

Performance Improvement of Data Fusion Based Real-Time Hand Gesture Recognition by Using 3-D Convolution Neural Networks With Kinect V2

S.Chandrasekhar^{1*} Dr.N.N.Mhala²

1.Ph.D. Scholar, BapuraoDeshmukh College of Engineering, Maharashtra State-442102 ,India

2.Professor &Principal, Government polytechnic college, Maharashtra state-400612,India

Abstract

Hand gesture recognition is one of the most active areas of research in computer vision. It provides an easy way to interact with a machine without using any extra devices. Hand gestures are natural and intuitive communication way for the human being to interact with his environment. In this paper, we propose Data Fusion Based Real-Time Hand Gesture Recognition using 3-D Convolutional Neural Networks and Kinect V2. To achieve the accurate segmentation and tracking with Kinect V2. Convolution neural network to improve the validity and robustness of the system. Based on the experimental results, the proposed model is accurate, robust and performance with very low processor utilization. The performance of our proposed system in real life application, which is controlling various devices using Kinect V2.

Keywords: Hand gesture recognition, Kinect V2, data fusion, Convolutional Neural Networks

DOI: 10.7176/IKM/9-1-02

1. Introduction

The most recent vision advances and the propelled PC equipment limit make continuous, exact and hearty hand following and motion acknowledgment promising. Hand signal acknowledgment is a standout amongst the most dynamic zones of research in PC vision [1]. A signal is a nonverbal correspondence in which noticeable body conveys a specific message. Motion gives an approach to PCs to comprehend human non-verbal communication [2]. Human motion acknowledgment framework dependent on a Convolution Neural Network (CNN) in which the skin shading model is enhanced and the hand present is aligned to expand acknowledgment exactnesses. Highlight extraction assumes a vital job in a human motion acknowledgment framework on the grounds that the data about shape, posture, and surface of a motion is useful [3]. The motion classifiers are shrouded Markov models contingent irregular fields and bolster vector machines (SVM) have been generally utilized.

Hand gesture recognition has been a promising topic and applied to many practical applications [4]. Gesture recognition algorithms can be divided into three different levels of static hand gesture recognition, dynamic gesture recognition; and 3D hand gesture recognition. To recognizing static hand gestures using convolutional neural networks. The 3D-based, gesture method gives depth information of images is explored to ensure more effective hand gesture recognition performance. To get depth information cannot be obtained from a single camera, needing a special device to obtain. hand gesture recognition deal with various kinds of challenges like rotation, scale changes, lighting changes, cluttered background, and distortion [4]. Kinect-based hand gesture recognition algorithms follow two classes they are feature based and learning based. Kinect V2 is used to recognize different human hand gestures. It is fused with many advanced visual technologies and has been widely used in various kinds of computer vision tasks, such as face recognition scene understanding and human gesture recognition [5]. Different kinds of gesture recognition are faced gesture recognition, sign language recognition, and hand gesture recognition.

2D Convolutional Neural Networks have been applied to the Gesture Recognition field in order to extract spatial features. 3D Convolutional Neural Networks were proposed to recognize isolated gestures[11]. To using 3D CNN's for user-independent continuous gesture recognition[5]. In this paper, Neural Network is used to recognize hand gesture. Neural Network creates new networks. Basic of Neural Network is Euclidean distance. There are three layers in the neural network. They are three types of layers Input layer, Inner layer, and an Output layer. The number of inner layers gives the perfection of the system. Properties of the neural network are Training and Testing.

