Hand gesture recognition

RA1911031010119 Gaurav Chaturvedi, RA1911031010088 Himayanth V
Batch ID NWC119088

Abstract: The area of gesture recognition is changing due to the development of new technology. With the introduction of technologies like the Kinect sensor, there are now more potential for human-computer interaction than with the conventional hand gesture recognition approaches, which are more dependent on gloves and sensors and less adaptable. The aim of working with picture datasets for improved recognition is to extract additional features to improve gesture recognition. It has been noted that adding more characteristics makes it much simpler to accurately recognise hand gestures, and accuracy may also be raised by optimizing the classification process. To improve the performance of the gesture recognition model, enhanced feature extraction and hybrid classification are used in this study.

Keywords: Gesture recognition, Hybrid classification, Feature extraction

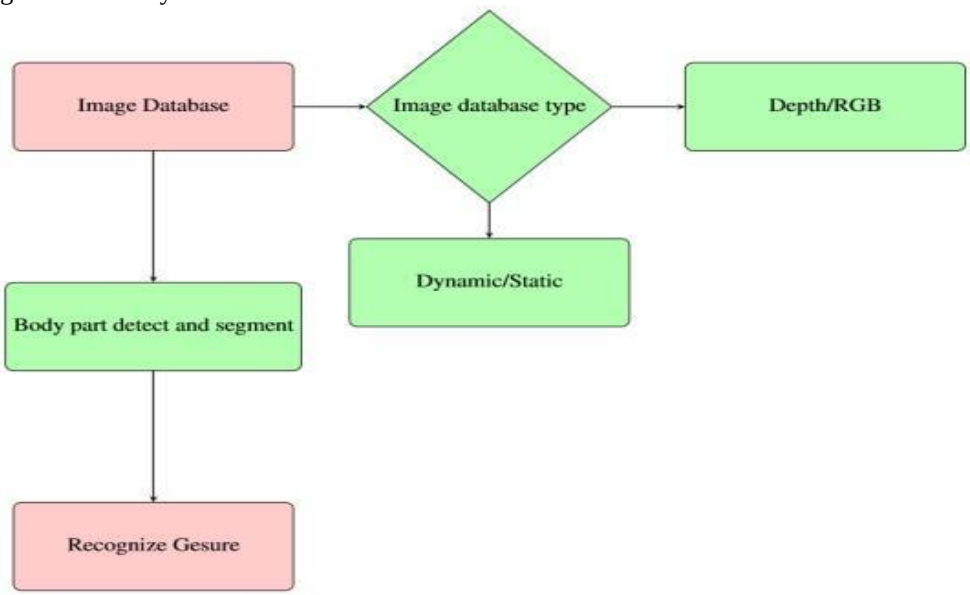
1. Introduction:

Traditional techniques, such as data gloves, attach sensors or markers to the fingers in order to detect hand motions using electro-mechanical or magnetic sensing. These techniques are efficient in measuring hand motions accurately and in real time, but they impede the natural mobility of the hand and cannot be used in non-contact contexts. The equipment also costs a lot for occasional usage and requires complicated calibration.

An alternate solution to the issues is provided by vision-based hand gesture detection techniques, which may be employed naturally in noncontact contexts. However, because of the optical sensors' limitations, the acquired pictures are susceptible to background clutter and poor illumination. As a result, these techniques frequently fail to identify and track the hand accurately.

An alternative possibility is to examine hand gesture detection in a new way with the advancement of depth cameras, such as the Kinect sensor. By combining the depth map and colour picture, the hand gesture may be identified and detected.

After the extraction of images successfully we follow the below framework:



The type of the pictures is determined by the image dataset; if the RGB image dataset, a further step is the conversion of the RGB photos into grayscale images. Following the preprocessing step, ROI (Region of Interest) segmentation is done to focus solely on the needed portion of the image.

