FOREX Publication
Open Access | Rapid and quality publishing

International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 11, Issue 3 | Pages 760-765 | e-ISSN: 2347-470X

Hand Gesture Recognition System based on 60 GHz FMCW Radar and Deep Neural Network

Daswini Nadar[1], Saista Anjum[2] and K.C. Sriharipriya[3]

[1]*Department of Embedded Technology, School of Electronics Engineering Vellore Institute of Technology, Vellore-632014, Tamil Nadu, India.; daswini.nadar2022@vitstudent.ac.in*
[2]*Department of Embedded Technology, School of Electronics Engineering Vellore Institute of Technology, Vellore-632014, Tamil Nadu, India.; saista.anjum2022@vitstudent.ac.in*
[3]*Department of Embedded Technology, School of Electronics Engineering Vellore Institute of Technology, Vellore-632014, Tamil Nadu, India.; sriharipriya.kc@vit.ac.in*

***Correspondence:** K.C. Sriharipriya; *sriharipriya.kc@vit.ac.in*

ABSTRACT- The proposed study provides a novel technique for recognizing hand gestures that use a combination of Deep Convolutional Neural Networks (DCNN) and 60 GHz Frequency Modulated Continuous Wave (FMCW) radar. The motion of a Human's hand is detected using the FMCW radar, and the various gestures are classified using the DCNN. Motion detection and frequency analysis are two techniques that the suggested system combines. The basis of the capability of motion detection in FMCW radars' is to recognize the Doppler shift in the received signal brought on by the target's motion. To properly identify the hand motions, the presented technique combines these two techniques. The system is analyzed using a collection of hand gesture photos, and the outcomes are analyzed with those of other hand gesture recognition systems which are already in use. A dataset of five different hand gestures is used to examine the proposed system. According to the experimental data, the suggested system can recognize gestures with an accuracy of 96.5%, showing its potential as a productive gesture recognition system. Additionally, the suggested system has a processing time of 100 ms and can run in real time. The outcomes also demonstrate the proposed system's resistance to noise and its ability to recognize gestures in a variety of configurations. For gesture detection applications in virtual reality and augmented reality systems, this research offers a promising approach.

General Terms: FMCW Radar, Gesture Classification, Signal Processing, Deep Neural Networks

Keywords: SCR, clutter cancellation, Range doppler heat map, feature extraction, accuracy

ARTICLE INFORMATION

Author(s): Daswini Nadar, Saista Anjum and K.C. Sriharipriya;
Received:24/04/2023; **Accepted**: 15/06/2023; **Published**: 20/09/2023;
E- ISSN: 2347-470X
Paper Id: IJEER240416;
Citation: 10.37391/ijeer.110319
Webpage-link:
https://ijeer.forexjournal.co.in/archive/volume-11/ijeer-110319.html

crossref member
CROSSREF.ORG THE CITATION LINKING BACKBONE

Publisher's Note: FOREX Publication stays neutral with regard to jurisdictional claims in Published maps and institutional affiliations.

1. INTRODUCTION

An important technology for many application sectors, including sign language recognition (SLR), is hand gesture recognition. Convolution neural networks (CNNs) are a tool in computer vision called hand gesture recognition that allows users to identify and categorize different hand movements. Complex hand gestures are used in sign languages, and even tiny hand movements can convey a wide range of meanings [1].

Numerous vision-based dynamic hand gesture detection algorithms were developed in response to this over the past ten years [1]. It is utilized in a variety of fields, including robots, virtual reality, and human-computer interaction (HCI). CNN makes it easy to find and identify hand motions in pictures and movies. The CNNs can be used to recognize and categorize various hand gestures after being trained using labeled data. Face recognition software that recognizes hand gestures can also be utilized for security and medical purposes. Systems for recognizing hand gestures with high accuracy and productivity can be developed using CNNs.

