ISSN: 2366-1313

AUTOMATED HAND GESTURE RECOGNITION USING A DEEP CONVOLUTIONAL NEURAL NETWORK MODEL

Dr. S. Gopi Krishna, Shaik Jasmin Taj

[1]Head of The Depertment of Computer Science & Engineering, Sri Mittapalli College of Engineering, Tummalapalem, Guntur, 522233, India.

[2]UG Sri Mittapalli College of Engineering, Tummalapalem, Guntur, 522233, India.

Abstract:

The recognition of gestures by hand is significant for interaction between humans and computers. In this painting we propose a new method of the recognition of hand movements. In our system hand location is removed from the past using the method of subtraction from heritage. The palms and palms are divided so it is possible to identify and comprehend the fingers. In the end the rule classifier is used to determine the meanings for hand movements. The tests on the statistical set of 1300 images demonstrate that our approach works as expected and is remarkably green. In addition, our approach provides better performance than a Kingdom-of-Art approach on a different collection of hand gestures.

Keywords: Artificial neural networks, convolutional neural network, empirical mode decomposition, hand gesture recognition, wavelet transform.

I INTRODUCTION

Currently, direct contact is the dominant form of interaction between the user and the machine. The interacting channel is based on devices such as the mouse, the keyboard, the remote control, touch screen, and other direct contact methods. Human to human interaction is achieved through more natural and intuitive noncontact methods, such as sound and physical movements. The flexibility and efficiency of noncontact interaction methods has led many researchers to consider exploiting them to support the human computer interaction. Gesture is one of the most important noncontact human interaction methods and forms a substantial part of the human language. Historically, wearable data gloves were usually used to obtain the angles and positions of each joint in the user's gesture. One of the recent year's trends is deep learning (DL) technology which is used in the digital image processing for solving complex problems (a classification, segmentation and image detection). DL techniques, such as convolutional neural networks (CNNs), have already influenced a wide range of signal processing activities within traditional and new advanced areas, including key aspects of machine learning and artificial intelligence [1]. In particular, CNNs showed superior performance in face detection applications [2, 3]. Furthermore, DL has made considerable progress in detection and classification of the hand gestures for implementation into the human computer interaction (HCI) technologies [4-6]. A gesture is a configuration and/or movement of a body part that expresses an emotion, intention or command. A set of gestures and their meanings form the gestures vocabulary. Gestures can be divided into two types: static and dynamic gestures. In static gestures the hand position does not change during the gesture demonstration. Static gestures mainly rely on the shape and flexure fingers angles. In the second case, hand position changes continuously so that dynamic gestures rely on the hand trajectories and orientations, in addition to the shape and fingers angles [5]. Intel Inc. created RealSense deep vision technology and developed (in collaboration with Microsoft) the tool for 3D face recognition, which provides access to Windows10 devices [7]. RealSense technology supported by an open source software development kit [8]. The second generation RealSense cameras and its stereo vision to calculate depth [9] were introduced in January 2018. Despite the fact that Intel RealSense D400 series devices appeared on the market only in recent years, they began to be used in many areas, for example, in security systems [10], robotics [11], medicine [12], and agriculture [13]. At the same time, there are no works that report on the RealSense D400 using for recognizing hand gestures that can be integrated into effective HCI systems. In this regard, the aim of this work is to develop a gesture recognition approach based on the combined use of the RealSense D435 depth sensor for gesture capture and a pre-trained convolutional neural network with VGG-16 architecture for features extraction and objects classifying. RealSense libraries from Intel, OpenCV and open-source DL frameworks Keras and TensorFlow are used for software implementation of the gesture recognition system on Python. To determine the performance of our approach we collected a database of 1,000 images, which consists of 40 different types for 5 gestures, which were presented to the sensor by 5 people, and tested the recognition system on the database..

II. LITERATURE SURVEY

Image processing is central to hand gesture recognition. Digital image processing is most relevant to our work where useful information related to hand gesture and movement need to be extracted from digital images. Segmentation is crucial in gesture recognition [4], [5]. Hand detection and

background removal are vital to the success of the gesture recognition algorithm. In previous work, a monocular camera was used in gesture recognition algorithms to filter out the background, which can be inconvenient in a real-world setting. Most methods used in hand detection are based on Harr features, colour, context information, or even shape. Such methods can provide accurate performance given the successful identification of the background and the hand in the image. However, there are limitations, e.g. a hand detection method relying on the skin colour will fail if the person is wearing a glove [6].

