Hand tracking and robotic imitation

Tihana Ujevic

Faculty of Science

University of Split

Ruđera Boškovića 33

Abstract

The field of motion recognition is advancing rapidly, demonstrating its utility across diverse domains. This paper explores the multifaceted applications of motion recognition technology in various sectors, including healthcare, film, robotics, and human-computer interaction. It delves into the potential of harnessing this technology to enhance user experiences and solve complex problems. Furthermore, this paper introduces a novel project aimed at developing a low-fidelity 3D hand model designed to imitate the movements of a human hand. This endeavor represents a step towards creating accessible and cost-effective solutions for gesture-based interactions for rehabilitation therapies, where precise motion mimicry plays a pivotal role in improving the user experience and overall effectiveness of such systems.

Keywords: *motion recognition, movement imitation, robot.*

1. Project Overview

The aim of this project is to develop a 3D printed hand that can imitate the movements of a human hand. The idea is to use that hand for the purpose of rehabilitation therapy; robotic hand would be used for showing exercises that pacient needs to perform and, while performing, developed application will return movement accuracy.

2. Problem Statement

Robot needs to perform given movements within its capabilities which are weakened due to mechanical reasons. Movements can be predefined in dataset or made in real time so robot should be able to imitate any given movement. An application should be developed in a way to display patient how well he did the exercise. Everything needs to be done using only web camera so that everyone could easily use it without any other equipment.

3. Related work

There are several ways in which robotic hands can imitate movements in real time. Using sensors to detect muscle signals: Engineers at MIT and Shanghai Jiao Tong University have designed a soft, lightweight, and potentially low-cost neuroprosthetic hand that can be controlled by detecting muscle signals from a residual limb. They used an algorithm to program the controller for their pneumatic system, which allows the amputee to perform daily activities, such as zipping a suitcase, pouring a carton of juice, and petting a cat, just as well as a person with a natural hand [1]. Mimicking human movements: A robot hand and arm can be controlled by demonstrating motions with a human hand. The robot observes the human operator via a single RGB camera and imitates their actions in real-time. This low-cost system does not require any special gloves, mocap markers, or even camera calibration and works from a single RGB camera [2]. Using mechatronics and robotics: Robotic studies have been developed as robot hands, robot legs, and humanoid robots. The new developments in mechatronics and robotics have led to many new possibilities [3].

There are several software options available for tracking hand movement: Ultraleap offers world-leading hand tracking software that generates a virtual model of your hand movements, including the joints and bones inside your hand. They also offer the Leap Motion Controller, an optical hand tracking module that captures the movements of your hands with unparalleled accuracy [4]. For hand tracking with no devices except camera, MediaPipe Hands is a solution, it detects the landmarks of the hands in an image. It can be used to localize key points of the hands and render visual effects over the hands. It operates on image data with a machine learning (ML) model as static data or a continuous stream [5].

As for dynamic hand gesture recognition, it can be based on 2D convolutional neural network and feature fusion [6]. The proposed method is accurate and effective on the Cambridge Hand Gesture dataset and Northwestern University Hand Gesture dataset. Other approach uses low complexity recurrent neural network (RNN) algorithms. One is based on video signal and employs a combined structure of a convolutional neural network (CNN) and an RNN. The other uses accelerometer data and only requires an RNN. Fixed-point optimization that quantizes most of the weights into two bits is conducted to optimize the amount of memory size for weight storage and to reduce the power consumption in hardware and software based implementations [7].

Overall, robotic hands can imitate movements in real time by detecting muscle signals, mimicking human movements, and using mechatronics and robotics. Tracking hands can be done using external devices such as sensors and controllers or using simply web camera with a help of AI libraries. And, for movement recognition, neural networks (CNN, RNN) are mostly used.

4. Application areas

One of the most famous and popular ways of practical use of this technology is a film industry. In today’s animation films, virtual character animation is an area of research within computer graphics that is progressing rapidly. Motion capture stands out as the key element responsible for a groundbreaking transformation in 3D animation production techniques. This technology enables animation producers to directly animate image models using the movements and expressions of actors, simplifying the animation production process significantly and elevating the overall quality of animations. [8]