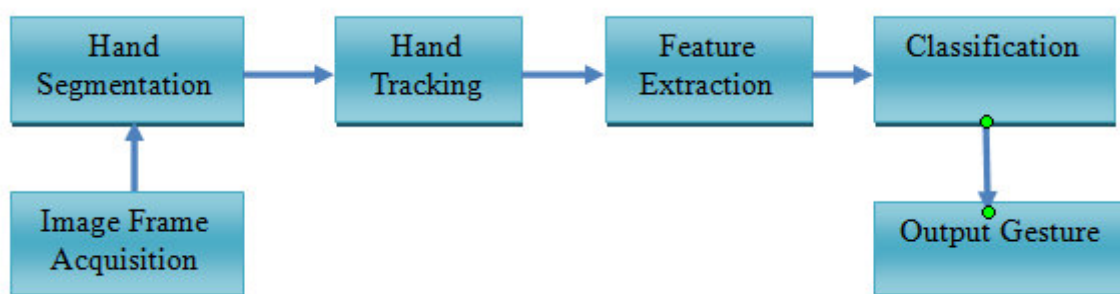


Figure 1: Basic Block Diagram of Hand Gesture Recognition

Figure 1 shows the block diagram of hand gesture recognition the image frame is acquired from the input video. The next step is segmentation which partitions an image into its constituent parts or objects. Hand tracking is a high-resolution technique that is employed to know the consecutive position of the hands of the user. After the successful tracking, there is a need to extract the important feature points from the available data points of the track path. In pattern recognition and in image processing, a feature extraction is a special form of dimensionality reduction. After feature extraction, classification mode plays a vital role in the gesture recognition process. The feature set as input and gives a class labeled output, which is required output gestures [8]. Gesture recognition aims to recognize meaningful movements of human bodies and is of utmost importance in intelligent human-computer interactions. There is a large variety of applications which involve hand gestures [9]. Hand gestures can be used to achieve natural human-computer interaction for virtual environments. This section an overview of a few application areas is given. Robotics, Sign Language, Vehicle interfaces and health care.

The paper is organized with six sections: The first section is an introduction and related work description of real-time hand gesture recognition. The second section gives detail of the literature review of some technologies. The third section offers information data fusion based hand gesture algorithm Section four provides experimental results of gesture recognition and performance analysis. The last section concludes the work.

2. Review of Related Research

In Paper [5] developed Real-Time Gesture Recognition Using Gaussian Mixture Model. In this approach gesture, reorganization is proposed by using the neural network and tracking to convert the sign language to voice/text format. Initially, they provide hand tracking of observed and recorded by typical video cameras. Most of the complete hand tracking systems comprise three layers: detection, tracking, and recognition. The detection layer is responsible for defining and extracting visual features and the tracking layer is responsible for performing temporal data association between successive image frames. Recognition layer is responsible for grouping the spatiotemporal data extracted in the previous layers and assigning the resulting groups with labels associated with particular classes of gestures. This method is reducing the required tracking time and further reducing the complexity in computation at the tracking phase. The system can detect and extract human hand from the complex image that is an image where a human body appears, modify the system so that it can work in any lighting condition and expand the system to recognize the hand tracking.

In paper [6], introduced implements recognition of hand gesture in real time. To recognize the gestures in real time is the forms the main objective of this method. The real-time recognition process uses three main steps they are background subtraction by using the codebook algorithm, contour, convex hull and convexity defect calculation, and final convexity defect calculation. An important challenge of this approach is robustness means because of lighting conditions and background noise its difficult recognize the different postures of gestures. The real-time video is converted into a fixed number of a number of frames. The frames are the input to the codebook algorithm. The code book algorithm converts the color image which is three channel images to a binary image which is single channel image. Background subtracted image is single channel image where hand image is white in color and background is black in color. The binary image is used for the calculation of counter, Convex hull and convexity defects and finally depending upon the calculation of defect points the fingers which are unfolded are counted.

In paper [11] proposed 3D Convolutional Neural Networks for large-scale user-independent continuous gesture recognition. The network performs three-dimensional convolutions to extract features related to both the appearance and motion from volumes of color frames. Depth and intensity information were combined into a single image and these in turn combined to form gesture volumes. The gesture volumes are then re-sampled to have fixed size before being used for training 3D CNN. Gestures can be defined as a time-based sequence of spatial configurations and disregarding either the spatial or the temporal information can result in poor performance in recognition. If layers must be initialized from scratch, careful choice of weight initialization can

also significantly improve performance. The 3D convolution and pooling layers help to learn the Spatio-temporal variations in the data.

In paper [12] introduce a hand gesture recognition system that utilizes depth and intensity channels with 3D convolutional neural networks. To reduce potential overfitting and improve generalization of the gesture classifier, we propose an effective spatiotemporal data augmentation method to deform the input volumes of hand gestures. Prevent overfitting and to increase the generalization performance of the classifier, we augmented the data online during training to propose data augmentation technique plays an important role in achieving superior performance. In this method classifier used for fused motion volume of normalized depth and image gradient values, and utilizes spatiotemporal data augmentation to avoid overfitting. It also employs spatiotemporal data augmentation for more training that is effective and to reduce potential overfitting.

In paper[14] developed static hand gesture recognition using Kinect V2.To apply different convolutional neural network architectures to this classification problem and evaluated the impact of kernel size on the recognition score. To increase the classification score, we decided to submit more samples to the network and then evaluate the results. To create our own gesture database using the *Kinect v2*, which has been used to capture not only color but also infrared and depth frames. On this database, we then tested different convolutional architectures and their impact on the convergence speed and the classification score. To achieve satisfactory results of classification score. The greatest success was achieved in recognizing color images. The functionality of the trained network was then verified by displaying the transition of a random color image through the individual convolutional layers.