Feature detection is a crucial part of 2-D and 3-D image processing [7]. Before any feature extraction technique is applied, the image data is pre-processed, and different pre-processing techniques are applied to it including thresholding, binarization, and normalization. Features are then extracted and used for classification purposes. The behaviour of an image is captured based on its features. A good feature set contains attributes with high information gain and can be used to effectively classify images into different groups. A method that utilizes the symmetric properties of visual data to detect spare and stable image features was presented by Huebner and Zhang [8]. Regional features were formed by using Qualitative Symmetry operator together with quantitative symmetry range information. Extraction and classification of local image structures are crucial to gesture recognition. Gevers et al. [9] proposed a method to classify the physical nature of local image structure using the geometrical and photometrical information.

To make the hand gesture recognition more accurate and thus ensuring a more natural user experience interacting with the machine interface, Bouchrika et al. [10], [11] applied a Wavelet Network Classifier (WNC) in a remote computer ordering application using hand gestures to place orders. Hands detection, tracking and gesture recognition techniques were applied. WNC was used for its effective classification results. An approach also proposed by Bouchrika et al. [12] made amendments to the Wavelet Network classification phase by making separated Wavelet Networks discriminating classes (n − 1) with the purpose of training each image. This resulted in less time required to complete the testing phase.

The proposed Wavelet Network architecture enables quick learning and recognition of actions by avoiding unnecessary hand movements. Another hand gesture recognition approach [13] was based on wavelet enhanced image pre-processing and supervised Artificial Neural Networks (ANN).

Contour segmentation was supported in the pre-processing. Reference points were used to provide 2-D hand gesture contour images to 1D signal conversion. Wavelet decomposition was used for 1D signals. Four statistical features were extracted from the wavelet coefficients. Six hand gestures were tested. An accuracy of 97% was achieved with fast feature extraction and computation. Murthy and Jadon [14] proposed a method for hand gesture recognition using Neural Networks. It is based on supervised feed-forward neural network net-based training and back-propagation technique to classify hand gesture in ten various categories including hand pointing up, down, left, right, front, etc. Convolution Neural Networks (CNNs) were used [15] to evaluate hand gesture recognition, where depth-based hand data was employed with CNN to obtain successful training and testing results. Another CNN method was proposed [16] that uses a skin model, hand position calibration and orientation to train and test the CNN..

III Methodology

Datasets

We prepared a new database which contains images with segmented static hand gestures shown in Figure. We selected these gestures, which are also included in alternative dataset [24, 25] and can be used to cross-validate proposed static hand gesture recognition system model. Depth camera Intel RealSense D400 placed over a table, the subjects sat close to it, and moved their right hand over the sensor at a distance between 10 and 75cm in front of it as shown in Figure 5. The gestures were performed by 10 different subjects (5 women and 5 men). Our dataset has a total of 2,000 pictures, including 1,000 RGB images and 1,000 depth maps, collected under different backgrounds in a few rooms with illumination changes. In order to increase the diversity of the database, we also tried image augmentation by using of flips, skews, rotations and etc. The static hand gestures database was used only for the training of the modified VGG-16. At the validation/testing stage RGB and depth images from RealSense D400 used directly to predict gesture. DCNN training is the fine-tuned process based on the pre-trained models VGG-16 on ImageSet. At the start, the learning rate is 0.01 and then decreases by 10 times every 2000 iterations. Weight decay is set to0.0004. At most 10,000 iterations were needed for the fine-tuning on training set. The experiments were done on a Intel(R) Core(TM) i7- 9750H CPU, NVIDIA GeForce GTX 1650, 16 GB RAM desktop. After performing transfer learning with a fine tuned last activation layer we were able to achieve an average accuracy of 95,5 % on the 5-class subset of the static hand gesture.

