Online Preprocessing of Gesture Signs using Background Substructure and Edge Detection Algorithms

Farah F. Alkhalid

Control and Systems Engineering, University of Technology, Iraq.
10352@uotechnology.edu.iq

***Abstract* - It is very important to communicate with the mute and deaf people, and listen to what they need. The gesture sign is one of the first methods of communication between the mutes and deaf on one hand, and other people on the other hand. This way has rules and basics that should be respected and be familiar with, at the same time; most people cannot translate sign language. For this reason, this research proposes a model to translate sign language by preprocessing sign images in order to get an accurate classification. In this paper, background subtraction and edge detection are used together to present images with high edge details for the gesture sign, which is the region of interest, this work is done using OpenCV in python3. The results show that using background subtraction and edge detection algorithms produced better quality images for the translation model with testing accuracy is 85.71%.**

Keywords - Sign language, OpenCV, Background Subtractor, Canny edge detection.

I. INTRODUCTION

Every day, disabled individuals are facing challenges to communicate with society. It is hard for the average citizens to comprehend the signals done by the mute individuals. By using deep learning models to create the necessary translation system, sign language may be available for the ordinary person to understand. The deep learning model must be trained to many signs in order to decide what the new sign is.

Before using the model, it is necessary to initiate a preprocessing phase to the captured images in order to make the classification more reliable.

Problem Statement: In this paper, hand motions are extracted from video images. To exract these signals from a live video sequence, we first need to isolate the local hand alone by deleting all the unnecessary parts in the video image. This process is carried out in two stages:

- Canny Edge Detection.
- Background Subtracteor and Threshold.

II. RELATED WORK

Pinto et al [1], investigated a hand recognition system utilizing Convolutional Neural Network (CNN). The system included the application of morphological filters, contrast generation, polygonal approximation, and segmentation during preprocessing, in which they prsented to a good feature extraction, but using skin color classification has problem like shade and shadow occlusions, resolution as well as skin tone variation between races.

In reference [2], the authors were interested in hand gesture and sign recognition so they anticipated the preprocessing steps of the comprising of picture securing, pre-handling, division, highlight extraction and classifiers, utilized procedures, and division. The paper tracked the timeline of these methods and there effect on the Indian sing language. However, no true calculations or results were found in this research that can conclude which method is the best to use in this area.

Priyanka and Dr. Kamalraj in [3] proposed calculation that utilizes the Non-neighborhood Mean channel for clamor expulsion utilizing Global Swarm Optimization (GSO) based Canny edge recognition for separating the edges. Highlights are separated utilizing two-dimensional Multi-goals Two Dimensional Discrete Wavelet Transform (2D-DWT) joined with Gray-level Co-event Matrix.

ASL Letters is presented In [4], in order to acknowledge the framework dependent on a multi-see expansion and deduction combination utilizing CNN. This technique recovered 3D information from a profundity picture and made an increasing viewpoint for viable preparation and decreased overfitting. Nevertheless, the decent variety of the picture could not permit perceiving the constant motion of a particular sign.

Otiniano Rodriguez et al. proposed an American Sign Language (ASL) acknowledgment framework utilizing the Kinect sensor [5]. They accomplished preferable outcomes over a single information framework, contrasted with Red, Green, and Blue (RGB) pictures, profundity pictures, and execution on both in this framework.

Rahim et al. also presented the Kinect sensor for Human Computer Interaction (HCI) [6]. The territory of hand and fingertips were related to shape boundaries and palm position, and the motion of the hand was perceived by estimating skeletal information.

The author of [7], provided an invariant framework that was able to detect occluded gestures. The work dealt with a hand-gesture recognition system using Kinect sensor skeleton data. However, there may still be concerns about

distance, and it is not clear how the presented algorithm recognized the “no gesture” situation.

In [8], the Kinect sensing device was used to perform the comprehensive application of continuous activity recognition, and it determined the sequence of activities from 3D skeleton information. However, the authors did not explain how these were conducted during the tracking activity.

