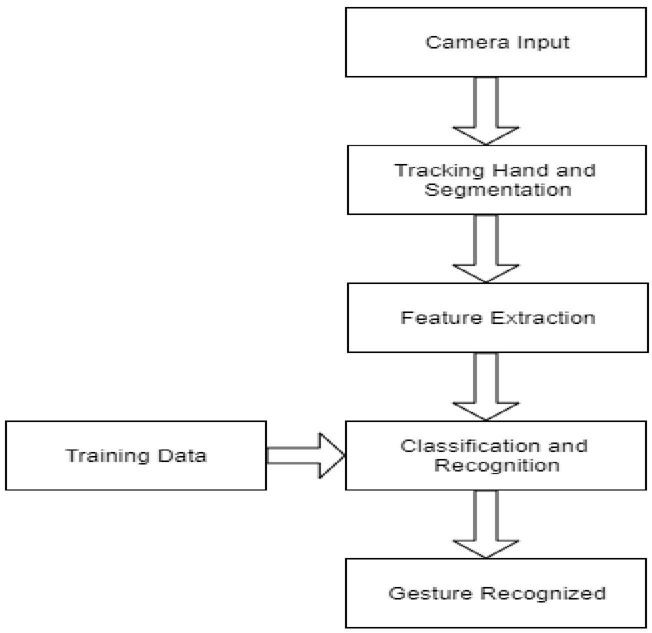

Fig -4: System Architecture

In our project, we utilised it to identify hand gestures and extract input for:

1. Multi-hand Monitoring
2. Facial Recognition
3. Segmentation
4. object tracking
5. object detection

In order to generate a better result, we have implemented a Hand Gestures Recognition System. The webcam is turned on while the software is running, and the kind of gesture used to detect the shape of the hand and give us the desired output is static. This project uses the curve of the hand to regulate loudness. The system receives input, captures the item, detects it, and then recognises hand gestures.

4.2 NumPy

A Python package called Open CV addresses the problem of PC vision. It is utilised for face detection, which is carried out utilising machine learning. It is a highly significant library that is used in several applications to identify various frames and detect faces. It also supports a number of programming languages. Additionally, it carries out motion and object detection. It may be used to recognise the faces of animals and supports a different operating system.

4.3 TensorFlow

Google developed TensorFlow, a framework that enables programmers to use "novel optimizations and training algorithms" for defining, developing, and using various machine learning models. The machine learning algorithms of TensorFlow may be thought of as directed or computational graphs. Each node in such a network denotes an operation, and the edges (tensors) indicate the data that moves between the operations. In order to generate a better result, we have implemented a Hand Gestures Recognition System. TensorFlow's model was modified in the library to only return necessary points on the body. PoseNet is used by the pose estimation programme to locate joints on the human body. The TensorFlow posture estimation library contains PoseNet, a pre-trained model that uses computer vision to predict a person's bodily joints [. Seven joints in the body are given coordinates with numbers ranging from 0 to 6.

TensorFlow has demonstrated the ability to offer solutions for object recognition using photos and may be used to train huge datasets to recognise specific things.

TensorFlow also features a library that the user might just save or reload as needed. This enables users to save the checkpoints with the highest evaluation score and makes it reusable for unsupervised learning or model fine-tuning.