3. Proposed Real Time Hand Gesture Recognition

Many researchers have been suggested on a gesture recognition system for different applications, with different recognition phases but they all agree with the main structure of the gesture recognition system. These phases are segmentation, features detection and extraction, and finally the classification or recognition phase. Fig 2 shows the gesture recognition system [28]. Hand gestures used for communication and conveying emotions can be acquired from images, video sequences, and datasets. Gestures may be 2D gestures (plane) or 3D gestures (free-form) or static gestures or dynamic gestures. Static gestures are configuration based on hand position, and dynamic gestures are motion-based hands movement. The introduction of high-quality depth sensors at a lower cost, such as the Microsoft Kinect, facilitated the development of many gesture recognition systems. The gesture recognition module runs in parallel and is able to recognize different gestures running at capture image.

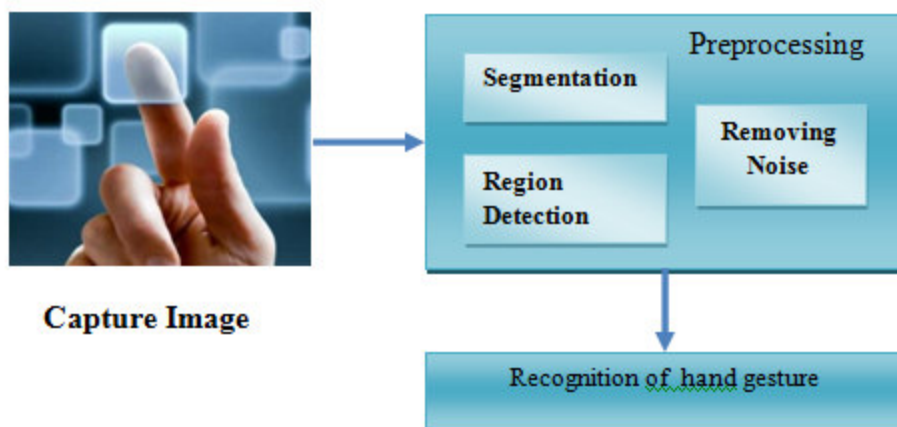


Figure 2: The flow of gesture recognition system

In order to satisfy the memory requirements and the environmental scene conditions, pre-processing of the raw video content is highly important [5]. Segmentation phase plays an important role in the system recognition process. Perfect segmentation effects on the accuracy of the recognition system [19], the segmentation process is to extract the hand region from the input image and isolate it from the background. Segmentation partitions an image into its constituent parts or objects [10]. In segmentation process, an input image is first converted into HSV or YCbCr color space, because in HSV color space Hue and Saturation are independent of luminance. These features are the useful information that can be extracted from the segmented hand object by which the machine can understand the meaning of that posture. The numerical representation of these features can be obtained from the visual perspective of the segmented hand object, which forms the feature extraction. In background subtraction method, one frame is subtracted from another to detect the regions in motion [5]. HSV is used to detect skin region. The range of the skin colors depends upon the lighting conditions. GMM is used for

noise removal. Gesture recognition systems are the detection of hands and the segmentation of the corresponding image regions [5]. Recognition or classification of hand gestures is the last phase of the recognition system. Hand gestures can be classified into two different approaches such as Rule-based Approaches and Machine learning based approaches.

Kernel filters are then commonly used in other CNN-based hand recognition methods are adopted [13]. The 3D CNN was introduced to this field, the results demonstrate its powerful ability to extract feature. The architecture of our convolutional neural network is consists of seven layers and takes pixel field as input. The CNN are typically trained like standard neural networks using back propagation [6]. The images acquired by Kinect the 3D CNN achieves good classification performance. In the discussed performer-independent test the results are slightly worse for the wrist positions determined automatically [9]. To reduce the processing time of CNN's, we extract low-level features using techniques from biologically inspired computer vision, which are then further processed hierarchically, and finally recognized

3.1 Real-time hand gesture recognition using Kinect v2

In this approach, Kinect v2 camera was used for hand recognition. Depth image was used for segmentation. Convex hull Open CV function was used to detect a number of defects (concavities) on hand and stored into the defect array. Based upon a number of defects, finger count was determined [20]. Kinect V2 to recognize different human hand gestures. Kinect is a kind of human-computer interaction facilities. It is fused with many advanced visual technologies and has been widely used in various kinds of computer vision tasks, such as face recognition, scene understanding, and human gesture recognition [4]. Kinect V2 can track at most six skeletons compared with two skeletons using Kinect. The contribution of this method to use Kinect V2 for hand gesture recognition. In the depth information of skeleton data, the proposed recognition model can achieve real-time performance, which is faster than some of the-state-of-the-art hand gesture recognition algorithms. Kinect-based hand gesture recognition algorithms can be roughly divided into two classes: 1) feature-based, and 2) learning-based [12]. The Graham Scan algorithm [6] is used to compute the convex hull of the detected hand clusters. Hand contours are detected by a modified Moore-Neighbor Tracing algorithm. After the K-means clustering, a group of hand points is found and stored. Figure 2 shows the hand gesture recognition using Kinect