Different steps involved in this process are explained as:

Image database: RGB and depth pictures are the two types of images in the collection. A regular camera can capture RGB photos, while depth cameras like Kinect and Leap Motion can capture RGB images and depth images at the same time. The ability to record some spatial information through the use of depth pictures makes it easier to identify and categorise motions.

Gestures are divided into static gestures and dynamic gestures, which determine whether get a single photo or a video

Body Part detection and segmentation: Problems with occlusion, variable light intensity, and light direction throughout the picture gathering process raise the bar for the algorithm's resilience. As the case for gesture recognition becomes more compelling, more and more algorithms are being developed to solve illumination invariance and occlusion issues.

The two popular methods of gesture segmentation involve CNN (Convolution Neural Network) and deep threshold. Both of these methods have been discussed below:

Gesture Segmentation Based on Convolutional Neural Networks:

Optimization using convolutional neural networks, based on FCN (Full Convolutional Neural Networks) or SegNet, is a component of convolutional neural network-based gesture segmentation. The last CNN layer is replaced by FCN and SegNet with a

deconvolution layer, the picture is up-sampled to its original size, and each pixel is forecasted. FCN and SegNet are superior than CNN in that they may take input pictures of any size, no longer need that all images be the same size, and do away with the issue of recurrent storage and convolution computations.FCN, however, also has blatant flaws. The image is not particularly clear when the upsampling factor is large, and the sensitivity to detail has to be increased; the relationship between various pixels is not completely used. The hand area extraction approach incorporates a convolutional neural network to anticipate the middle finger joint of the palm, greatly enhancing the division impact.. Hand area may be divided using SegNet. The toughness of the division is increased by taking a screenshot of the region close to the hand, but the resulting image comprises the edge area, which reduces the identification accuracy.

The segmentation method based on convolutional neural networks is flexible, and a variety of different methods can be combined to perform gesture segmentation.

Gesture Segmentation Based on Depth Threshold Method:

The depth threshold technique extracts an image whose distance is within a preset range by measuring the distance between each pixel and the camera in accordance with the distance between the object and the camera in the depth picture. The depth range for the hand on the depth picture is defined, or the hand is immediately taken into account as the item nearest to the camera, in order to better extract the hand range. With this technique, the preprocessing effect is enhanced, a more precise hand area is obtained, and gesture detection accuracy is increased.However, it has restrictions on the recognition process and a restricted recognition range. Although the depth threshold approach can segment images quickly and effectively, user behavior is severely constrained, and there is not much room for improvement.

Recognition and classification of gestures: There are two types of gesture recognition: dynamic recognition and static recognition. While dynamic gestures are variations in gesture motions over time, or several successive static gestures, static gestures are gestures made in a single image.

Three types of images are used for gesture recognition: depth maps, RGB maps, and RGB-D maps. The distance between the camera and the object may be accurately represented by the depth map. The grey picture resembles the depiction of the depth map. The depth map differs because each pixel displays the separation between the object and the camera. The RGB-D picture consists of depth images as well as RGB three-channel images.

Although the two images look different, there is a one-to-one correspondence between the pixels.

2. Related Work:

Xing Guo, et.al (2019) in order to solve the problems of low recognition rate and less recognition gesture categories caused by incomplete artificial feature extraction information in traditional static gesture recognition methods. a deep CNN framework is designed by using the principle of convolution neural network (Convolutional Neural Network, CNN) to recognize static gesture movements. Combined with a variety of optimal structures of convolution neural network in deep learning, a model with independent static gesture recognition function is realized. The model method can not only ensure the high accuracy and robustness of the recognition results, but also achieve the speed of smooth recognition.[1]

Roman Golovanov, et.al (2020) presents a combined hand gesture recognition system that uses a hand detector to detect hand in the frame and then switches to gesture classifier if a hand was detected. The paper illustrates the proposed combined algorithm. Descriptions of used hand detector and gesture recognition algorithms also are given. Equations for the evaluation of potential performance increase and experimental results are presented. The proposed system is tested on publicly accessible gesture bases and on video sequences prepared by the authors. The experimental results are consistent with theoretical estimates and demonstrate the benefits of the proposed gesture recognition system design.[2]

