Ms. Kirti Sahu2023, 11:6
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology

An Open Access Journal

Real Time Hand Gestures Recognition Using Convolutional Neural Network

M.Tech Scholar Ms. Kirti Sahu, Prof.Dr. Ashish Kumar Khare
Department of Computer Science &Engg.
Lakshmi Narain College of Technology & Science, Bhopal,MP,India

Abstract-HCI (Human-Computer Interaction) is a huge area that encompasses a variety of interactions, including gestures. In HCI, gesture recognition refers to nonverbal motions that are employed as a way of communication. A system could be used to detect and recognize gestures and use them to communicate information for device control. This is a significant field in HCI that deals with device interfaces and users. Gesture recognition is the process of recording gestures that are created in a specific way and then being detected by a device such as a camera. Hand gestures can be used for a variety of purposes as a means of communication. People with various disabilities, such as those with hearing impairments, speech impairments, and stroke patients, can utilize it to communicate and meet their fundamental needs. Hand gestures have been the subject of numerous studies in the past. Different strategies for implementing hand gesture tests were proposed in some publications. There are a variety of methods for extracting characteristics from photos, as well as Artificial Intelligence (AI), which offers a variety of classifiers for classifying various sorts of data. This study examines the problem using a variety of algorithms. This study used image processing technologies including Wavelet Transforms and Empirical Mode Decomposition to extract picture features in order to detect 2D hand movements.

Keywords- Python, NumPy, TensorFlow, Tflearn,Keras, Convolutional Neural Network, Training, Classification.

I. INTRODUCTION

A Gesture is defined as the physical movement of the hands, fingers, arms and other parts of the human body through which the human can convey meaning and information for interaction with each other [1]. There are two different approaches for human–computer interactions, the data gloves approach and the vision-based approach. The vision-based approach was investigated in the following experiments including, the detection and classification of hand gestures. A Hand gesture is one of the logical ways to generate a convenient and

high adaptability interface between devices andusers. Applications such as, virtual object manipulation, gaming and gesture recognition be used in HCI systems. Hand tracking, as a theory aspect, deals with three fundamental elements of computer vision: hand segmentation, hand part detection, and hand tracking. The best communicative technique and the common concept used in a gesture recognition system is hand gestures. Hand gestures can be detected by one of these following techniques: posture is a static hand shape ratio without hand movements, or a gesture is dynamic hand motion with or without hand movements. Using any type of camera will detect any type of hand gesture;

© 2023Ms. KirtiSahu. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited.

keeping in mind that different cameras will yield different resolution qualities. Two- dimensional cameras have the ability to detect most finger motions in a constant surface called 2D [2].

A Human Computer Interaction (HCI) has several types of interaction and one of those is called gestures. One simple definition of a gesture is a non-verbal method of communication utilised in HCI interfaces. The high target of gesture is to design a specific system that can identify human gestures a designedly and use these gestures to convey information for device control. Recently, HCI has increased in relevance as its usage increases across different applications including human motion acquisition. Initially, it must define the idea of human motion acquisition which records the movements of a human or an object and convey them as 2D image data. For instance, Disney is one of most famous companies which uses human motion capture as a modern technology in the cartoon world production.

II. AIM AND OBJECTIVES

The aim of the research is to develop a system for hand gesture recognition using any type of camera, background, illuminations or position of hand, by finding the most appropriate algorithms to implement the system and test the validation of system. This system helps individuals with special needs and people who have experienced stroke to communicate accurately.

Using WT and EMD algorithms for feature extraction and AI for classification provides different results while CNN provides an accurate result. The objectives of the research include investigations, experimentation and development of appropriate algorithms for hand gesture recognition. The main objectives are highlighted below: Study the concept of gesture recognition, detection phases and algorithms used in gesture recognition detection such as WT and EMD for feature extraction and AI and CNN for classification.

III. LITERATURE REVIEW

This presents some current studies regarding different techniques used for gesture recognition. It provides a list of journal and conference papers which proposed the problems faced by researchers and the proposed solutions.

Professor Gabriel M. Lippmann
Proposed the use of micro lenses array at the image surface [3]. He presented this concept to the French Academy of Sciences as La Photography Integral. The spatial image with full parallax in all directions recorded completely by Professor Lippmann and is known as a fly's-eye lens array. Essentially, the display system was a screen containing a number of small lenses. During the 1920s, some scientists, such as Herbert Ives, started to think about simplifying Lippmann's concept by joining a lenticular lens sheet, which contains a signaller array of spherical lenses known as lenticules. It was designed to view different angles of images to provide a pixel from each micro picture.

Wang et al. [4]
Combined integral imaging and augmented reality and named it 3D augmented reality micro integral imaging display system. It had greatly enhanced the viewer's perception of reality. There exist two advantages of this system; First one is it facilitates 3D augmented reality display, and second it has a compact design.