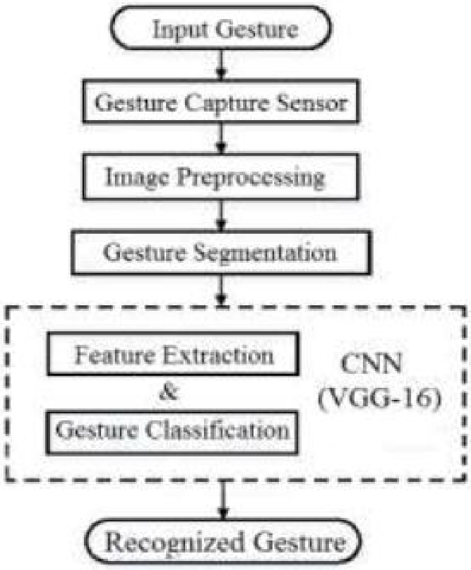

Figure : Block diagram of proposed static hand gesture recognition system

Performance metrics

The confusion matrix (CM) has been used for obtaining quantitative results. The precision, recall, and F-Score measures are used as metrics for evaluation. CM of size n n associated with a classifier shows the predicted and actual classification, where n is the number of different classes [25]. We use an alternative construction of CM: each column represents the percentage of hand gestures that belongs to each class, and along the first diagonal are the correct classifications, whereas all the other entries show misclassifications [24]. From CM matrix two measurements can be directly observed, the precision and the recall, which can be expressed as follows:

Precision = d/d+b

Recall = 100*d/d+c

IV Proposed Method

A. Dataset

Hand Gestures Input

Hand gestures represent the input to different gesture detection methods evaluated in this study.ten 2-D and 3-D hand gestures with plain backgrounds. They are recorded within long distances and used in the study's experimental work. The implementation framework illustrating the extraction and the classification steps is shown in Fig. 2. Using an iPhone 6 Plus camera with resolution 4k at 30 fps, the hand motions shown in Fig. 1 are recorded. Each recording lasts 10 seconds and the resolution of the recorded video is 3840 × 2160. The first system is created using optical flow object by estimating and displaying the optical flow of objects in the video. The length of videos is between 15 to 65 frames. Each video has a different number of frames, which depends on the first section of motion.

Computing Platform Specification

The experiment was performed using a Dell laptop XPS 15 9550 with 6 th processor Generation Intel Quad Core i7, memory type DDR4 16 GB, speed 2133 MHz, 512 GB storage hard drive, 15.6-inch Ultra-HD 15.6" IPS 1920×1080 RGB Optional 3840×2160 IGZO IPS display w/Adobe RGB colour space and touch. Windows 10 (64 bits) operating system was used and the system is implemented using MATLAB R2017bV language.

Implementing Wavelet Transform with ANN

The system is implemented using the db8 WT tool following the steps outlined below:

1. Read each video using a video reader function.

2. Create an 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 IMF

V. RESULT ANALYSIS

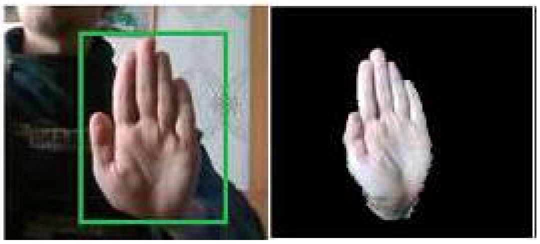

Figure: Hand gesture segmentation 1

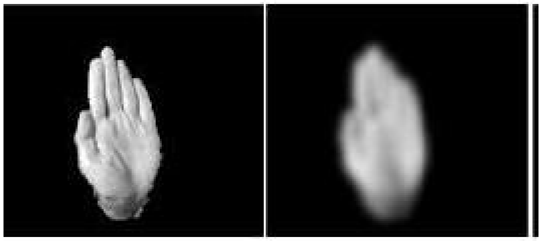

Figure: Hand gesture segmentation 2

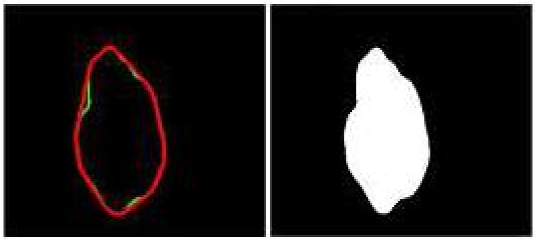

Figure: Hand gesture segmentation 3

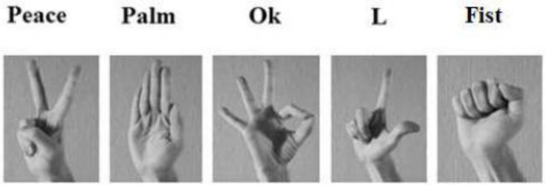

Figure: Data Set evaluation

REFERENCES

[1] S. R. Sree, et al., “Real-World Application of Machine Learning and Deep Learning,” in 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT), pp. 1069-1073, 2019.