Qinglian Yang, et.al (2020) proposes a new gesture recognition system based on Deep Neural Network (DNN) and Leap Motion [10]. The palm model is reconstructed to obtain the feature data, and then the feature data of all experimenters are obtained by the Leap Motion controller. The data are finally put into the DNN model for training to implement the recognition of specific hand gesture after being normalized. By testing with 30000 gesture frames of five volunteers, it is found that the recognition accuracy of the proposed scheme can reach 98\%, which indicates that the scheme can complete the specific gesture recognition with a high average recognition rate.[3] Depeng Zhu, et.al (2019) a soldier identification intelligent recognition system was designed. First, the system performs data normalization and endpoint detection on the collected gesture information. Then, feature extraction of the processed data

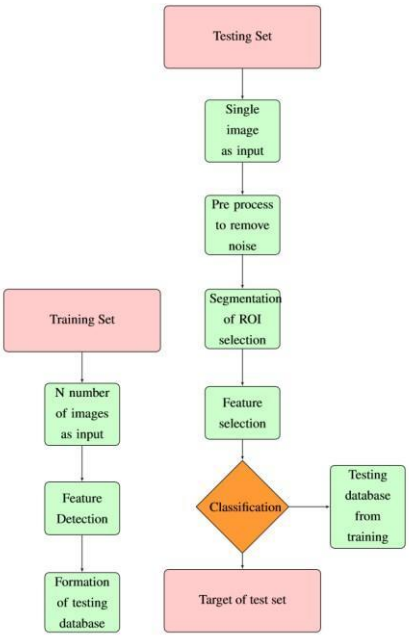
such as mean, peak-to-peak and root mean square values in the time domain. Finally, the dynamic time warping (DTW) algorithm is used to calculate the similarity between the test gesture and the template gesture for the extracted feature parameters, and then the recognition result is obtained. The experimental results show that the system has the characteristics of high recognition accuracy. Therefore, it is suitable for gesture recognition and communication when performing combat

missions.[4]

Xunlei Zhang, et.al (2020) paper adopts 3D separable convolution as an alternative solution. To further extract semantic and action information of gestures, we combine attentional mechanisms and long- and short-term memory. To further extract semantic and action information of gestures, we combine attentional mechanisms and long- and short-term memory (LSTM) networks for gesture recognition in this paper. We conducted experiments on the ChaLearn Large-Scale Gesture Recognition Dataset (IsoGD), and the experimental results validate the effectiveness of our method.[5] Xuexiang Zhang, et.al (2019) the use of gesture recognition in industrial production to control robots is mostly a gesture-oriented teaching method, lacking a systematic description of dynamic gestures and static gestures, making it difficult for the robot to understand the complete intention expressed by the operator. The static gesture real-time recognition is centered and the computer vision control in the complex environment realizes the space movement and posture movement of the robot. This way of using the human body language to directly control the robot movement, the human experience, the intention and the manipulator's high efficiency The combination of sustainability and other advantages can accomplish tasks that cannot be accomplished by people or robots alone.[6] Liying Wang, et.al (2019) a novel method of high

precision fine-grained gesture recognition is proposed based on a terahertz radar, which is able to sense any gesture movement when its range of motion is greater than 5mm. First, High Resolution Range Profile (HRRP) sequences are extracted from radar echoes. Then, the HRRP features are fed to Random Forest classifier after dimensionality reduction by PCA. In order to verify the proposed method is effective to detect fine-grained gestures, four kinds of similar gestures with multi-fingers are designed for experiments. The results indicate the recognition rate exceeds 99.7%, which demonstrates a great prospect in fine-grained gesture recognition using a terahertz radar.[7] Xingxiu He, et.al (2020) designs a number gesture recognition system based on Kinect. The system uses the depth image captured by Kinect sensor to extract gesture information for recognition. In the process of recognition, the AdaBoost algorithm in machine learning is adopted firstly, by the help of Visual Gesture Builder to train the number gesture classifier, and export the classifiers to the recognition program to recognize the current gesture. Finally, numbers are outputted according to the recognized gesture, and the basic functions of number gesture recognition system are realized. Through the experimental verification, the recognition system has enough accuracy.