[23]. Kinect can capture depth information by projecting an infrared dots pattern and its subsequent capture by an infrared camera. The depth information that is captured will be converted into a grayscale image. The image does not contain any color information. When someone operates the Kinect, the person's hands should be in the front, so that the Kinect can extract the hands by judging the depth.



Fig 3: Hand Gesture Recognition Using Kinect

Hand gesture recognition concerns two challenging problems they are hand detection and gesture recognition. To robust and detect the hand how to effectively and accurately recognize the gesture of the hand [24]. This system has several key features. Capable of capturing images in the dark, Identifying up to two hands, under all reasonable rotations of the hands, Translating and displaying gestures in real time and Allowing the user to choose different scenarios [13]. Present an efficient and accurate hand gesture recognition system using the Kinect sensor as the input device. Both the depth and color information obtained from the Kinect sensor is used for hand detection gesture recognition module provides an effective mechanism for recognizing hand shapes with input variations and distortions.

3.1.1 Convex Hull

Convex Hull is a region based structural method for shape representation. The prime objective of using Convex Hull on segmented hand is to get a convex deficiency of an image [18]. Convex Hull needs to be calculated by boundary tracing or using morphological operation. Polygon approximation was used for extracting Convex Hull to reduce computational time. The extraction of convex hull found significant convex deficiencies along the boundary. The hand shape was represented by a defect array of concavities. Algorithm (1) describes the working using convex Hull. The main objective of using 3-D Kinect was to simplify the task of pre-processing by

capturing depth image for hand segmentation. The use of Kinect camera made this algorithm easy for pre-processing such as subtraction of complex background and skin color detection.

Advantages:

- i) Use of 3 D camera simplified the task of segmentation.
- ii) Dynamic background subtraction.
- iii) Convex Hull method gives accurate defect of hand shape.
- iv) Can be used for motion tracking and detection.

The links in the convex hull are used to find the fingertips. This is done by going around each point in a convex hull and calculating the angles at those points. The convex hull method is fast but it is not robust [1].

Algorithm 1: Real-time Hand Gesture recognition using Kinect V2 by Convex Hull

```

procedure FINGERDETECTIONBYKINECT(ImageI).
Depth Image = GetDepthRGBImage(I);
Smooth Image(DepthImage);
Segmented Image = PerformThreshold(Depth Image,OTSU Method)
Contour = Get Contour(Segmented Image); Contour1 = ApproxPoly(Contour);
Convex Hull= GetConvex Hull(Contour1);
Defect=Get ConvexityDefects(Contour1,ConvexHull);
Interpret FingerCount(Defect);
end procedure
    
```

3.1.2 Real-time hand gesture recognition using 3D Convolutional neural network

To introduce a hand gesture recognition system that utilizes depth and intensity channels with 3D convolutional neural networks. To reduce potential overfitting and improve generalization of the gesture classifier, we propose an effective spatiotemporal data augmentation method to deform the input volumes of hand gestures [11]. Gestures can be defined as a time-based sequence of spatial configurations and disregarding either the spatial or the temporal information can result in poor performance in recognition. These descriptors are then tracked through the time domain using approaches such as graphical models to represent the temporal aspect of the gesture. The network architecture consists of 8 3Dconvolutional layers, five 3D max-pooling layers, two fully connected layers, and a softmax classification layer. This spatiotemporal helps to provide robustness to temporal variations; distinctive patterns are encoded in the same manner, regardless of where they occur within a local region of the spatiotemporal volume [11]. To remove the noise from the final predictions of the network, we have applied 2-stage majority filtering. The majority filter sizes were chosen empirically to maximize the frame-based accuracy.

The key component of 3D CNN is the convolutional layers. They are formed by grouping neurons in a rectangular grid. The training process aims to learn the parameters of the convolution. Pooling layers are usually placed after a single or a set of serial or parallel convolutional layers and take small rectangular blocks from the convolutional layer and subsample them to produce a single output from each block [22]. Finally, dense layers it's called as "fully-connected" layers perform classification using the features that have been extracted by the convolutional layers and then have been subsampled by the pooling layers. Every node of a dense layer is connected to all nodes of its previous layer.