[2] T. V. Janahiraman and P. Subramaniam, “Gender Classification Based on Asian Faces using Deep Learning,” in 2019 IEEE 9th International Conference on System Engineering and Technology (ICSET), pp. 84-89, 2019.

[3] R. I. Bendjillali, et al., “Illumination-robust face recognition based on deep convolutional neural networks architectures,” Indonesian Journal of Electrical Engineering and Computer Science, vol. 18, no. 2, pp. 1015-1027, 2020.

[4] A. K. H. AlSaedi and A. H. H. AlAsadi, “A new hand gestures recognition system,” Indonesian Journal of Electrical Engineering and Computer Science, vol. 18, no. 1, pp. 49-55, 2020.

[5] P. K. Pisharady and M. Saerbeck, “Recent methods and databases in vision-based hand gesture recognition: A review,” Computer Vision and Image Understanding, vol. 141, pp. 152-165, 2015.

[6] B. K. Chakraborty, et al., “Review of constraints on vision-based gesture recognition for human–computer interaction,” IET Computer Vision, vol. 12, no. 1, pp. 3-15, 2017.

[7] L. Keselman, et al., “Intel R RealSense TM Stereoscopic Depth Cameras,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 1-10, 2017.

[8] Intel® RealSense™ SDK 2.0. https://www.intelrealsense.com/developers/.

[9] Intel RealSense D400 Series Product Family. Datasheet. 2019 Intel Corporation. Document Number: 337029-007. https://www.intel.com/.

[10] R. D. Bock, “Low-cost 3D security camera. Autonomous Systems: Sensors, Vehicles, Security, and the Internet of Everything,” International Society for Optics and Photonics, vol. 10643, pp. 106430E, 2018.

[11] Q. Fang, et al., “RGB-D Camera based 3D Human Mouth Detection and Tracking Towards Robotic Feeding Assistance,” in Proceedings of the 11th Pervasive Technologies Related to Assistive Environments Conference, pp. 391-396, 2018.

[12] H. Aoki, et al., “Study on Non-Contact Heart Beat Measurement Method by Using Depth Sensor,” in World Congress on Medical Physics and Biomedical Engineering, pp. 341-345, 2019.

[13] T. N. Syed, et al., “Seedling-lump integrated non-destructive monitoring for automatic transplanting with Intel RealSense depth camera,” Artificial Intelligence in Agriculture, vol. 3, pp. 18-32, 2019.

[14] B. Liao, et al., “Hand gesture recognition with generalized hough transform and DC-CNN using realsense,” in 2018 Eighth International Conference on Information Science and Technology (ICIST), pp. 84-90, 2018.

[15] M. B. Holte, et al., “Fusion of range and intensity information for view invariant gesture recognition,” in 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp. 1-7, 2008.

[16] M. Van den Bergh, et al., “Real-time 3D hand gesture interaction with a robot for understanding directions from humans,” in 2011 Ro-Man, pp. 357-362, 2011.

[17] Z. Ren, et al., “Robust hand gesture recognition based on finger-earth mover's distance with a commodity depth camera,” in Proceedings of the 19th ACM international conference on Multimedia, pp. 1093-1096, 2011.

[18] D. Wu, et al., “One shot learning gesture recognition from rgbd images,” in 2012 IEEE Computer Society Conference on Computer Vision

and Pattern Recognition Workshops, pp. 7-12, 2012.

[19] C. Keskin, et al., “Randomized decision forests for static and dynamic hand shape classification,” in 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp. 31-36, 2012.

[20] V. Chernov, et al., “Integer-based accurate conversion between RGB and HSV color spaces,” Computers & Electrical Engineering, vol. 46, pp. 328-337, 2015.