3. Research Methodology:



The plan for this research includes following steps:

Pre-processing phase: - The pre-processing stage of gesture recognition is the initial stage. The picture dataset, which is gathered from an authentic data source, is used as input during this step. Images are transformed to gray scale for further processing after being taken as input.

Segmentation: - Image segmentation is the process of breaking up a digital image into multiple pieces. The primary goal of segmentation is to recognise objects or extract data from photos. The duty of photo examination is made simpler by this method. The process of image segmentation is used to identify objects and the boundaries of images. Differentiated characteristics are shared by pixels with the same label portion in order to label every pixel in an image.

Feature Extraction: - The region of interest is the outcome as of right now. In order to extract the characteristics from this area of interest, this phase is implemented. Feature extraction is the process of removing from an image a group of values referred to as features. For further processing, these attributes offer information about the image. To identify an infection within a plant, a variety of characteristics are often used, including colour, texture, morphology, and colour coherence vector. Different strategies can be used to extract features. A system may be created using these techniques. These techniques include the histogram-based feature extraction approach, colour co-occurrence method, spatial grey-level dependency matrix, and gray-level co-occurrence matrix (GLCM). A statistical method for classifying textures is the GLCM method.

Classification of Data: - The third process, model creation, involves splitting the entire dataset into training and test sets. When compared to the test set, the training set will be larger. The classification model is used, with training and test sets as input. The entire dataset will be split into ratios of 70 and 30. For the gesture recognition system, 70% will be training sets and 30% will be test sets. For the purpose of recognising gestures, the hybrid classification method—a mix of KNN, SVM, and random forest—is used.

4. Proposed Work:

The identification of gestures is the foundation of this research. There are several phases involved in gesture recognition algorithms, including pre-processing, segmentation, form representation, feature matching, and classification. The underlying study uses the deep map approach for pre-processing, threshold-based segmentation for manual segmentation, and the DWT feature matching technique for feature matching. Last but not least, neural network classification approach is used. Low accuracy is provided by the gesture recognition algorithms. The high accuracy gesture recognition method needs to be improved.

Following are the various objectives of this research work:

4.1 To research and evaluate different gesture recognition methods

4.2 Use an approach for gesture recognition based on neural networks.
4.3 To suggest a hybrid classification method for gesture recognition
Use the suggested method and compare it to what is already available in terms of accuracy, precision, and recall.