A variety of features, including hand-crafted spatiotemporal descriptors and implemented models, were utilized to classify gestures, in addition to gesture classifiers, hidden Markov models, conditional random fields, and support vector machines (SVM) [1]. Sensor-based activity recognition collects data from the accelerometer, gyroscope, radar, acoustical sensors, and some other sensors, and explores through a vast quantity of low-level sensed data for the deep high-level knowledge of human behaviors. Due to the following factors, radar-based human activity recognition (HAR) has gained a lot of interest as one of the sensor-based techniques. First of all, radar can be used in tough areas because it is resistant to light and weather conditions. Secondly, radar might secure visual privacy. The returning signals modified by the target carry a variety of time-varying range and velocity information of activities instead of recording the visual shape of the target. Thirdly, radar-based HAR is applicable in more settings since it can detect people through walls. Last but not least, radar systems are more user-friendly because they don't require any tags to be linked to people's bodies. As a result, radar has recently been used to detect human activity more frequently [2]. Recognition of hand gestures involves identifying significant motions performed by the fingers, hands, and arms. While creating a smart and effective human-computer interface, it is crucial. When it comes to remote control of electronic appliances using gesture recognition, radar sensors have several advantages over optical sensors.

Since the start of the twentieth century, touch controls have gradually substituted conventional control strategies based on interactive hardware like mice and keyboards. Touch is not the sole means of communication, though. Other media, like gestures and sounds are also feasible and tempting ways to finish conversations because they ask for no additional hardware or physical contact. Gesture recognition depends on both hardware devices and data processing algorithms. The radar system is inexpensive and resistant to ambient brightness. Interest in gesture recognition using radar has increased. However, the high computer performance and power resources of the devices must also be considered by the radar system.

This technique tracks the movements of a person's hands using Doppler radar sensors and then utilizes deep learning algorithms to extract useful information from the data. The gesture can then be recognized and translated into text or speech using this information. The usage of this technology may make it easier for those who have hearing loss to interact with hearing persons. It could also be applied to give users a more organic way of engaging with computers and other electronic gadgets [4].

A new procedure that enables the recognition and detection of several gestures in real-time at 60 GHz frequency-modulated continuous wave (FMCW), multiple-input multiple-output (MIMO), single-chip radar (SCR), is called real-time multi-gesture recognition [5]. As a result of this technology, various movements can be recognized from a single device without physical contact. It is a reliable method of understanding complex motions and has a wide range of uses in robotics, gaming, and other fields [5].

2. METHODS

Estimating an object's range, angle, and velocity will enable position tracking with regard to radar. As a result, the method of gesture recognition mostly relies on the position and movement of the hand. A radar with high resolution and high angle capture resolution is required to record such minute motions. For this reason, MIMO radars are the best solution since they include many antennas that gather the signals and add the inputs in an incoherent manner before turning the combined data into range-doppler heat maps using 2D-FFT. Many features needed for the recognition of gestures can be extracted from these heat maps.

The extraction of features is the method used most frequently to solve classification problems. The Doppler range heatmap that was acquired using 2D FFT will be utilized to generate a collection of features from each frame for a radar output image. Each frame simply means that one value will be generated for one feature with each frame the radar returns. Thus, each feature of the heatmap will be provided by all these frames when they are applied to a dynamic time window. These characteristics could include Doppler spread, doppler average, or range. Following the collection of these features, an artificial neural network is used to identify and compute the predicted gesture. A total of 5 gestures were considered for this experiment. A flowchart has been provided which condenses all the basic steps of the proposed in *fig.1*.