From the extracted set of skeletal joints, we select a subset of them that is typically involved in hand gestures. Fig. 4. Note the visual difference of these images, which may not be significant, yet allows the CNN to learn the differences between the two classes [21]

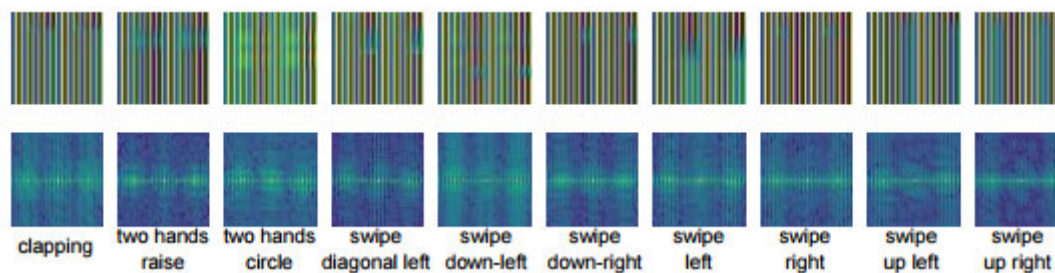


Fig 4: Sample images from all 10 gestures used throughout our experiments: upper row: signal images; lower row: activity images

The 3D CNN for hand gesture recognition. The system consists of two networks, a high-resolution network, and a low-resolution network – the predictions are multiplied during testing. The network architectures are quite simple; they perform thorough data augmentation during training. Fortunately, they detail their training and data

augmentation approaches. For data augmentation, they use reverse ordering of the frames and horizontal mirroring, partial elastic deformation, and temporal elastic deformation. 3D convolution process is decomposed into two parts: depth-wise and point-wise [20]. However, due to the decomposition, the network is deepened, which lead to gradient dispersion and renders the network difficult to train. In order to solve that problem, two methods are used: skip connection and layer-wise learning rate. Skip connection alleviates the problem of gradient dispersion, but it does not completely solve it, so the layer-wise learning rate is needed, such that the network can be trained. It is worth noting that skip connection can improve the accuracy of the network.

Computational cost of a 3D convolution process is

$$D_k \cdot D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F \cdot D_f$$

Where D_k is the number of kernels in the temporal dimension and D_f is the number of frames, M is the number of channels of input, N is the number of channels of output. The Convolutional neural network's structure has the following characteristics i) In order to reduce the number of weights ii)most convolutional layer can have more feature maps iii)network can learn the down sampling process iv)before reducing the size of the feature maps the network will be able to extract more failure information.

Algorithm 2 describes the recognition process with FD descriptor. The descriptors improve the in terms of accuracy and time required for recognition. The Recognition accuracy was calculated by Equation (1) and the average accuracy is represented in equation (2), [18].

Algorithm 2: Real-time Hand Gesture recognition using Fourier Transform descriptor

```

procedure REALTIMEFD(ImageI)
    Neighbourhood Classification by Fourier descriptor
    CurrentSymbol.FD= ExtractFD(I);
    M = GetMeanSymbolList();
    for i ← 1, M.length do
        ComputeEuclidianDist(CurrentSymbol,M[i])
    end for
    DetectedSymbol=GetMinimumDistanceSymbol()
end procedure
    
```

The Recognition accuracy was calculated by Equation (1) where currently 10 sample test cases are considered for each class in real time. Equation (2) gives an average accuracy of image system for a total number of classes [18]. 3D convolutional networks are more expensive in the computation efficiency. 3D matrix needs more memories in the computer. 3D convolutional operations need more calculations than 2D convolutional operations [19].

$$accuracy(\%) = \frac{CorrectClasssification}{TotalTests} * 100 \quad (1)$$

$$avg.accuracy = \sum_{i=1}^n \frac{accuracy}{n} \quad (2)$$

Where n is the total number of classes

The hand gesture recognition system also needs to meet the requirements including real-time performances, accuracy, and robustness [6]. Kinect v2 Sensor provides a depth sensor and an infrared sensor to provide the depth information and infrared information. To using white areas of two resultant images caused by the side effect of sensor limitations.

4. EXPERIMENTAL RESULTS

In this section, the proposed hand gesture recognition model is tested to verify its robustness and efficiency. Fig. 5 shows a typical depth histogram corresponding to the capture image frame shown The histogram of the extracted foreground from the image. To bring in more clarity, we carry out histogram equalization. The gesture is parameterized using depth variation and motion information content of each cell of the grid. This can be visualized as segmenting the histogram of each cell into the 10 specified levels and storing the normalized count of pixels. Fig. 6 and fig .7 illustrates this part. The motion information is extracted by noting the variation in depth between each pair of consecutive frames. It is obtained by subtracting a depth image from the preceding depth image [22]. The difference image gives the path of motion of the body part. Each frame of the video was represented by a row of the matrix. The columns represent the feature points.