With the advancement of modern computer technology, computer animation has developed rapidly. In particular, the introduction of 3D concepts improved the perception of depth and realism, getting more closely with the way people perceive the world today. In contemporary animated films, virtual human animation is a rapidly evolving field within computer graphics, with motion capture playing a main role. Furthermore, it represents an advanced technology that enables real-time capture and digital analysis of the 3D motion of humans, animals, or objects. This is represented by the use of motion capture in European and American 3D animated films like Avatar, The Adventures of Tintin, Puss in Boots, A Monster in Paris, and Ice Age etc. [8]

Animation production is a complex and time-consuming task. To make easier the workload for animation producers in the realm of computer 3D animation, performance animation has been extensively researched and implemented, with motion capture technology being the crucial and indispensable component. Motion capture technology involves the recording of human body movements and translating them into digital patterns. Essentially, it records the motion and spatial changes of humans or animals and uses this data to animate virtual characters, resulting in smoother and more precise outcomes compared to manual animation. In its early stages, motion capture solely recorded the body movements of actors, with facial expressions being added later by animators. However, the film "Rise of the Apes" marked a significant advancement in facial capture technology, introducing the "face muscular tissue simulation technique" for the first time. This innovation allowed the acting skills of actors to be faithfully conveyed to viewers. Moreover, there's a growing recognition that "motion capture technology" should increasingly be referred to as "performance capture technology." The portrayal of the orangutan Caesar in "Rise of the Apes" through motion capture technology is a true classic. The film masterfully captures the discrete and complicated expressions and body language of Caesar, preserving both his primal instincts and emerging human qualities with remarkable authenticity. This achievement shows the art of seamlessly blending fiction with reality. For instance, when Caesar stands before Doctor Will with a dignified yet independent gesture, the film's title, "Rise of the Apes," becomes unequivocally clearly. In essence, it is the technology of "motion capture" is responsible for the success of this film, demonstrating that Hollywood's cutting-edge science and technology not only deliver spectacular stunts during action sequences but also integrate technology with storytelling. [8]

In film and television performance animation system, based on computer graphic principles, we record motion conditions of animated objects in the form of graphs with the use of several video capture devices, and then deal with graphic data with the use of computer. Working processes of motion capture technology system are as follows: write character scripts, draw shooting scripts, digitalize role style design, and make animation model according to requirements for scripts. [8]

Simultaneously, performers are tasked with executing motion performances, capturing the motion characteristics of animated objects through video capture devices, and subsequently processing this graphical data using computers to generate animation sequences. To facilitate later editing, performers are required to put on blue clothes and attach special markers or luminous points at critical positions on their bodies, such as joints, hips, elbows, and wrists. A vision system is then employed to identify and process these markers. Real-time video capture is achieved using multiple video cameras, allowing the motion conditions of these marked points to be observed in each frame from the images obtained. Consequently, a continuous motion path for a specific point, with variations over time, can be obtained. Subsequently, 3D reconstruction technology is employed to convert these motion paths of these points into skeleton patterns.

In the early days, the production of 3D animation relied heavily on manual adjustments by animators frame by frame, which was time-consuming, complicated, and inefficient. This led to significant time consumption, especially when creating movements using 3D animation software, resulting in somewhat awkward and stiff animations in certain cases. However, motion capture technology has revolutionized the creation of lifelike animated characters by capturing the movements of actors, athletes, and dancers. These actors perform with fluid and natural motions, reducing the need for animators to closely adjust each frame. For instance, in the 2011 film "Rise of the Apes," none of the gorillas are actual animals; they are animated performances captured through motion capture technology by actors behind these furry creatures. Similarly, the robotic fighters in "Real Steel" execute their actions naturally and smoothly, with the actors themselves wearing blue clothes and stilts. In "The Adventures of Tintin," the characters exhibit cartoon-like appearances with highly natural movements and expressions, all achieved through the motion capture of real actors. [8]

One of the most iconic applications of motion capture in films is seen in "The Lord of the Rings," where the character Gollum, who appeared quite unusual, earned an Oscar nomination for "Best Actor in a Supporting Role." During the filming of Gollum's scenes, a three-step process was employed: first, real-life shooting with Andy Serkis (dressed in white) performing alongside other actors; second, digitally removing Andy Serkis to insert Gollum into the scenes; and third, Andy Serkis performing independently while wearing motion-capture attire with markers. Animators initially modeled Andy Serkis' movements and later transformed the data into the Gollum character model. This production method, although effective, was time-consuming and expensive. [8]

The film "The Polar Express" in 2004 played a significant role in popularizing motion capture technology, making it a standard technique in the industry.