Aggoun et al. [5]
Developed 3D holoscopic video systems for 3D TV applications. They used a field lens and a square aperture. Self- Similarity (SS) coding approach outperforms the standard High Efficiency Video Coding (HEVC) scheme and they also explore that search and retrieval effectiveness relies on the depth map's quality. A combined state build-up estimation was extended to perceive a larger number of motions dependent on their momentary directions.

Fatah et al. [6]
Targeted the digital refocusing where they proposed a method. This method uses Michelson

contrast formula to extract all-in-focus images. The highest contrast values at different points in space can return the focused points where the objects are initially positioned, which make it possible to obtain all-in-focus image. . It requires a new system which effectively separates the equivalent situated pixels under each Exposure Index. Oliveira presented in their thesis a 2D image extraction technique for holoscopic 3D images. They named it Disparity Assisted Patch Blending which outperforms existing methods. The second contribution in their research work is the identification of potential non-reference image quality assessment metrics. These metrics are able to measure 2D image extraction to extractions to compare with human perception.

Rash et al. [7]

Were among the first to investigate 3D hand motions. They performed 3D video motion analysis in order to measure hand motions. The goal was to illustrate the validity of this technique by contrasting it with a two-dimensional movement, considered the 'highest quality level'. Their investigation was performed to decide if (1) markers put on the dorsal part of the hand and fingers precisely measure joint edges, and (2) the 3-D method for evaluating finger movements is accurate by utilizing a standard movement investigation framework.

Hue et al. [8]

Then tracked fingertip positions using two stereo cameras in order to minimise the error between mapped and measured fingertip positions, they constrained an Inverse Kinematics (IK) solver. The user wears the data glove and moves fingers openly while the wrist is fixed on the table. The vision framework records a progression of finger movements and measures the specific fingertip places of each finger. Simultaneously, the uncalibrated crude sensor estimations of the data glove are additionally recorded. Teleoperation tests show that the human hand model offers adequate precision for teleoperation task. Their experimental results show a degree of error of less than 5 mm.

Sridhar et al. [9]

Used a different approach to recover poses from depth sensor data. First, they extracted the hands by filtering the data and then applying principal component analysis to resolve the orientation. Signaller SVM classifier was used to classify the fingertip locations. After all those steps, a pose estimation algorithm was used. The algorithm was able to match fingertip locations to poses in a database with similar fingertip locations.

Majkowska et al. [10]

Capturing body and finger motion is a challenging task due to the size of motion differences and markers. They suggested two sessions for capturing hand and body motions, where the finger motion was recorded in a smaller area where the subject remained standing or seated. Four markers were placed at the hand, wrist and forearms. The position of the markers allows for later alignment of the hand and body motions. A three-step algorithm was proposed. In the first step, stroke, hold and retraction was matched using acceleration, a velocity profile based DTW. In the second, step frames were aligned to the frames of the full body motion. In the last step, a smoothing of the resulting motions for seamlessly fitting together was applied.

Van den Noort et al. [11]

Proposed a system they named PowerGlove. This new system has multiple miniature inertial sensors with an opt-electronic marker system. This 3D measurement system can measure the finger motion while performing finger tasks. The subjects perform different finger tasks such as flexion, fast flexion, tapping, hand open/close and circular pointing. The median root and mean square difference for all the finger task presented above was then calculated. It revealed that fast and circular pointing tasks have the largest differences while the smallest differences were observed in flexion tasks. This system measured the 3D hand and finger kinematics and their position in an ambulatory setting. These results may help in hand function and quantifying hand motor symptoms in clinical practice.

Krupicka et al. [12]
Also developed a measurement tool for objective measurement of the finger tapping test. A contactless 3D capture system using two cameras and wireless reflexive markers were used in the measurement system. An algorithm for extracting, matching and tracking markers was proposed. They compared the performance of their system with OptiTrack, a commercial motion capture system. Human beings communicate with each other through voice communication or speaking certain languages. Beside voice communication, hands are another means of communication between hearing impaired people. Hand gesture recognition is an interesting topic among researchers from different fields, such as computer vision, human computer interaction and image processing.

IV.METHODOLOGY

Users interact with computers through the provided interfaces, motions or vocal. These different interactions need to be such that information retrieval is easier and Human Computer Interaction (HCI) is concerned with the way humans interact with technology. It deals with how humans work with computers and how computer systems can be designed to best facilitate the users in achieving their goals. With the advent of third and fourth generation languages, the user interfaces have improved quite dramatically. In future days, Human Computer Interaction HCI will become a field with a variety of sectors that need to characterize it. Users will be able to use any type of interaction which is a potential part of HCI, Interaction can be body movements, facial features and vocals.