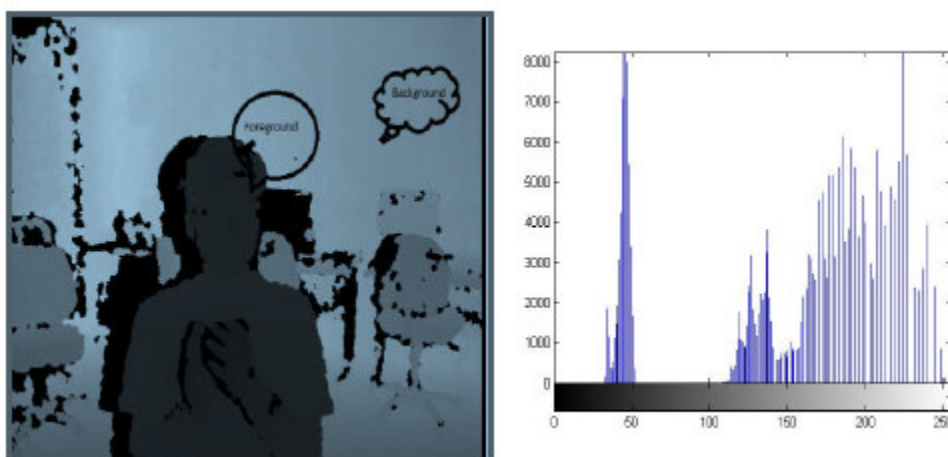


Fig 5: Real-time recognition for depth image to the histogram

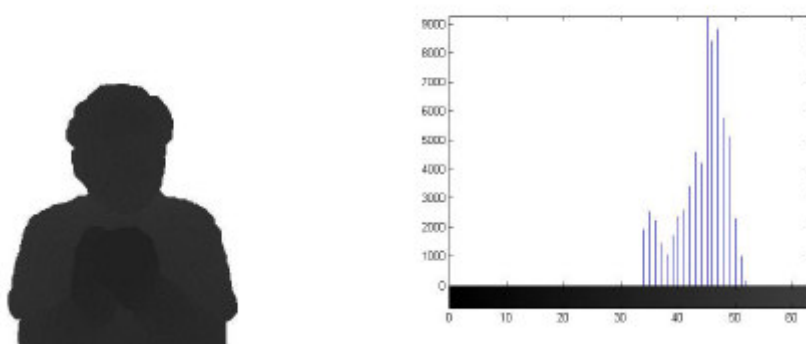


Fig 6: Extract the image into a histogram

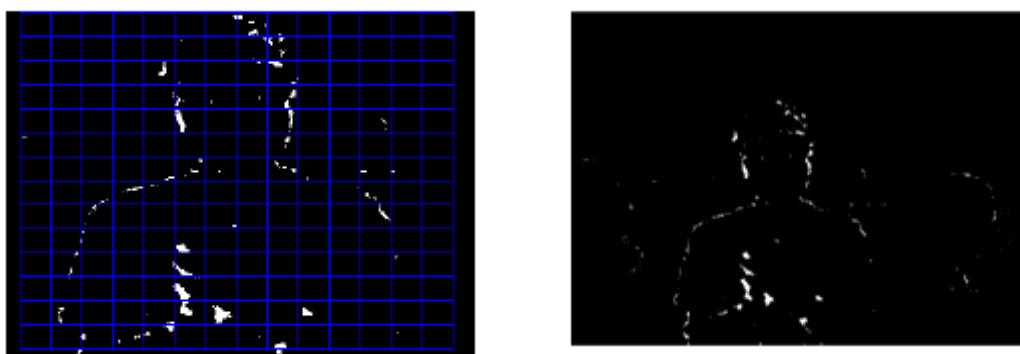


Fig 7. Image segmentation and noise removal

Figure 8 shows us the accuracies across the iterations. The 3D neural network accuracy had a fast-rising trend until it achieved a peak score. 3D Convolutional Neural Networks to the problem of large-scale continuous user-independent gesture recognition [22]. 3D CNN was used to train the system for the classification of the 8 gestures. A matrix was generated for the entire training data set. Each frame of the video was represented by a row of the matrix. The columns represent the feature points.