Figure 1. Andy Serkis, motion-capture (source: https://blog.castac.org/2019/03/out-of-body-workspaces-andy-serkis-and-motion-capture-technologies/)

"The Polar Express" stands out as a pioneering CG animation film that exclusively relies on motion capture technology from start to finish. What makes it particularly noteworthy is that it marks the first instance of combining body motion capture and facial performance capture in a single production. This achievement earned the film a place in the Guinness World Records in 2006. During the filming of "The Polar Express," an impressive array of over 60 video cameras were strategically positioned to simultaneously capture the body movements and facial expressions of several actors. This approach resulted in animated characters whose motions closely resembled those of real individuals. As motion capture technology continued to mature, each passing year saw the creation of remarkable works employing this technique. For instance, "King Kong" (2005) featured towering versions of the titular character and menacing dinosaurs, both brought to life through CG animation. "Pirates of the Caribbean II" (2006) and "Beowulf" (2007), directed by Robert Zemeckis, showcased CG human characters that pushed the boundaries of realism. However, the top of CG character performance using motion capture technology was achieved with "The Curious Case of Benjamin Button" (2008), which featured a CG character of a real person. This outstanding accomplishment earned recognition from the Oscars as the only performance capture of its kind approved by the Academy. [8]

Moreover, the motion capture technology employed in "Rise of the Apes" represents a seamless integration of the advancements seen in "Avatar" and motion capture technology. It can be stated that the motion capture techniques employed in "Rise of the Apes" have effectively addressed all the challenges typically associated with motion capture. This includes real-time body capture, live on-set demonstrations, simultaneous facial capture, and the elimination of the need for additional performance time. Additionally, the team behind the film has made significant improvements to the identification system. They have achieved near-perfect capture of actors' performances by installing approximately 50 light sources on their faces. This combination of advanced technology utilizing Solver software and skin line manufacturing, has been instrumental in creating lifelike animated gorilla characters. "Rise of the Apes" is not merely a confirmation to the advancements in motion capture technology; it also serves as a means of showcasing the immense potential and value of this technology in the world of filmmaking.

With continuous development in digital film production technology, it is required to enhance motion capture technology continuously. Motion capture technology has realized a visualization way simulating technical motions in 3D way, so that animation producers can further perfect computer vision technology for effectiveness and practicability of system. [8]

Furthermore, motion capture has been used in the medical industry for a while now. Interestingly, motion capture technology was initially conceived for clinical purposes. Healthcare professionals employ motion capture to assess the movements of individuals with mobility-limiting conditions, such as cerebral palsy. The patient's movements are recorded and subsequently replicated, enabling doctors to closely analyze their gait and walking patterns. This detailed examination aids in formulating appropriate treatment strategies for the respective condition. [9]

Today, companies are taking innovative steps towards preventive healthcare solutions by harnessing the power of motion capture technology. An example of this is Velocity EHS, a U.S.-based company specializing in environmental, health, and safety (EHS) software. They have developed sensorless motion-capture assessment solutions designed to evaluate the ergonomic quality of workspaces. Following this assessment, users receive guidance on how to reduce the risk of musculoskeletal disorders (MSD), which are the most prevalent category of workplace injuries. This guidance is based on extensive data collected from millions of data points related to MSD risks, providing valuable insights for improving workplace safety and health. [9]

Motion capture and tracking technologies, when integrated with virtual reality (VR), have found valuable applications in the military sector, particularly in the assessment and treatment of injuries among military personnel. A notable example is The Stanford Hall Rehabilitation Centre in the UK, where researchers employ motion capture to precisely study human movements, helping patients understand how their injuries affect their mobility. Using this information, patients can continue their rehabilitation even after leaving the center. Another institution, the Defence Medical Rehabilitation Centre (DMRC) located in Loughborough, England, utilizes 3D motion capture cameras in conjunction with the Computer Assisted Rehabilitation Environment (Caren) machine. This setup allows for the simulation of patients' movements across various terrains. The motion capture cameras play a crucial role in enabling the medical team to analyze muscle engagement, providing patients with a comprehensive visual representation of their weight distribution and other critical factors during their rehabilitation process. This technology enhances the effectiveness of rehabilitation efforts for military personnel recovering from injuries. [9]

Sports is another industry where motion capture technology is extremely useful. Sports science research uses inertial motion capture technology that combines—an accelerometer (measures the speed of movement), a gyroscope (measures the change in rotational angle) and a magnetometer (measures magnetic field). [9]