5. RESULT

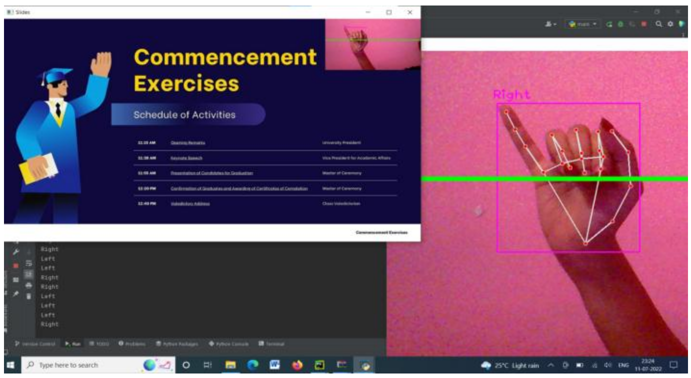

Fig -5: Hand Gesture to move on to the next slide

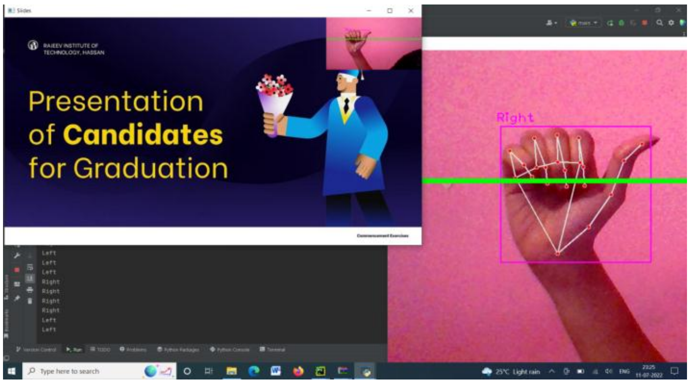

Fig -6: Hand Gesture for going back to previous slide

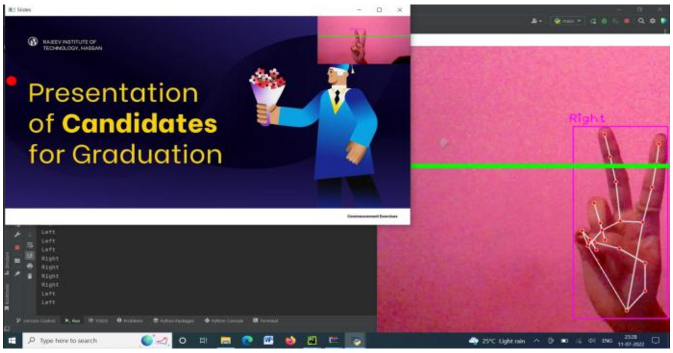

Fig -7: Getting a pointer on slide

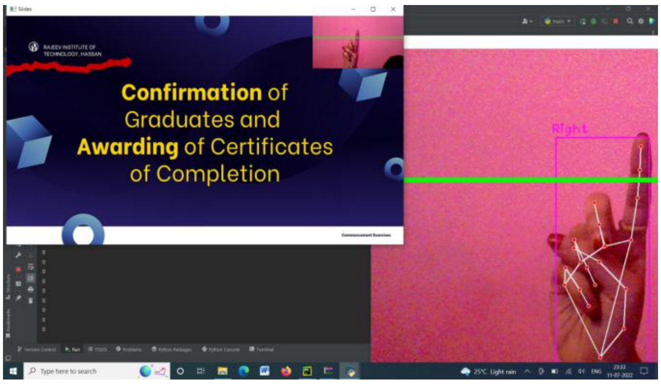

Fig -8: Draw using that pointer

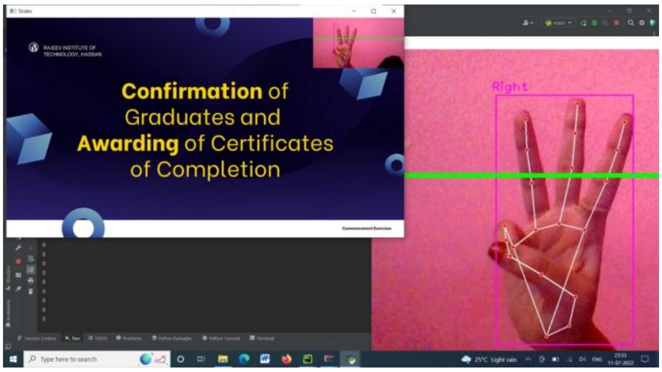

Fig -9: Erase the drawing on slide

6. CONCLUSION

This project showcases a programme that enables hand gestures as a practical and simple method of software control. A gesture-based presentation controller doesn't need any special markers, and it can be used in real life on basic PCs with inexpensive cameras since it doesn't need particularly high quality cameras to recognise or record the hand movements. The method keeps track of the locations of each hand's index finger and counter tips. This kind of system's primary goal is to essentially automate system components so that they are easy to control. As a result, we have employed this method to make the system simpler to control with the aid of these applications in order to make it realistic.