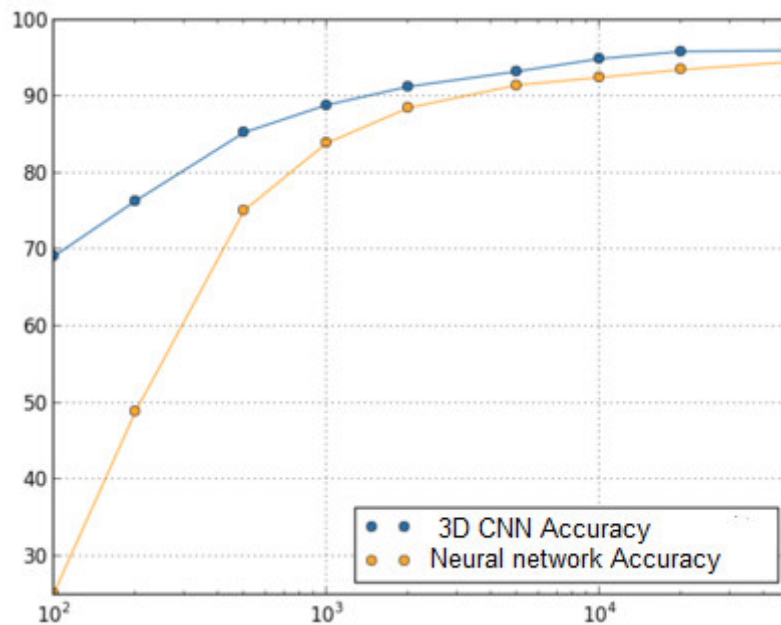


Fig 8: 3D Convolutional Neural Network Accuracy

Figure 9 shows the effect of recognition time performance. We get access to these two types of data using the Kinect SDK. [4] After acquiring depth images from Kinect V2, we obtain hand skeleton points including wrist and central palm. Then, these two points are mapped into depth images and their depth values are saved in the system. To get the accurate recognition Kinect V2, the model can achieve real-time performance, which is faster than the Kinect and 3D CNN hand gesture recognition methods.

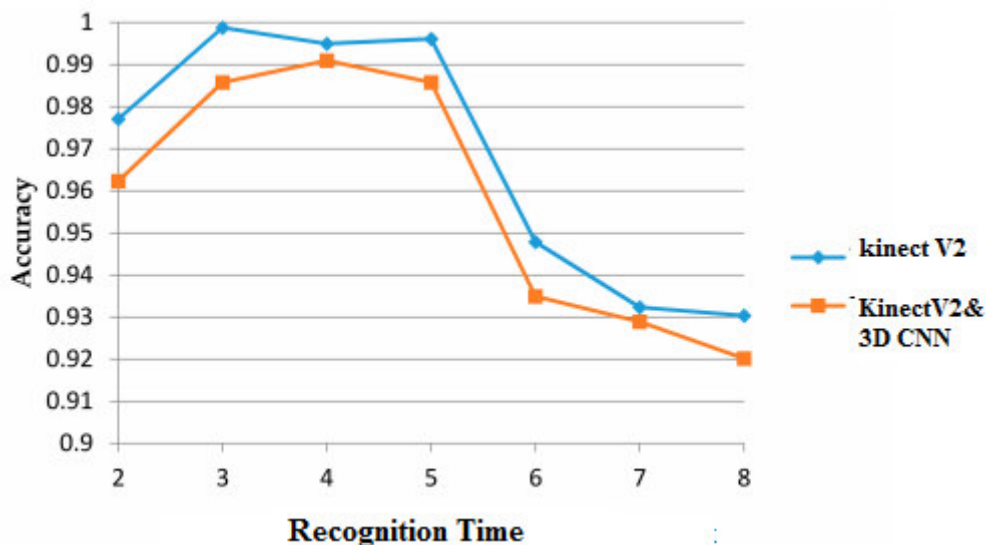


Fig 9: Recognition Time Analysis

Hand gesture recognition system also needs to meet the requirements including real-time performances, accuracy, and robustness. The performance was measured in terms of time.

5. Conclusion

In this paper, we propose a data fusion-based real-time hand gesture recognition model by fusing depth information and hand data. To get the accurate segmentation and tracking with Kinect V2, the model can achieve real-time performance. To applying 3D Convolutional Neural Networks to the problem of large-scale continuous user-independent gesture recognition. The 3D convolution and pooling layers help to learn the spatiotemporal variations in the data. According to the experimental results, accuracy was maintained high. The advantage of using the 3D convolutional neural networks is high processing speed so that the results in a real-time manner. Based on the experimental results, the proposed method is an accurate and robust performance with very low processor utilization. The accuracy and the robustness make this system a versatile component that can be integrated into a variety of applications in daily life. In the future, we will try to dynamic recurrent neural

networks can be used for continuous gesture recognition.