1. Gesture Recognition

In the present day, HCI is assuming greater significance in our daily lives. Gesture recognition can be named as a method a long this path. Therefore, what is gesture recognition? In the previous section Gesture Recognition were defined as non-verbal motions used as a method of communication in HCI interfaces. Gestures are one of the significant aspects of HCI in both interpersonally and in the device interfaces. Another definition of gestures is physical movements or positions of a human's fingers, hands, arms or full body used to convert information. Gestures, in a virtual reality system can be used to navigate, control or interact with a computer. The process by which gestures are formed in certain ways by a person, are made known to a system, is the main principle of gesture recognition. Signs can be expressed in a multitude of ways by gestures, for example, sign language used by hearing impaired people. Other examples of gestures developed outside the computer field can be seen in use by traffic police, construction labours, and airport ground controllers. Gestures can be static, which means that the user adopts a pose, or dynamic where the motion is a gesture by itself.

2. Types of Gesture Recognition

Gesture recognition has been introduced briefly in the previous sections. The gestures are made by the user then are recognised by the receiver. It is for the meaningful body motions including movements of the fingers, hands, arms, head, face, or body to convey meaningful information. In this section, some essential types of gesture recognition are addressed briefly. Hand gesture recognition is one of the understandable ways to generate a convenient, high adaptability interface between devices and users. Using a series of finger and hand movements through the operation of complex machines is allowed by hand gesture recognition.

3. Hand Gesture Recognition

The hand is often well known as the most natural and instinctive interaction for humans' interaction. In the HCI world, an appropriate hand tracking is the first phase to develop instinctive HCI systems that may be used in applications such as, virtual object manipulation, gaming and gesture recognition. Moreover, hand tracking is an interesting principle point which deals with three main parts of computer vision which are segmentation of hand, detection of hand parts, and tracking of the hand. Hand gestures are frequently the most expressive way and the most used in gesture recognition system involving a posture is a static finger shape ration without hand motion and a gesture which is dynamic hand motion with or without finger movements.

A hand gesture requires tracking of 27 degrees of freedom of hand including two major categories, a hand posture is a static hand pose without any movements; while hand motion is any movement of the hand, either the full hand or fingers. A hand movement consists of three major types are data-glove based, vision based and electrical field sensing. Measuring the human body or body parts requires electrical field sensing, and this device is used officially to measure the distance of human hand or other body part from a device. Currently, most of the significant types almost all researchers are interested in studying, are data-glove-based and vision-based technologies. The data glove based is simply a glove that has multi variety of sensors used to detect hand and finger motions. There are many styles of data glove and each one has its uses, such as MIT Data Glove, CyberGlove III, CyberGlove II, Fifth Dimension Sensor Glove Ultra, X-IST Data Glove and P5 Glove.

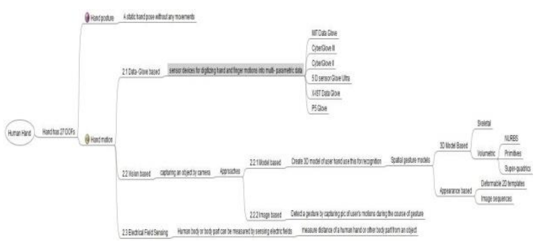


Figure 1: Hand Gesture Recognition Map

4. Data Glove

The history of hand gesture recognition began with the invention of data gloves. Some researchers understood that sign language inspires gestures and that it may be used to suggest simple instructions for a computer. A data glove is a special wired glove with tactile switches or sensors which attach to the fingers or joints of the glove and is worn by a human. Optical goniometers and the tactile switches or resistance sensors estimate the twisting of dissimilar joints when present with basic measurements and that determines if a hand is opened or closed or some finger joints are straight or twisted. A computer is provided results which are mapped to exceptional gestures and interpreted.

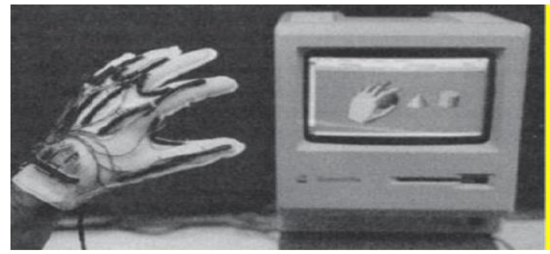


Figure 2: Data Glove

5. Vision Based Systems

Gestures recognition is one of the most natural communicative methods between human and computers in virtual environments. Camera techniques are used to identify hand gestures. It started laterally with the early development of the first data gloves. The first computer vision gesture recognition system was reported in the 1980s. Moreover, vision-based recognition is normally natural and comfortable. As shown in Figure 3.3, a flow diagram of a normal gesture recognition plan.