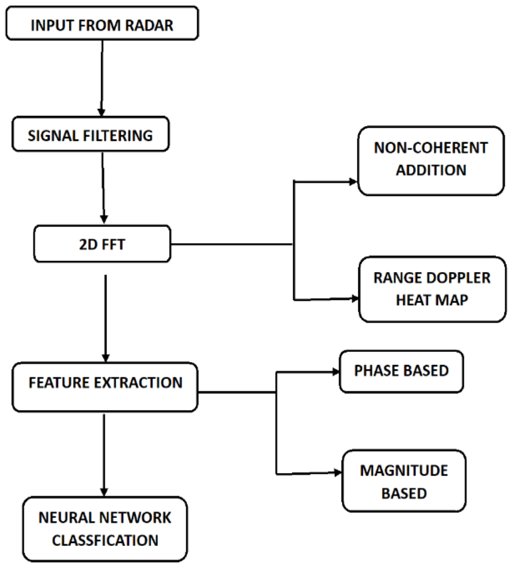

Figure 1: Flowchart depicting the working of gesture recognition in the proposed model.

2.1 Pre-Processing

1.1 Pre-processing the raw data obtained from the radar is a crucial step that must be followed in order to increase the classification process' accuracy. As a result, the signals are gathered and pre-processed to lessen the noise in the input and enhance the input's overall signal-to-noise ratio. The undesirable signals in the gathered radar input are reduced by applying signal filtering techniques like clutter cancellation. This helps to increase the radar's overall accuracy so that it can detect even the smallest movements of any object.

Other methods, such as adaptive notch filtering, pulse banking, and time-frequency analysis, can be employed to denoise radar signals. Time-frequency analysis is the most effective signal-filtering technique for addressing radar signal interference in the system. Although a highly effective technique, it has a high level of complexity and processing time due to the use of wavelets and Fourier transformations to identify and eliminate undesirable signals.

The input radar signals are filtered before being subjected to a 2D FFT. The fundamental 2D FFT is made up of a series of complex values that have changed over time and are grouped in the range and Doppler dimensions. When used on the gathered inputs, the dimension of range and Doppler heatmap is created, allowing the extraction of features for additional categorization, which will be covered in the following section. An example of a heat map of one of the gestures has been depicted in *fig.2*.

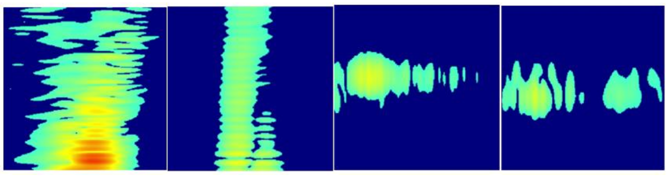

Figure 2: The range Doppler heat map of the 4 main gestures. "Up", "Down", "Left to Right", "Right to Left"

2.2 Feature Extraction

The A series of feature extractions are then applied to the range doppler heat map produced by the 2D FFT. The three primary characteristics of radar-based images are based on magnitude, phase, and statistics. In this study, features that are magnitude-based and phase-based, or based on angle, will be derived. More details about these two attributes are ahead.

2.2.1 Feature Based on Magnitude

It is possible to calculate features depending on magnitude using mathematical methods. The range-Doppler heat maps were produced using 2D FFT on each frame and non-coherently summing all the FFT generated to construct a matrix and then the heat map image is used to calculate these magnitudes. Take the RDI image produced by noncoherent addition (Range Doppler images). The range-Doppler heat map's magnitude value Z_i will be determined by the index of these RDI images, which will be given as í. The range and Doppler values of the indexed range Doppler images will be taken as F_i and x_i.

Weight Range: The position of the hand can be determined by using this feature, also called range centroid. This can be calculated using *equation (1).*

$$R = \frac{\Sigma_i z_i x_i}{\Sigma_i z_i} \quad \text{...... (1)}$$

Weight Doppler: Detection of the velocity centroid of the movement of the hand is done by calculating its weight Doppler shown in *equation (2)*.

$$v = \frac{\Sigma_i z_1 F_i}{\Sigma_i z_i} \quad \text{...... (2)}$$

Instantaneous Energy: For recognizing hand gestures, the hand must be detected first. This feature helps in its detection as showcased below *equation (3)*.