These performance metrics derived from motion capture technology serve as invaluable tools for athletes and coaches alike. Athletes can use these data to understand precisely how they can push themselves beyond their limits and strive for excellence in their respective sports. It provides them with insights into areas where they can improve and fine-tune their skills to become the best in their field. From a coach's standpoint, motion capture offers a strategic advantage by providing a deeper understanding of each player's strengths and weaknesses. This knowledge allows coaches to develop more effective strategies for positioning players on the field. By capitalizing on their strengths and addressing their weaknesses, coaches can optimize team dynamics to achieve the best possible results. Ultimately, motion capture technology contributes to enhancing the overall performance and competitiveness of athletes and teams. [9]

One of the companies working within this space is the Australian startup VR Motion Learning. The company has created a VR tennis application called Tennis Esports that will be capable of tracking players' movements and giving them feedback on how they can improve their performance. Another noteworthy company contributing to the improvement of athletic performance is Playermaker, a UK-based football tech company. Playermaker has designed specialized sensors that can be installed on footwear. These sensors offer coaches valuable insights into an athlete's skills, including areas for improvement. Additionally, they provide information on injury risk assessment and strategies for quicker recovery process in case of injury. This technology aims to optimize athlete performance while prioritizing their safety and well-being. [9]

As tracking and motion capture technology advances and becomes cheaper and more accessible, it will become even more useful in understanding human movement. This can give us high-quality experiences in films, TV shows and video games. We are already seeing the advent of devices, like Stretch Sense's motion capture gloves to measure accurate hand movement for character animation in games. But as our understanding of human movement improves, we would be better able to use our physical abilities to their maximum potential. [9]

5. Analysis

In this chapter, I will make analysis of technologies used for achiveing the project goal.

5.1. Hand tracking

Hand tracking is made using CVZone, computer vision package that uses OpenCV and MediaPipe libraries and makes it easy to run image processing and AI functions [8]. MediaPipe is a framework mainly used for building audio, video, or any time series data. With the help of the MediaPipe framework, we can build very impressive pipelines for different media processing functions. Basically, the MediaPipe uses a single-shot palm detection model and once that is done it performs precise key point localization of 21 3D palm coordinates in the detected hand region[11].

The MediaPipe pipeline utilizes multiple models like, a palm detection model that returns an oriented hand bounding box from the full image. The cropped image region is fed to a hand landmark model defined by the palm detector and returns high-fidelity 3D hand key points[11].

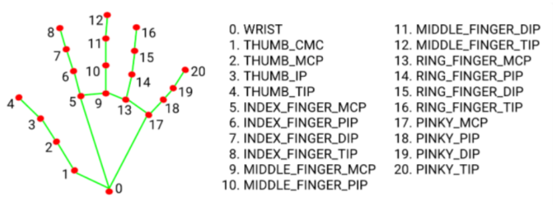

Figure 2. Hand keypoints (source: https://google.github.io/mediapipe/solutions/hands.html)

5.1.1. Implementation

```
import cvzone
import cv2

cap = cv2.VideoCapture(0)
detector = cvzone.HandDetector(maxHands=1, detectionCon=1)

while True:
    success, img = cap.read()
    img=detector.findHands(img)
    lmList, bbox = detector.findPosition(img)

    cv2.imshow("image", img)
```

CVZone and CV2 are imported. The code initializes a video capture object „cap" using „cv2.VideoCapture(0). This sets up the webcam as the video source, where 0 refers to the default camera index. An instance of the „cvzone.HandDetector" class is created (finds hands using MediaPipe library), assigned to the variable „detector". The „maxHands" parameter is set to 1, indicating that the code expects to detect only one hand. The „detectionCon" determines the confidence treshold for detecting a hand. A code enters a „while" loop to continously read frames from a webcam, detect and track hand, and display the processed frames. Inside the loop, „cap.read" reads the current frame from camera and stores it in „img" variable. The „success" variable indicates whether the frame was successfully read. The „detector.findHands(img)“ method is called to detect and draw hand landmarks on the „img“. This method overlays the detected landmarks on the input image and returns the modified image. The „detector.findPosition(img)“ method is called to obtain the coordinates of the hand landmarks („lmList“) and the bounding box („box“) surrounding the hand in the image. The modified „img“ with detected landmarks and bounding box is displayed using „cv2.imshow(„image“, img)“.