Shin et al. proposed Japanese and English character input systems based on the hand tapping of gestures [9]. However, the system required large computational time, and the input characters were hard to remember by the user’s hand tapping gestures.

In [10], the author introduced a 3D movement sensor-based ASL acknowledgment framework. They utilized deep learning to group the 26 letters of the English letter set in ASL which are got from the tangible information.

In this research, we used an edge detection algorithm with a background subtraction algorithm in order to specify the edges because edges have many details, and ignore the background, which has unwanted features and showed satisfactory results.

III. THE PROPOSED SYSTEM

Since all datasets of sign language are captured with empty background, the focus will be on the Region of Interest (ROI), which refers to gesture sign. Therefore, to make the training model successful it is important to provide gesture signs with an empty background. Since the camera returns video streams, so it must be divided into frames to be dealt with as digital images.

The operation sequence of the proposed model is shown in Figure (1). The fixed camera is shooting video stream, which has to be transferred into frames in order to treat them as a digital image after that applying the edge detection algorithm and background subtraction algorithm, the two algorithms will dispose the background and detect the edges efficiently to show more details of the gesture sign.

After preprocessing the input image, the deep learning model-based CNN will recognize the letter represented by the gesture sign, and then the model will write this letter and read it, repeat the operation to the next letter and so on.

The newly dealt image will be the input to the recognition model designed and modeled using deep learning after training. After completing the recognition phase, the system should split the letters and then write them. If the input image is not recognized, then the system will try to capture another frame and repeat the steps from the beginning.

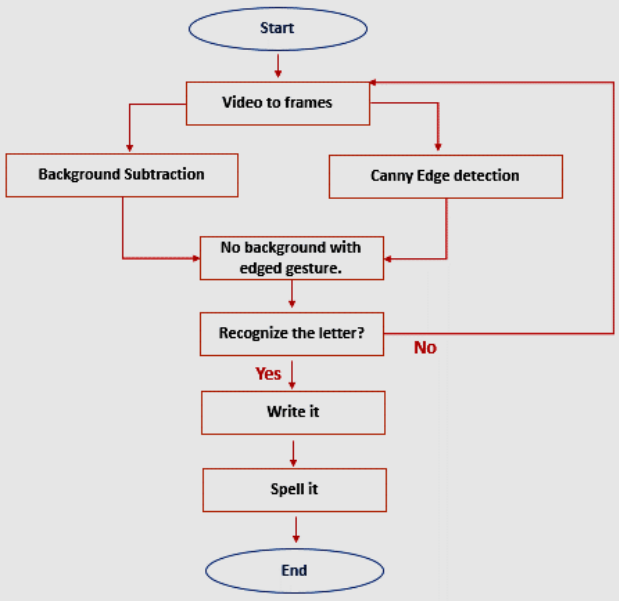

Figure 1. Proposed System Diagram.

A. Canny Edge Detection

The procedure of edge detection includes recognizing sharp edges in the picture and creating a paired picture as they appear. Commonly, the procedure draws white lines on a dark foundation to demonstrate those edges. Each edge location is considered as a high pass sifting activity. A high pass channel permits high-recurrence substances to go through and obstructs the low-recurrence content. edges are high-recurrence content. In edge discovery, it is needed to hold these edges and dispose of everything else. Consequently, a part that is what could be compared to a high pass channel, should be assembled [11].

Edge recognition has a great deal of unmistakable quality in PC vision. It manages the forms of a picture generally signified as a framework of a particular article in a picture. Figure (2) shows this method [12].

Canny's edge detection algorithm is the most widely used because of its ease of use and accuracy levels. Canny edge detection is a multi-step algorithm that can detect edges with noise suppressed at the same time.