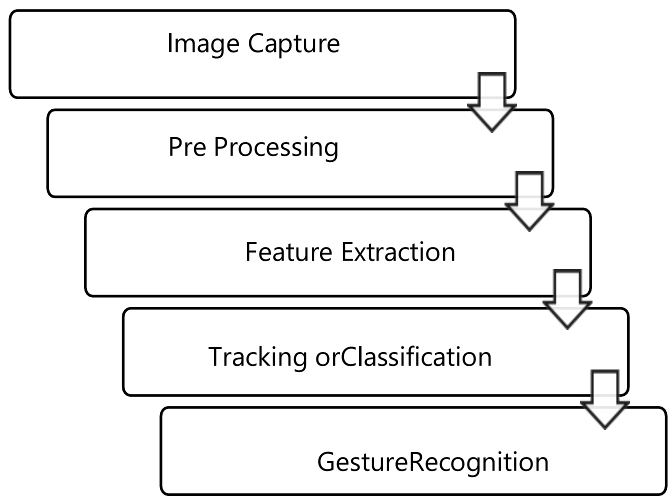


Figure 3: Typical computer vision-based gesture recognition approach

Using vision-based techniques require contending with other issues related to occlusion of user's body parts. Although tracking devices had the ability to detect movements of hands quickly while the human's body moving. Vision-based devices could grasp properties such as colour and texture for analysing a gesture, while typical tracking devices may not handle these Vision-based techniques may also differ between themselves in the number of cameras used, their speed and latency, the structure of the environment, like speed of movement and lighting, user requirements –each user must wear something unique– the low-level features used, such as region, edges, histogram, silhouette and moment; and nor 2D or 3D representation is used and either time is represented.

6. Image Processing and Recognition

Image processing is a way to transform a photo into a digital form and perform other operations on it in order to produce an improved image or gain useful information from it. The input is an image, such as a video frame and the output can also be an object or picture. This is a type of signal propagation. The system of image processing contains different tools such as image acquisition, image enhancement, image restoration, colour image processing, wavelet and multiresolution processing, segmentation and object recognition. Image restoration is storing specific parts of an image using a Point Spread Function. The colour image processing tool is used when the image is black and white. Wavelets and multiresolution processing are used if the images are to be rendered in different degrees or wavelets of resolution. Compression reduces the size of images using specific function. Morphological processing is an external structure of the image using dilation and erosion. Segmentation is implemented by splitting an image into different parts. Object recognition used to recognise and save the image description.

7. Video Processing

Video processing is based on video data analysis of an allocated time within a video to complete a desired process. Video processing is also a key part of signal processing that transforms the video signal's characteristics, enhances or degrades the video quality, converts the video to a communication channel and storage media etc. Compressed, encoded, and modulated video signal is received from the channel of communication or from the storage device. The signal is extracted, separated, amplified and shifted in the frequency domain if necessary, from other channel or storage signals. It is then demodulated and converted into a virtual stream. There may be one or more compressed video streams, audio channels and data channels in the optical stream.

8. Artificial Intelligence

AI is the ability of a machine to perform cognitive tasks and act intelligently. The field of AI tries to understand intelligent entities. AI is a new discipline that began in 1956. With a help of AI, it is possible for machines to learn from their own experience, adapt to new inputs and perform human-like tasks. AI is widely used in finance, education, healthcare, transportation fields and in other industries such as computer vision, medical diagnosis, robotics and remote sensing. The father of computer science and AI is Alan Turing who proposed a 'Turing test' in 1950 which was designed to provide an operational definition of intelligence. If a machine passes this Turing test, it is said to be intelligent. But no machines have completely passed this test as of yet. There are other indicators of intelligence such as Intelligence Quotient (IQ) tests and brain size, but none of them convey intelligence in machines. According to Daniel Gilbert, there is one fundamental element in which our minds differ from the minds of animals and computers; it can experience something that has not yet happened.

9. Artificial Neural Network

ANN is defined as an interconnected assembly of nodes like the neural structure of the human brain and can solve different types of problems in an easy manner. The brain works by learning from experiences. ANN is a system that processes information in a similar manner to the biological nervous system. The key aspect of this system is the unique structure of the information processing system. The system is composed of a large number of unified processing elements working together to solve certain issues. It is specifically configured for data classification or pattern recognition applications via learning processes. The architecture of an ANN is composed of three main layers including an input layer, the hidden layer (one layer or more) and the output layer. ANN can be trained using a supervised or unsupervised approach. In a supervised approach, ANN is simply trained by matched input and output while the unsupervised approach is an attempt to obtain the ANN to realize the structure of input data. There are several benefits associated with using ANN such as self-learning and large data handling. The advantage of using an ANN is ANN has the ability to learn and train data models for non-linear and complicated relationships. Different applications may be used by an ANN such as image processing, object detection and forecasting.