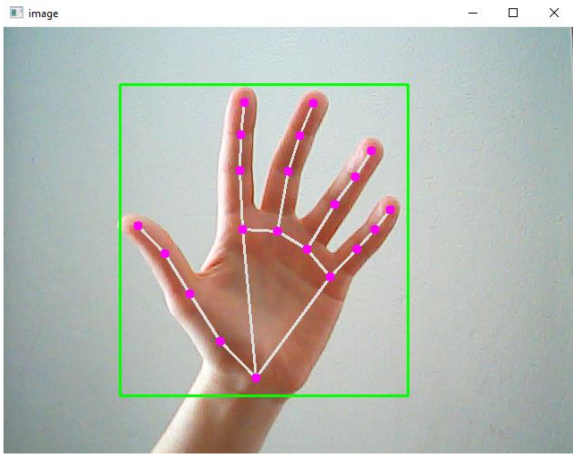

Figure 3. Code output

5.2. Processing obtained data

After detecting hand landmarks, we need to process them in a way to simply present the movement performed. I will present two approaches: using neural networks and by calculating distance between fingers.

5.2.1. Action recognition using LSTM Deep Learning Model

Action recognition using LSTM deep learning model is a technique that uses Long Short-Term Memory (LSTM) network models, which are a type of recurrent neural network, to learn and remember long sequences of input data. LSTM models have been shown to provide state-of-the-art results on challenging activity recognition tasks with little or no data feature engineering [13].

Firstly, a labeled dataset of videos is required, where 3 different actions are presented with 30 videos each. For this project, I used two types of actions presenting two exercises for a patient. Videos are captured from web camera and every video contains 30 frames. Movement estimation is made with MediaPipe Holistic model.

```
model = Sequential()
model.add(tf.keras.layers.LSTM(64,
return_sequences=True,
activation='relu',
input_shape=(30,1662)))
model.add(tf.keras.layers.LSTM(128,
return_sequences=True,
activation='relu'))
model.add(tf.keras.layers.LSTM(64,
return_sequences=False,
activation='relu'))
model.add(tf.keras.layers.Dense(64,
activation='relu'))
model.add(tf.keras.layers.Dense(32,
activation='relu'))
model.add(tf.keras.layers.Dense(action
s.shape[0], activation='softmax'))

model.compile(optimizer='Adam',
loss='categorical_crossentropy',
metrics=['categorical_accuracy'])
model.fit(X_train,            y_train,
epochs=2000)
```

Sequential model from Tensorflow allows you to build a neural network by adding layers in a sequential order. Six layers are added. An LSTM layer with 64 units, configured to return sequences. It uses ReLu activation function, input shape of the layer is 30 sequnces length and 1662 features in each input. Second LSTM layer has 64 units, also configured to return sequences and uses ReLu. Third LSTM layer with 128 units is configured not to return sequences. Fourth and fifth layer are fully connected dense layers with 64 and 32 units and apply ReLu activation function. Last layer is output layer, which is a fully connected dense layer with a number of units equal to the number of actions (in this case 2). It uses the softmax activation function, which produces a probability distribution over the different action classes. Model compiles with Adam optimizer and categorical crossentropy loss function (suitable for multi-class classification problems). The metrics parameter is set to track categorical accuracy during training. Once the model is compiled, we can proceed with training it on labeled dataset using „fit()“ function.

The model turned out to be imprecise with accuracy about 50%. For this dataset, it equals to guessing the action performed. That's why this approach won't be used to realize this project.

5.2.2. Finger distance calculation

For now, we have only an array containing 21 arrays that present coordinates of every hand keypoint. The idea is to calculate distance between finger tips and wrist. Distance would be categorized in 3 classes: 0 – finger is closed, 1 – finger is partially closed and 2 – finger is open.

```
def Distance(p1,p2, lmList):
        x1, y1 = lmList[p1][1], lmList[p1][2]
        x2, y2 = lmList[p2][1], lmList[p2][2]
    length = math.hypot(x2 - x1, y2 - y1)
    return length
```

This is a function that returns distance between two keypoints by subtracting x and y coordinates of two keypoints. For example, calculating distance between index finger and wrist would be done by passing into Distance function: 0 (for wrist), 8 (for index tip) and, already mentioned, „lmList" list of landmarks. Then, we need to set approximate values for separation into classes e.g. if the distance is greater than 260, then finger is open (class 2), else if distance is greater than 130, then finger is partially closed (class 1), else finger is closed (class 0). Same process goes for all fingers. Finally, this will result with one array containing five numbers of zeros, ones or twos.

```
lengthIndex = Distance(0,8,lmList)
        if lengthIndex > 260:
            fingers.append(2)
        elif lengthIndex > 130:
            fingers.append(1)
        else:
            fingers.append(0)
```