$$I_E = \Sigma_i z_i \quad \text{...... (3)}$$

Dispersion of Range: When the weight range is calculated, it has a lot of dispersion which needs to be collected for proper recognition of gestures. The calculation is shown below *equation (4).*

$$RD = \sqrt{\left(\frac{\Sigma_i z_i \left(x_i - R\right)^2}{\Sigma_i z_i}\right)} \quad \text{...... (4)}$$

Dispersion of Doppler: In the same way as the dispersion of range, the calculation of the doppler range can also get dispersed. This dispersion is calculated using the following formula *equation (5).*

$$DD = \sqrt{\left(\frac{\Sigma_i z_i \left(F_i - v\right)^2}{\Sigma_i z_i}\right)} \quad \text{...... (5)}$$

2.2.2 Feature Based on Phase

The phase features are extracted from the range doppler heat maps as the next step in the feature extraction process after the magnitude features have been determined. The FFTs of the strongest detected spots are taken into consideration and tallied during phase feature extraction. This is done in order to extract features like azimuth angle and elevation angle by applying a 3D FFT to the aggregated points. It is important to keep in mind that for successful feature extraction, clutter, and undesirable signals must be ignored or removed from the spectrum. The collection of feature vectors is then updated to include these angles in order to process the classification data further.

2.3 Classification

Following the extraction of all features, including instantaneous energy, weight range, doppler, and range dispersion, For the classification process, these features are added to a feature vector. A convolutional network layer performs the classification in the work that is being suggested. The classification of the five primary dominant gestures—"No gesture," "Up to down," "Down to up," "Left to right," and "Right to left"—is done using the convolutional network layer. The resnet50 convolutional network layer was selected for this procedure. To determine the best hyperparameters for the current study, an almost 6000-image dataset of range-Doppler heat maps was used. This dataset was then divided into training data and validation data. While the classifier model's accuracy was verified by adding test data in real-time via the radar.

3. HARDWARE SPECIFICATION

DCA 1000 EVM: For two and four-lane low-voltage differential signaling (LVDS) traffic, the DCA1000 evaluation module (EVM) enables real-time data collection and streaming from TI AWR and IWR radar sensor EVMs. A PC running the MMWAVE-STUDIO program can send the data out in real-time via 1-Gbps Ethernet to record and visualize the data before passing it to a preferred application for data processing and algorithm development.

IWR6843ISK: A simple-to-use 60 GHz mm-Wave sensor evaluation kit based on the IWR6843 device with a long-range antenna is called IWR6843ISK. To assess the IWR6843 device, utilize the IWR6843ISK. Accessibility to data sets and power *via* USB interface is achieved by this board.

3.1 FMCW Chirp and Other Parameters Configuration

The hardware used for this experiment was the Texas Instruments IWR6843, as was previously mentioned. The configuration of chirp transmission and reception uses chirp parameters that are connected to the radar. Table 1 displays the condensed tabular column of parameters utilized with the radar.

Table 1: Chirp Parameters of 60GHz FMCW radar

Chirp Bandwidth	~4GHz
Chirp Rep Interval	20ms
Chirp Rep Interval	20ms
Range Resolution	0.01m
Velocity Resolution	0.05m/s

4. RESULTS

TI's IWR6843 FMCW radar was used to test the suggested model. Preprocessing involved cleaning up the signals' clutter before using 2D FFT to create the range-Doppler heat map that was already shown in *fig 2*. Following that features from the heatmap pictures are extracted.

As can be seen in *fig. 3*, the magnitude features, such as range, doppler, dispersion, and energy, were computed, while the phase-based features, which consider the different types of gestures shown, include azimuth angle and elevation.

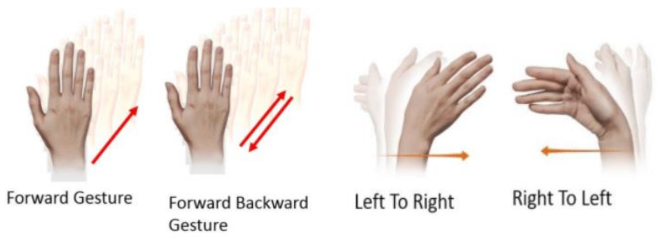

Fig. 3: Illustration of different gestures whose features were extracted.