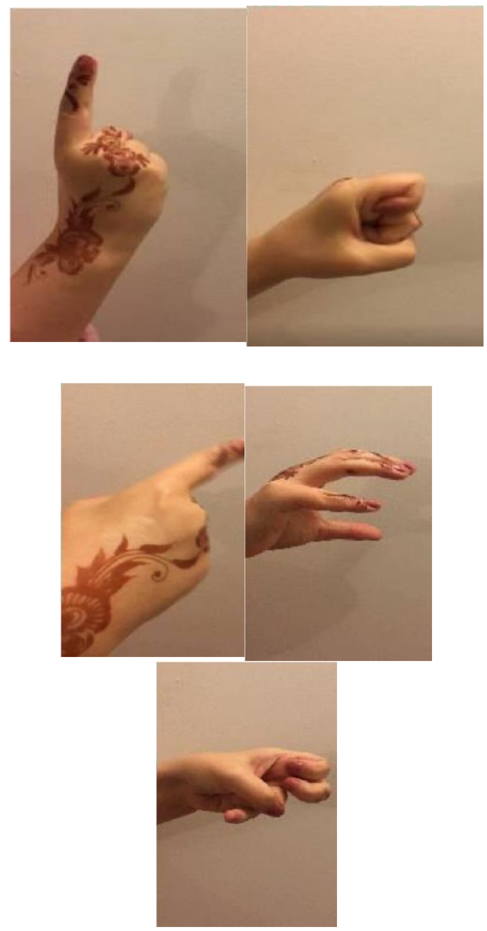
10. Convolutional Neural Network

A Convolutional neural network (CNN) is a type of artificial neural network specifically designed for image recognition. A neural network following the activity of human brain neurons is a patterned hardware and/or software system. CNN is also defined as a different type of multi-layer neural network and each layer of a CNN converts one amount of activations to another through a function. CNN is a special architecture used for deep learning. CNN can be categorised in two phases, namely Training and Inference. To build a CNN- based architecture, it applies three key types of layers: Convolutional Layer, Pooling Layer and the Fully Connected Layer. The first layer is a convolutional layer which is the main block of CNN. It takes many filters that are applied to the given image and creates different activation features in the picture. The second layer is pooling which is used to down sample. It will obtain input from non-linear activation and the output will depend on the window size. The last layer is fully connected where a target is identified to determine the category of final output. Due to the three layers, which removes the necessity for feature extraction by using image processing tools, the image data is learned directly by CNN. CNN causes the recognition results to be unique and it might be retrained easily for new recognition missions while it is allowed to build on the pre-existing network. All the following factors have made the usage of CNN significant in the last few years.

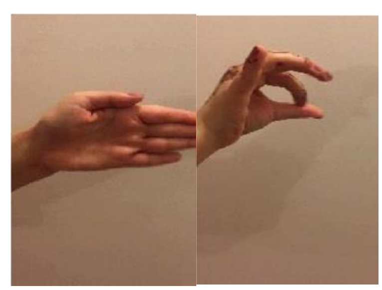
V. IMPLEMENTATIONS

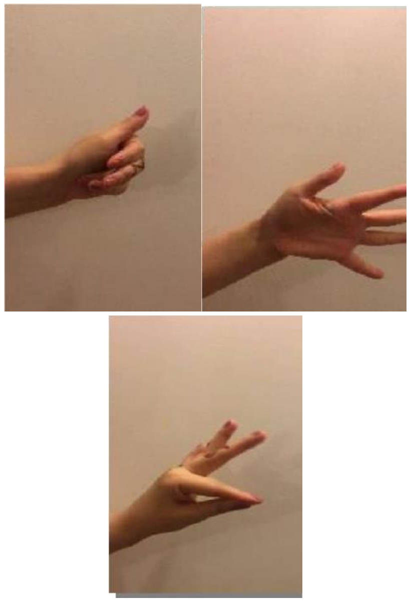
1. Hand Gestures Input

hand gestures are the input to different gesture detection algorithms. Figure shows 10 various hand gestures which are recorded in short distance with a plain background that is used in this experiment. Some motions are 2D while others are 3D.



1. Sweep motion 2. Shrink motion 3. Circular motion 4. Squeeze motion 5. 2 Fingers Shrink





1.Back/Forth 2. Rub motion 3. Click motion 4. Dance motion 5. Pinch motion

Figure 4: Different hand gestures.