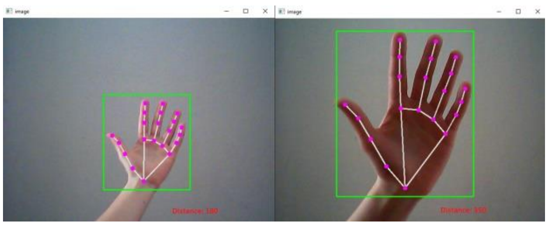

Figure 4. Distance difference

Problem of this approach is that the length between keypoints depends on distance between hand and camera. The farther the hand is, the shorter the length is. So, for this prototype, I will mark rectangle within the specified values will be valid if the hand fits right in.

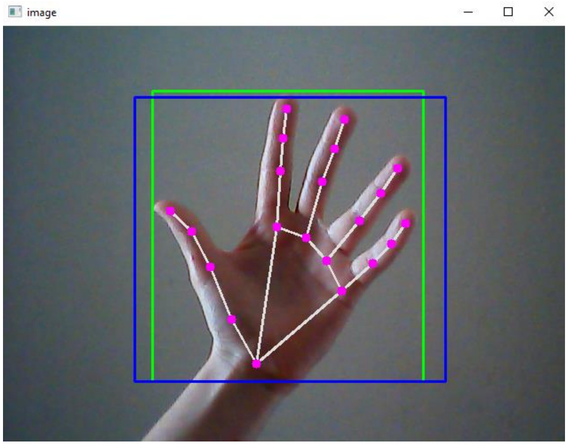

Figure 5. Hand needs to fit blue rectangle

6. Robotic hand

InMoov is an open-source 3D printed project developed by Gael Langevin, a French sculptor and designer. It is a humanoid robotic hand that can be built using 3D printing technology and off-the-shelf electronic components. The goal of the InMoov project is to provide a low-cost, accessible platform for individuals interested in robotics and prosthetics [12].

InMoov hand will be used for this project together with Arduino (for communication between program and hand) and servos (for fingers movement). Hand comes with certain limitations: no finger rotation is possible and it can't move precisely as the human hand (that's why we use only 3 classes of movement).

3D model is downloaded from the official InMoov page.

6.1. Sending data to robot

CVZone has a SerialObject class that allows to transmit data to a serial device (in this case Arduino).

```
mySerial= cvzone.SerialObject("COM3", 9600, 1)
mySerial.sendData(fingers)
```

„COM3" is a parameter that specifies the port or communication channel to which the serial device is connected. In this case, it is set to „COM3", which is a common notation for a serial port on Windows systems. The actual port name may vary depending on specific hardware and operating system. Parameter „9600“ sets the baud rate for the serial communication. Baud rate refers to the rate at which data is transferred over the serial connection. In this case, the baud rate is set to 9600 bits per second, which is a commonly used standard baud rate. The last parameter sets the timeout value for reading from the serial port. It represents the maximum time (in seconds) that program will wait for data to be received from the device. Here, it is set to one second, indicating that if no data is received within that timeframe, the read operation will time out. „sendData“ sends array of fingers based on which the robot will make moves.

6.2. Receiving data

```
void receiveData() {
  int i = 0;
  while (Serial.available()) {
   char c = Serial.read();

    if (c == '$') {
      stringCounterStart = true;
    }
    if (stringCounterStart == true )
    {
              if    (stringCounter    <
stringLength)
      {
                  myReceivedString    =
String(myReceivedString + c);
        stringCounter++;
      }
             if    (stringCounter    >=
stringLength) {
                stringCounter   =   0;
stringCounterStart = false;
                      servoPinky     =
myReceivedString.substring(1,
2).toInt();
                       servoRing     =
myReceivedString.substring(2,
3).toInt();
                     servoMiddle     =
myReceivedString.substring(3,
4).toInt();
                      servoIndex     =
myReceviedString.substring(4,
5).toInt();
                      servoThumb     =
myReceviedString.substring(5,
6).toInt();
        myReceivedString = "";
      }
    }
  }
}
```

Arduino receives string of five numbers, those numbers are then separated to each finger variable. If the value for a finger is 0, rotate the servo to 90 degrees to close the finger. If the value is 1, rotate servo to 30 to partially close finger. Else (value is 2), finger is open and rotate servo to 0 degrees.

```
if (servoIndex == 0){
  servoindex.write(90);
}
else if (servoIndex == 1){
  servoindex.write(30);
}
else{
  servoindex.write(0);
}
```

7. Improvements

This is a low cost and a low fidelity prototype that could be upgraded both in sofware and hardware part. Program could be modified to work on every hand and at any distance from camera. As for robot goes, the model could be done with more precise movements so that we don't have only three types of movement but any possible.