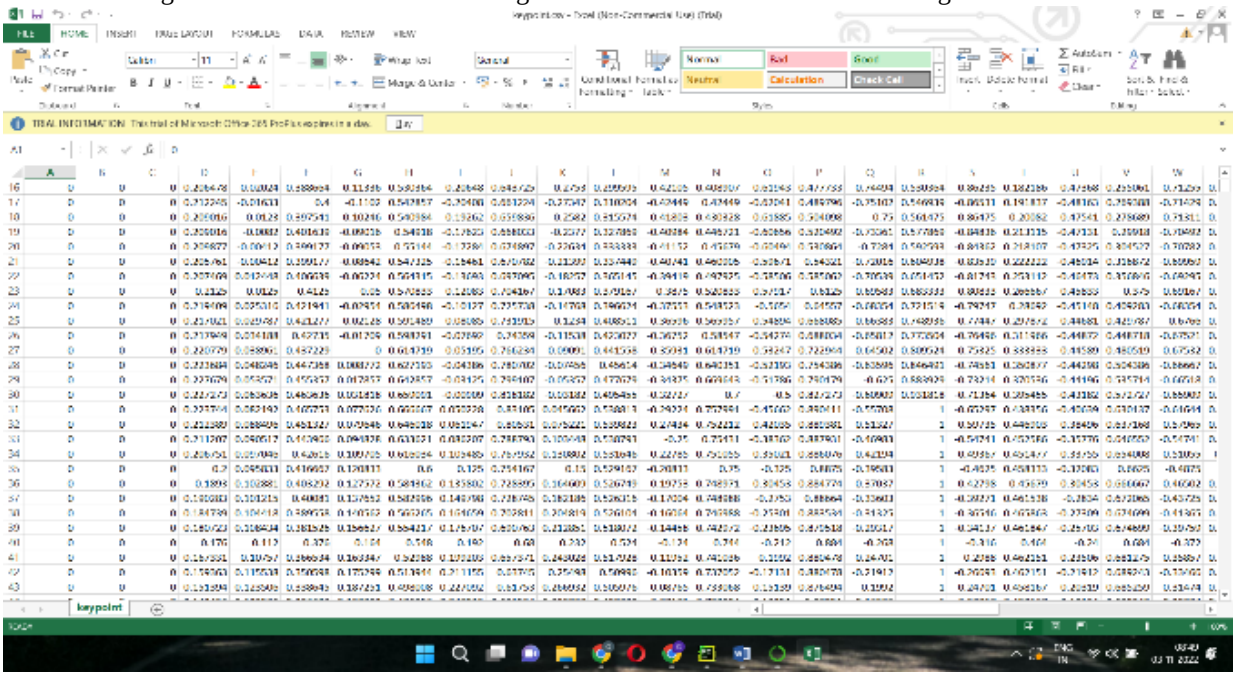
5. Gesture Recognition:

5.1 Dataset:

The fist and palm photos that make up the data set I used are utilised to train the model and to test it later using the same data set.

Testing database:

A testing database with features and a target set will be used to record the training results in a csv file:



4.1 Platform and IDE:

Anaconda: Anaconda is a platform for the distribution of computer languages used in scientific computing, such as Python (machine and deep learning applications, large scale data processing). To execute Python code for analysis and prediction, we utilised the Spyder IDE.

Modules and libraries used:

MediaPipe: Google created a framework for adaptable machine learning solutions called MediaPipe. It is a lightweight, cross-platform framework that is available under an open-source licence. Some pre-trained ML solutions are included with MediaPipe, including face detection, pose estimation, hand recognition, object detection, etc.

TensorFlow: The Google Brains team created the open-source TensorFlow library for machine learning and deep learning. While it may be used to many different tasks, deep neural networks are its main emphasis.

glob: Using precise pattern matching, this is used to locate all file paths.

csv (Comma Separated Values): Databases and spreadsheets may be read and written using these module objects.

skimage: Open source Python tool for image preprocessing that includes features like the greycomatrix (converts RGB image to grey scale)

numpy: This is a core Python library used for complex calculations, big mathematical operations, and the processing of enormous amounts of data.

sklearn: Python library for predictive analysis and classification of data.

pandas:It is a free software programme for manipulating data that is used to filter data frames.

matplotlib:This Python package is used to produce various charts and visualisations..

sys: This Python module gives users access to several variables and functions that are used to modify code in the runtime environment.

5.5 Code flow:

Training: Reading every picture, extracting features including contrast, dissimilarity, homogeneity, ASM, energy, and correlation, translating those features into data frames, creating a target using the random module, and putting the output in a CSV file (testing.csv) for testing.

Testing: Reading a picture, doing any necessary preprocessing, and then calculating its contrast, dissimilarity, homogeneity, ASM, energy, and correlation, before reading the CSV file produced during training and forecasting the outcomes. Voting classifiers are employed to calculate accuracy in this case; hybrid classification, which included GaussianNB, RandomForestClassifier, and LogisticRegression, is proposed.