Feature Detection using Wavelet Transforms Algorithm

The system is implemented using the db8 WT tool into the following stages:

1. Read each video using video reader function
2. Create optical flow object that spreads the object velocities in an image.
3. Estimate and display the optical flow of objects in the video.
4. Divide a video into certain frames; each frame contains 8 IMFs.
5. Apply appcoef2 function which is used to compute an approximation coefficient of 2D signals.
6. Extract each level using wrcoef function to reconstruct the coefficients of each level in the video.
7. The execution time of WT is estimated only once.
8. The image data is trained and tested using a Neural Network system
9. The execution time of image data training and testing are also calculated.

2. Empirical Mode Decomposition Algorithm

The implementation of EMD is similar to WT with some variances. The same 4 steps of WT are used, but with different functions: reshape function (returns the M-by-N matrix) whose elements takes column-wise from X, ceemd function (a noise improved data analysis algorithm) complementary collaborates with EMD.

3. Implement Convolutional Neural Network (CNN)

Deep learning has an intelligent method such as CNN, which is used to train data without requiring any image processing tool. In our experiment, we made a new directory for each video. 10 images are generated to transfer the image frame RGB to grey and resize it to 48×27. All videos have 70 frames. The image's data is split into training and testing datasets. The CNN topology is created in 7 layers; each layer has the following functionality and size: ImageInputLayer Input size, Convolution2DLayer Filter size [5,5], ReLULayer (Rectified Linear Unit), MaxPooling2DLayer Pool size [2,2], FullyConnectedLayer Input size [auto] and Output size, SoftmaxLayer and ClassificationOutputLayer Output size [auto]. The hyperparameters of the CNN is generated inside training options function. The value of max epochs parameter is set to 200 epochs.

4. Short Distance Results

The system is implemented ten times to obtain the mean of ten-hand motions. Standard deviation is a measure of how extensively values are different from the average value of the group. Two different results in training, testing represented and compared by finding the best tool to detect micro gestures. The training accuracy is achieved by implementing a model on the training data and obtaining accuracy of the algorithm, whereas the testing accuracy is accuracy for the testing data. The three provided algorithms (WT, EMD and CCN) in this chapter are characterized by different measures. The total time execution for WT in four stages (Data input, Pre-processing, Image segmentation, Feature extraction) is less than the total time execution of EMD.Table shows the summary of the values acquired against the parameters when being in training mode. The total time execution for WT, EMD and CNN in training. The execution time of WT is less than the total time execution of EMD and CNN. The accuracy results of CNN exceeded WT and EMD by acquiring the highest value.

Table 1: Comparison between WT, EMD and CNN for Training

Methods	WT+ANN	EMD+ANN	Proposed Method
Accuracy	92.75	96.50	98.25

The parameter values of CNN are constant for all categories. Its execution time is approximate 714 second which is significant and not preferred in experiments. Table compares the three algorithms performance while they were being training and tested for this study. CNN had the total execution time, higher than WT and EMD. In accuracy factor, WT achieved a value which is lower than EMD and CNN. The accuracy results of CNN surpassed WT and EMD with a high value.

Table 2: CNN for Testing and Testing approach

Epochs	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
20	81.25	0.224	75.25	0.289
40	88.61	0.107	84.25	0.206
60	89.15	0.085	90.25	0.116
80	96.33	0.038	96.00	0.098
100	99.51	0.042	98.25	0.071

5. Stroke Recognition Systems

Convolution Neural Networks (CNNs) are used to evaluate hand gesture recognition, where depth-based hand data was employed with CNN to obtain successful training and testing results. In this study, Soodtoetong and Gedkhaw present their study in process and methods related to sign language recognition using Deep Learning. 3D CNN was applied for recognising images received through Kinect sensor. The method using 3D CNN was found to be very effective and the highest accuracy was found to be 91.23% .

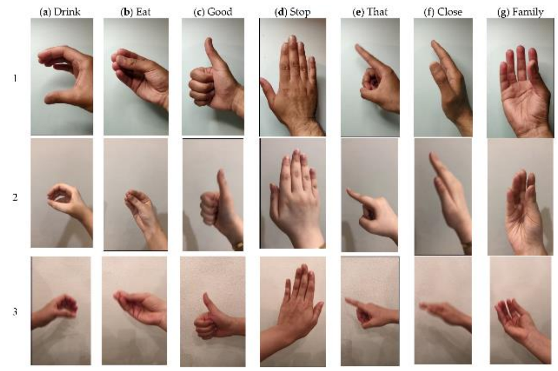


Figure 5: Three examples of seven universal hand gestures for three different hands

6. Convolutional Neural Network Implementation

CNN forms an integral part of deep learning as is applied to train data without using any image processing technique. In this experimental work, a new directory is created for each video. Hundred and forty videos are read to generate 24,698 image frames. The method to convert the image frame from RGB colour to grey and resize it to 227×227 from the original image size is shown in Figure 6.2. Each recorded video has a various number of frames between 3394 to 3670 frames. The data of images is divided into training and testing datasets. The number of training frames is 2485 which is 70%. The topology of CNN is generated in seven layers with each layer having the following functionality and size: Image Input Layer size [227,227,1], Convolution2DLayer Filter size [5,20], Rectified Linear Unit (ReLULayer), MaxPooling2Dlayer Pool size [2,2], Fully Connected Layer size [auto] and Output size[7], Softmax Layer and Classification Output Layer Output size [auto]. The CNN hyper parameters are produced inside the training options function.