8. Results

We have a functional prototype of robotic hand that can imitate human's hand movements in real time. Movements are classified in 3 categories, so servos rotate in three different degrees. Classes are gained by calculating distance between finger tips and wrist. For now, it works good on hands of similar size and when on approximately equal distance from camera.

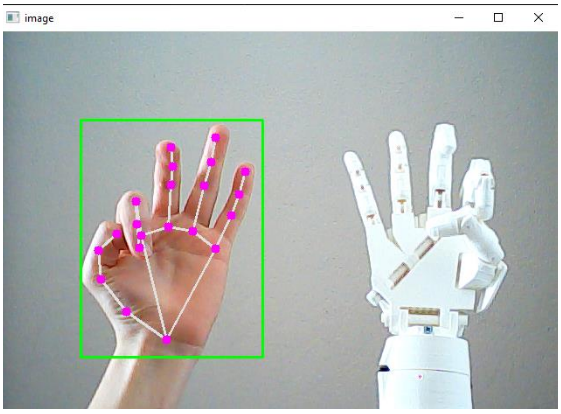

Figure 6. Robot hand imitates human hand

References

[1] Massachusetts Institute of Technology (MIT) News. "MIT researchers develop an inflatable robotic hand with tactile sensors." [Online]. Available: https://news.mit.edu/2021/inflatable-robotic-hand-tactile-0816

[2] A. Sivakumar, K. Shaw, D. Pathak(2022). Robotic Telekinesis: Learning a Robotic Hand Imitator by Watching Humans on YouTube. Available: https://arxiv.org/pdf/2202.10448.pdf.

[3].S. Suvitha, C. Lillygrace , M.Nandhini, S. Nithyashree (2019). Implementation of Robotic Hand Imitating the Real Time Motions. Available: http://www.ijirset.com/upload/2019/march/81_JOURNAL%20PAPER%20ROBOTIC_NEW.pdf.

[4] Weichert, F.; Bachmann, D.; Rudak, B.; Fisseler, D. Analysis of the Accuracy and Robustness of the Leap Motion Controller. Sensors (2013).Available: https://www.mdpi.com/1424-8220/13/5/6380.

[5] "MediaPipe Hand Tracking." [Online]. Available: https://developers.google.com/mediapipe/solutions/vision/hand_landmarker.

[6] Yu, J., Qin, M. & Zhou, S. Dynamic gesture recognition based on 2D convolutional neural network and feature fusion (2022). Available: https://www.nature.com/articles/s41598-022-08133-z.

[7] IEEE Xplore. "A Low-Cost Real-Time 3D Hand Tracking System for Cyber-Physical Systems." [Online]. Available: https://ieeexplore.ieee.org/document/7539037.

[8] M. Zhang Proceedings of the 2nd InternationalSymposium on Computer, Communication, Control and Automation (2013).

Available: https://www.atlantis-press.com/proceedings/isccca-13/5878.

[9] Jumpstart Magazine. "The Unique Uses of Motion Capture Technology." [Online]. Available: https://www.jumpstartmag.com/the-unique-uses-of-motion-capture-technology/.

[10] GitHub. "cvzone - Computer Vision Library for OpenCV." [Online]. Available: https://github.com/cvzone/cvzone.

[11] Analytics Vidhya. "Building a Hand Tracking System Using OpenCV." [Online]. Available: https://www.analyticsvidhya.com/blog/2021/07/building-a-hand-tracking-system-using-opencv/.

[12] InMoov. "InMoov - The First Open Source 3D Printed Life-Size Robot." [Online]. Available: https://inmoov.fr/.

[13] Machine Learning Mastery. "How to Develop RNN Models for Human Activity Recognition Time Series Classification." [Online]. Available: https://machinelearningmastery.com/how-to-develop-rnn-models-for-human-activity-recognition-time-series-classification/.