Figure 2. Canny Edge Detection

The Process of Canny's edge detection algorithm starts by Applying a Gaussian filter to smooth the image in order to remove the noise; a Gaussian filter kernel is convolved with the image. The equation for a Gaussian filter kernel of size $(2k+1)\times(2k+1)$ is given by:

$$H_{ij}=\frac{1}{2\pi\sigma^2}exp\left(-\frac{(i-(k+1))^2+(j-(k+1))^2}{2\sigma^2}\right)$$
$$1 \le i,j \le (2k+1) \quad (1)$$

Where σ =1.4.

After filtering, the intensity gradient of the image must be found by using four channels to identify flat, vertical and corner-to-corner edges in the obscured picture. The edge identification administrator restores an incentive for the main subordinate in the flat heading (G_x) and the vertical bearing (G_y). From this, the edge slope and course can be resolved:

$$G=\sqrt{G_x^2+G_y^2} \quad (2)$$
$$\Theta= \text{atan2}(G_y,G_x) \quad (3)$$

The edge heading is adjusted to one of four edges which are vertical, flat and the two diagonals (0°, 45°, 90°, and 135°). An edge course falling in each shading area will be set to a particular edge esteems, for example, θ in [0°, 22.5°] or [157.5°, 180°] maps to 0° [13].

B. Background Subtractor and Threshold Algorithm

The background subtraction strategy performs well for situations where there is a need to distinguish moving items in a static scene. As the name demonstrates, this calculation works by recognizing the foundation and subtracting it from the present edge to acquire the forefront [14]. It is a significant preprocessing step in numerous vision-based applications. One significant component of this calculation is that it chooses the fitting number of Gaussian dissemination for every pixel [12]. Figure 3 shows the effect of this algorithm.

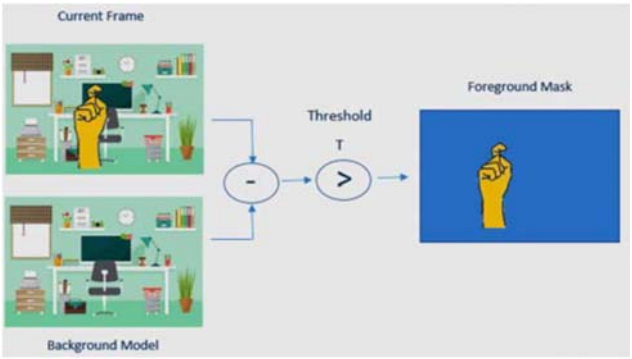

Figure 3. Background Subtractor and threshold algorithm

On the off chance that there is a picture or a casing of video that simply has the static foundation and no moving guests, it is a clear errand since it should simply subtract the new picture from the foundation to remove the frontal area alone [11,12].

$$Id2\ (x,y) = \begin{cases} 1, & if\ Id[x,y] > T \\ 0, & Otherwise \end{cases} \quad (4)$$

Where:
Id2 = the new pixel value.
(x,y) = Image pixel coordinates.
T = Threshold Value.

IV. RESULTS AND DISCUSSION

The main goal of this research is to make a real-time image preprocessing for streaming videos captured by a fixed camera to present the ideal state of the input image of a hand sign language to be translated to the corresponding letter.
Two algorithms are used together, Canny's edge detection and background subtraction with a threshold.

A. Canny's Edge Detection Results

Using Canny's edge detection is very useful in image recognition of features; hence, the edges can give more details, because any change in the finger location refers to a different letter, still the image is crowded and has unwanted

features that lie in the background, so by disposing of the background, the region of interest will be more efficient which leads to a more accurate result, as in Figure 4.