7. Experimental Results

The performance of CNN algorithm is compared between training and testing using several parameters including execution time, which is the duration taken by the software to implement the task. Sensitivity measures the percentage of positives that are appropriately identified.

The experiments were executed ten times to acquire the mean of seven-hand gestures. Two different training and testing modes were presented and compared to find the best result. Training accuracy is accomplished by applying a prototype on the training data and determining the accuracy of the algorithm. A summary of the values obtained for various parameters in training and testing approach is listed in Table 6.1. It can be noticed that the execution times for training and testing are the same. The accuracy result of training is 98.25%. Accuracy in training is 99.51% whereas in testing is 98.25. Training loss is 0.042 and validation loss is 0.071. In table 6.2 there is a comparison between WT with ANN, EMD with ANN and CNN. While using CNN accuracy is greater than other approaches. The training parameter values in CNN are fixed for all categories. The execution time is approximate 16,458 seconds, which is duration to train and test the system using seven hand gestures which are used in the experiment. Overall, training has the best values in most parameters. The implementation of CNN algorithm in this study has many advantages. Firstly, CNN able to capture the features of image without any human intervention. It is capable to learn the image or video faster than ANN. CNN surpass ANN on conventional image recognition. On the other hand, the execution time of CNN is longer than ANN.

Table 3: CNN Training and testing Approach

Epochs	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
20	81.25	0.224	75.25	0.289
40	88.61	0.107	84.25	0.206
60	89.15	0.085	90.25	0.116
80	96.33	0.038	96.00	0.098
100	99.51	0.042	98.25	0.071

Table 4: comparison between WT, EMD, ANN and CNN

Methods	WT+ANN	EMD+ANN	Proposed Method
Accuracy	91.69	95.76	99.25

VI. CONCLUSION

This thesis shows two essential experiments in the field of hand gesture. The first experiment is 2D video detection which consist of detecting ten different gestures in short distance. The aim of these two experimental works is to compare different algorithms in training and testing approaches to discover the best algorithm to extract and classify hand gesture recognition. These algorithms were evaluated in terms of different parameters such as execution time, accuracy, training loss, validation loss, training accuracy and validation accuracy. After the pre-processing phase, both studies were implemented using two image processing tools which are WT and EMD.

WT is one image processing technique which performs signal analysis with one signal frequency differing at the end of time. An innovative technology used in both non- stationary and non-linear data namely EMD. The primary function of this method is decomposing a signal into intrinsic mode functions consistently through the domain. For classification, ANN is used for both experiments which is defined as a system that processes information and has structure much like that of the biological nervous system. CNN is a multilayers neural network which is one of the deep learning techniques used efficiently in the field of gesture recognition. In system implementation, WT and EMD algorithms are used to extract image features which are later fed into ANN for gesture classification. Applying CNN in both experiments reduces two phases which are image extraction and classification to one phase only. Comparing the results showed that CNN is clearly the most appropriate method to be used in hand gesture system.

VII. FUTURE WORK

After performing all the experiments and reviewing the results, the following are suggestions for future work: Extend the number of hand gestures to cover all universal common gestures like victory/peace, hungry, cold, luck and more to build a strong model for people who have experienced a stroke and people with hearing impairments. These gestures could be learnt easily to communicate better with other people. Extend gestures to include different parts of the body such as hands and lips for gesture recognition to cover universally common gestures. For instance, developing a gesture to represent drinking using both a hand and the lips.