Following feature extraction, the classification model was implemented, and its accuracy was validated in real-time using the test dataset after the model had been trained using the training data.

For choosing hyperparameters for optimum accuracy, plots of accuracy and loss on the validation dataset were created and are displayed in *figures 4 and 5*.

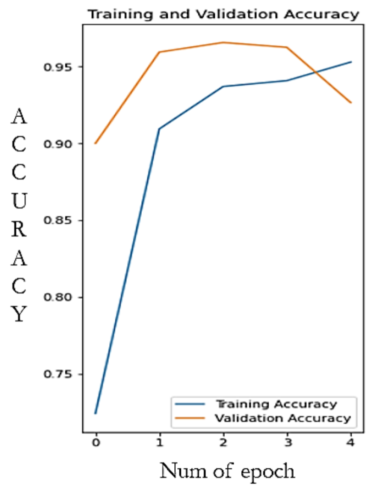

Figure 4: Illustrate the average accuracy of 96.7% in the proposed model

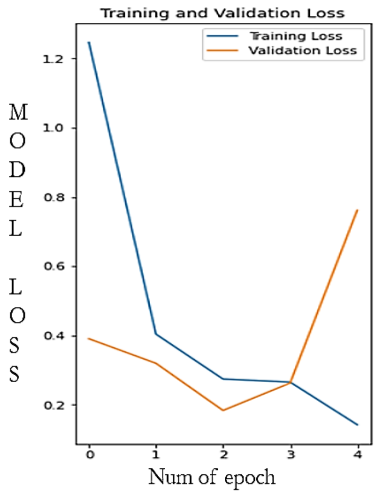

Figure5: Illustrate the average loss in training in the proposed model to be 0.3.

A random heatmap from the test dataset was used to evaluate the validity of the hand gesture recognition system, and the model correctly predicted categorization results. *Figures 6* and *7* show this outcome with two inputs.

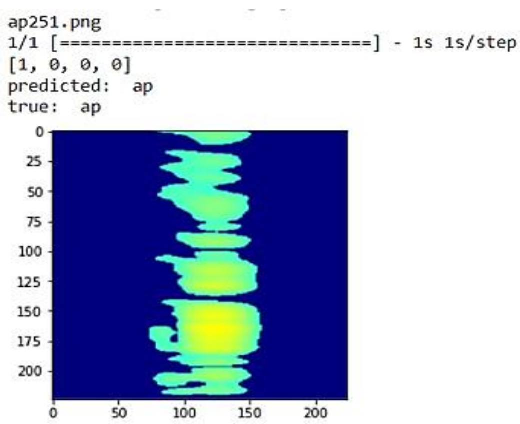

Figure 6: True gesture input: 'Up'; Predicted Gesture output: 'Up'

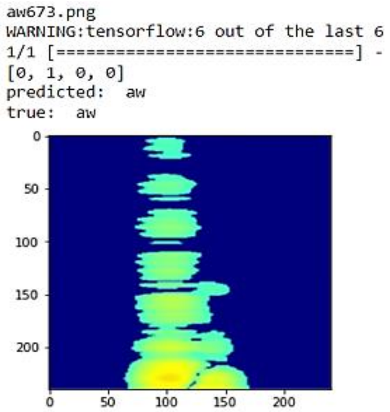

Figure 7: True gesture input: 'Down'; Predicted Gesture Output: 'Down'

A confusion matrix was created to graphically display the accurate readings of the proposed model's ability to recognize each gesture in order to demonstrate this accuracy. Five gestures, labeled -"No gesture," "Up to down," "Down to up," "Left to right," and "Right to left," were included in the confusion matrix.