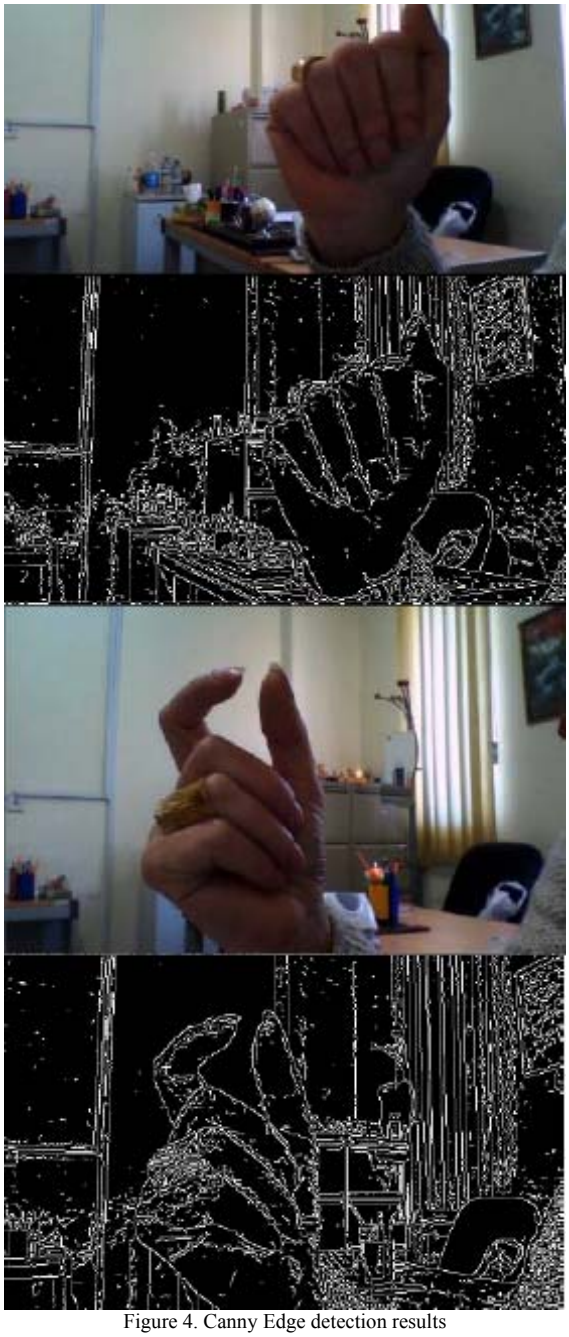

Figure 4. Canny Edge detection results

B. Background Subtraction Results

The main idea of background subtraction is to remove all static elements and show the moving one. Since the camera is fixed so all fixed things will be determined as background and should be removed, and the moving hand that represents different signs will appear, as shown in figure 5.

Figure 5. Background Subtraction Results

C. Combination of the Two Algorithms Results

The effects of using the previous two algorithms can present high focusing on hand gesture which is the region of interest in the Arabic sign language translation model, the results are as denoted in figure 6. The code is available on [15].

Figure 6. Two algorithms result.

D. Test Results

The next process is to feed the image to CNN model to make recognition, the accuracy testing is 85.71% whith 14 images as test which is very accepted accuracy, Figure 7 shows the accuracy testing of image regarding epochs.

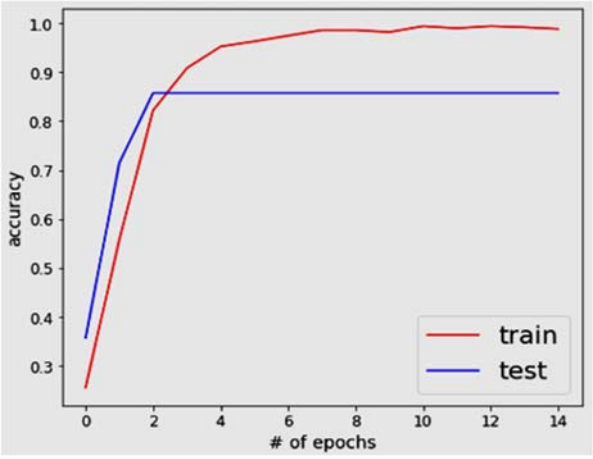

Figure 7. accuracy testing of image

V. CONCLUSION

The gestures presented in diverse backgrounds have to be accurately processed and segmented so that it can be classified precisely by the hand gesture recognition system. This study compares the performance of the proposed Image Segmentation Algorithm with a standard Canny's Edge Detection Algorithm and Background Subtraction Algorithm. These two techniques have worked together to submit a new image with a gesture sign with a high edge to represent the pause of the sign. The two algorithms are free, fast, and supports all operating systems. The image will enter to the sign recognition model using deep learning which is the next step of this research in order to dispose of the unwanted features, the results are satisfying and have details that are needed to be fed to the next recognition model, where the testing accuracy is =85.71% when using the two algorithms, the source code is available in [15].