REFERENCES

1. Y.Li, J. Huang, F. Tian, H.-A. Wang, and G.-Z. Dai, "Gesture interaction in virtual reality,"Virtual Reality & Intelligent Hardware, vol. 1, no. 1, pp. 84–112, Jan. 2019.
2. V.Pavlovic, R. Sharma, and T. Huang, "Visual interpretation of hand gestures for human-computer interaction: a review," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 677–695, 1997.
3. H.Hasan and S. Abdul-Kareem, "Human–computer interaction using vision-based hand gesture recognition systems: a survey," Neural Computing and Applications, vol. 25, no. 2, pp. 251–261, 2013.
4. A.Aggoun, E. Tsekleves, M. R. Swash, D. Zarpalas, A. Dimou, P. Daras, P. Nunes, and L.
5. D. Soares, "Immersive 3D Holoscopic Video System," IEEE MultiMedia, vol. 20, no. 1, pp. 28–37, 2013.
6. M.G. Lippmann, "La photographieintegrale," Comptes-Rendus Acad. Sci., vol. 146, pp. 446–551, 1908.
7. J.Wang, X. Xiao, H. Hua, and B. Javidi, "Augmented Reality 3D Displays With Micro Integral Imaging," Journal of Display Technology, vol. 11, no. 11, pp. 889–893, 2015.
8. [11] O.A. Fatah, "Generating stereoscopic 3D from Holoscopic 3D," in 2013 3DTV Vision Beyond Depth (3DTV-CON), 2013, pp. 1–3.O. A. Fatah, P. Lanigan, A. Aggoun, and M. R. Swash, "Digital Refocusing: All-in-Focus Image Rendering Based on Holoscopic 3D Camera," Journal of Computer and Communications, vol. 04, no. 06, pp. 24–35, May 2016.J. F. Oliveira, "2D Image Rendering for 3D Holoscopic Content using Disparity-Assisted Patch Blending," 2013.
9. G.S. Rash, P. Belliappa, M. P. Wachowiak, N. N. Somia, and A. Gupta, "A demonstration of the validity of a 3-D video motion analysis method for measuring finger flexion and extension," Journal of Biomechanics, vol. 32, no. 12, pp. 1337–1341, 1999.
10. H.Hu, X. Gao, J. Li, J. Wang, and H. Liu, "Calibrating human hand for teleoperating the HIT/DLR hand," IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA 04. 2004, Apr. 2004.
11. S.Sridhar, A. Oulasvirta, and C. Theobalt, "Interactive Markerless Articulated Hand Motion Tracking Using RGB and Depth Data," 2013 IEEE International Conference on Computer Vision, 2013.
12. A.Majkowska, V. Zordan, and P. Faloutsos,
13. "Automatic splicing for hand and body animations," ACM SIGGRAPH 2006 Sketches on - SIGGRAPH 06, 2006.
14. J.C. V. D. Noort, H. G. Kortier, N. V. Beek, D. H. E. J. Veeger, and P. H. Veltink, "Measuring 3D Hand and Finger Kinematics—A Comparison between Inertial Sensing and an Opto-Electronic Marker System," Plos One, vol. 11, no. 11, Mar. 2016.
15. R.Krupicka, Z. Szabo, S. Viteckova, and E. Ruzicka, "Motion capture system for finger movement measurement in Parkinson disease," Radioengineering, vol. 23, no. 2, pp. 659–664, 2014.
16. R.Pradipa and M. S. Kavitha, "Hand Gesture Recognition – Analysis of Various Techniques, Methods and Their Algorithms," 2014 Int. Conf. Innov. Eng. Technol., vol. 3, no. 3, pp. 2003–2010, 2014.
17. R.P. K. Poudel. "3d Hand Tracking,",ProQuest Dissertations Publishing, 2014.
18. S. Park, S. Yu, J. Kim, S. Kim, and S. Lee, "3D hand tracking using Kalman filter in depth

space," EURASIP Journal on Advances in Signal Processing, vol. 2012, no. 1, 2012.

19. C.Manresa, J. Varona, R. Mas, and F. J. Perales, "Hand Tracking and Gesture Recognition for Human-Computer Interaction," ELCVIA Electronic Letters on Computer Vision and Image Analysis, vol. 5, no. 3, p. 96, Jan. 2005.
20. P.Krejov, "Real Time Hand Pose Estimation for Human Computer Interaction," no. September 2015.
21. Ramachandra and N. Shrikhande. Hand gesture recognition by analysis of codons. Sep 9, 2007, Available: http://dx.doi.org/10.1117/12.733193. DOI: 10.1117/12.733193.
22. M.Kitagawa and B. Windsor, MoCap for artists: Workflow and techniques for motion capture. Elsevier/Focal Press: Taylor&Francis,2008.Available:https://books.google.co.uk/books/about/MoCap_for_Artists.html?id=pJFowfd5EtkC.Accessed: Oct. 30, 2016.
23. M.Ye, Q. Zhang, L. Wang, J. Zhu, R. Yang, and J. Gall, "A Survey on Human Motion Analysis from Depth Data," Lecture Notes in Computer Science Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications, pp. 149–187, 2013.
24. N.A. Ibraheem and R. Z. Khan. "Survey on various gesture recognition technologies and techniques". International Journal of Computer Applications 50(7), 2012. Available: http://search.proquest.com/docview/1032039156. DOI: 10.5120/7786-0883.
25. P.Premaratne, Human Computer Interaction Using Hand Gestures, 2014th ed. Singapore: Springer, 2014.