REFERENCES

[1] Research Gate, Google.

[2] Cevikcan, Ustunug A, Industry 4.0: Managing The Digital Transformation, Springer Series in Advanced Manufacturing, Switzerland. 2018. the following DOI: 10.1007/978-3-319-57870-5.

[3] Pantic M, Nijholt A, Pentland A, Huanag TS, Human-Centered Intelligent Human-Computer Interaction (HCI2): How Far We From Attaining It?,International Jounal of Autonomous and Adaptive Communications Systems (IJAACS), vol.1 no.2, 2008. pp 168-187. DOI: 10.1504/IJAACS.2008.019799.

[4] Hamed Al-Saedi A.K, Hassin Al-Asadi A, Survey of Hand Gesture Recognition System. IOP Conferences Series: Journal of Physics: Conferences Series 1294 042003. 2019.

[5] Z.Ren, J.Meng, Yuan J. Depth Camera Based Hand Gesture Regconition and its Application in Human-Computer-Interaction. In Processing of the 2011 8th International Conference on Information, Communication and Signal Processing (ICICS). Singapore. 2011.

[6] S.Rautaray S, Agrawal A. Vision Based Hand Gesture Recognition for Human Computer Interaction: A Survey. Springer Artificial Intelligence Review. 2012. DOI: https://doi.org/10.1007/s10462-012-9356-9.

[7] Lugaresi C, Tang J, Nash H, McClanahan C, et al. MediaPipe: A Framework for Building Perception Pipelines. Google Research. 2019. https://arxiv.org/abs/2006.10214.

[8] Z.Xu, et.al, Hand Gesture Recognition and Virtual Game Control Based on 3D Accelerometer and EMG Sensors, In Processing og IUI'09, 2009, pp 401-406.

[9] Lee H, Kim J. An HMM-Based Threshold Model Approach for Gesture Recognition. IEEE Trans on PAMI vol.21. 1999. pp 961-973.

[10] Wilson A, Bobick A, Parametric Hidden Markov Models for Gesture Recognition. IEEE Trans. On PAMI vol.21, 1999. pp.884-900.

[11] Wu Xiayou, An Intelligent Interactive System Based on Hand Gesture Recognition Algorithm and Kinect, In 5th International Symposium on Computational Intelligence and Design.2012.

[12] Wang Y, Kinect Based Dynamic Hand Gesture Recognition Algorithm Research, In 4th International Conference on Intelligent Human Machine System and Cybernetics. 2012.

[13] Galveia B, Cardoso T, Rybarczyk, Adding Value to The Kinect SDK Creating a Gesture Library, 2014.

[14] Lugaresi C, Tang J, Nash H et.al, MediaPipe: A Framework for Perceiving and Processing Reality. Google Research. 2019.

[15] Zhag F, Bazarevsky, Vakunov A et.al, MediaPipe Hands: On – Device Real Time Hand Tracking, Google Research. USA. 2020. https://arxiv.org/pdf/2006.10214.pdf

[16] MediaPipe: On-Device, Real Time Hand Tracking, In https://ai.googleblog.com/2019/08/on-device real-time-hand-tracking-with.html. 2019. Access 2021.

[17] Grishchenko I, Bazarevsky V, MediaPipe Holositic – Simultaneoue Face, Hand and Pose Prediction on Device, Google Research, USA, 2020, https://ai.googleblog.com/2020/12/mediapipe holistic-simultaneous-face.html, Access 2021.

[18] MediaPipe Github: https://google.github.io/mediapipe/solutions/hands. Access 2021.