Table 2: DNN Classifier Confusion Matrix

Predicted / Actual	NO GESTURE	UP TO DOWN	DOWN TO UP	LEFT TO RIGHT	RIGHT TO LEFT
NO GESTURE	**98.23%**	0.15%	0.45%	0.33%	0.09%
UP TO DOWN	1.60%	**97.62%**	0.43%	0.17%	0.00%
DOWN TO UP	0.32%	0.19%	**96.78%**	0.12%	0.24%
LEFT TO RIGHT	0.12%	0.65%	0.00%	**97.16%**	0.35%
RIGHT TO LEFT	1.61%	0.10%	0.15%	0.45%	**96.56%**

Two types of radar systems, FMCW and Doppler, are employed in a variety of scenarios. Some of the main variations between the two are as follows:

The Doppler effect, which monitors the frequency change in the radar signal brought on by the velocity of the target object, underlies the operation of Doppler radar. Contrarily, FMCW radar measures the target object's distance and speed using frequency-modulated continuous-wave signals.

Range: While FMCW radar can be used for both short-range and long-range applications, including automobile radar and surveillance, Doppler radar is primarily utilized for short-range applications, such as speed radar and traffic monitoring.

Precision: FMCW radar is renowned for its excellent resolution and accuracy, enabling it to measure distance and velocity with higher precision.

Cost: Doppler radar is generally less expensive than FMCW radar, making it more accessible for certain applications such as weather monitoring.

Interference: FMCW radar is less susceptible to interference from other radar sources compared to Doppler radar, which can be affected by nearby radar signals.

5. CONCLUSION

In order to improve the accuracy of recognizing and discriminating hand gestures in real-time utilizing a radar-based system, an unique hand gesture recognition process was described in this research. This process was trained on a deep neural classifier. Due to its quick processing and excellent accuracy, the work done with a 60 GHz FMCW radar for gesture recognition has the potential to replace camera-based gesture recognition systems in the market for certain applications like automotive audio control and the handling of medical equipment, among others.

REFERENCES

[1] Chen, Min, Francisco Herrera, and Kai Hwang. "Cognitive computing: architecture, technologies and intelligent applications." IEEE Access 6 (2018) 19774-19783.

[2] Chen, S., et al., A vision of IoT: Applications, challenges, and opportunities with china perspective. IEEE Internet of Things journal, 2014. 1(4): p. 349-359.

[3] Costa, F.; Genovesi, S.; Borgese, M.; Michel, A.; Dicandia, F.A.; Manara, G. A Review of RFID Sensors, the New Frontier of Internet of Things. Sensors 2021, 21, 3138.

[4] Friess, P., Internet of things: converging technologies for smart environments and integrated ecosystems. 2013: River Publishers.

[5] Kahlert, M., Understanding customer acceptance of Internet of Things services in retailing: an empirical study about the moderating effect of degree of technological autonomy and shopping motivations. 2016, University of Twente.

[6] Li, Bangy, and Tieniu Tan. "Online text-independent writer identification based on temporal sequence and shape codes." In 2009 10th International Conference on Document Analysis and Recognition, pp. 931-935. IEEE, 2019.

[7] Lu Tan and Neng Wang, "Future internet: The Internet of Things (IOT) in IEEE Access, vol. 8, pp. 188082-188134, 2018.

[8] Ning, H. and S. Hu, Technology classification, industry, and education for Future Internet of Things. International Journal of Communication Systems, 2012. 25(9): p. 1230-1241.

[9] Aniruddha Bhattacharya and Madhusudan Singh (2021), Implementation of GF-HOG Technique for Effective Commercial and Industrial Load Clustering and Classification for Better Demand Response . IJEER 9(3), 66-75. DOI: 10.37391/IJEER.090307.

[10] Sumari, Arwin Datumaya Wahyudi, Rosa Andrie Asmara, Dimas Rossiawan Hendra Putra, and Ika Noer Syamsiana. "Prediction Using Knowledge Growing System: A Cognitive Artificial Intelligence Approach.", pp. 15-20. IEEE, 2021.