6. REFERENCES

- [1] Farooq, Javeria, and Muhaddisa Barat Ali. "Real-time hand gesture recognition for computer interaction." *Robotics and Emerging Allied Technologies in Engineering*, 2014 International Conference on. IEEE, 2014.
- [2] Lin, Hsien-I., Ming-Hsiang Hsu, and Wei-Kai Chen. "Human hand gesture recognition using a convolution neural network." *Automation Science and Engineering (CASE)*, 2014 IEEE International Conference on. IEEE, 2014.
- [3] Molchanov, Pavlo, et al. "Hand gesture recognition with 3D convolutional neural networks." *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2015.
- [4] Lan, Yuhai, Jing Li, and Zhaojie Ju. "Data fusion-based real-time hand gesture recognition with Kinect V2." *Human System Interactions (HSI)*, 2016 9th International Conference on. IEEE, 2016.
- [5] Meethian, Aisha, and B. M. Imran. "Real Time Gesture Recognition Using Gaussian Mixture Model." *International Journal of Scientific & Engineering Research* 4.8 (2013).
- [6] Ding, Zhen, et al. "sEMG-Based Gesture Recognition with Convolution Neural Networks." *Sustainability* 10.6 (2018): 1865.
- [7] Kim, Min-Soo, and Choong Ho Lee. "Hand Gesture Recognition for Kinect v2 Sensor in the Near Distance Where Depth Data Are Not Provided." *International Journal of Software Engineering and Its Applications* 10.12 (2016): 407-418.
- [8] Zhu, Guangming, et al. "Multimodal gesture recognition using 3-D convolution and convolutional LSTM." *IEEE Access* 5 (2017): 4517-4524.
- [9] Simion, G., V. Gui, and M. Ottesteanu. "Vision based hand gesture recognition: A review." *international journal of circuits, systems and signal processing* 6 (2012): 275-282.
- [10] Choudhury, Ananya, Anjan Kumar Talukdar, and Kandarpa Kumar Sarma. "A review on vision-based hand gesture recognition and applications." *Intelligent Applications for Heterogeneous System Modeling and Design*. IGI Global, 2015. 256-281.
- [11] Camgoz, Necati Cihan, et al. "Using convolutional 3d neural networks for user-independent continuous gesture recognition." *Pattern Recognition (ICPR)*, 2016 23rd International Conference on. IEEE, 2016.
- [12] Li, Yi. "Hand gesture recognition using Kinect." *Software Engineering and Service Science (ICSESS)*, 2012 IEEE 3rd International Conference on. IEEE, 2012.
- [13] Ren, Zhou, et al. "Robust part-based hand gesture recognition using kinect sensor." *IEEE transactions on multimedia* 15.5 (2013): 1110-1120.
- [14] Goga, Jozef, and Slavomir Kajan. "Hand gesture recognition using 3D sensors." *ELMAR*, 2017 International Symposium. IEEE, 2017.
- [15] John, Vijay, et al. "Deep learning-based fast hand gesture recognition using representative frames." *Digital Image Computing: Techniques and Applications (DICTA)*, 2016 International Conference on. IEEE, 2016.
- [16] Duan, Jiali, et al. "Multi-modality fusion based on consensus-voting and 3d convolution for isolated gesture recognition." *arXiv preprint arXiv:1611.06689* (2016).
- [17] More, Sagar P., and Abdul Sattar. "Hand gesture recognition system using image processing." *Electrical, Electronics, and Optimization Techniques (ICEEOT)*, International Conference on. IEEE, 2016.
- [18] Ghotkar, Archana S., and Gajanan K. Kharate. "Vision based real time hand gesture recognition techniques for human computer interaction." *Int J Comput Appl* 70.16 (2013): 1-8.
- [19] Khan, Rafiqul Zaman, and Noor Adnan Ibraheem. "Survey on gesture recognition for hand image postures." *Computer and information science* 5.3 (2012): 110.
- [20] Hu, Zhongxu, et al. "3D separable convolutional neural network for dynamic hand gesture recognition." *Neurocomputing* 318 (2018): 151-161.
- [21] Molchanov, Pavlo, et al. "Hand gesture recognition with 3D convolutional neural networks." *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2015.
- [22] Biswas, Kanad K., and Saurav Kumar Basu. "Gesture recognition using microsoft kinect®." *Automation, Robotics and Applications (ICARA)*, 2011 5th International Conference on. IEEE, 2011.
- [23] Du, Heng, and T. To. "Hand gesture recognition using Kinect." *Techical Report*, Boston University (2011).
- [24] Ren, Zhou, et al. "Robust hand gesture recognition with kinect sensor." *Proceedings of the 19th ACM international conference on Multimedia*. ACM, 2011.