6. Analysis and Results:

6.1 Result of base paper implementation

Base paper python implementation gives 55% accuracy on the image dataset

```
 (26, 5)
(26,)
[1]
Input image is of Plam Gesture
train_set_x shape: (18, 5)
train_set_y shape: (18,)
test_set_x shape: (8, 5)
test_set_y shape: (8,)
Voting Classifier accuracy is
55.0
              precision    recall  f1-score   support

           0       0.00      0.00      0.00         6
           1       0.25      1.00      0.40         2

    accuracy                           0.25         8
   macro avg       0.12      0.50      0.20         8
weighted avg       0.06      0.25      0.10         8
```

Result of new implementation new implementation gives higher accuracy with proposed changes on the same dataset

Result Analysis

```
Classification Report
              precision    recall  f1-score   support

           0       0.99      0.97      0.98       392
           1       0.97      0.96      0.96       383
           2       0.95      0.97      0.96       340
           3       0.94      1.00      0.97        77
           4       1.00      1.00      1.00        65

    accuracy                           0.97      1257
   macro avg       0.97      0.98      0.97      1257
weighted avg       0.97      0.97      0.97      1257
```

The implementation results shown in the images above demonstrate how utilizing hybrid classification, the base paper's implementation results on the same dataset may be improved.

Simply modifying the method and using the suggested classifier combinations resulted in a significant gain in accuracy in the same context.

By adhering to the research plan, the results were improved. It was discovered that capturing more features improves classification and provides better accuracy than working with minimal features, so more features like contrast, dissimilarity, homogeneity, ASM, energy, and correlation were targeted for capture.

6. Conclusion:

It has been noted that gesture detection is utilised in many industries and creates a bridge between human and computer connection and comprehension; as a result, it has to be applied to new domains in order to reap its advantages.

More effective object identification and detection techniques can be utilised to minimise the complexity in the future since gesture recognition uses advanced machine learning algorithms, which are quite difficult in terms of time complexity and efficiency.

Gaming is a very common trend among young people today, and since most popular games use gesture recognition, there is a significant chance that they will be used in the future. If we can improve accuracy by using various techniques, as I have suggested in my thesis, it could elevate user experience to a whole new level.

References:

1. Xing Guo, Wu Xu, Wen Quan Tang, Cong Wen,
2. "Research on Optimization of Static Gesture Recognition Based on Convolution Neural Network", 2019, 4th International Conference on Mechanical, Control and Computer Engineering (ICMCCE).
3. Roman Golovanov, Dmitry Vorotnev, Darina Kalina,
4. "Combining Hand Detection and Gesture Recognition Algorithms for Minimizing Computational Cost", 2020, 22th International Conference on Digital Signal Processing and its Applications (DSPA).
5. Qinglian Yang, Weikang Ding, Xingwen Zhou,
6. Dongdong Zhao, Shi Yan, "Leap Motion Hand Gesture Recognition Based on Deep Neural Network", 2020, Chinese Control and Decision Conference (CCDC).
7. Depeng Zhu, Ranran Wei, Weida Zhan, Ziqiang Hao,
8. "Individual Soldier Gesture Intelligent Recognition
9. System", 2019, IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS).
10. Xunlei Zhang, Yun Tie, Lin Qi, "Dynamic Gesture Recognition Based on 3D Separable Convolutional
11. LSTM Networks", 2020, IEEE 11th International Conference on Software Engineering and Service Science (ICSESS).
12. Xuexiang Zhang, Xuncheng Wu, "Robotic Control of
13. Dynamic and Static Gesture Recognition", 2019, 2nd World Conference on Mechanical Engineering and Intelligent Manufacturing (WCMEIM).
14. Liying Wang, Zongyong Cui, Zongjie Cao, Shengping Xu, Rui Min, "Fine-Grained Gesture Recognition Based on High Resolution Range Profiles of Terahertz Radar", 2019, IEEE International Geoscience and Remote Sensing Symposium.
15. Xingxiu He, Jia Zhang, "Design and Implementation of
16. Number Gesture Recognition System Based on Kinect", 2020, 39th Chinese Control Conference (CCC).