ACKNOWLEDGMENT

I would like to extend my sincere thanks to “Dr. Shibly Ahmed Hameed” in the Control and Systems Engineering Department / the University of Technology-Iraq for his assistance in checking the plagiarism of this research.

REFERENCES

[1] Raimundo F. Pinto, Carlos D. B. Borges,1 Antônio M. A. Almeida and Iális C. Paula, “Static Hand Gesture Recognition Based on Convolutional Neural Networks”, Journal of Electrical and Computer Engineering, Hindawi, Volume 2019, |Article ID 4167890, page 12,

[2] Divyashree B, Manjushree K, “Image Processing Techniques for Hand Gesture and Sign Recognition”, International Research Journal of Engineering and Technology, Vol 06 Issue 01, Jan 2019.

[3] Priyanka Parvathy, Dr. Kamalraj Subramaniam, “Performance comparison of segmentation algorithms for hand gesture recognition”, International Journal of Engineering & Technology, 7 (3) (2018) 1227-1232, 2018.

[4] Tao, W.; Leu, M.C.; Yin, Z. American Sign Language alphabet recognition using Convolutional Neural Networks with multiview augmentation and inference fusion. Eng. Appl. Artif. Intell. 2018, 76, 202–213.

[5] Rodriguez, K.O.; Chavez, G.C. Fingerspelling recognition from RGBD information using kernel descriptor. In Proceedings of the IEEE XXVI Conference on Graphics, Patterns and Images, Arequipa, Peru, 5–8 August 2013; pp. 1–7.

[6] Rahim, M.A.; Shin, J.; Islam, M.R. Human-Machine Interaction based on Hand Gesture Recognition using Skeleton Information of Kinect Sensor. In Proceedings of the 3rd International Conference on Applications in Information Technology, Aizu-Wakamatsu, Japan, 1–3 November 2018; pp. 75–9

[7] Kumar, P.; Saini, R.; Roy, P.P.; Dogra, D.P. A position and rotation invariant framework for sign language recognition (SLR) using Kinect. Multimed. Tools Appl. 2018, 77, 8823–8846.

[8] Saini, R.; Kumar, P.; Roy, P.P.; Dogra, D.P. A novel framework of continuous human activity recognition using Kinect. Neurocomputing 2018, 311, 99–111.

[9] Shin, J.; Kim, C.M. Non-touch character input system based on hand tapping gestures using the Kinect sensor. IEEE Access 2017, 5, 10496–10505.

[10] Chuan, C.H.; Regina, E.; Guardino, C. American sign language recognition using the leap motion sensor. In Proceedings of the IEEE 13th International Conference on Machine Learning and Applications, Detroit, MI, USA, 3–6 December 2014; pp. 541–544.
[11] Gabriel Garrido, Prateek Joshi, "OpenCV 3.x with Python By Example Second Edition", Packta, BIRMINGHAM – MUMBAI, 2018.
[12] Sunila Gollapudi , "Learn Computer Vision Using OpenCV With Deep Learning CNNs and RNNs", ISBN-13 (electronic): 978-1-4842-4261-2, Hyderabad, Telangana, India, Apress, 2019
[13] https://opencv-python-tutroals.readthedocs.io/en/latest/ py_tutorials/py_imgproc/py_canny/py_canny.html, [accessed Jan 08, 2020]
[14] https://docs.opencv.org/3.4/d1/dc5/ tutorial_background_subtraction.html [accessed Jan 08 2020]
[15] https://drive.google.com/file/d/ 1XIyh29CTgsQXQKMghWs2FrQPZDmeRYFM/view?usp=sharing.